6.1 The Equation Error method

The function to be minimized is given by

$$W_e(n) = E\{[\hat{A}_n(q)d(n) - \hat{B}_n(q)x(n)]^2\} \tag{70}$$

Only to introduce the EE method, consider

- $\nu(n)$ is the measurement noise such that $\boldsymbol{\nu}(n) = [\nu(n-1), \ldots, \nu(n-N)]^T$.
- $\boldsymbol{\theta} = [a_1, \ldots, a_N, b_0, \ldots, b_N]^T$, i.e., the (monic constrained) parameter vector.
- $\boldsymbol{\varphi}(n) = [d(n-1), \ldots, d(n-N), x(n), \ldots, x(n-N)]^T$, the regressor.
- The ideal model associated: $\boldsymbol{\theta}_0 = [a_1^o, \ldots, a_{n_a}^o, b_0^o, \ldots, b_{n_b}^o]^T$
- $e_e(n) = d(n) - \boldsymbol{\varphi}^T(n)\boldsymbol{\theta}(n)$ is the equation error: **linear in the parameters!**.

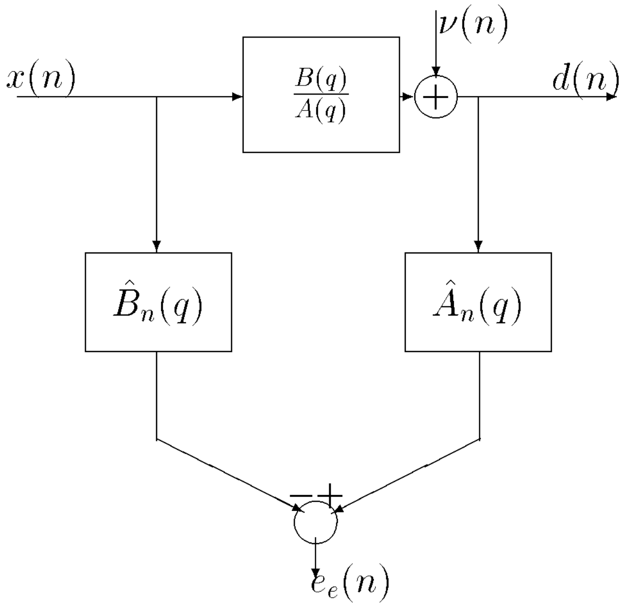


Figure 39: Equation Error Method

The stochastic gradient version

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{\varphi}(n)e_e(n)$$

or the *Gauss-Newton* version

$$\begin{aligned}\boldsymbol{\theta}(n+1) &= \boldsymbol{\theta}(n) + \mu\boldsymbol{P}(n+1)\boldsymbol{\varphi}(n)e_e(n) \\ \boldsymbol{P}(n+1) &= \left(\frac{1}{1-\mu}\right)\left(\boldsymbol{P}(n) - \frac{\boldsymbol{P}(n)\boldsymbol{\varphi}(n)\boldsymbol{\varphi}^T(n)\boldsymbol{P}(n)}{\frac{1-\mu}{\mu} + \boldsymbol{\varphi}^T(n)\boldsymbol{P}(n)\boldsymbol{\varphi}(n)}\right)\end{aligned}$$

For the strictly sufficient order case and $\nu(n) = 0$, the mean behavior of the LMSEE algorithm can be analyzed using

$$E\left\{\tilde{\boldsymbol{\theta}}(n+1)\right\} = (\boldsymbol{I} - \mu\boldsymbol{R})E\left\{\tilde{\boldsymbol{\theta}}(n)\right\}$$

where $\boldsymbol{R} = E\{\boldsymbol{\varphi}(n)\boldsymbol{\varphi}^T(n)\}$.

Assuming that $\boldsymbol{R}$ is positive definite, it can be decomposed as $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is diagonal, formed with the eigenvalues of $\boldsymbol{R}$.

Premultiplying both sides of the previous equation by $\boldsymbol{Q}^T$, it can be shown that the resulting system converge *in the mean* to the solution of

$$\boldsymbol{\theta}^* = \boldsymbol{R}^{-1}E\left\{\boldsymbol{\varphi}(n)d(n)\right\}$$

when $n \to \infty$, if the convergence factor μ satisfy

$$0 < \mu < \frac{2}{\lambda_N}$$

where λ_N is the maximum eigenvalue of $\boldsymbol{R}$.

6.2 Generic stability properties

- Even for $\nu(n) \neq 0$ under certain conditions the stability of the estimate can be guaranteed.

- Define the signal-to-noise ratio by $\mathcal{S} = \frac{E\left\{\left(\frac{B(q^{-1})}{A(q)}x(n)\right)^2\right\}}{E\{\nu^2(n)\}} = \frac{E\left\{y^2(n)\right\}}{E\{\nu^2(n)\}}$ and $\boldsymbol{\theta} = [\boldsymbol{a}\ \boldsymbol{b}]^T$ and $\boldsymbol{\theta}_0 = [\boldsymbol{a_0}\ \boldsymbol{b_0}]^T$.

- Using (70), $\left[\boldsymbol{R}_\nu + \boldsymbol{R}_{y/x}\right]\boldsymbol{a} = -\boldsymbol{r}_\nu + \boldsymbol{R}_{y/x}\boldsymbol{a_0}$, where

$$\begin{aligned}
\boldsymbol{R}_\nu &= E\begin{bmatrix}\nu(n-1)\\ \vdots \\ \nu(n-n_a)\end{bmatrix}[\nu(n-1)\ \cdots\ \nu(n-n_a)\,] \\
\boldsymbol{r}_\nu &= E\begin{bmatrix}\nu(n-1)\\ \vdots \\ \nu(n-n_a)\end{bmatrix}\nu(n) \\
\boldsymbol{R}_{y/x} &= \boldsymbol{R}_y - \boldsymbol{R}_{xy}^T\boldsymbol{R}_x^{-1}\boldsymbol{R}_{xy}
\end{aligned}$$

- Then, for large $\mathcal{S}$,

$$\boldsymbol{a} \;=\; \boldsymbol{a_0} - \boldsymbol{R}_{y/x}^{-1}[\boldsymbol{r}_\nu + \boldsymbol{R}_\nu\boldsymbol{a_0} + O(\|R_\nu\boldsymbol{R}_{y/x}^{-1}\|^2) = \boldsymbol{a_0} + O(1/\mathcal{S})$$

- On the other hand, for small $\mathcal{S}$, we get

$$\begin{aligned}
\boldsymbol{a} &= -\boldsymbol{R}_\nu^{-1}\boldsymbol{r}_\nu + \boldsymbol{R}_\nu^{-1}\boldsymbol{R}_{y/x}[\boldsymbol{a_0} + \boldsymbol{R}_\nu^{-1}\boldsymbol{r}_\nu] + O(\|R_\nu^{-1}\boldsymbol{R}_{y/x}\|^2) \\
&= -\boldsymbol{R}_\nu^{-1}\boldsymbol{r}_\nu + O(\mathcal{S})
\end{aligned}$$

- Then, to summarize,

  Lemma: For a signal-to-noise ratio given by $S = \frac{E\left\{\left(\frac{B(q))}{A(q)}x(n)\right)^2\right\}}{E\{\nu^2(n)\}}$, $A_n(q)$ has zeros inside the stability region if some of the following conditions is satisfied

  – S is sufficiently high.

  – S is sufficiently low.

6.3 Stationary points and ODE associated

6.3.1 The sufficient order case

$$E\left[\boldsymbol{\varphi}(n)\, e_e(n)\right] = 0$$

or alternatively

$$E\left\{\left[\begin{array}{c}\frac{B(q)}{A(q)}x(n-i)+\nu(n-i)\\ x(n-j)\end{array}\right][\tfrac{(\overline{A}(q)B(q)-\overline{B}(q)A(q))}{A(q)}x(n)+\overline{A}(q)\nu(n)]\right\}=0$$

$$E\left[\begin{array}{c}\frac{B(q)}{A(q)}x(n-i)\\ x(n-j)\end{array}\right][\tfrac{(\overline{A}(q)B(q)-\overline{B}(q)A(q))}{A(q)}x(n)]+E\left[\begin{array}{c}\boldsymbol{\nu}(n)\\ 0\end{array}\right]\nu(n)$$

$$-E\left[\begin{array}{c}\boldsymbol{\nu}(n)\\ 0\end{array}\right]\left[\begin{array}{c}\boldsymbol{\nu}(n)\\ 0\end{array}\right]^T\overline{\boldsymbol{\theta}}=0$$

for $i = 1, ..., N$, e $j = 0, ..., N$.

We can conclude that, even when $\nu(n)$ is white noise, the stationary points are not well defined, i.e., **the estimates are biased**.

Assuming $\nu(n) = 0$, and using the theorem and notation introduced in chapter 4, we can rewrite the previous equation as follows

$$\boldsymbol{\mathcal{S}}(B, A)\boldsymbol{\mathcal{P}}(A, A, m_1, m_2)\, \boldsymbol{h} = 0$$

where

- $\boldsymbol{\mathcal{S}}(B, A)$ is a non singular Sylvester matrix of rank $m_1 = n_a + n_b$,
- $\boldsymbol{h}$ is a vector of dimension $m_2 = max(N + n_b, N + n_a)$, with components defined by the coefficients of

$$\overline{A}(q)B(q) - A(q)\overline{B}(q)$$

For the strictly sufficient-order case, $m_2 = m_1$, $\boldsymbol{\mathcal{P}}(A, A, m_1, m_1)$ is positive definite, and the unique solution of (6.3.1) is

$$\overline{A}(q)B(q) - A(q)\overline{B}(q) = 0$$

then, in this case, the system can be identified and the solution of the method is unique.

The ODE associated, for the stochastic gradient version algorithm, is given by

$$\frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} = E[\boldsymbol{\varphi}(n)e_e(n)]$$

and for the Gauss-Newton version

$$\begin{aligned} \frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} &= \boldsymbol{\varrho}^{-1}(t)\mathbf{R}\left(\boldsymbol{\theta_0} - \boldsymbol{\vartheta}(t)\right) \\ \frac{\partial \boldsymbol{\varrho}(t)}{\partial t} &= \mathbf{R} - \boldsymbol{\varrho}(t) \end{aligned} \tag{71}$$

where

$$\mathbf{R} = E[\,\boldsymbol{\varphi}(n)\boldsymbol{\varphi}^T(n)]$$

Related to equations (71), a suitable Liapunov function is the following

$$V((\boldsymbol{\vartheta}(t) - \boldsymbol{\theta_0}), \boldsymbol{\varrho}(t)) = (\boldsymbol{\vartheta}(t) - \boldsymbol{\theta_0})^T \boldsymbol{\varrho}(t)(\boldsymbol{\vartheta}(t) - \boldsymbol{\theta_0})$$

such that

$$(dV/dt) = -(\boldsymbol{\vartheta}(t) - \boldsymbol{\theta_0})^T(\boldsymbol{R} + \boldsymbol{\varrho}(t))(\boldsymbol{\vartheta}(t) - \boldsymbol{\theta_0}) \leq 0$$

then $V((\boldsymbol{\vartheta}(t) - \boldsymbol{\theta_0}), \boldsymbol{\varrho}(t))$ is a Liapunov function for (71) that shows that $\boldsymbol{\vartheta}(t) = \boldsymbol{\theta_0}$ is a unique stable solution of (71).

Assuming no measurement noise and a unit norm constraint to define the coefficient vector, such that

$$
\begin{aligned}
E\{e_e^2(n)\} &= [-\boldsymbol{a}^T \ -\boldsymbol{b}^T] E\{\boldsymbol{\varphi}(n)\boldsymbol{\varphi}^T(n)\} \begin{bmatrix} -\boldsymbol{a} \\ -\boldsymbol{b} \end{bmatrix} \\
&= [-\boldsymbol{a}^T \ -\boldsymbol{b}^T] \begin{bmatrix} \boldsymbol{R}_d & \boldsymbol{R}_{xd}^T \\ \boldsymbol{R}_{xd}^T & \boldsymbol{R}_x \end{bmatrix} \begin{bmatrix} -\boldsymbol{a} \\ -\boldsymbol{b} \end{bmatrix}
\end{aligned}
$$

A factorization of the covariance matrix will be useful

$$
\begin{bmatrix} \boldsymbol{R}_d & \boldsymbol{R}_{xd}^T \\ \boldsymbol{R}_{xd}^T & \boldsymbol{R}_x \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{N+1} & \boldsymbol{R}_x^{-1}\boldsymbol{R}_{xd} \\ & \boldsymbol{I}_{N+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_d - \boldsymbol{R}_{xd}^T\boldsymbol{R}_x^{-1}\boldsymbol{R}_{xd} & \\ & \boldsymbol{R}_x \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{N+1} & \\ \boldsymbol{R}_{xd}^T\boldsymbol{R}_x^{-1} & \boldsymbol{I}_{N+1} \end{bmatrix}
$$

then, defining $\boldsymbol{R}_{d/x} = \boldsymbol{R}_d - \boldsymbol{R}_{xd}^T\boldsymbol{R}_x^{-1}\boldsymbol{R}_{xd}$, and pre and post-multiplying by the parameter vector

$$
E\{e_e^2(n)\} = \boldsymbol{a}^T\boldsymbol{R}_{d/x}\boldsymbol{a} + [\boldsymbol{b} - \boldsymbol{R}_x^{-1}\boldsymbol{R}_{xd}\boldsymbol{a}]^T\boldsymbol{R}_x[\boldsymbol{b} - \boldsymbol{R}_x^{-1}\boldsymbol{R}_{xd}\boldsymbol{a}]
$$

in particular, minimizing with respect to $\boldsymbol{b}$

$$
E\{e_e^2(n)\} = \boldsymbol{a}^T\boldsymbol{R}_{d/x}\boldsymbol{a} \geq 0, \ \forall \boldsymbol{a} \neq 0
$$

Theorem: If $x(n)$ is persistently exciting of sufficient order, then $\boldsymbol{R}_{d/x}$ has rank $M \leq N$, if and only if $deg\ H(z) = N$.

Now, considering $x(n)$ white noise, it is possible to relate the previous results with the decomposition theorem of chapter 4. To do this, consider

$$E\{d(n)d(n-k)\} \;=\; \sum_{l=0}^{\infty} h_l h_{k+l}$$

then

$$\boldsymbol{R}_d \;=\; \begin{bmatrix} h_0 & h_1 & \cdots & h_N & h_{N+1} & \cdots \\ 0 & h_0 & \ddots & \vdots & h_N & \cdots \\ \vdots & \cdots & \cdots & h_1 & \vdots & \cdots \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots \end{bmatrix}$$

$$\boldsymbol{R}_{xd}^T \;=\; \begin{bmatrix} h_0 & h_1 & \cdots & h_N \\ 0 & h_0 & \ddots & \vdots \\ \vdots & \cdots & \cdots & h_1 \\ 0 & \cdots & 0 & h_0 \end{bmatrix}$$

and $\boldsymbol{R}_x = \boldsymbol{I}$. Thus, if $\boldsymbol{J}$ is an $(N+1) \times (N+1)$ exchange matrix with ones in the antidiagonal,

$$\boldsymbol{J}\boldsymbol{R}_{d/x}\boldsymbol{J} \;=\; \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ h_{N+1} & h_{N+2} & h_{N+3} & \cdots \end{bmatrix} [.]^T$$

Finally

$$E\{e_e^2(n)\} \;=\; \boldsymbol{a}^T\boldsymbol{R}_{d/x}\boldsymbol{a} = [\boldsymbol{a}^T\boldsymbol{J} \;\; \boldsymbol{0}^T]\boldsymbol{\Gamma}_H^2 \begin{bmatrix} \boldsymbol{J}\boldsymbol{a} \\ \boldsymbol{0} \end{bmatrix}$$

This vanishes if only if

$$\boldsymbol{\Gamma}_H \begin{bmatrix} \boldsymbol{J}\boldsymbol{a} \\ \boldsymbol{0} \end{bmatrix} \;=\; \boldsymbol{0}$$

as advanced when discussed rational approximation theory and Hankel forms.

6.3.2 The insufficient order case

Following with the analysis without measurement noise,

- The monic constraint over the coefficient vector is $a_0 = 1$, then the optimal choice of $\boldsymbol{a}$ is

$$\boldsymbol{R}_{d/x}\boldsymbol{a} = \begin{bmatrix} \sigma_e^2 \\ \mathbf{0}_N \end{bmatrix}$$

  where σ_e^2 is the equation error variance under the monic constraint.

- The unit norm constraint over the coefficient vector is $\boldsymbol{a}^T\boldsymbol{a} = 1$. Then the optimal solution to

$$E\{e_e^2(n)\} = \boldsymbol{a}^T\boldsymbol{R}_{d/x}\boldsymbol{a}$$

  in this case is given by

$$\boldsymbol{R}_{d/x}\boldsymbol{a} = \lambda_{min}(\boldsymbol{R}_{d/x})\boldsymbol{a}$$

  and the minimized equation error variance becomes

$$\boldsymbol{a}^T\boldsymbol{R}_{d/x}\boldsymbol{a} = \lambda_{min}(\boldsymbol{R}_{d/x})$$

Some related results

- If $x(n)$ is an AR process of order not exceding N, then the estimate obtained for both, unit and monic constrained methods gives a minimum phase (stable) polynomial.

- Let $\hat{H}(z)$ be the N order transfer function obtained by minimizing the equation error variance with unit norm constraint and $x(n)$ is white noise, then

$$\begin{aligned}
\hat{h}_k &= h_k \qquad k = 0, 1, ..., N \\
\sum_{l=0}^{\infty} \hat{h}_l \hat{h}_{k+l} &= \sum_{l=0}^{\infty} h_l h_{k+l} \qquad k = 1, ..., N \\
\sum_{l=0}^{\infty} h_l^2 - \sum_{l=0}^{\infty} \hat{h}_l^2 &= \lambda_{min}(\boldsymbol{R}_{d/x}) \\
\lambda_{min}(\boldsymbol{R}_{d/x}) &\leq E\{x^2(n)\}\sigma_{N+1}^2(\Gamma_H)
\end{aligned}$$

  where $\sigma_{N+1}(\Gamma_H)$ is the $N+1$ singular value.

- Let $\hat{H}(z)$ be the N order transfer function obtained by minimizing the equation error variance with monic constraint and $x(n)$ is white noise, then

$$\begin{aligned}
\hat{h}_k &= h_k \qquad k = 0, 1, ..., N \\
\sum_{l=0}^{\infty} \hat{h}_l \hat{h}_{k+l} - \sum_{l=0}^{\infty} h_l h_{k+l} &= \frac{\sigma_e^2}{2\pi} \int_{-\pi}^{\pi} \frac{e^{jkw}}{|A(e^{jw})|^2} dw \qquad k = 0, 1, ..., N
\end{aligned}$$

6.3.3 Example

The performance of the EE method for a system identification application is shown in the following example.

Example 1: The plant

$$d(n) = \frac{q^{-1}}{1 \;-\; 1.236\, q^{-1} \;+\; 0.382\, q^{-2}} x(n) + \nu(n)$$

where $x(n)$ is an stochastic process generated by

$$x(n) = u(n) \;-\; 0.764\, u(n-2) \;+\; 0.146\, u(n-4)$$

where $u(n)$ and $\nu(n)$ are uncorrelated white noise of zero mean.

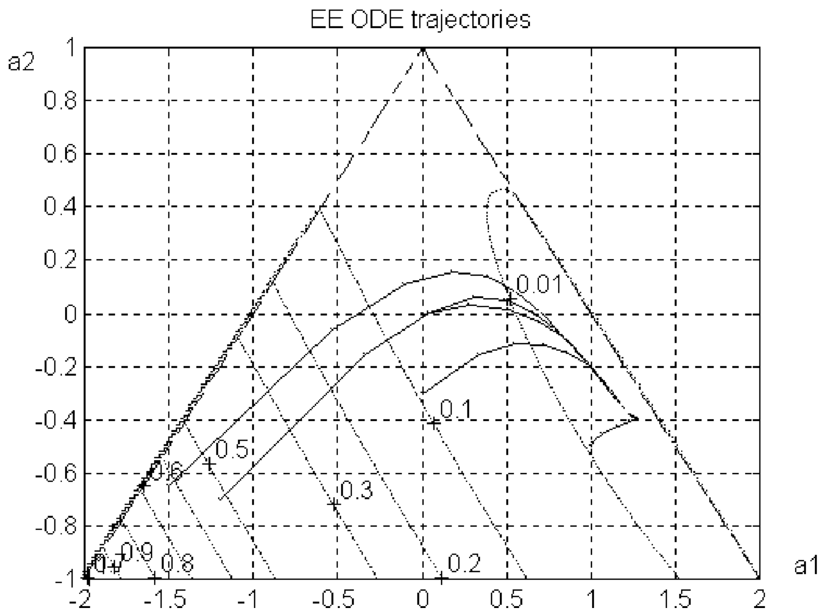


Figure 40: ODE trayectories for example 1, without noise

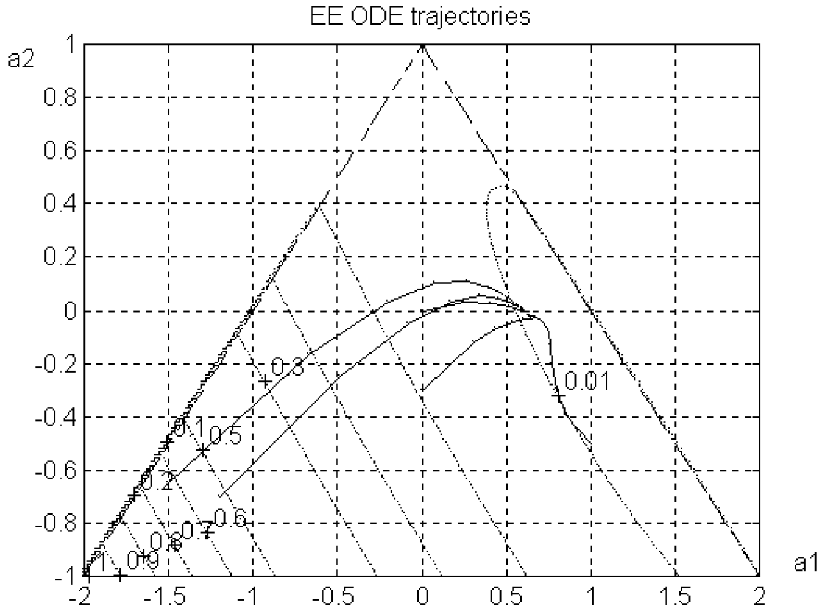


Figure 41: ODE trajectories for example 1, variance of $\nu(n)$ 1.0

6.4 Instrumental Variable Methods (IV)

- The IV method was idealized in order to avoid the biased estimates given by the EE method.
- The regressor of *instrumental variables*, $\boldsymbol{\zeta}(n)$, is chosen to be uncorrelated with $\nu(n)$ but not independent of $x(n)$.

$$E\{\boldsymbol{\zeta}(n)e_e(n)\} = E\{\boldsymbol{\zeta}(n)(d(n) - \boldsymbol{\varphi}^T(n)\boldsymbol{\theta}(n))\} = 0$$

where $e_e(n)$ is the equation error, $\boldsymbol{\zeta}(n)$ is the IV regressor and $\boldsymbol{\varphi}(n)$ is the common regressor defined for the EE method.

The stochastic gradient version

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{\zeta}(n)e_e(n)$$

or the Gauss-Newton version

$$\begin{aligned}
\boldsymbol{\theta}(n+1) &= \boldsymbol{\theta}(n) + \mu\boldsymbol{P}(n+1)\boldsymbol{\zeta}(n)e_e(n) \\
\boldsymbol{P}(n+1) &= \left(\frac{1}{1-\mu}\right)\left(\boldsymbol{P}(n) - \frac{\boldsymbol{P}(n)\boldsymbol{\varphi}(n)\boldsymbol{\zeta}^T(n)\boldsymbol{P}(n)}{\frac{1-\mu}{\mu} + \boldsymbol{\varphi}^T(n)\boldsymbol{P}(n)\boldsymbol{\zeta}(n)}\right)
\end{aligned}$$

Defining $\hat{y}(n) = \frac{\hat{B}_n(q)}{\hat{A}_n(q)}x(n)$, we assume the following selection of instrumental variables $\boldsymbol{\zeta}(n)$

$$\boldsymbol{\zeta}(n) = [\hat{y}(n-1), ..., \hat{y}(n-N), x(n), ..., x(n-M)]^T$$

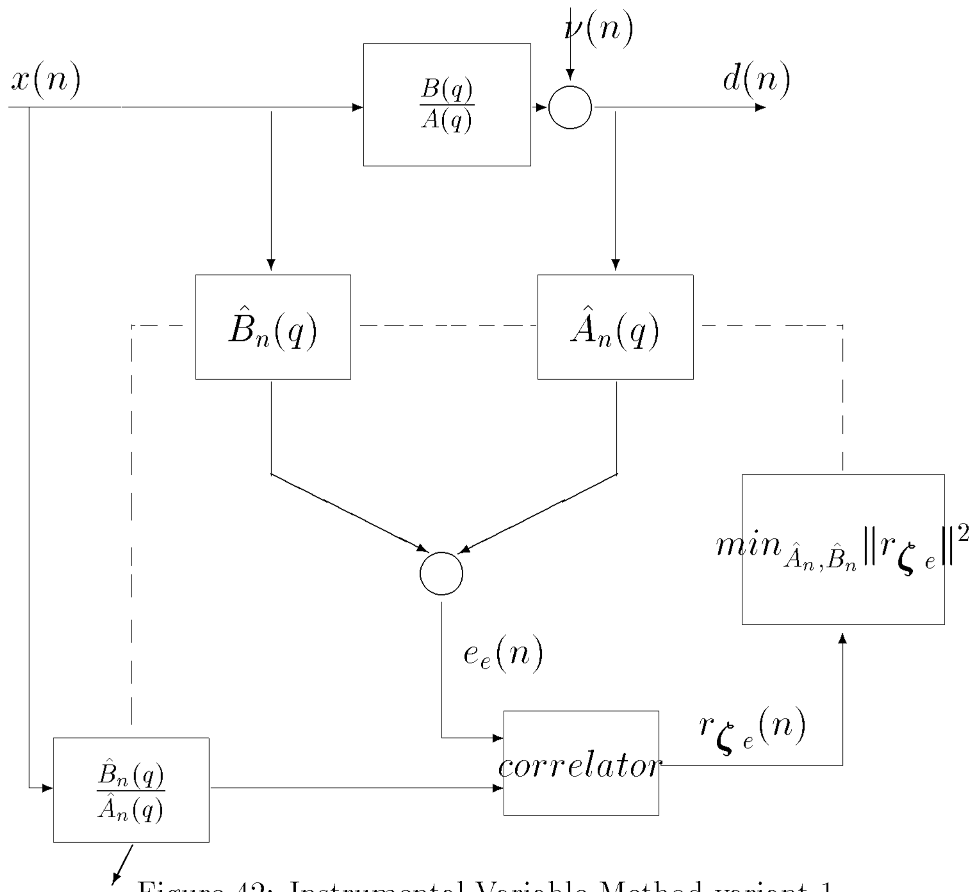


Figure 42: Instrumental Variable Method variant 1.

6.4.1 ODE associated

For this variant we can only give local convergence conditions (indefined Liapunov function).

The ODE associated is given by

$$\begin{aligned}\frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} &= \boldsymbol{\varrho}^{-1}(t)\boldsymbol{G}(t)\left(\boldsymbol{\theta}_0 - \boldsymbol{\vartheta}(t)\right) \\ \frac{\partial \boldsymbol{\varrho}(t)}{\partial t} &= \boldsymbol{G}(t) - \boldsymbol{\varrho}(t)\end{aligned}$$

where $\boldsymbol{G}(t) = E\{\boldsymbol{\zeta}(n)\boldsymbol{\varphi}^T(n)\}$.

For the gradient version

$$\frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} = E\{\boldsymbol{\zeta}(n)e_e(n)\}$$

The stationary points of the IV method can be analyzed using

$$\boldsymbol{\mathcal{S}}(\overline{B}, \overline{A})\boldsymbol{\mathcal{P}}(\overline{A}, A, m_1, m_2)\boldsymbol{h} = 0$$

where **h** is a vector of dimension $m_2 = max(N+n_b, M+n_a)$, with components given by the coefficients of $\overline{A}(q)B(q) - \overline{B}(q)A(q)$,

If $\overline{A}(q)$ and $\overline{B}(q)$ are coprime, $\boldsymbol{\mathcal{S}}(\overline{B}, \overline{A})$ is non singular. For sufficient-order $m_1 = m_2$, $\boldsymbol{\mathcal{P}}(\overline{A}, A, m_1, m_1)$ is non singular.

Then, the unique solution of (6.4) is given by

$$\overline{A}(q)B(q) - \overline{B}(q)A(q) = 0$$

as a consequence, $h_i = 0$, para $i = 0, ..., m_1$, i.e., $\boldsymbol{\theta}(n) = \boldsymbol{\theta}_0$ is the unique possible solution.

Note that, independently of the stationary points, $\boldsymbol{\mathcal{P}}(\overline{A}, A, m_1, m_1)$ can be singular in certain points in the parameter space .This is called *generic consistency* of the method.

Since this points can exist, it is possible to find stationary points close to them where the behavior of the method is not suitable.

6.4.2 Example

Example 2 Consider the plant

$$d(n) = \frac{q^{-1}}{1 - \; 2r\, q^{-1} \; + \; r^2\, q^{-2}} x(n) + \nu(n)$$

where r is a constant to be determined, and $x(n)$ is given by

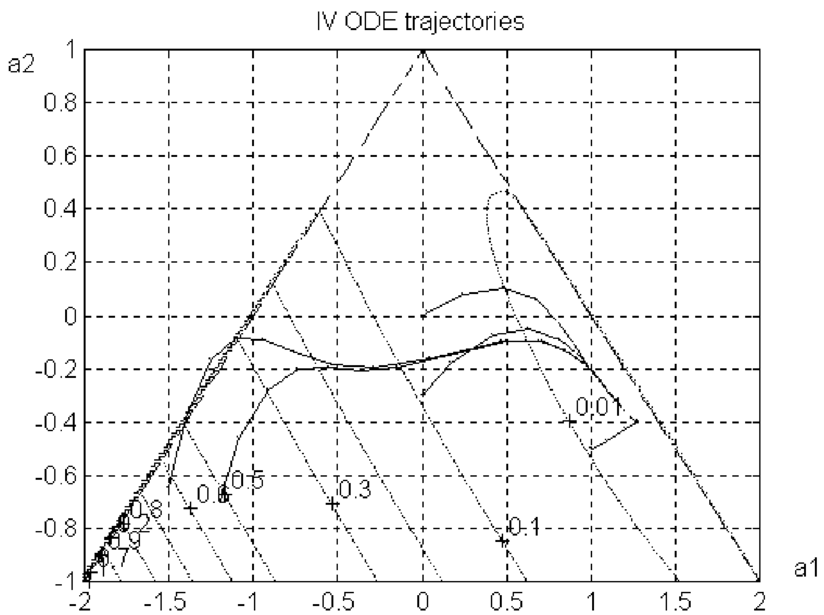
$$x(n) = u(n) \; - \; 2r^2\, u(n-2) \; + \; r^4\, u(n-4)$$

where $u(n)$ and $\nu(n)$ are white noise, zero mean, unit variance and uncorrelated. Note that the input do not satisfy the theorem of chapter 4.

The IV regressor

$$\boldsymbol{\zeta}(n) = [\hat{y}(n-1) \;\; \hat{y}(n-2) \;\; x(n-1)]^T$$

$\boldsymbol{\mathcal{P}}(\overline{A}, A, m_1, m_1)$ is singular for this particular example if for the polynomial $\overline{A}(q) = (1 + 2r\, q^{-1} + r^2 q^{-2})$, the constant r is equal to 0.618.



6.5 The Bias Remedy LMSEE (BRLE)

Objective: Characterization of the bias related to the LMSEE in order to introduce an algorithm for bias reduction.
Some definitions:

- The system identification problem is of sufficient order.
- The model to identify is given by $y(n) = \boldsymbol{\varphi}_0^T(n)\boldsymbol{\theta_0}$ where:
  - $\boldsymbol{\theta_0} = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}]^T$
  - and $\boldsymbol{\varphi}_0(n) = [y(n-1), \ldots, y(n-n_a), x(n), \ldots, x(n-n_b)]^T$.

  i.e., we use the monic constraint.
- $e_0(n) = d(n) - \hat{y}(n)$ is the output error, where:
  - $d(n) = y(n) + \nu(n)$
  - $\hat{y}(n) = \hat{\boldsymbol{\varphi}}(n)^T\boldsymbol{\theta}(n)$.
- $\boldsymbol{\varphi}(n) = \boldsymbol{\varphi}_0(n) + \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}$ where, $\boldsymbol{\nu}(n) = [\nu(n-1), \ldots, \nu(n-N)]^T$.
- $e_0(n) - e(n) = \boldsymbol{\theta}(n)^T \begin{bmatrix} \boldsymbol{e_0}(n) \\ 0 \end{bmatrix}$, where, $\boldsymbol{e_0}(n) = [e_0(n-1), \ldots, e_0(n-N)]^T$.

In order to analyze the bias of the LMSEE, consider

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{\varphi}(n)[d(n) - \boldsymbol{\varphi}(n)^T\boldsymbol{\theta}(n)]$$

that can be rewritten as

$$\begin{aligned} \boldsymbol{\theta}(n+1) &= \boldsymbol{\theta}(n) + \mu\boldsymbol{\varphi}(n)[-\boldsymbol{\varphi}(n)^T\boldsymbol{\theta}(n) + \nu(n) + \boldsymbol{\varphi}_0^T(n)\boldsymbol{\theta_0}] \\ &= \boldsymbol{\theta}(n) + \mu[-\boldsymbol{\varphi}(n)\boldsymbol{\varphi}(n)^T\boldsymbol{\theta}(n) + \boldsymbol{\varphi}(n)\nu(n) + \boldsymbol{\varphi}(n)\boldsymbol{\varphi}_0^T(n)\boldsymbol{\theta_0}] \end{aligned} \tag{72}$$

Assuming convergence in the mean, i.e.,

$$\lim_{n\to\infty} E\{\boldsymbol{\theta}(n+1)\} = \lim_{n\to\infty} E\{\boldsymbol{\theta}(n)\} \tag{73}$$

then equation (72) becomes

$$\begin{aligned} E\{\boldsymbol{\theta}(n+1)\} &= E\{\boldsymbol{\theta}(n)\} - \mu[E\{\boldsymbol{\varphi}(n)\boldsymbol{\varphi}(n)^T\boldsymbol{\theta}(n)\} \\ &- E\{\boldsymbol{\varphi}(n)\nu(n)\} - E\{\boldsymbol{\varphi}(n)\boldsymbol{\varphi}_0^T(n)\boldsymbol{\theta_0}\}] \\ &= E\{\boldsymbol{\theta}(n)\} - \mu[B1 - B2 - B3] \end{aligned} \tag{74}$$

Since was assumed that $\boldsymbol{\varphi}(n)$ is uncorrelated with $\boldsymbol{\theta}(n)$, we obtain

$$\begin{aligned} B1 &= E\left\{[\boldsymbol{\varphi}_0(n) + \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}][\boldsymbol{\varphi}_0(n) + \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}]^T\right\} E\{\boldsymbol{\theta}(n)\} \\ &= (\boldsymbol{\Omega} + \boldsymbol{\Sigma})E\{\boldsymbol{\theta}(n)\} \\ B2 &= E\left\{[\boldsymbol{\varphi}_0(n) + \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}]\nu(n)\right\} = E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}\nu(n)\right\} \\ B3 &= E\left\{[\boldsymbol{\varphi}_0(n) + \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}]\boldsymbol{\varphi}_0^T(n)\right\}\boldsymbol{\theta_0} = \boldsymbol{\Omega}\boldsymbol{\theta_0} \end{aligned}$$

where: $\boldsymbol{\Omega} = E\{\boldsymbol{\varphi}_0(n)\boldsymbol{\varphi}_0^T(n)\}$, $\boldsymbol{\Sigma} = E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}[\boldsymbol{\nu}^T(n)0]\right\}$.

In this way, equation (74) can be rewritten as follows

$$E\{\boldsymbol{\theta}(n+1)\} = E\{\boldsymbol{\theta}(n)\} - \mu[(\boldsymbol{\Omega}+\boldsymbol{\Sigma})E\{\boldsymbol{\theta}(n)\} - E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}\nu(n)\right\} - \boldsymbol{\Omega}\boldsymbol{\theta_0}] \tag{75}$$

where, by (73)

$$\lim_{n\to\infty} E\left\{\boldsymbol{\theta}(n)\right\} = (\boldsymbol{\Omega}+\boldsymbol{\Sigma})^{-1}(E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}\nu(n)) + \boldsymbol{\Omega}\boldsymbol{\theta_0}\right\}$$

In order to eliminate the bias it is necessary that $\boldsymbol{\Sigma} = E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}[\boldsymbol{\nu}(n)^T 0]\right\} =$ and $E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}\nu(n)\right\} = 0$. In this case

$$\lim_{n\to\infty} E\{\boldsymbol{\theta}(n)\} = \boldsymbol{\theta_0}$$

In a general case it is necessary to eliminate in equation (75) the factor C_0 defined by

$$\begin{aligned} C_0 &= -\boldsymbol{\Sigma} E\left\{\boldsymbol{\theta}(n)\right\} + E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix} \nu(n)\right\} \\ &= -E\left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix} \left\{\begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}^T \boldsymbol{\theta}(n) - \nu(n)\right\}\right\} \end{aligned}$$

to achieve the objective of bias elimination.

Assuming convergence of the LMSEE, we must have $\nu(n) \approx e_0(n)$, then a possible choice of the bias compensation factor is

$$\begin{aligned} C &= \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix} \left\{\nu(n) - \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}^T \boldsymbol{\theta}(n)\right\} \\ &\cong \begin{bmatrix} \mathbf{e_0}(n) \\ 0 \end{bmatrix} \left\{e_0(n) - \begin{bmatrix} \mathbf{e_0}(n) \\ 0 \end{bmatrix}^T \boldsymbol{\theta}(n)\right\} \\ &= \begin{bmatrix} \mathbf{e_0}(n) \\ 0 \end{bmatrix} e(n) \end{aligned}$$

where, by making $\mu_c = \mu\tau$, we obtain

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu e(n)(\boldsymbol{\varphi} - \tau \begin{bmatrix} \mathbf{e_0}(n) \\ 0 \end{bmatrix}) \tag{76}$$

that define the **Bias Remedy Least Mean Squares Equation Error** (BRLE) method.

6.5.1 Analysis of convergence in the mean

Objective: to which point the stability of the LMSEE method can be maintained while the bias reduction of BRLE is achieved.

The analysis is based in the coefficient error defined by

$$E\{\tilde{\boldsymbol{\theta}}(n+1)\} = E\{\boldsymbol{\theta}(n+1) - \boldsymbol{\theta}_0\} \tag{77}$$

Using (76), we can rewrite this equation as

$$E\{\tilde{\boldsymbol{\theta}}(n+1)\} \;=\; E\{\boldsymbol{\theta}(n) + \mu\boldsymbol{\varphi}_c(n)(d(n) - \boldsymbol{\varphi}^T(n)\boldsymbol{\theta}(n))\} - \boldsymbol{\theta}_0 \tag{78}$$

and considering that

$$\begin{aligned} d(n) - \boldsymbol{\varphi}^T(n)\boldsymbol{\theta}(n) &= \boldsymbol{\varphi}_0^T(n)\boldsymbol{\theta}_0 + \nu(n) - \boldsymbol{\varphi}^T(n)\boldsymbol{\theta}(n) \\ &= \boldsymbol{\varphi}^T(n)(\boldsymbol{\theta}_0 - \boldsymbol{\theta}(n)) + \nu(n) - \begin{bmatrix} \boldsymbol{\nu}(n) \\ 0 \end{bmatrix}^T \boldsymbol{\theta}(n) \end{aligned} \tag{79}$$

and defining

$$\boldsymbol{R} = (\boldsymbol{I} - \mu E\{\boldsymbol{\varphi}(n)\boldsymbol{\varphi}^T(n)\}) \tag{80}$$

equation (78) can be rewritten as

$$\begin{aligned} E\{\tilde{\boldsymbol{\theta}}(n+1)\} &= \left(\boldsymbol{R} - \mu E\{\tau(n)\begin{bmatrix} \boldsymbol{e}_0(n) \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{\varphi}^T(n)\}\right) E\{\tilde{\boldsymbol{\theta}}(n)\} \\ &\quad + \mu E\{(1-\tau(n))\begin{bmatrix} \boldsymbol{\nu}(n) \\ \boldsymbol{0} \end{bmatrix}(\nu(n) - \begin{bmatrix} \boldsymbol{\nu}(n) \\ \boldsymbol{0} \end{bmatrix}^T \boldsymbol{\theta}_0)\} \\ &= (\boldsymbol{R} - \boldsymbol{A}_1(n))E\{\tilde{\boldsymbol{\theta}}(n)\} + \boldsymbol{B}_1(n) \end{aligned} \tag{81}$$

where

$$\begin{aligned} \boldsymbol{A}_1(n) &= \mu E\{\tau(n)\begin{bmatrix} \boldsymbol{e}_0(n) \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{\varphi}^T(n)\} \\ \boldsymbol{B}_1(n) &= \mu E\{(1-\tau(n))\begin{bmatrix} \boldsymbol{\nu}(n) \\ \boldsymbol{0} \end{bmatrix}(\nu(n) - \begin{bmatrix} \boldsymbol{\nu}(n) \\ \boldsymbol{0} \end{bmatrix}^T \boldsymbol{\theta}_0)\} \end{aligned} \tag{82}$$

The asymptotic convergence of (81) can be demonstrated using the quasi-invariant system theorem introduced in chapter 4.

Theorem: The equation $E\{\tilde{\boldsymbol{\theta}}(n+1)\}$ is asymptotically stable if the following conditions are satisfied

1. $0 \leq \tau(n) \leq min(1, \frac{\epsilon}{\|\boldsymbol{e}_0(n)\|})$, for a constant $\epsilon > 0$;
2. $0 < \mu < min(\mu_1, \frac{2}{\lambda_N})$, where λ_N is the maximum eigenvalue of $E\{\boldsymbol{\varphi}(n)\boldsymbol{\varphi}^T(n)\}$, and μ_1 is a positive constant.

Outline of the Proof: It is necessary to show

1. That the homogeneous part of (81) is asymptotically stable.

   Since the eigenvalues of $\boldsymbol{R}$ are inside the unit circle, it is possible to define a transition matrix $\boldsymbol{\Phi}(n)$ that satisfy

$$\|\boldsymbol{\Phi}(n)\| < c\beta^n \tag{83}$$

   where $c > 0$ e $0 < \beta < 1$. Assuming that $0 < E\{\|\boldsymbol{\varphi}(n)\|\} < p$, is easy to show that

$$\|\boldsymbol{A}_1(n)\| \;\leq\; \epsilon\mu p = \delta \tag{84}$$

   then by chosen $0 < \mu < \mu_1 = \frac{1-\beta}{c\,\epsilon p}$, we have $0 < \beta + c\delta < 1$. Then, following the quasi-invariant system theorem, the homogeneous part defined by

$$\boldsymbol{u}_1(n+1) = (\boldsymbol{R} + \boldsymbol{A}_1(n))\boldsymbol{u}_1(n) \tag{85}$$

   is asymptotically stable.

2. That the disturbance part of (81) is bounded.

   For $0 \leq \tau(n) \leq 1$, this can be shown using (82), since $\|\boldsymbol{B}_1(n)\| \leq \mu r + \mu r\|\boldsymbol{a}\|$, where $\boldsymbol{a}^T = [a_1 ... a_N]^T$ and r is an upper bound of the noise variance.

Finally, note that asymptotic convergence is achieved if, considering the bounded disturbance term $\boldsymbol{B}_1(n)$, the bias reduction parameter $\tau(n)$ satisfies, $\tau(n) \to 1$, i.e., $\boldsymbol{B}_1(n) \to 0$, or

$$E\{\tilde{\boldsymbol{\theta}}(n+1)\} \to 0 \qquad for\; n \to \infty$$

6.5.2 Example

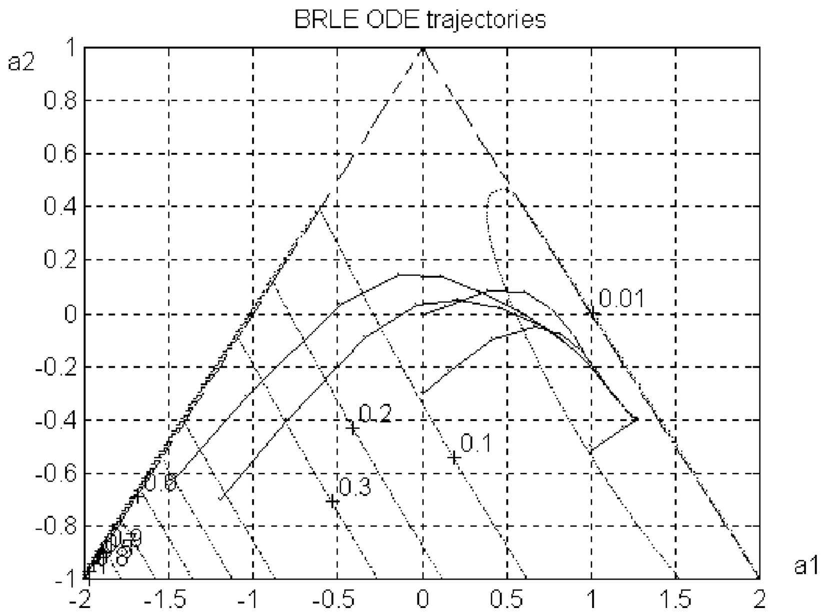


Figure 44: ODE trajectories for example 1, variance of $\nu(n)$ 1.0

7 Hyperstable adaptive filter

HARF, an stable but incomplete solution

- Hyperstability theorem application.
- Stationary points and ODE associated.
- SHARF algorithm.
- Discussion of the insufficient order case.

7.1 Introduction

- In this chapter is shown the application of stability theory concepts related to a non linear feedback system previously introduced in chapter 4.

- The main aspect of the algorithms is its *theoretical* asymptotic stability, that in real world applications is severely constrained by a **positive real condition**.

- This stability property, or *hyperstability*, has been useful in many control applications, where bounded variables are more important than convergence speed or MSE performance of a parameter updating algorithm.

- An important concern with the practical utilization of this family of algorithms is in the undermodelled case, where convergence properties are not well defined.

7.2 Review of the LMS with a posteriori error

- Consider the use of **a posteriori prediction error** in a FIR filter,

$$\overline{y}(n) = \overline{\boldsymbol{\phi}}^T(n)\boldsymbol{\theta}(n+1)$$

where $\hat{y}(n) = \boldsymbol{\phi}^T(n)\boldsymbol{\theta}(n)$ and $\boldsymbol{\phi}(n) = [\hat{y}(n-1) \ldots \hat{y}(n-N)\; x(n) \ldots x(n-N)]^T$.

- In the FIR case, the use of the a posteriori error can lead to faster convergence and improved estimate variance.

- Then, we can write

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{x}(n)e(n)$$

where $\boldsymbol{\theta}(n) = [\hat{\theta}_0(n) \ldots \hat{\theta}_N(n)]^T$, $\boldsymbol{x}(n) = [x(n) \ldots x(n-N)]^T$, and $e(n) = d(n) - \boldsymbol{x}^T\boldsymbol{\theta}(n)$.

- And the a posteriori prediction error algorithm can be obtained as follows

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\frac{\partial}{\partial\boldsymbol{\theta}(n+1)}\left\{\frac{1}{2}\overline{e}^2(n)\right\} \tag{86}$$

with $\overline{e}(n) = d(n) - \sum_{j=0}^{N} b_j(n+1)x(n-j) = d(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n+1)$, then

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{x}(n)[d(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n+1)]$$

- But

$$\begin{aligned}
\overline{e}(n) &= d(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n+1) \\
&= d(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n) + \boldsymbol{x}^T(n)\boldsymbol{\theta}(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n+1) \\
&= e(n) - \boldsymbol{x}^T(n)[\boldsymbol{\theta}(n+1) - \boldsymbol{\theta}(n)] \\
&= e(n) - \boldsymbol{x}^T(n)\mu\boldsymbol{x}(n)\overline{e}(n) \\
&= e(n)/[1 + \boldsymbol{x}^T(n)\mu\boldsymbol{x}(n)]
\end{aligned}$$

such that

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{x}(n)\frac{[d(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n+1)]}{[1 + \boldsymbol{x}^T(n)\mu\boldsymbol{x}(n)]}$$

similarly to the normalized LMS algorithm.

- For the IIR adaptive filter case, we can obtain, through a similar procedure

$$\overline{e}(n) = \frac{d(n) - \overline{\boldsymbol{\phi}}^T(n)\boldsymbol{\theta}(n)}{1 + \overline{\boldsymbol{\phi}}^T(n)\mu\overline{\boldsymbol{\psi}}(n)}$$

- Since $\overline{\boldsymbol{\phi}}^T(n)\mu\overline{\boldsymbol{\psi}}(n)$ do not have, in general, a positive definite form, we introduce the *slow convergence approximation* in order to obtain

$$1 + \overline{\boldsymbol{\phi}}^T(n)\mu\overline{\boldsymbol{\psi}}(n) \approx 1$$

- Then

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\overline{\boldsymbol{\psi}}(n)[d(n) - \overline{\boldsymbol{\phi}}^T(n)]$$

7.3 Stable adaptive filter based in a nonlinear system

- Note that the model for the a posteriori prediction error for the FIR case can be written

$$\overline{y}(n) = \sum_{j=0}^{N} b_j(n+1)x(n-j)$$

where was assumed that

$$d(n) = \sum_{j=0}^{N} b_j^o x(n-j)$$

- Then

$$\begin{aligned} \overline{e}(n) &= d(n) - \overline{y}(n) = d(n) - \boldsymbol{x}^T(n)\boldsymbol{\theta}(n+1) \\ &= \boldsymbol{x}^T(n)\tilde{\boldsymbol{\theta}}(n+1) \end{aligned}$$

with $\tilde{\boldsymbol{\theta}}(n+1) = \boldsymbol{\theta}_o - \boldsymbol{\theta}(n+1)$.

- For the IIR adaptive filter, with

$$d(n) = \frac{B(q)}{A(q)}x(n)$$

the model takes the form

$$\begin{aligned} \overline{e}(n) &= d(n) - \overline{y}(n) \\ &= \sum_{i=1}^{N} a_i^o[d(n-i) - \overline{y}(n-i)] + \sum_{i=1}^{N}[a_i^o - a_i(n+1)]\overline{y}(n-i) \\ &\quad + \sum_{j=0}^{N}[b_j^o - b_j(n+1)]x(n-j)] \\ &= \sum_{i=1}^{N} a_i^o \overline{e}(n-i) + \overline{\boldsymbol{\phi}}^T(n)\tilde{\boldsymbol{\theta}}(n+1) \end{aligned}$$

or

$$\overline{e}(n) = \frac{1}{A(q)}[\overline{\boldsymbol{\phi}}^T(n)\tilde{\boldsymbol{\theta}}(n+1)]$$

7.3.1 Stability of the homogeneous error system

- For the FIR case

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu \boldsymbol{x}(n)\boldsymbol{x}^T(n)\tilde{\boldsymbol{\theta}}(n+1)$$

- Then

$$\tilde{\boldsymbol{\theta}}(n+1) = \tilde{\boldsymbol{\theta}}(n) - \mu \boldsymbol{x}(n)\boldsymbol{x}^T(n)\tilde{\boldsymbol{\theta}}(n+1)$$

or

$$\tilde{\boldsymbol{\theta}}(n+1) = [\boldsymbol{I} - \mu \boldsymbol{x}(n)\boldsymbol{x}^T(n)]^{-1}\tilde{\boldsymbol{\theta}}(n)$$

- This is the state equation formulation of an homogeneous system with transition matrix given by

$$[\boldsymbol{I} - \mu \boldsymbol{x}(n)\boldsymbol{x}^T(n)]^{-1}$$

- It is not hard to shown, using the previous formulation, that

$$\tilde{\boldsymbol{\theta}}^T(n+1)\mu^{-1}\tilde{\boldsymbol{\theta}}(n+1) - \tilde{\boldsymbol{\theta}}(n)\mu^{-1}\tilde{\boldsymbol{\theta}}(n) = -(2 + \boldsymbol{x}^T(n)\mu\boldsymbol{x}(n))^2\overline{e}^2(n) < 0$$

- If the condition of persistent excitation over $\boldsymbol{x}(n)$ is satisfied, this system is globally asymptotically stable.
- Note that, without considering a noise term, a similar equation can be obtained for the Equation Error method (in fact a FIR filter under this point of view!).

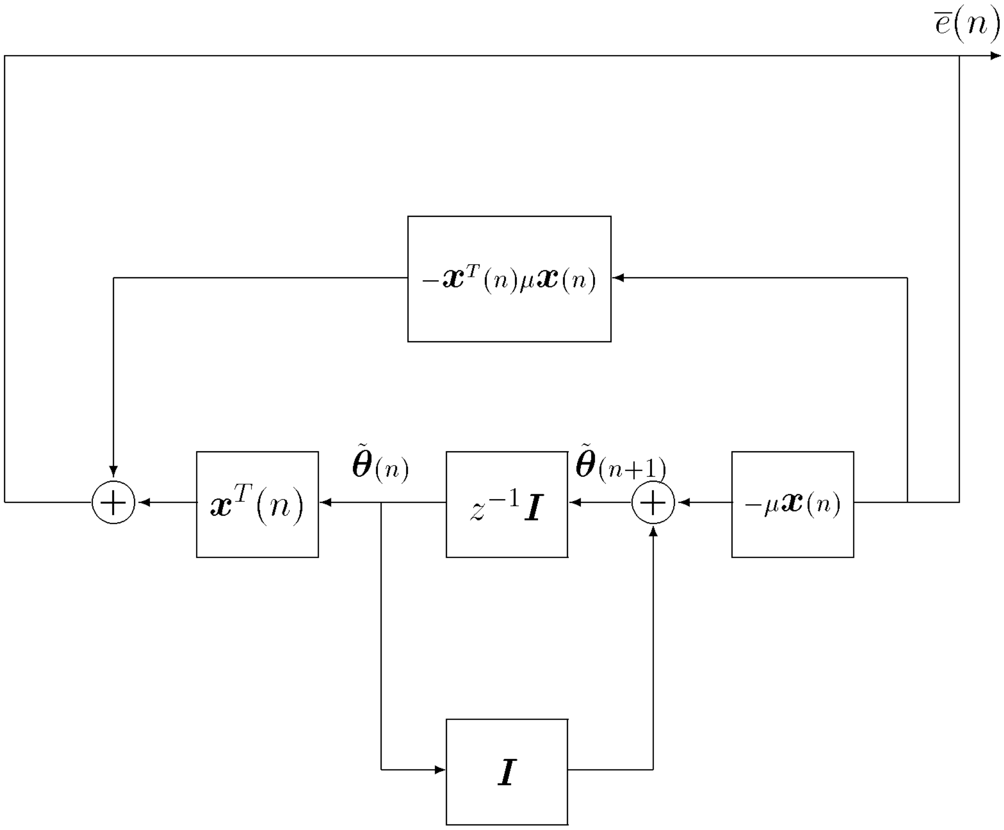


Figure 45: Homogeneous error system for a FIR adaptive algorithm

- Objective, work with the IIR adaptive filter case, but using the **output error**.
- Consider an algorithm as follows [Feintuch, 1976]

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\overline{\boldsymbol{\phi}}(n)\overline{e}(n)$$

then, the homogeneous error system associated is

$$\tilde{\boldsymbol{\theta}}(n+1) = \tilde{\boldsymbol{\theta}}(n) - \mu\overline{\boldsymbol{\phi}}(n)\left\{\frac{1}{A(q)}\overline{\boldsymbol{\phi}}^T(n)\tilde{\boldsymbol{\theta}}(n+1)\right\}$$

that represent a non linear (also time-variant) homogeneous error system.

Under which conditions this system maintain the stability of the FIR case?

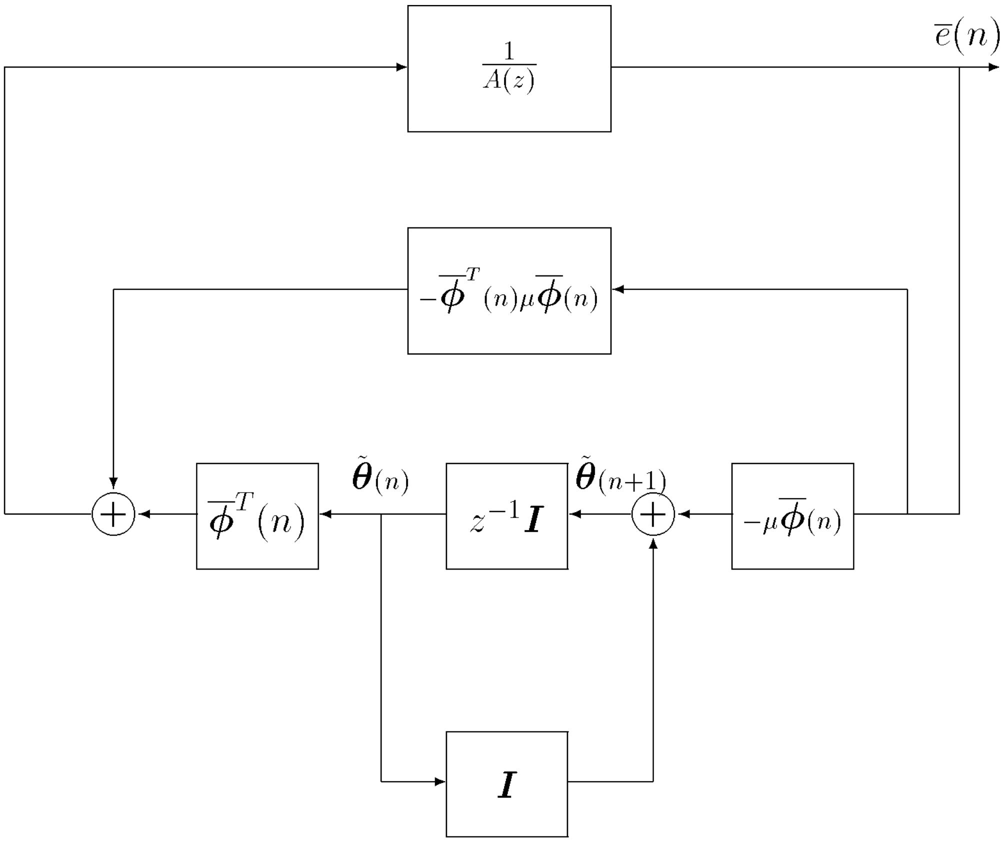


Figure 46: Homogeneous error system for a IIR adaptive algorithm

Rewritten the theorem of section 4.5.2, we can present general conditions of stability,

> Theorem: Consider the general homogeneous error system of the figure. This system is globally asymptotically stable if the following conditions are satisfied:
>
> - $H(z)$ is strictly positive real (i.e., $Re[H(z)] > 0$, for all $|z| = 1$).
> - $\frac{1}{2}\overline{\phi}^{T}(n)\mu\overline{\boldsymbol{\phi}}(n) - \lambda \geq 0$ for all n (uniform observability).

In terms of the error $\epsilon(n)$ of the figure, the global stability implies that:

$$\epsilon(n) \rightarrow 0, \quad for \;\; n \rightarrow \infty$$

and also that the states of $H(z)$ are bounded.

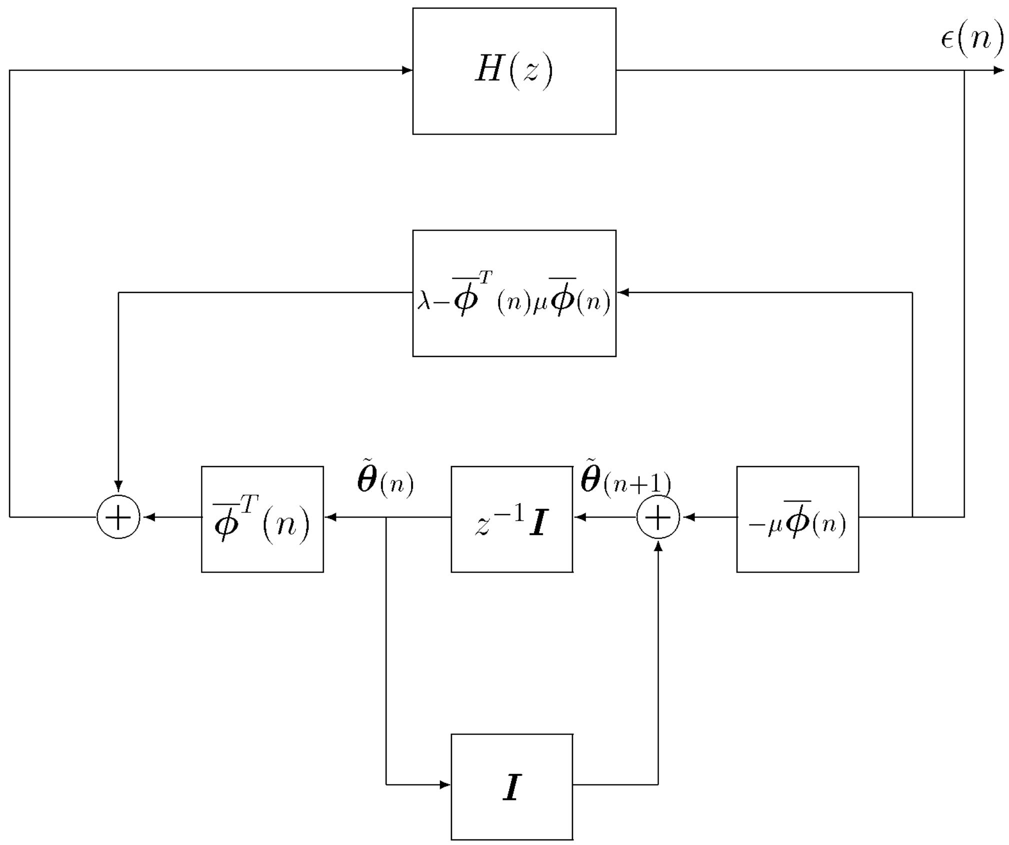


Figure 47: General homogeneous error system for an IIR adaptive algorithm

- These conditions are trivially verified for the FIR adaptive filter homogeneous error system model, since $H(q) = 1$ and $\lambda = 0$.
- For the homogeneous error system related to the IIR adaptive filter the SPR condition of the theorem introduce a condition to the systems where stability of the IIR adaptive filter can be guaranteed, i.e.

$$Re\left[\frac{1}{A(z)}\right] > 0$$

- For a second order system, for example the SPR conditions is satisfied for the plant inside the dashed region of the unit circle.

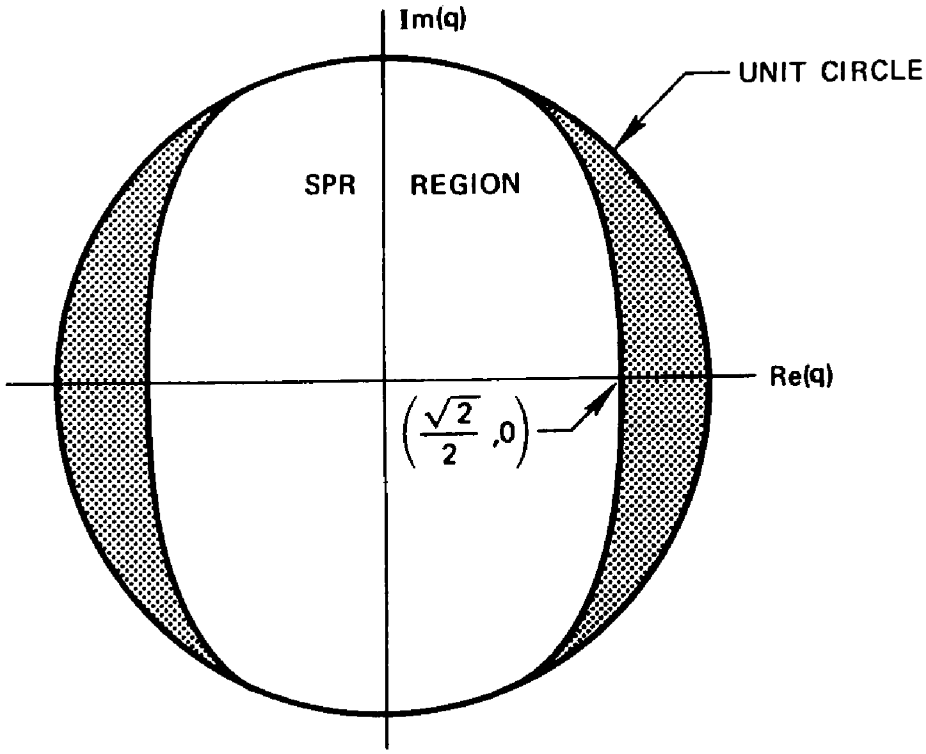


Figure 48: SPR condition of a second order system.

- A form to overcome this limitation is the introduction of a filtering or *smoothing error*, $\epsilon(n)$, defined by

$$\epsilon(n) = \overline{e}(n) + \sum_{i=1}^{N} d_i \overline{e}(n-i) = D(q)\overline{e}(n)$$

- This smoothing error used in the previous equation determines the following algorithm

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\overline{\boldsymbol{\phi}}(n)\epsilon(n)$$

- Note that

$$\begin{aligned}
\epsilon(n) &= d(n) - \overline{\boldsymbol{\phi}}^T(n)\boldsymbol{\theta}(n+1) + \sum_{i=1}^{N} d_i \overline{e}(n-i)] \\
&= d(n) - \overline{\boldsymbol{\phi}}^T(n)\boldsymbol{\theta}(n) + \overline{\boldsymbol{\phi}}^T(n)[\boldsymbol{\theta}(n) - \boldsymbol{\theta}(n+1)] \\
&\quad + \sum_{i=1}^{N} d_i \overline{e}(n-i)] \\
&= d(n) - \overline{\boldsymbol{\phi}}^T(n)\boldsymbol{\theta}(n) - \overline{\boldsymbol{\phi}}^T(n)\mu\overline{\boldsymbol{\phi}}(n) \\
&\quad + \sum_{i=1}^{N} d_i \overline{e}(n-i)] \\
&= [\boldsymbol{I} - \overline{\boldsymbol{\phi}}^T(n)\mu\overline{\boldsymbol{\phi}}(n)]^{-1}\left[d(n) - \overline{\boldsymbol{\phi}}^T(n)\boldsymbol{\theta}(n) + \sum_{i=1}^{N} d_i \overline{e}(n-i)\right]
\end{aligned}$$

- The Hyperstability theorem in this model implies that, for $\epsilon(n) \to 0$ (i.e., $\overline{e}(n) \to 0$), then $\tilde{\boldsymbol{\theta}}(n+1)$ is bounded.

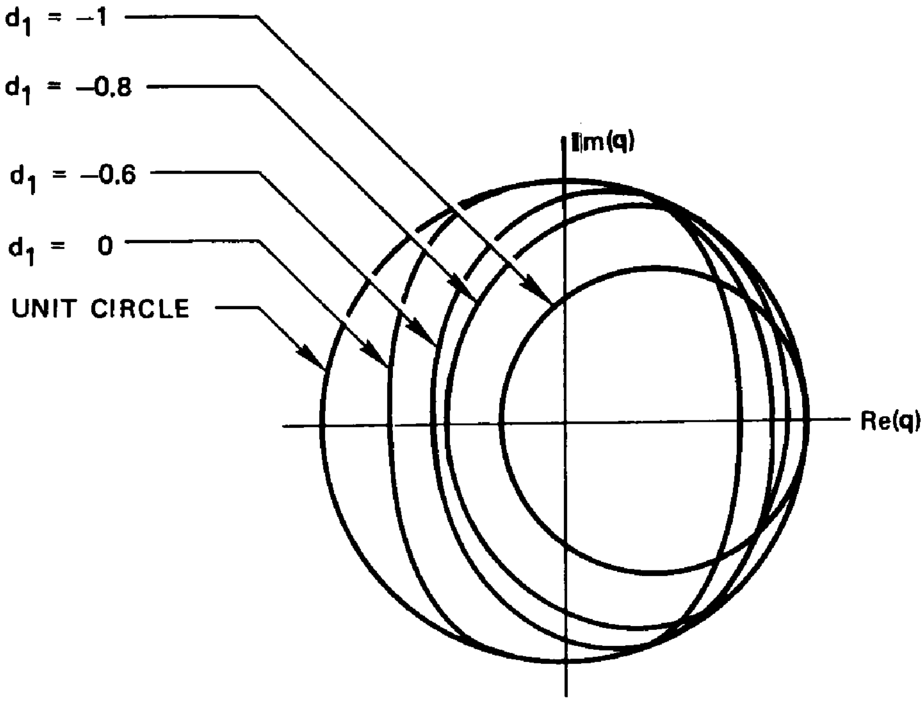


Figure 49: SPR condition of a second order system with different compensators.

- A simplified version of this algorithm, called the **Simple Hyperstable Adaptive Filter** (SHARF) is the following

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu \boldsymbol{\phi}(n) \epsilon(n)$$

  where $\epsilon(n) = e(n) + \sum_{i=1}^{N} d_i e(n-i)$.

- In particular, if $D(q)$ is time varying (adjusted at each iteration) such that

$$\hat{d}_i(n) = a_i(n)$$

  the proposed algorithm can be seen as a variant of the Instrumental Variable methods. Under the Hyperstability theory this algorithm was proposed initially by Landau (1978).

7.3.2 Forced error system

- Assuming now that measurement noise exist,

$$e(n) = d(n) - \hat{y}(n) + r(n)$$

- A possible extension of the previous discussion is to use an ARMAX model

$$d(n) = (1 - A(q))d(n) + B(q)x(n) + C(q)\nu(n)$$

  where $\nu(n)$ is zero mean, white noise uncorrelated with $x(n)$.

- Following similar steps than with the homogeneous error system we can obtain the following model for the forced error system

$$\overline{e}(n) = \frac{1}{C(q)}[\overline{\boldsymbol{\varphi}}(n)^T(n)\tilde{\boldsymbol{\theta}}(n+1)] + \nu(n)$$

  where $\overline{\boldsymbol{\varphi}}(n) = [d(n-1) \ldots d(n-N)\, x(n) \ldots x(n-N)\, \overline{e}(n-1) \ldots \overline{e}(n-P)]^T$.

- The SPR condition now must be verified over $\frac{1}{C(z)}$.

- On the other hand, the existence of the term $\nu(n)$ in the previous equation indicates the noisy convergence of this algorithm.

7.4 ODE associated

The ODE associated to the HARF method, and in particular to the SHARF algorithm using the slow convergence approximation, is the following

$$\begin{aligned} \frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} &= E\{\boldsymbol{\phi}(n)\boldsymbol{\psi}^T(n)(\boldsymbol{\theta}(n) - \boldsymbol{\theta}_o)\} \\ &= \boldsymbol{R}\tilde{\boldsymbol{\vartheta}}(t) \end{aligned}$$

where

$$\boldsymbol{\psi}(n) = [\frac{D(q)}{A(q)}d(n-1), \cdots, \frac{D(q)}{A(q)}d(n-N), x(n), \cdots, x(n-N)]^T \qquad (87)$$

and

$$\boldsymbol{R} = E\left\{\boldsymbol{\phi}(n)\boldsymbol{\psi}^T(n)\right\}$$

To study a Liapunov function related to the stability analysis of the ODE consider the following

> Lemma: Suppose $deg\ \hat{H}(z) = N$. If $\frac{D(z)}{A(z)}$ is SPR then $\boldsymbol{R} + \boldsymbol{R}^T$ is positive definite.

With this result, choose as Liapunov function the following

$$V(\boldsymbol{\vartheta}(t)) = \frac{1}{2}\|\boldsymbol{\vartheta}(t) - \boldsymbol{\theta}_o\|^2$$

such that

$$\begin{aligned} (dV/dt) &= \tilde{\boldsymbol{\vartheta}}^T(t)\frac{\partial \tilde{\boldsymbol{\vartheta}}(t)}{\partial t} + \left(\frac{\partial \tilde{\boldsymbol{\vartheta}}(t)}{\partial t}\right)^T \tilde{\boldsymbol{\vartheta}}(t) \\ &= -\tilde{\boldsymbol{\vartheta}}^T(t)(\boldsymbol{R} + \boldsymbol{R}^T)\tilde{\boldsymbol{\vartheta}}(t) \\ &\leq 0 \end{aligned}$$

7.5 The insufficient order case

The forced error system can be used in the insufficient order case,

$$d(n) = d_M(n) + d_U(n)$$

where $d_M(n)$ is the modelable part of the plant. In such situation

$$\begin{aligned} \overline{e}(n) &= \sum_{i=1}^{N} a_i[d_M(n-i) - \overline{y}(n-i)] + \overline{\boldsymbol{\phi}}^T(n)\tilde{\boldsymbol{\theta}}(n+1) + d_U(n) \\ &= \frac{1}{A(q)}[\overline{\boldsymbol{\phi}}^T(n)\tilde{\boldsymbol{\theta}}(n+1)] + d_U(n) \end{aligned}$$

Then, using the idea of error smoothing we obtain

$$\tilde{\boldsymbol{\theta}}(n+1) = \tilde{\boldsymbol{\theta}}(n) - \mu\overline{\boldsymbol{\phi}}\left\{\frac{D(q)}{A(q)}\overline{\boldsymbol{\phi}}^T(n)\tilde{\boldsymbol{\theta}}(n+1) + D(q)d_U(n)\right\}$$

This is meaningful if the modelable part $d_M(n) = \hat{H}(z)x(n)$, represent a stationary point of the algorithm.

Consider the ODE associated to the SHARF

$$\begin{aligned} \mathbf{0}_{N+1} &= E\left\{\begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N) \end{bmatrix}\epsilon(n)\right\} = \left\langle\begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-N} \end{bmatrix}, \mathcal{S}_x(z)D(z)[H(z) - \hat{H}(z)]\right\rangle \\ \mathbf{0}_N &= E\left\{\begin{bmatrix} \hat{y}(n-1) \\ \vdots \\ \hat{y}(n-N) \end{bmatrix}\epsilon(n)\right\} = \left\langle\begin{bmatrix} z^{-1} \\ \vdots \\ z^{-N} \end{bmatrix}\hat{H}(z), \mathcal{S}_x(z)D(z)[H(z) - \hat{H}(z)]\right\rangle \end{aligned}$$

If $f(z) = \sum_{k=-\infty}^{\infty} f_k z^{-k} = \mathcal{S}_x(z)D(z)[H(z) - \hat{H}(z)]$, the first equation is satisfied if and only if: $f_k = 0$, for $k = 0, 1, ..., N$, or $[f(z)]_+ = z^{-(N+1)}g(z)$.

Then, in the second equation

$$
\begin{aligned}
\mathbf{0}_N &= \left\langle \begin{bmatrix} z^{-1} \\ \vdots \\ z^{-N} \end{bmatrix} \hat{H}(z), \mathcal{S}_x(z)D(z)[H(z) - \hat{H}(z)] \right\rangle \\
&= \left\langle \begin{bmatrix} z^{-1} \\ \vdots \\ z^{-N} \end{bmatrix} \hat{H}(z), [f(z)]_+ \right\rangle \\
&= \left\langle \begin{bmatrix} z^{-1}/A(z) \\ \vdots \\ z^{-N}/A(z) \end{bmatrix} B(z), z^{-(N+1)}g(z) \right\rangle \\
&= \left\langle \begin{bmatrix} 1/A(z) \\ \vdots \\ z^{-N+1}/A(z) \end{bmatrix}, z^{-N}B(z^{-1})g(z) \right\rangle
\end{aligned}
$$

since the second operand is a causal function, and assuming no pole-zero cancellations, by using the decomposition theorem it must be $g(z) = V(z)Q(z)$ for some $Q(z) \in \mathcal{H}_2$.

Theorem: Suppose $\hat{H}(z)$ has no pole-zero cancellations, then $\hat{H}(z)$ is a stationary point if and only if $f_0 = 0$ and

$$
[f(z)]_+ = z^{-(N+1)}V(z)Q(z),
$$

for some $Q(z) \in \mathcal{H}_2$.

Corollary: Suppose $\hat{H}(z)$ has no pole-zero cancellations and the input is white, then $\hat{H}(z)$ is a stationary point if and only if

$$H(z) - \hat{H}(z) \;=\; z^{-(N+1)}V(z)Q'(z),$$

for some $Q'(z) \in \mathcal{H}_2$.

then this leads to a stationary point similar to an $N+1$-sample Padé approximant to $H(z)$.

Note that, since $V(z^{-1})\hat{H}(z)$ is an anticausal function, then multiplying the previous equation by $V(z^{-1})$ results in

$$[V(z^{-1})H(z)]_+ \;=\; z^{-(N+1)}Q'(z)$$

which can be written as $\boldsymbol{q}' = \boldsymbol{\Gamma}_H \boldsymbol{v}$, or explicitly

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ q_0' \\ q_1' \\ \vdots \end{bmatrix} = \boldsymbol{\Gamma}_H \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix}$$

in the sufficient order case, if $\boldsymbol{v}$ is orthogonal to the first N rows of $\boldsymbol{\Gamma}_H$ then it is orthogonal to all rows, giving $\boldsymbol{\Gamma}_H \boldsymbol{v} = 0$.

In the undermodelled case, $< H(z), z^{-k}V(z) >= 0$, for $k = 1, 2, ..., N$, need not correspond to *good* approximation of $H(z)$.

7.6 Some examples

- $x(n)$ unit variance white noise (no measurement noise).
- $H(z) = \frac{1-2z^{-1}}{1-z^{-1}+0.25\,z^{-2}}$.

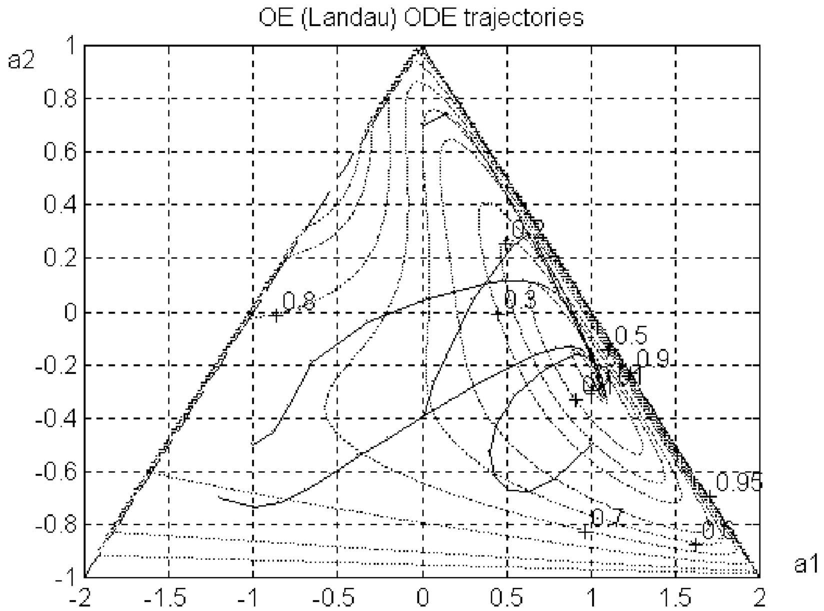


Figure 50: ODE trajectories of Output error (Landau) method.

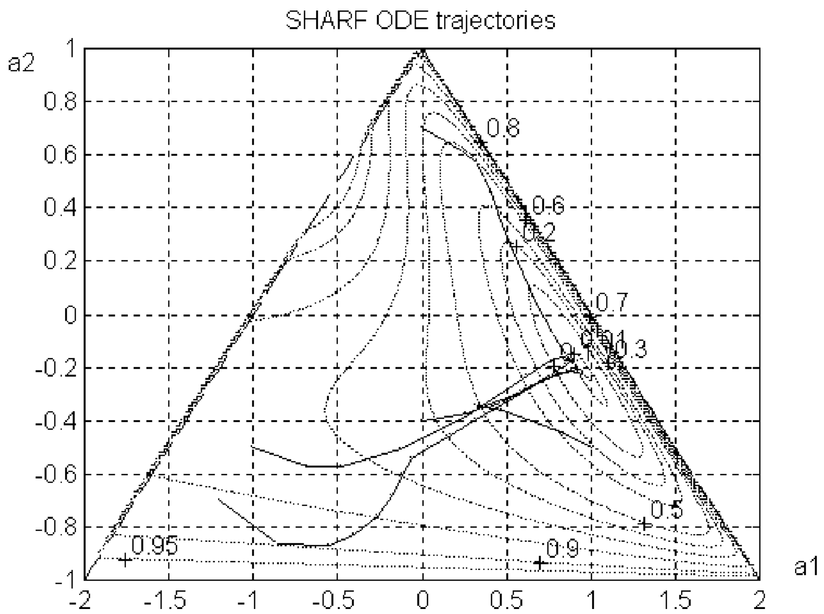


Figure 51: ODE trajectories of SHARF algorithm ($D(z)$ ideal).

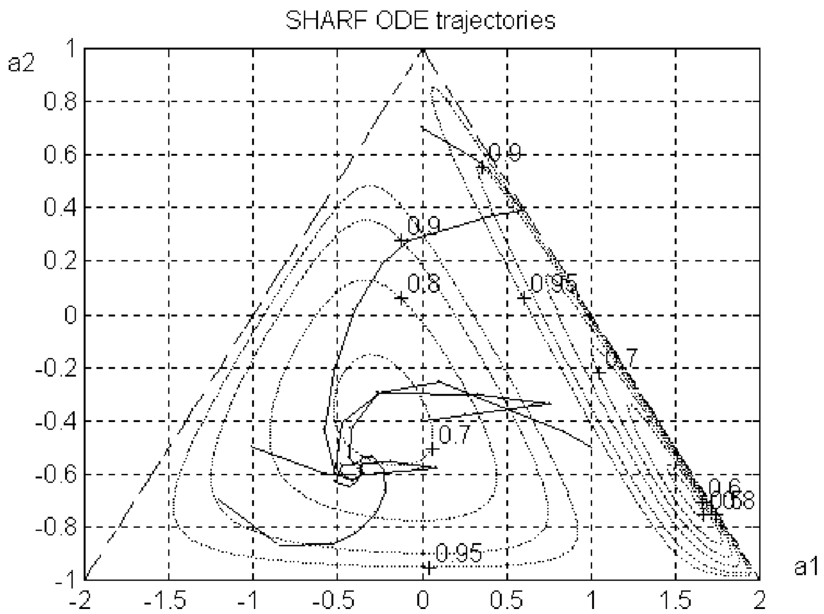


Figure 52: ODE trajectories of SHARF algorithm, insufficient order (order zero numerator).

8 Steiglitz-McBride method

The closest approximation to the global minimum

- Stationary points and ODE associated (local convergence).
- The reduced order case.
- Bound on the MSE related to the MSOE.
- Direct-form realization.
- Other realizations: lattice, orthonormal.