The ODE Method and the ELS Algorithm

See Aström and Ljung.

<u>System</u> $\quad y(k) = \phi_0^T(k-1)\theta_0 + w(k)$

$$\theta_0^T = (-a_1, \ldots, -a_n, b_1, \ldots, b_m, c_1, \ldots, c_p).$$

$$\phi_0^T(k-1) = (y(k-1), \ldots, u(k-1), \ldots, w(k-1), \ldots).$$

<u>ELS</u>

$$\hat\theta(k+1) = \hat\theta(k) + R^{-1}(k)\phi(k)e(k+1)$$

$$e(k+1) = y(k+1) - \hat\theta(k)^T\phi(k)$$

$$R(k+1) = R(k) + \phi(k)\phi(k)^T$$

where $\phi(k)$ is identical to $\phi_0(k)$, except $w(j)$ is replaced by

$$\hat w(j) = y(j) - \phi(j-1)^T\hat\theta(j), \quad j \geq 1.$$

<u>Step 1</u> <u>Normalization</u> $R(k)$ grows linearly if $\{\phi(k)\}$ is bounded. To obtain a bounded signal, set

$$R^N(k+1) = \frac{1}{k+1}R(k+1) = \frac{1}{k+1}\sum_1^{k+1}\phi(i)\phi(i)^T + \frac{R(0)}{k+1}$$

$$= \frac{1}{k+1}\left(R(k) + \phi(k+1)\phi(k+1)^T\right)$$

$$= \frac{1}{k+1}\left(kR^N(k) + \phi(k+1)\phi(k+1)^T\right)$$

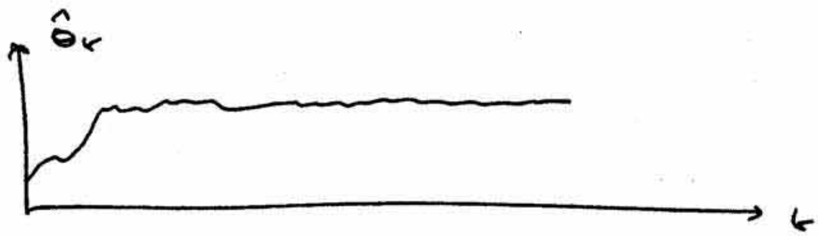
$$= R^N(k) + \frac{1}{k+1}\left(\phi(k+1)\phi(k+1)^T - R^N(k)\right)$$

This leads to a time-scale separation ...

$$\left.y(t) = \phi_0(t-1)^T \theta_0 + w(t)\right\} \quad \underline{\text{FAST}}$$

$$\left.\begin{aligned}\hat{\theta}(t+1) &= \hat{\theta}(t) + \frac{1}{t}\left[R^N(t)\right]^{-1} \phi(t)\, e(t+1) \\ \tilde{R}(t+1) &= R^N(t) + \frac{1}{t+1}\left[\phi(t+1)\phi(t+1)^T - R^N(t)\right]\end{aligned}\right\} \quad \underline{\text{SLOW}}$$

<u>Step 2</u> <u>Averaging</u>



Consider $\hat{\theta}(t) \cong \theta$ for a long time interval

Replace,

$$\begin{cases}\phi(t)\left(y(t+1) - \phi^T(t)\hat{\theta}(t)\right) & \text{by} \quad E_{\theta}\left[\phi(t)\left(y(t+1) - \phi^T(t)\theta\right)\right] = G(\theta). \\ \phi(t)\phi(t)^T & \text{by} \quad E_{\theta}\left[\phi(t)\phi(t)^T\right] = M(\theta)\end{cases}$$

$\Rightarrow$ Approximate equations,

$$\hat{\theta}(t+1) \cong \hat{\theta}(t) + \frac{1}{t} R^N(t)^{-1}\, G(\hat{\theta}(t))$$

$$R^N(t+1) \cong R^N(t) + \frac{1}{t}\left[M(\hat{\theta}(t)) - \tilde{R}^N(t)\right].$$

<u>Resembles a differential equation!</u>

<u>Note</u> Expectation defining G and M are in steady state, assuming θ is <u>fixed</u>.

Step 3 Let $t_\phi = 0$, $t_{k+1} = t_k + \frac{1}{k}$, $k \geq 1$.

Then we can formalize derivative interpretation:

$$\frac{\hat{\theta}(k+1) - \hat{\theta}(k)}{t_{k+1} - t_k} = \left[R^N(k)\right]^{-1} E_{\theta=\hat{\theta}(k)}\left[\phi(k)\, e(k+1)\right]$$

$$\frac{R^N(k+1) - R^N(k)}{t_{k+1} - t_k} = E_{\theta=\hat{\theta}(k)}\left[\phi(k+1)\phi(k+1)^T\right] - R^N(k).$$

Considering the ODEs,

$$\frac{d\bar{\theta}(t)}{dt} = \bar{R}^{-1}(t)\; E_{\bar{\theta}(t)}\left[\phi(k)\left(y(k+1) - \phi^T(k)\bar{\theta}(t)\right)\right]$$

$$\frac{d\bar{R}(t)}{dt} = E_{\bar{\theta}(t)}\left[\phi(k)\phi(k)^T\right] - \bar{R}(t)$$

we can hope for an approximation of the form

$$\boxed{\begin{aligned} \bar{\theta}(t_k) &\cong \hat{\theta}(k) \\ \bar{R}(t_k) &\cong R^N(k) \end{aligned}}$$

True under some strong assumptions.

The next step is to make sense of these expectations!

Step 4   What is $G(\bar{\theta}) = E_{\bar{\theta}}\left[(y(t) - \phi^T(t-1)\bar{\theta})\, \phi(t-1) \right]$?

With $\bar{\theta}$ fixed we have,

$\hat{w}(t) = y(t) - \phi^T(t-1)\bar{\theta}$,   so with $\tilde{w}_t = w_t - \hat{w}_t$,

$$
\begin{aligned}
-\tilde{w}(t) &= y(t) - \phi^T(t-1)\bar{\theta} - w(t) \\
&= \phi_0^T(t-1)\,\theta_0 - \phi^T(t-1)\,\bar{\theta} - w(t) \\
&\qquad + w(t) \\
&= [\phi_0(t-1) - \phi(t-1)]^T\theta_0 + [\theta_0 - \bar{\theta}]^T\phi(t-1) \\
&= \left[\begin{pmatrix} y(t-1) \\ \vdots \\ u(t-1) \\ \vdots \\ w(t-1) \\ \vdots \end{pmatrix} - \begin{pmatrix} y(t-1) \\ \vdots \\ u(t-1) \\ \vdots \\ \hat{w}(t-1) \\ \vdots \end{pmatrix} \right]^T \theta_0 + \tilde{\theta}^T\phi(t-1). \\
&= c_1\tilde{w}(t-1) + \dots + c_p\tilde{w}(t-p) + \tilde{\theta}^T\phi(t-1).
\end{aligned}
$$

Thus, $\boxed{\tilde{\theta}^T\phi(t-1) = -C(z^{-1})\,\tilde{w}(t).}$

So, $\tilde{\theta}^T\phi(t-1) = C(z^{-1})\left[y(t) - \phi^T(t-1)\bar{\theta} - w(t) \right]$

$$
\begin{aligned}
\Rightarrow \quad y(t) &= \phi^T(t-1)\bar{\theta} + w(t) + \frac{1}{C(z^{-1})}\left[\tilde{\theta}^T\phi(t-1) \right] \\
&= \phi^T(t-1)\bar{\theta} + w(t) + \tilde{\theta}^T\phi_f(t-1)
\end{aligned}
$$

where $x_f = \frac{1}{C(z^{-1})}x$ for any sequence $\{x(t)\}$.

Step 4 continued

$$E_{\bar\theta}\left[\phi(t-1)\left(y(t)-\bar\theta^T\phi(t-1)\right)\right]$$
$$= E_{\bar\theta}\left[\phi(t-1)\left(w(t)+\phi_f(t-1)^T\tilde\theta\right)\right]$$
$$= E_{\bar\theta}\left[\phi(t-1)\,\phi_f(t-1)^T\right]\tilde\theta$$
$$= G(\bar\theta).$$

Write this as $G(\bar\theta) = M_f(\bar\theta)\,\tilde\theta$, where

$$M(\bar\theta) = E_{\bar\theta}\left[\phi(t)\phi(t)^T\right]$$
$$M_f(\bar\theta) = E_{\bar\theta}\left[\phi(t)\,\phi_f(t)^T\right].$$

The ODE becomes,

$$\frac{d}{dt}\bar\theta(t) = \bar R^{-1}(t)M_f(\bar\theta(t))\,\tilde\theta(t)$$
$$\frac{d}{dt}\bar R(t) = M(\bar\theta(t)) - \bar R(t).$$

Step 5 stability of the ODE.

Lyapunov function: $V(t) = \tilde\theta^T(t)\,\bar R(t)\,\tilde\theta(t).$

The product rule gives,

$$\dot V = \dot{\tilde\theta}^T\bar R\,\tilde\theta + \tilde\theta^T\dot{\bar R}\,\tilde\theta + \tilde\theta^T\bar R\,\dot{\tilde\theta}$$

Substituting, ...

$$\dot{V} = -\tilde{\theta}^T M_f^T \bar{R}^{-1} \bar{R} \tilde{\theta} + \tilde{\theta}^T [M - \bar{R}]\tilde{\theta} - \tilde{\theta}^T \bar{R} \bar{R}^{-1} M_f \tilde{\theta}$$

$$= -\tilde{\theta}^T \{ M_f^T + M_f - M + \bar{R} \} \tilde{\theta}$$

$$= -\underbrace{\tilde{\theta}^T \bar{R} \tilde{\theta}}_{V} - \tilde{\theta}^T \{ \underbrace{M_f^T + M_f - M}_{\oplus ?} \} \tilde{\theta}$$

For stability we search for conditions under which the following matrix is positive semi definite:

$$(M_f^T + M_f - M)(\bar{\theta})$$

$$= E_{\bar{\theta}}\left[\phi_f(t)\phi(t)^T + \phi(t)\phi_f(t)^T - \phi(t)\phi(t)^T \right].$$

Step 6: Positive-Real Condition

Consider any $x \in \mathbb{R}^N$, let $z(t) = x^T\phi(t)$

$z_f(t) = x^T \phi_f(t), \quad t \geq 0.$

$$x^T (M_f^T + M_f - M) x$$

$$= E_{\bar{\theta}}\left[x^T \phi_f(t)\phi(t)^T x + x^T\phi(t)\phi_f(t)^T x - x^T\phi(t)\phi(t)^T x \right]$$

$$= E_{\bar{\theta}}\left[2 z(t) z_f(t) - z^2(t) \right].$$

$$\geq 0 \ ?$$

Recall $\phi_f = \frac{1}{c(z^{-1})}\phi$, so $z_f = \frac{1}{c(z^{-1})} z$.

Expectation becomes,

$$2\,E\left[z(k) \left\{ \left(\frac{1}{C(z^{-1})} - \frac{1}{2} \right) z(k) \right\} \right] = 2E[z(k)\,\xi(k)]$$

Positivity is a form of passivity:

$$z \longrightarrow \boxed{H} \xrightarrow{\ \xi\ } \qquad \langle z, \xi \rangle > 0 \quad \text{all } z.$$

We can compute:

$$E[z(k)\,\xi(k)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_z(\omega)\, H(e^{-j\omega})\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_z(\omega)\, \mathrm{Re}\left(H(e^{-j\omega}) \right) d\omega$$

To ensure positivity we require

$$\mathrm{Re}\left(H(e^{-j\omega}) \right) = \mathrm{Re}\left(\frac{1}{C(e^{-j\omega})} - \frac{1}{2} \right) \geq 0, \qquad \underline{\underline{\text{all } \omega}}$$

<u>Conclusions</u>: ① Under the SPR condition we have $\frac{d}{dt} V(t) \leq -V(t)$, so $V(t) \leq V(0)\, e^{-t}$, $t \geq 0$.

$\rightarrow 0$ exp. fast.

② If in addition $\bar{R}(t) \geq \varepsilon I$ for some $\varepsilon > 0$, and all $t \geq 0$, then

$$V(t) \geq \varepsilon \|\tilde{\theta}(t)\|^2 \rightarrow 0 \quad \text{exp. fast.}$$

Recall, $\bar{R}(t_k) \cong R^N(k) = \frac{R(k)}{k}$, giving the condition

$$R(k) \geq \varepsilon k I, \qquad k \geq 1.$$

Bounds for stochastic model : See notes

It is not true that $\hat{\theta}(t) = \bar{\theta}(t)$

However, we can show that

$$\hat{w}(t) - w(t) = \frac{1}{C(z^{-1})} \phi^T(t-1)\, \tilde{\theta}(t)$$

Here with $V_n = \tilde{\theta}_n^T R_{n-1} \tilde{\theta}_n$ we can obtain a useful recursion

$$V_{n+1} = V_n - 2(\phi_n^T \tilde{\theta}_{n+1}) \left[\left(\frac{1}{C(z^{-1})} - \frac{1}{2} \right) \phi_n^T \tilde{\theta}_{n+1} \right]$$

$$- \phi_n^T R_{n-1} \phi_n w_{n+1}^2 + 2 \frac{\phi_n^T R_{n-1} \phi_n}{1 + \phi_n^T R_{n-1} \phi_n} w_{n+1}^2$$

$$+ 2 \left(\phi_n^T \tilde{\theta}_n + \frac{\phi_n^T R_{n-1} \phi_n}{1 + \phi_n^T R_{n-1} \phi_n} (e_{n+1} - w_{n+1}) \right)$$

Similar to previous recursion.

Positive real condition $\Rightarrow$ negative drift.

Conclusions: As before, for white noise, under P.R condition,

1) $$\|\tilde{\theta}_{n+1}\|^2 = O\left(\frac{\log(\lambda_{max}(R_n))}{\lambda_{min}(R_n)} \right)$$

2) $$\sum_0^n \left(\tilde{\theta}_{k+1}^T \phi_k\right)^2 = O\left(\log\left(\lambda_{max}(R_n)\right)\right)$$

Since $\hat{w}_n - w_n = \frac{1}{C(z^{-1})}\left(\phi_{n-1}^T \tilde{\theta}_n\right)$, and $\frac{1}{C}$ is stable,

3) $$\sum_0^n \left(\hat{w}_{k+1} - w_{k+1}\right)^2 = O\left(\log\left(\lambda_{max}(R_n)\right)\right)$$

Crucial point is, these bounds hold for *any* sequence $\{\phi_n\}$. Hence we have the system description

$$y(t+1) = \hat{\theta}(t+1)^T \phi(t) + \hat{w}(t+1)$$
$$= \hat{\theta}(t)^T \phi(t) + e(t+1)$$

where $\{e(t)\}$ or $\{\hat{w}(t)\}$ is *almost* noise!

Other Algorithms: Similar results apply.

Ex Stochastic Gradient.

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{\phi(k)}{r(k)} e(k+1)$$

$$r(k+1) = r(k) + \|\phi(k+1)\|^2$$

If $C(z^{-1})$ is strictly positive real:

$$Re\,(C(e^{-j\omega})) > 0, \qquad \omega > 0,$$

then it can be shown that

$$\frac{1}{r(n)} \sum_{1}^{n} (w(k) - \hat{w}(k))^2 \longrightarrow 0, \quad n \to \infty.$$

So, if $\phi(k)$ is bounded,

$$\frac{1}{n} \sum_{1}^{n} (w(k) - \hat{w}(k))^2 \to 0, \quad n \to \infty.$$

This is a milder condition.

It is not known if the SG algorithm outperforms ELS in this sense – perhaps the Lyapunov function $V(t) = \tilde{\theta}^T R \tilde{\theta}$ is not optimal.

The ODE method in design,

or, the prediction error method revisited.

See Ljung "System ID"

Consider the model,

$$y(k) = H(z^{-1}, \theta_0)\, u(k-1) + n(k)$$

and generic predictor,

$$\hat{y}(k|\theta) = H(z^{-1}, \theta)\, u(k-1)$$

We take the cost criterion,

$$\Gamma(\theta) = \tfrac{1}{2} E\left[(y(k) - \hat{y}(k|\theta))^2 \right]$$

expectation in steady-state.

We assume all processes are w.s.s., zero-mean, and eventually we assume $E[n(k)\,u(k-i)] = 0$, $i \geq 1$.

Consider the ODE, $\boxed{\dfrac{d}{dt}\bar{\theta}(t) = -\gamma \nabla_\theta \Gamma(\bar{\theta}(t))}$

where $\gamma > 0$, and $\nabla_\theta \Gamma$ is the gradient: can be computed as follows

$$\nabla_\theta \Gamma(\theta) = \tfrac{1}{2} E\left[2(y(k) - \hat{y}(k|\theta))^2 (-\nabla_\theta \hat{y}(k|\theta)) \right]$$

$$= -E\left[e(k|\theta)\, \phi(k-1|\theta) \right]$$

where $e(k|\theta) = y(k) - \hat{y}(k|\theta)$,

$$\phi(k-1|\theta) = \nabla_\theta \hat{y}(k|\theta) = \left[\nabla_\theta H(z^{-1}, \theta)\right] u(k-1).$$

...

<u>Claim</u> If $n(k)$, $\{u(k-i) : i \leq 1\}$ are uncorrelated as described above, then θ_0 is a stationary point:

$$\nabla_\theta \Gamma(\theta_0) = 0 \in \mathbb{R}^N$$

<u>proof</u>

$$\Gamma(\theta) = \frac{1}{2} E\Big[\Big\{\big[H(z^{-1},\theta_0) - H(z^{-1},\theta)\big] u(k-1) + n(k)\Big\}^2\Big]$$

$$= \frac{1}{2} E\Big[\Big(\big[H(z^{-1},\theta_0) - H(z^{-1},\theta)\big] u(k-1)\Big)^2\Big] + \frac{1}{2}\sigma_n^2$$

where we have used the uncorrelated-assumption.

Clearly $\theta = \theta_0$ is a local and global minimum since $\Gamma(\theta_0) = \frac{1}{2}\sigma_n^2$. ∎

Coming back to the ODE:

$$\boxed{\begin{aligned} \frac{d}{dt}\bar\theta(t) &= -\gamma \nabla_\theta \Gamma(\bar\theta(t)) \\ &= \gamma E\Big[\underbrace{e(k \mid \bar\theta(t))}_{y - \hat{y}_{\bar\theta}}\ \underbrace{\phi(k-1 \mid \bar\theta(t))}_{\nabla_\theta \hat{y}_{\bar\theta}}\Big] \end{aligned}}$$

Recall prediction-error method!

<u>Goal</u> Construct an algorithm whose associated ODE is equal to (*)

<u>Solution</u> Remove expectation — that is, reverse steps in construction of ODE.

Let $\Phi(z^{-1}, \theta) = \nabla_\theta H(z^{-1}, \theta)$, so

$$\phi(t|\theta) = \Phi(z^{-1}, \theta)\, u(t)$$

Then we define,

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{\gamma}{k}\left(e(k+1)\, \phi(k)\right)$$

$$\phi(k) = \Phi(z^{-1}, \hat{\theta}(k))\, u(k)$$

$$e(k+1) = y(k+1) - \hat{y}(k+1 \mid \hat{\theta}(k))$$

$$= y(k+1) - H(z^{-1}, \hat{\theta}(k))\, u(k)$$

This parametrization leads to a specific algorithm. In this case it is of the <u>output-error form</u>.

Suppose instead we use the parametrization,

$$y(k) = (1 - A(z^{-1}, \theta_0))\, y(k) + B(z^{-1}, \theta_0)\, u(k-1) + n(k)$$

$$\hat{y}(k|\theta) = (1 - A(z^{-1}, \theta))\, y(k) + B(z^{-1}, \theta_0)\, u(k-1).$$

We then obtain an algorithm in <u>equation-error form</u> by again considering $\frac{1}{2}E\left[(y(k) - \hat{y}(k|\theta))^2\right] = \Gamma(\theta)$.

The equation-error form leads to simpler models and analysis because the gradient $\nabla_\theta \hat{y}(k|\theta)$ is more easily computed.

Question: When is the ODE asymptotically stable?

Answer: Consider quadratic approximation,

$$\nabla_\theta \Gamma(\theta) \cong \tfrac{1}{2}(\theta - \theta_0)^T \Sigma (\theta - \theta_0), \qquad \theta \sim \theta_0,$$

where $\Sigma = \Delta \Gamma(\theta_0)$. The ODE is a. stable if $\Sigma > 0$. This can be computed:

$$\nabla_\theta \Gamma(\theta) = -E[e(k|\theta)\phi(k-1|\theta)]$$

So $$\Delta \Gamma(\theta) = -E[(\nabla_\theta e(k|\theta))\phi(k-1|\theta)^T + e(k|\theta)\nabla\phi(k-1|\theta)]$$

at $\theta = \theta_0$ we have $e(k|\theta) = n(k)$, so

$$\Delta \Gamma(\theta_0) = -E[(\nabla_\theta e(k|\theta))\phi(k-1|\theta)^T].$$

Finally, $$\nabla_\theta e(k|\theta) = \nabla_\theta (y(k) - \hat{y}(k|\theta))$$
$$= -\nabla_\theta \hat{y}(k|\theta) = -\phi(k-1|\theta).$$

So, $$\Sigma = \Delta\Gamma(\theta_0) = E[\phi(k-1|\theta)\phi(k-1|\theta)^T]$$

positivity of Σ is a form of persistence of excitation.

Output error method for ARMAX model

Consider now the parametrization,

$$H(z^{-1}, \theta)\, z^{-1} = \frac{b_1 z^{-1} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + \cdots + b_n z^{-n}} = \frac{B(z^{-1}, \theta)}{A(z^{-1}, \theta)}$$

$\hat{y}(k|\theta)$ is the solution to

$$A(z^{-1}, \theta)\, \hat{y}(k|\theta) = B(z^{-1}, \theta)\, u(k)$$

or $$\hat{y}(k|\theta) + a_1 \hat{y}(k-1|\theta) + \cdots + a_n \hat{y}(k-n|\theta)$$
$$= b_1 u(k-1) + \cdots + b_m u(k-m), \quad k \geq 0.$$

To compute $\nabla \hat{y}$ consider

$$\hat{y}(k|\theta) = \frac{B(z^{-1}, \theta)}{A(z^{-1}, \theta)}\, u(k)$$

$$\therefore \quad \nabla \hat{y}(k|\theta) = \left[\frac{A \nabla B - B \nabla A}{A^2}\right] u(k)$$

and $$\begin{cases} \nabla B(z^{-1}, \theta) = (0, \ldots, 0, z^{-1}, \ldots, z^{-m})^T \\ \nabla A(z^{-1}, \theta) = (z^{-1}, \ldots, z^{-n}, 0, \ldots, 0)^T \end{cases}$$

So $$\nabla \hat{y}(k|\theta) = \frac{1}{A(z^{-1})}\left\{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ u(k-1) \\ \vdots \\ u(k-m) \end{bmatrix} - \frac{B(z^{-1})}{A(z^{-1})}\begin{bmatrix} u(k-1) \\ \vdots \\ u(k-n) \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right\}$$

$$= \frac{1}{A(z^{-1})}\left(-\hat{y}(k-1|\theta), \ldots, -\hat{y}(k-n|\theta), u(k-1), \ldots, u(k-m)\right)^T$$

$$= \phi(k-1|\theta)$$

stability of A seems crucial.

Algorithm:

$$\begin{cases} \hat{y}(k) = -\hat{a}_1(k)\,\hat{y}(k-1) - \cdots - \hat{a}_n(k)\,\hat{y}(k-n) \\ \qquad + \hat{b}_1(k)\,u(k-1) + \cdots + \hat{b}_m(k)\,u(k-m). \\ \phi_0(k-1) = (-\hat{y}(k-1), \ldots, -\hat{y}(k-n), u(k-1), \ldots, u(k-m))^T \\ \phi(k) + \hat{a}_1(k)\,\phi(k-1) + \cdots + \hat{a}_n(k)\,\phi(k-n) = \phi_0(k) \\ e(k+1) = y(k+1) - \hat{\theta}(k)^T \phi(k) \end{cases}$$

Given these, we can use RLS, SG, etc.

SG:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{1}{r(k)}\phi(k)\,e(k+1)$$

$$r(k+1) = r(k) + \|\phi(k)\|^2, \qquad k \geq 0.$$

Question: Is it necessary to use $\phi(k)$ instead of $\phi_0(k)$??

Answer: Suppose $A(z^{-1})$ for plant is s. passive (SPR)

$$\mathrm{Re}\,(A(e^{-j\omega})) > 0 \quad \text{for all } \omega \in \mathbb{R}.$$

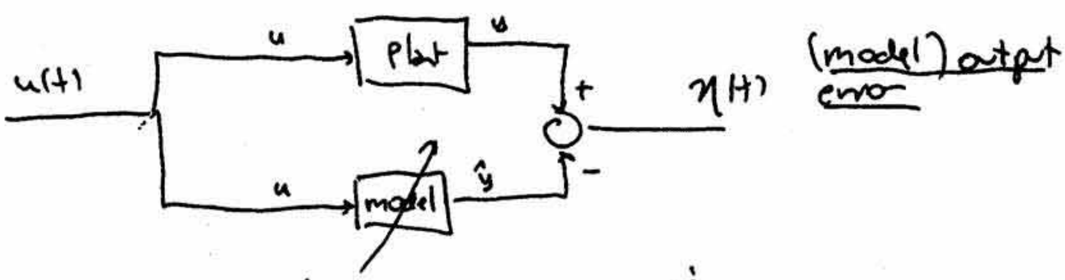
Then the SG algorithm is stable when $(\phi(k_0))$ is wide sense stationary, and convergent if $E[\phi(k)\phi(k)^T] = Q > 0$.

For LS we require SPR property for $\frac{1}{A} - \frac{1}{2}$.

Convergence of output error methods

First note we have seen this in <u>example 2</u>:

$$\dot{y} = ay + bu \qquad \text{plant}$$
$$\dot{\hat{y}} = \hat{a}\hat{y} + \hat{b}u \qquad \text{model}$$



Choosing $\dot{\hat{a}} = \gamma_1 \hat{y}\eta$, $\dot{\hat{b}} = \gamma_2 u \eta$

$\Rightarrow \quad \int_0^\infty \eta^2(s)\,ds < \infty$ provided $a < 0$ (stable plant).

Contrast to <u>homework #1</u>: (generalization)

$$e = \text{equation error} = \dot{y} - (\hat{a}y + \hat{b}u)$$
$$= \hat{A}(s)y - \hat{B}(s)u$$

$$\dot{\hat{a}} = \gamma_1 y e \qquad \dot{\hat{b}} = \gamma_2 u e$$

Today we will look at an output-error algorithm in greater detail to see where passivity arises in the analysis.

(steady-state analysis)

Example

$$y(k) = a\,y(k-1) + b\,u(k-1) + n(k) = \theta_0^T \phi_0(k-1) + n(k)$$
$$\hat{y}(k) = \hat{a}\,\hat{y}(k-1) + \hat{b}\,u(k-1) = \hat{\theta}^T \phi(k-1)$$

$\phi(k) = (\hat{y}(k), u(k))^T$; $\phi_0(k) = (y(k), u(k))^T$.

Will L.S. work when using ϕ in place of ϕ_0 ??

Assume u is a white-noise sequence, independent of n.

$$E[\hat{y}(k)\,u(k)] = E[(\hat{a}\,y(k-1) + \hat{b}\,u(k-1))\,u(k)]$$
$$= 0$$

$$E[y(k)\,u(k)] = 0$$

$$E[\hat{y}^2(k)] = \hat{a}^2 E[\hat{y}(k-1)^2] + \hat{b}^2 E[u(k-1)^2]$$
$$= \hat{a}^2 E[\hat{y}(k-1)^2] + \hat{b}^2 \sigma_u^2.$$

$$\therefore\ E[\hat{y}^2(k)] \longrightarrow \frac{\hat{b}^2 \sigma_u^2}{1-\hat{a}^2}, \quad k \to \infty, \text{ if } |\hat{a}| < 1.$$

$$E[y(k)^2] = a^2 E[y(k-1)^2] + b^2 E[u(k-1)^2] + \sigma_n^2$$

if $n(k)$ is ind. of $(u(k-1), y(k-1))$

$$\therefore\ E[y(k)^2] \to \frac{\sigma_n^2 + b^2\sigma_u^2}{1-a^2} \quad \text{if } |a| < 1.$$

$$E[\hat{y}(k)\,y(k)] = a\hat{a}\,E[\hat{y}(k-1)\,y(k-1)] + b\hat{b}\,\sigma_u^2$$

$$\to \frac{b\hat{b}\,\sigma_u^2}{1-a\hat{a}}$$

Assume all is stationary ...

Normalize $\sigma_u^2 = \sigma_n^2 = 1$

L.S. estimates will satisfy,

$$\hat{\Theta} = E[\phi\phi^T]^{-1}\left(E[\phi\phi_0^T]\,\theta_0 + \cancelto{0}{E[\phi\phi_0^T w]}\right)$$

$$= \begin{bmatrix} \frac{\hat{b}^2}{1-\hat{a}^2} & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{b\hat{b}}{1-a\hat{a}} & 0 \\ 0 & 1 \end{bmatrix} \theta_0$$

$$= \begin{bmatrix} \frac{1-\hat{a}^2}{\hat{b}^2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{b\hat{b}}{1-a\hat{a}} & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$= \begin{pmatrix} \frac{(1-\hat{a}^2)(b\hat{b})}{\hat{b}^2(1-a\hat{a})}\,a \\ b \end{pmatrix}$$

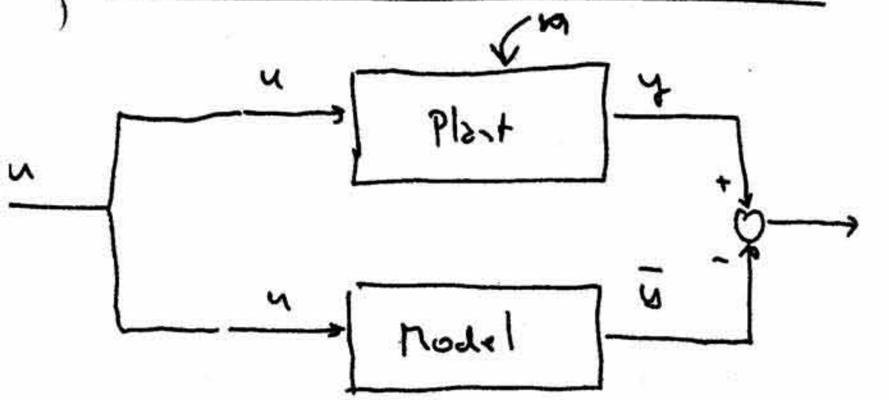
So, $\hat{b} = b$. But then,

$$\hat{a} = \frac{(1-\hat{a}^2)\,\cancel{b^2}}{\cancel{b^2}(1-a\hat{a})}\,a$$

$$\hat{a} - a\hat{a}^2 = (1-\hat{a}^2)\,a$$

$$\hat{a} = a$$

Construction of an O.E. Algorithm



$$\eta(t) = y(t) - \hat{H}_t u(t)$$

$$\hat{H} = \frac{\hat{B}(z^{-1})}{\hat{A}(z^{-1})}$$

$$\hat{A}_t \bar{y}(t) = \hat{B}_t u(t)$$

Let $\bar{y}$ denote the <u>aposteriori</u> <u>model</u> output

$$\bar{y}(t) = \hat{\theta}(t)^T \bar{\phi}(t-1), \quad \eta(t) = y(t) - \bar{y}(t).$$

$$\hat{\theta}(t) = [-\hat{a}_1(t), \dots, -\hat{a}_n(t), \hat{b}_1(t), \dots, \hat{b}_m(t)]^T$$

$$\bar{\phi}(t-1) = [\bar{y}(t-1), \dots, \bar{y}(t-n), u(t-1), \dots, u(t-m)]^T.$$

We let $\hat{y}(t)$ = apriori model output,

$$\hat{y}(t) = \hat{\theta}(t-1)^T \bar{\phi}(t-1),$$

and
$$e(t) = y(t) - \hat{y}(t)$$

The most general results are obtained via filtering
Let $D(z^{-1})$ be stable, and stably-invertible

$$D(z^{-1}) = 1 + \sum_{1}^{\infty} d_i z^{-i}$$

Generalized aposteriori output-error:

$$\bar{\eta} = D(z^{-1})\eta$$

Generalized apriori output-error

$$\bar{n} = e + \left[D(z^{-1}) - 1\right]\eta$$

Given these definitions we may use LS, LMS, etc. with ϕ replaced by $\bar{\phi}$.

<u>Consider NLMS</u> $\hat{\theta}(t) = \hat{\theta}(t-1) + \dfrac{\bar{\phi}(t-1)}{1 + \|\bar{\phi}(t-1)\|^2}\bar{n}(t)$.

Goal: 1) Show $\sum_{t=0}^{\infty} e^2(t) < \infty$ in noise-free case.

2) Show ODE is stable in wss. case.

Under a p.e. condition it will then follow that

$$\hat{\theta}(t) \to \theta_0, \quad t \to \infty.$$

Consider noise-free case...

<u>Step 1</u> Connect $\eta(t)$ and $\bar{\phi}^T(t-1)\tilde{\theta}(t)$.

Polynomial notation: $\begin{cases} A(z^{-1})y = B(z^{-1})u \\ \hat{A}(z^{-1})\bar{y} = \hat{B}(z^{-1})u \end{cases}$

$$A(z^{-1})[y - \bar{y}] = B(z^{-1})u - A\bar{y} + \underbrace{(\hat{A}\bar{y} - \hat{B}u)}_{\text{zero.}}$$

$$= \theta_0^T\bar{\phi}(t) - \hat{\theta}^T(t)\bar{\phi}(t).$$

That is, $A(z^{-1})\eta(t) = \tilde{\theta}(t)^T \bar{\phi}(t-1)$
$= \bar{\phi}(t-1)^T \tilde{\theta}(t)$.

Applying $D(z^{-1})$ to both sides then gives,

$$A(z^{-1})\bar{\eta}(t) = D(z^{-1})\left\{ \bar{\phi}(t-1)^T \tilde{\theta}(t) \right\}.$$

Similar to ELS!

We will require a passivity condition of $\frac{D}{A}(z^{-1})$.

<u>Step 2</u>: Relate aposteriori & apriori errors.

$$\bar{\phi}^T(t-1)\left\{ \hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\bar{\phi}(t-1)}{1+\|\bar{\phi}(t-1)\|^2}\bar{n}(t) \right\}$$

$$\bar{y}(t) = \hat{y}(t) + \frac{\|\bar{\phi}(t-1)\|^2}{1+\|\bar{\phi}(t-1)\|^2}\bar{n}(t).$$

<u>Recall</u> $\eta(t) = y(t) - \bar{y}(t)$; $e(t) = y(t) - \hat{y}(t)$

$$\eta(t) = e(t) - \frac{\|\bar{\phi}(t-1)\|^2}{1+\|\bar{\phi}(t-1)\|^2}\bar{n}(t).$$

<u>Recall</u> $\bar{n}(t) = e(t) + [D(z^{-1})-1]\eta(t) = e(t) + \bar{\eta}(t) - \eta(t)$

$$e(t) + \bar{\eta}(t) - \bar{n}(t) = e(t) - \frac{\|\phi(t-1)\|^2}{1+\|\phi(t-1)\|^2}\bar{n}(t).$$

That is,

$$\bar{\eta}(t) = \frac{1}{1 + \|\bar{\phi}(t-1)\|^2}\,\bar{n}(t)$$

So NLMS becomes

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\bar{\phi}(t-1)}{1 + \|\bar{\phi}(t-1)\|^2}\,\bar{n}(t)$$

$$= \hat{\theta}(t-1) + \bar{\phi}(t-1)\,\bar{\eta}(t).$$

<u>Step 3:</u> Lyapunov function, $V(t) = \frac{1}{2}\|\tilde{\theta}(t)\|^2$.

$$\tilde{\theta}(t) = \tilde{\theta}(t-1) - \bar{\phi}(t-1)\,\bar{\eta}(t)$$

$$\tilde{\theta}(t) + \bar{\phi}(t-1)\,\bar{\eta}(t) = \tilde{\theta}(t-1)$$

$$V(t) + \tilde{\theta}(t)^T\bar{\phi}(t-1)\,\bar{\eta}(t) + \tfrac{1}{2}\|\bar{\phi}(t-1)\|^2\,\bar{\eta}(t)^2 = V(t-1).$$

So,

$$V(t) = V(0) - \sum_{1}^{t}\left[\bar{\phi}(k-1)^T\tilde{\theta}(k)\right]\bar{\eta}(k) - \frac{1}{2}\sum_{1}^{t}\|\bar{\phi}(k-1)\|^2\,\bar{\eta}(k)^2.$$

Now, $\bar{\eta} = \frac{D}{A}\left[\bar{\phi}^T\tilde{\theta}\right]$

$$\bar{\phi}^T\tilde{\theta} \longrightarrow \boxed{D/A} \longrightarrow \bar{\eta}$$

Assume $\frac{D}{A}$ is SPR: $\mathrm{Re}\left(\frac{D}{A}(e^{j\omega})\right) > 0$ all $\omega > 0$.

Then we can find $\varepsilon_0 > 0$ s.t.

$$0 \le V(t) \le V(0) - \varepsilon_0\sum_{1}^{t}\|\bar{\phi}(k-1)^T\tilde{\theta}(k)\|^2 - \frac{1}{2}\sum_{1}^{t}\|\bar{\phi}(k-1)\|^2\,\bar{\eta}(k)^2.$$

Consequences : $\sum_1^\infty (\bar\phi(k-1)^T \tilde\theta(k))^2 < \infty$

$$\sum_1^\infty \bar\eta(k)^2 < \infty$$

and $\sum_1^\infty \bar n(k)^2 < \infty$ if $\{\phi(k)\}$ is bounded,

since $\bar n = (1 + \|\phi\|^2)\bar\eta$.

Also, since $e(t) = \bar n(t) + \bar\eta(t) - \eta(t)$,

$\sum_1^\infty e(k)^2 < \infty$ if $\{\phi(k)\}$ bounded

See Landau, mid 1970s.

What if there is noise? We then can examine an ODE if $\{\phi(k)\}$ is w.s.s., and $E[n(k)\,\phi(k-1)] = 0$.

ODE $\quad \frac{d}{dt}\bar\theta(t) = G(\bar\theta(t))$

$$G(\theta) = E_\theta\left[\frac{\bar\phi(k-1)}{1 + \|\phi(k-1)\|^2}\bar n(k)\right]$$

$$= E_\theta\left[\bar\phi(k-1)\,\bar\eta(k)\right]$$

This ODE is a. stable via identical argument:

$V(t) = \frac{1}{2}\|\tilde\theta(t)\|^2$; $\quad \frac{d}{dt}V(t) = \tilde\theta(t)^T\dot{\bar\theta}(t)$

$$= E_{\bar\theta(t)}\left[\left(\tilde\theta(t)^T\bar\phi(k-1)\right)\bar\eta(k)\right].$$