Parameter Convergence in PAAs

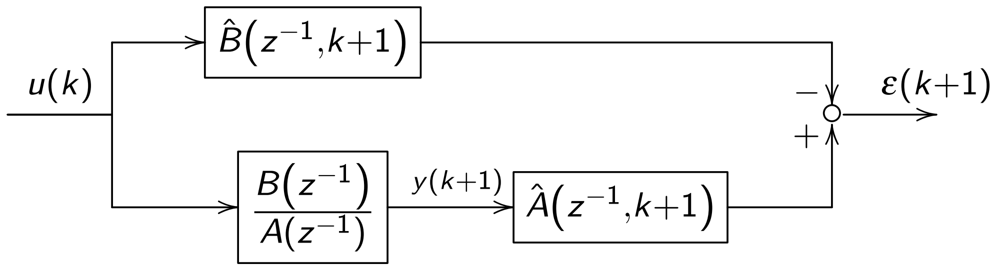
Outline

1. Big picture
2. Parameter convergence conditions
3. Effect of noise on parameter identification
4. Convergence improvement in the presence of stochastic noises
5. Effect of deterministic disturbances

Big picture

why are we learning this:

Consider a series-parallel PAA



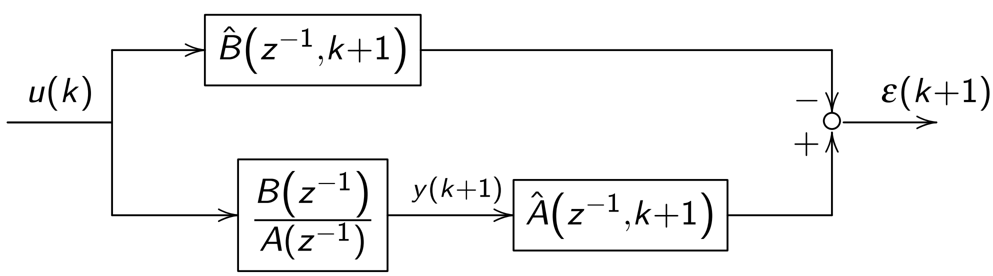
where the plant is stable.
(Hyper)stability of PAA gives

$$\lim_{k\to\infty}\varepsilon(k)=\lim_{k\to\infty}\left\{-\tilde{\theta}^T(k)\phi(k-1)\right\}=0$$

But this does not guarantee

$$\lim_{k\to\infty}\tilde{\theta}(k)=0\Longleftrightarrow\lim_{k\to\infty}\hat{\theta}(k)=\theta$$

Parameter convergence condition



$\varepsilon(k)\to 0$ means

$$\hat{A}(z^{-1},k+1)\frac{B(z^{-1})}{A(z^{-1})}u(k)-\hat{B}(z^{-1},k+1)u(k)\to 0$$

$$\Rightarrow\left[\hat{A}(z^{-1},k+1)B(z^{-1})-A(z^{-1})\hat{B}(z^{-1},k+1)\right]u(k)\to 0$$

$$\Leftrightarrow\left[\hat{A}(z^{-1})B(z^{-1})\pm A(z^{-1})B(z^{-1})-A(z^{-1})\hat{B}(z^{-1})\right]u(k)\to 0$$

$$\Leftrightarrow\left[\tilde{A}(z^{-1})B(z^{-1})-A(z^{-1})\tilde{B}(z^{-1})\right]u(k)\to 0$$

where $\tilde{A}(z^{-1})=\hat{A}(z^{-1})-A(z^{-1})$.

Parameter convergence condition

Consider

$$\overbrace{\left[\tilde{A}\left(z^{-1}\right)B\left(z^{-1}\right)-A\left(z^{-1}\right)\tilde{B}\left(z^{-1}\right)\right]}^{\text{a new polynomial } \alpha_0+\alpha_1 z^{-1}+\cdots+\alpha_{m+n}z^{-m-n}}u(k)\to 0$$

$$\tilde{B}\left(z^{-1}\right)=\tilde{b}_0+\tilde{b}_1 z^{-1}+\cdots+\tilde{b}_m z^{-m} \qquad B\left(z^{-1}\right)=b_0+b_1 z^{-1}+\cdots+b_m z^{-m}$$
$$A\left(z^{-1}\right)=1+a_1 z^{-1}+\cdots+a_n z^{-n} \qquad \tilde{A}\left(z^{-1}\right)=\tilde{a}_1 z^{-1}+\cdots+\tilde{a}_n z^{-n}$$

Two questions we are going to discuss for assuring $\tilde{\theta}=0$:

- is $\alpha_i=0$ true iff $\tilde{a}_i=0,\ \tilde{b}_i=0$ (i.e., $\{\alpha_i\}=0\Leftrightarrow\tilde{\theta}=0$)?
- if $\alpha_i\neq 0$, can $\left[\alpha_0+\alpha_1 z^{-1}+\cdots+\alpha_{m+n}z^{-m-n}\right]u(k)=0$?

Parameter convergence condition

Qs 1: $\alpha_i=0\iff\tilde{a}_i=0,\ \tilde{b}_i=0$?

Ans: yes if $B\left(z^{-1}\right)$ and $A\left(z^{-1}\right)$ are coprime (usually the case)

$$\alpha_0+\alpha_1 z^{-1}+\cdots+\alpha_{m+n}z^{-m-n}=\tilde{A}\left(z^{-1}\right)B\left(z^{-1}\right)-A\left(z^{-1}\right)\tilde{B}\left(z^{-1}\right)$$

- the right hand side is composed of terms of $\tilde{a}_i b_j$ and $a_p\tilde{b}_q$
- comparing coefficients of z^{-k} gives

$$\begin{bmatrix}\alpha_0\\ \alpha_1\\ \vdots\\ \vdots\\ \vdots\\ \alpha_{m+n}\end{bmatrix}=S\begin{bmatrix}\tilde{b}_0\\ \tilde{b}_1\\ \vdots\\ \tilde{b}_m\\ \tilde{a}_1\\ \vdots\\ \tilde{a}_n\end{bmatrix},\ S:\text{ Sylvester matrix composed of }\{-a_i,b_j\}$$

- S is non-singular if and only if $B\left(z^{-1}\right)$ and $A\left(z^{-1}\right)$ are coprime

Parameter convergence condition

Qs 2: if $\alpha_i \neq 0$, can $\left[\alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_{m+n} z^{-m-n}\right] u(k) = 0$?

Simple example with $n + m = 2$, $u(k) = \cos(\omega k) = \text{Re}\left\{e^{j\omega k}\right\}$:

$$\left[\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2}\right] u(k) \to 0$$
$$\Leftarrow \left[\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2}\right] e^{j\omega k} \to 0$$

which can be achieved either by $\alpha_0 = \alpha_1 = \alpha_2 = 0$ (the desired case) or by

$$\left(1 - e^{-j\omega} z^{-1}\right)\left(1 - e^{j\omega} z^{-1}\right) e^{j\omega k}$$
$$= \left[1 - 2\cos(\omega) z^{-1} + z^{-2}\right] e^{j\omega k} \to 0$$

Parameter convergence condition

Qs 2: if $\alpha_i \neq 0$, can $\left[\alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_{m+n} z^{-m-n}\right] u(k) = 0$?

If, however,

$$u(k) = c_1 \cos(\omega_1 k) + c_2 \cos(\omega_2 k) = \text{Re}\left\{c_1 e^{j\omega_1 k} + c_2 e^{j\omega_2 k}\right\}$$

then

$$\left[\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2}\right] u(k) \to 0$$

can only be achieved by $\alpha_0 = \alpha_1 = \alpha_2 = 0$ (the desired case).

Observations:

- complex roots of $\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2}$ always come as pairs
- impossible for $\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2}$ to have four roots at $e^{\pm j\omega_1}$ and $e^{\pm j\omega_2}$
- if the total number of parameters $n + m = 3$, $u(k)$ should contain at least $2 \left(= \frac{n+m+1}{2}\right)$ frequency components

Parameter convergence condition

general case:

$$\alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_{m+n} z^{-m-n} = 0$$

- number of the pairs of roots $= (m+n)/2$, if $m+n$ is even
- number of the pairs of roots $= (m+n-1)/2$ if $m+n$ is odd

Theorem (Persistant of excitation for PAA convergence)

For PAAs with a series-parallel predictor, the convergence

$$\lim_{k\to\infty} \hat{\theta}_i(k) = \theta_i(k)$$

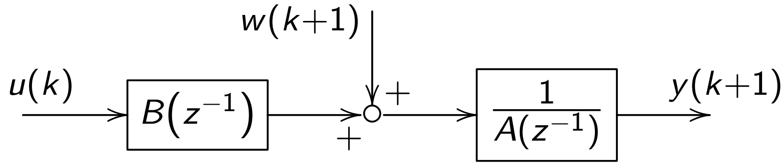
is assured if
1, the plant transfer function is irreducible
2, the input signal contains at least $1+(m+n)/2$ *(for* $n+m$ *even) or* $(m+n+1)/2$ *(for* $m+n$ *odd) independent frequency components.*

Outline

1. Big picture
2. Parameter convergence conditions
3. Effect of noise on parameter identification
4. Convergence improvement in the presence of stochastic noises
5. Effect of deterministic disturbances

Effect of noise on parameter identification

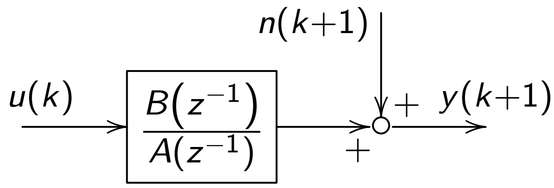
Noise modeling:



i.e.

$$A\left(z^{-1}\right)y(k+1) = B\left(z^{-1}\right)u(k) + w(k+1)$$
$$y(k+1) = \theta^T\phi(k) + w(k+1)$$

or



i.e.

$$y(k+1) = \theta^T\phi(k) + A\left(z^{-1}\right)n(k+1)$$

which is equivalent to $w(k+1) = A\left(z^{-1}\right)n(k+1)$ in the first case

Effect of noise on parameter identification

plant output: $y(k+1) = \theta^T\phi(k) + w(k+1)$

predictor output: $\hat{y}(k+1) = \hat{\theta}^T(k+1)\phi(k)$

a posteriori prediction error:

$$\varepsilon(k+1) = y(k+1) - \hat{y}(k+1) = \overbrace{-\tilde{\theta}^T(k+1)\phi(k)}^{\underline{\varepsilon}(k+1):\text{ error without noise}} + w(k+1)$$

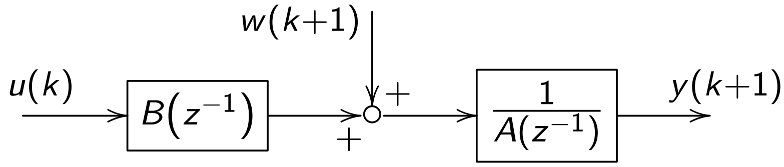
PAA: $$\hat{\theta}(k+1) = \hat{\theta}(k) + F(k)\phi(k)\varepsilon(k+1)$$
$$= \hat{\theta}(k) + F(k)\phi(k)\underline{\varepsilon}(k+1) + F(k)\phi(k)w(k+1)$$

- $F(k)\phi(k)w(k+1)$ is integrated by PAA
- need: $\boxed{\mathrm{E}[\phi(k)w(k+1)] = 0}$

and a vanishing adaptation gain $F(k)$:

$F^{-1}(k+1) = \lambda_1(k)F^{-1}(k) + \lambda_2(k)\phi(k)\phi^T(k)$, $\lambda_1(k) \xrightarrow{k\to\infty} 1$ and $0 < \lambda_2(k) < 2$

Series-parallel PAA convergence condition



$$\hat{\theta}(k+1) = \hat{\theta}(k) + F(k)\phi(k)\underline{\varepsilon}(k+1) + F(k)\phi(k)w(k+1)$$

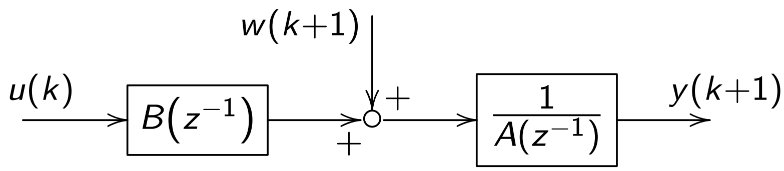
In series-parallel PAA:

$$\phi(k) = [-y(k), -y(k-1), \ldots, -y(k-n+1), \\ u(k), u(k-1), \ldots, u(k-m)]^T$$

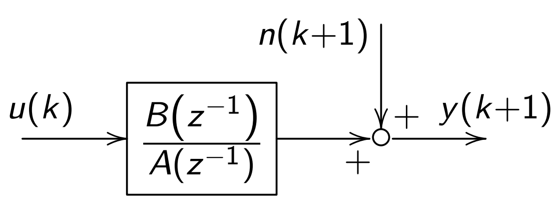
$\boxed{\mathrm{E}[\phi(k)w(k+1)] = 0}$ is achieved if

- $w(k+1)$ is a white noise, and
- $u(k)$ and $w(k+1)$ are independent

Series-parallel PAA convergence condition



Issues: $w(k+1)$ is rarely white, e.g.,



where the output measurement noise $n(k+1)$ is usually white but

$$y(k+1) = \theta^T\phi(k) + \overbrace{A(z^{-1})\,n(k+1)}^{w(k+1)}$$

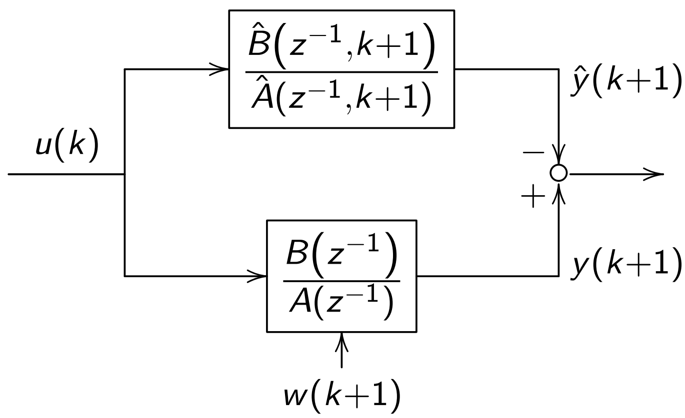
so $w(k+1)$ is not white.

Parallel PAA convergence condition

In parallel PAA:

$$\phi(k) = [-\hat{y}(k), -\hat{y}(k-1), \ldots, -\hat{y}(k-n+1), \\ u(k), u(k-1), \ldots, u(k-m)]^T$$

$\boxed{E[\phi(k)w(k+1)] = 0}$ does not require $w(k+1)$ to be white as $\hat{y}(k)$ does not depend on $w(k+1)$ by design



Summary

Theorem (Series-parallel PAA convergence condition)

When the predictor is of series-parallel type, the PAA with a vanishing adaptation gain has unbiased convergence when
i. $u(k)$ *is rich in frequency (persistent excitation) and is independent from the noise* $w(k+1)$
ii. $w(k+1)$ *is white*

Theorem (Parallel PAA convergence condition)

When the predictor is of parallel type, the PAA with vanishing adaptation gain has unbiased convergence when
i. $u(k)$ *satisfies the persistent excitation condition*
ii. $u(k)$ *is independent from* $w(k+1)$

Note: parallel predictors have more strict stability requirements

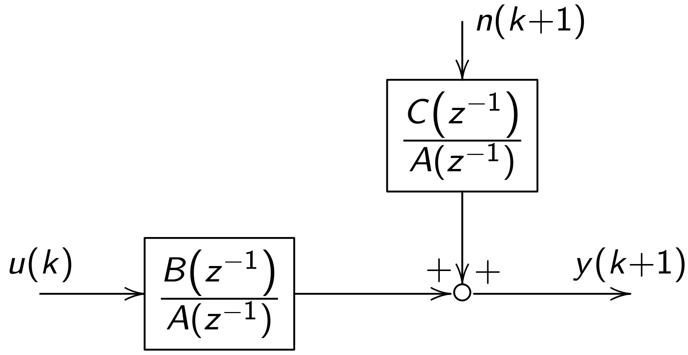
Outline

1. Big picture
2. Parameter convergence conditions
3. Effect of noise on parameter identification
4. Convergence improvement in the presence of stochastic noises
5. Effect of deterministic disturbances

Convergence improvement when there is noise

extended least squares

If the effect of noise can be expressed as



i.e. $w(k+1) = C(z^{-1})n(k+1) = \left[1 + c_1 z^{-1} + \ldots c_{n_C} z^{-n_C}\right] n(k+1)$

where $n(k+1)$ is white, then

$$y(k+1) = \theta^T \phi(k) + C\left(z^{-1}\right) n(k+1) = \theta_e^T \phi_e(k) + n(k+1)$$

$$\theta_e^T = \left[\theta^T, c_1, \ldots, c_{n_C}\right]$$

$$\phi_e^T(k) = \left[\phi^T(k), n(k), \ldots, n(k - n_C + 1)\right]$$

Convergence improvement when there is noise

extended least squares

a posteriori prediction

$$\hat{y}(k+1) = \hat{\theta}_e^T(k+1)\phi_e(k)$$
$$\phi_e^T(k) = \left[\phi^T(k), n(k), \dots, n(k-n_C+1)\right]$$

but $n(k), \dots, n(k-n_C+1)$ are not measurable. However, if $\hat{\theta}_e$ is close to θ_e, then

$$\varepsilon(k+1) = y(k+1) - \hat{y}(k+1) \approx n(k+1)$$

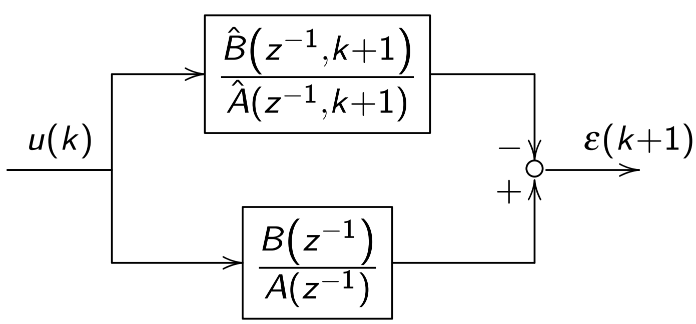
extended least squares uses

$$\hat{y}(k+1) = \hat{\theta}_e^T(k+1)\phi_e^*(k)$$
$$\phi_e^*(k) = \left[\phi^T(k), \varepsilon(k), \dots, \varepsilon(k-n_C+1)\right]^T$$

where $\varepsilon(k) = y(k) - \hat{y}(k)$

Convergence improvement when there is noise

output error method with adjustable compensator



If $A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_n z^{-n}$, let $\hat{C}(z^{-1}) = 1 + \hat{c}_1 z^{-1} + \cdots + \hat{c}_n z^{-n}$ and

$$\nu(k+1) = \hat{C}\left(z^{-1}, k+1\right)\varepsilon(k+1)$$
$$\nu^o(k+1) = \varepsilon^o(k+1) + \sum_{i=1}^{n} \hat{c}_i(k)\varepsilon(k+1-i)$$

construct PAA with $\theta_e^T = \left[\theta^T, a_1, \dots, a_n\right]$ and $\nu(k+1)$ as the adaptation error.

Convergence improvement when there is noise

output error method with adjustable compensator

$$\hat{\theta}_e(k+1) = \hat{\theta}_e(k) + \frac{F_e(k)\phi_e(k)}{1+\phi_e^T(k)F_e(k)\phi_e(k)}\nu^o(k+1)$$

$$\hat{\theta}_e^T(k) = \left[\hat{\theta}^T(k), \hat{c}_1(k), \ldots, \hat{c}_n(k)\right]$$

$$\phi_e^T(k) = \left[\phi^T(k), -\varepsilon(k), \ldots, -\varepsilon(k+1-n)\right]$$

$$F_e^{-1}(k+1) = \lambda_1(k)F_e^{-1}(k) + \lambda_2(k)\phi_e(k)\phi_e^T(k)$$

Stability condition:

$$1 - \frac{\lambda}{2} \text{ is SPR; } \lambda = \max_k \lambda_2(k) < 2$$

Different recursive identification algorithms

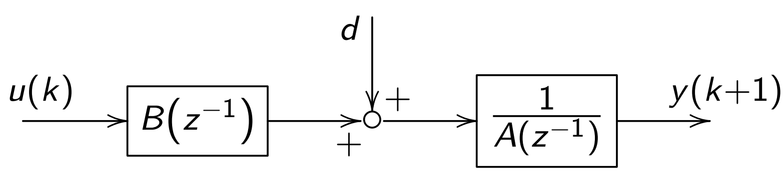
- there are more PAAs for improved convergence
- each algorithm suits for a certain model of plant + disturbance

Outline

1. Big picture
2. Parameter convergence conditions
3. Effect of noise on parameter identification
4. Convergence improvement in the presence of stochastic noises
5. Effect of deterministic disturbances

Effect of deterministic disturbances

Intuition: if the disturbance structure is known, it can be included in PAA for improved performance.
Example (constant disturbance):



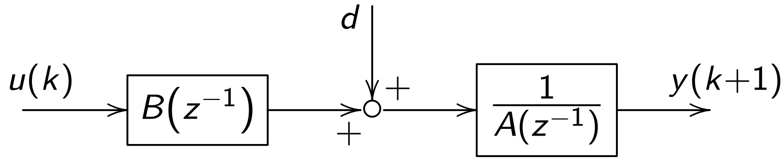
$$y(k+1) = -\sum_{i=1}^{n} a_i y(k+1-i) + \sum_{i=0}^{m} b_i u(k-i) + d = \theta^T \phi(k) + d$$

Approach 1: enlarge the model as

$$y(k+1) = \left[\theta^T, d\right] \begin{bmatrix} \phi(k) \\ 1 \end{bmatrix} = \theta_e^T \phi_e(k)$$

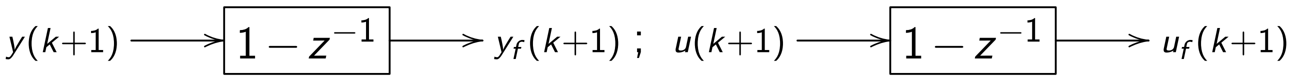
and construct PAA on θ_e.

Effect of deterministic disturbances



$$y(k+1) = -\sum_{i=1}^{n} a_i y(k+1-i) + \sum_{i=0}^{m} b_i u(k-i) + d = \theta^T \phi(k) + d$$

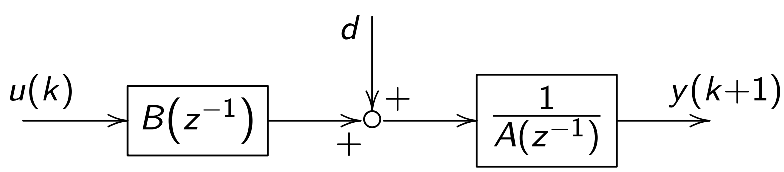
Approach 2: notice that $\left(1-z^{-1}\right) d = 0$, we can do



and have a new "disturbance-free" model for PAA:

$$y_f(k+1) = -\sum_{i=1}^{n} a_i y_f(k+1-i) + \sum_{i=0}^{m} b_i u_f(k-i)$$

Effect of deterministic disturbances



Similar considerations can be applied to the cases when d is sinusoidal, repetitive, etc