3. IDENTIFICATION AND ADAPTIVE ESTIMATION

3.1 Identification problem

We consider a single input single output system (SISO) with following assumptions:

A1) PLANT. The plant is a SISO system described by the transfer function:

$$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \hat{P}(s) = k_p \frac{\hat{n}_p(s)}{\hat{d}_p(s)}$$

where $\hat{r}(s)$ and $\hat{y}_p(s)$ are the Laplace transforms of the input $r(t)$ and output $y_p(t)$ of the plant, $\hat{n}_p(s)$

and $\hat{d}_p(s)$ are monic (i.e. coefficient of highest power in s is 1), co-prime (i.e. no common roots) polynomials of degree m and n respectively, m is unknown but $m \leq n-1$ (i.e. $\hat{P}(s)$ is strictly proper).

A2) Reference input. The input $r(t)$ is piecewise continuous and bounded on $\mathbb{R}_+$.

Problem Estimate (or identify) the parameters k_p and the coefficient of $\hat{n}_p(s)$ and $\hat{d}_p(s)$ from the measurements of $r(t)$ and $y_p(t)$ only.

3.2 Parameter identification

The identifier structure presented here is known as "equation error identifier".

First, write $\hat{P}(s)$ as

$$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \hat{P}(s) = \frac{\alpha_n s^{n-1} + \dots + \alpha_1}{s^n + \beta_n s^{n-1} + \dots + \beta_1} \qquad (I)$$

with the $2n$ coefficients $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_n$ are unknown.

Introduce a monic n-th order polynomial denoted

$$\hat{\lambda}(s) = s^n + \lambda_n s^{n-1} + \dots + \lambda_1$$

assumed to be Hurwitz : i.e. the roots are all in $\mathbb{C}^-$.

From the definition of $\hat{P}(s)$ and (I)

$$\hat{\lambda}(s)\hat{y}_p(s) = k_p \hat{n}_p(s)\hat{r}(s) + (\hat{\lambda}(s) - \hat{d}_p(s))\hat{y}_p(s)$$

and then

$$\hat{y}_p(s) = \frac{\alpha_n s^{n-1} + \dots + \alpha_1}{\hat{\lambda}(s)}\hat{r}(s) + \frac{(\lambda_n - \beta_n)s^{n-1} + \dots + (\lambda_1 - \beta_1)}{\hat{\lambda}(s)}\hat{y}_p(s). \qquad (II)$$

Now define

$$\hat{a}^*(s) \triangleq \alpha_n s^{n-1} + \dots + \alpha_1 = k_p \hat{n}_p(s)$$

$$\hat{b}^*(s) \triangleq (\lambda_n - \beta_n)s^{n-1} + \dots + (\lambda_1 - \beta_1) = \hat{\lambda}(s) - \hat{d}_p(s)$$

From (II)

$$\hat{y}_p(s) = \frac{\hat{a}^*(s)}{\hat{\lambda}(s)}\hat{r}(s) + \frac{\hat{b}^*(s)}{\hat{\lambda}(s)}\hat{y}_p(s)$$

and $$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \frac{\hat{a}^*(s)}{\hat{l}(s) - \hat{b}^*(s)} \qquad \text{(III)}$$

Moreover $\hat{a}^*(s)$ and $\hat{b}^*(s)$ are the unique polynomials for which

$$\hat{P}(s) = \frac{\hat{a}^*(s)}{\hat{l}(s) - \hat{b}^*(s)} = \frac{\hat{y}_p(s)}{\hat{r}(s)}$$

when $\hat{n}_p(s)$ and $\hat{d}_p(s)$ are coprime.

Indeed if there exist $\hat{a}^{*\prime}(s) = \hat{a}^*(s) + \delta\hat{a}(s)$ and $\hat{b}^{*\prime}(s) = \hat{b}^*(s) + \delta\hat{b}(s)$ such that

$$\hat{P}(s) = \frac{\hat{a}^{*\prime}(s)}{\hat{l}(s) - \hat{b}^{*\prime}(s)}$$

then

$$\frac{\delta\hat{a}(s)}{\delta\hat{b}(s)} = \underbrace{-K_p\frac{\hat{n}_p(s)}{\hat{d}_p(s)}}_{-\frac{\hat{a}^*}{\hat{l}(s) - \hat{b}^*(s)}} = -\hat{P}(s) \qquad \text{(IV)}$$

But (IV) has no solution since the degree of $\hat{d}_p$ is n, $\hat{n}_p, \hat{d}_p$ are coprime and the degree of $\hat{\delta b}$ is $\leq n-1$.

It is possible to give a state space representation of (III) in controllable canonical form:

$$y_p(t) = a^{*T} w_p^{(1)}(t) + b^{*T} \hat{w}_p^{(2)}(t) \quad (V)$$

where $a^* = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$, $b^* = \begin{pmatrix} \lambda_1 - \beta_1 \\ \vdots \\ \lambda_n - \beta_n \end{pmatrix}$

and

$$\begin{cases} \mathring{w}_p^{(1)} = \Lambda w_p^{(1)} + b_\lambda z \\ \mathring{w}_p^{(2)} = \Lambda w_p^{(2)} + b_\lambda y_p \end{cases}$$

where $\Lambda \triangleq \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \\ -\lambda_1 & -\lambda_2 & -\lambda_3 & \cdots & -\lambda_n \end{pmatrix}$, $b_\lambda \triangleq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$

Indeed, in Laplace domain

$$(sI-\Lambda)^{-1}b_\lambda = \frac{1}{\hat{\lambda}(s)}\begin{pmatrix} 1 \\ s \\ \vdots \\ s^{n-1} \end{pmatrix}$$

and

$$\hat{w}_p^{(1)}(s) = (sI-\Lambda)^{-1}b_\lambda \hat{r}(s) + (sI-\Lambda)^{-1}w_p^{(1)}(0)$$

$$\hat{w}_p^{(2)}(s) = (sI-\Lambda)^{-1}b_\lambda \hat{y}_p(s) + (sI-\Lambda)^{-1}w_p^{(2)}(0)$$

from which, under zero initial conditions,

$$\hat{y}_p(s) = a^{*T}\hat{w}_p^{(1)}(s) + b^{*T}\hat{w}_p^{(2)}(s)$$

$$= \frac{\hat{a}^*(s)}{\hat{\lambda}(s)}\hat{r}(s) + \frac{\hat{b}^*(s)}{\hat{\lambda}(s)}\hat{y}_p(s)$$

i.e. ($\underline{\underline{III}}$).

By setting $\theta^* \triangleq \begin{pmatrix} a^* \\ b^* \end{pmatrix} \in \mathbb{R}^{2n}$

$$w_p(t) = \begin{pmatrix} w_p^{(1)}(t) \\ w_p^{(2)}(t) \end{pmatrix} \in \mathbb{R}^{2n}$$

$$\Rightarrow \boxed{y_p(t) = \theta^{*T} w_p(t)} \qquad (\overline{\underline{VI}})$$

3.2.1 Identifier structure

Define the estimator:

$$\dot{w}^{(1)} = \Lambda w^{(1)} + b_\lambda r$$

$$\dot{w}^{(2)} = \Lambda w^{(2)} + b_\lambda y_p$$

to reconstruct the states $w_p^{(1)}, w_p^{(2)}$.

Define $w(t) \triangleq \begin{pmatrix} w^{(1)}(t) \\ w^{(2)}(t) \end{pmatrix}$.

From ($\underline{\overline{VI}}$) we get

$$\boxed{\begin{aligned} y_p(t) &= \Theta^{*} w(t) + \varepsilon(t) \\ \text{where } \varepsilon(t) &\triangleq \Theta^{*T}(w_p(t) - w(t)) \end{aligned}}$$

The term $\varepsilon(t)$ is exponentially decaying since

$$\dot{\overbrace{w_p(t) - w(t)}} = \Lambda (w_p(t) - w(t))$$

and Λ is Hurwitz. The vector $w(t)$ is the <u>regressor vector</u>.

The output of the identifier $\theta(t)$ of θ^* is

$$y_I(t) \triangleq \theta^T(t)\, w(t)$$

where $\theta(t)$ is designed later according to different strategies.

The <u>parameter error</u> is defined as

$$\phi(t) \triangleq \theta(t) - \theta^*$$

and the <u>identifier error</u> as

$$e_I(t) \triangleq y_I(t) - y_p(t) = \phi^T(t)\, w(t) - \varepsilon(t)$$

N.B. In what follows we ignore $\varepsilon(t)$! ← exponentially decaying term !

$\boxed{\text{Identifier type I : gradient algorithm}}$

The update law for the identifier in this first case is chosen as

$$\dot{\theta} = -g\, e_I\, w \,, \quad g > 0 \qquad (\overline{\text{VII}})$$

g is called ADAPTATION GAIN

76

The right-hand term is proportional to

$$\frac{\partial}{\partial \theta}\left(e_I^2(\theta)\right) = 2e_I(\theta)\,w$$

where $e_I(\theta)$ is seen as a function of the parameter vector θ. Therefore the update law (VII) can be seen as a steepest descent method.

$\boxed{\text{Identifier type II : normalized gradient}}$

Another similar update law is the normalized gradient :

$$\dot{\theta} = -g\,\frac{e_I w}{1+\gamma w^T w} \qquad g, \gamma > 0$$

the regressor is normalised. The advantage is that even if w is unbounded $\dot{\theta}$ is bounded w.r.t. w:

$$\dot{\theta} = -g\boxed{\frac{w w^T}{1+\gamma w^T w}}\,\phi$$

Identifier type III : normalised gradient with projection

If the nominal parameter θ^* is known a priori to lie in a set $\mathcal{X} \subset \mathbb{R}^{2n}$ (closed, convex and with smooth boundary) we choose:

$$\dot{\theta} = \begin{cases} -g \dfrac{e_I W}{1+\gamma W^T W} & \theta \in \text{int}(\mathcal{X}) \\ P_z\left[-g \dfrac{e_I W}{1+\gamma W^T W}\right] & \text{if } \theta \in \partial\mathcal{X} \text{ and } e_I W^T \mathcal{X}_{perp} < 0 \end{cases}$$

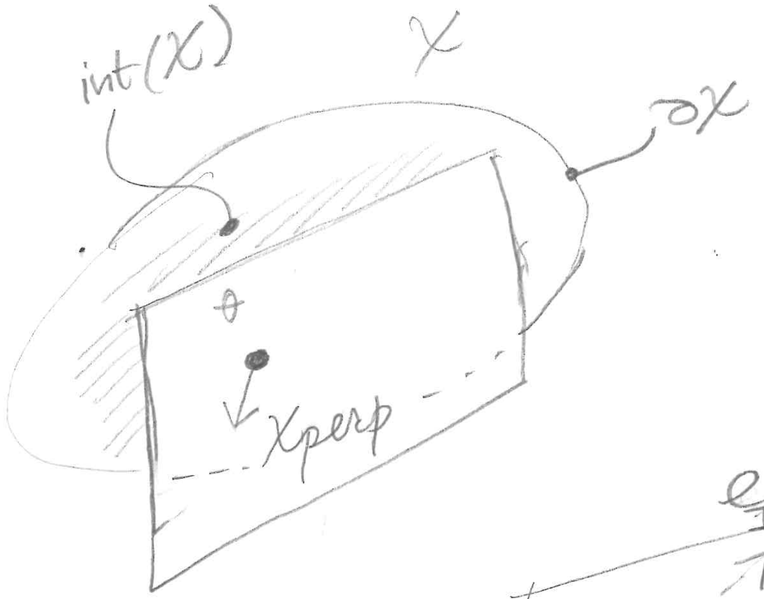
where

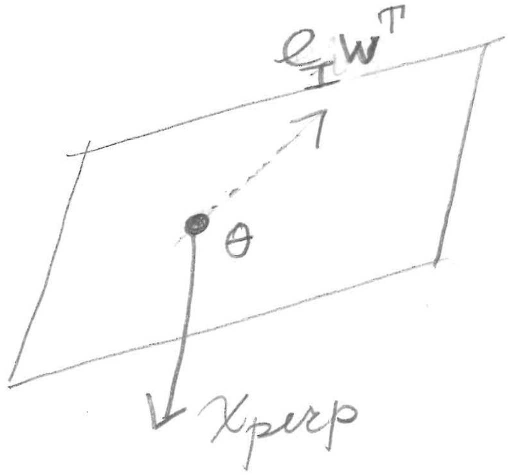
$\partial\mathcal{X} \triangleq$ boundary of $\mathcal{X}$

$\text{int}(\mathcal{X}) \triangleq$ interior of $\mathcal{X}$

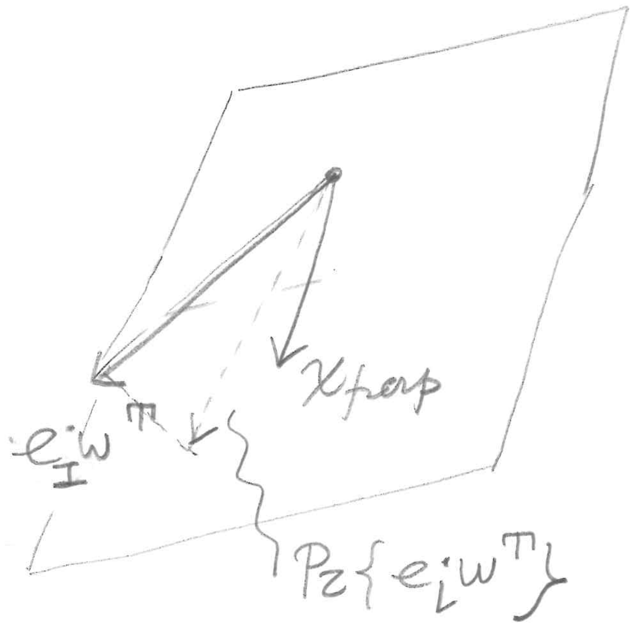
$P_z(z) \triangleq$ projection of z onto the hyperplane tangent to $\partial\mathcal{X}$ at θ

$\mathcal{X}_{perp} \triangleq$ the unit vector perpendicular to the hyperplane (pointing outward)





a)



b) :

If bounds p_i^-, p_i^+ are known for each θ_i^* the update law can be modified at the boundary as follows:

$$\dot{\theta}_i = 0 \quad \text{if} \quad \theta_i = p_i^- \ \& \ \dot{\theta}_i < 0$$

$$\text{or} \quad \theta_i = p_i^+ \ \& \ \dot{\theta}_i > 0 .$$

Identifier type IV: Least squares

The parameter θ^* can be considered to be the unknown state of the system:

$$\dot{\theta}^* = 0 \qquad (VIII)$$

with output

$$y_p(t) = w^T(t)\,\theta^*(t) \qquad (IX)$$

Assuming (VIII)-(IX) are perturbed by zero mean white gaussian noises of spectral intensity $Q \in \mathbb{R}^{2n \times 2n}$ and $1/g \in \mathbb{R}$, respectively, the least-squares estimator (or Kalman-Bucy filter) is

$$\dot{\theta} = -g P w e_I$$

$$\dot{P} = Q - g P w w^T P, \quad P(0) > 0.$$

Q and g are design parameters. The update for θ is similar to the gradient update law.

The parameter θ^* is assumed to be constant so that usually $Q = 0$:

$$\mathring{P} = -gPww^TP$$

or

$$\mathring{P}^{-1} = gww^T$$

(since $\mathring{\widehat{P^{-1}}} = -P^{-1}\mathring{P}P^{-1}$). The last equation implies that P^{-1} may grow without bound and P becomes arbitrarily small so that the adaptation also becomes very slow.

This problem is usually prevented with a "forgetting factor":

$$\mathring{P} = -g(-\lambda P + Pw^Tw P)$$

$$\mathring{\widehat{P^{-1}}} = -g(\lambda P^{-1} + ww^T)$$

The normalized least squares identifier is defined as:

$$\dot{\theta} = -g \frac{P w e_1}{1 + \gamma w^T P w}, \quad g, \gamma > 0$$

$$\dot{P} = -g \frac{P w w^T P}{1 + \gamma w^T P w}$$