Final Version OCT 13 2005

# Lecture 4

We discuss deterministic identification algorithms and associated convergence properties in a style similar to our discussion of MRAS. We begin with a basic example.

<u>Example 1.</u> Consider the plant $P(s) = \dfrac{k_p}{s+a_p}$ with unknown parameters $k_p$ and ~~$a_p$~~ $a_p$. From the equation

$$\dot{y}_p = -a_p y_p + k_p r$$

it follows that measurements at two distinct times $t_1$ and $t_2$ determine the parameters by the formula,

$$\begin{pmatrix} -a_p \\ k_p \end{pmatrix} = \begin{pmatrix} y_p(t_1) & r(t_1) \\ y_p(t_2) & r(t_2) \end{pmatrix}^{-1} \begin{pmatrix} \dot{y}_p(t_1) \\ \dot{y}_p(t_2) \end{pmatrix}$$

<u>provided the inverse exists</u>. Also, one needs derivatives of plant output. To avoid such derivative measurement use <u>filtering</u>. Let $(s+a_p)\,\hat{y}_p(s) = k_p\,\hat{r}(s)$.

Then
$$\frac{(s+a_p)}{(s+\lambda)}\hat{y}_p(s) = \frac{k_p}{(s+\lambda)}\hat{r}(s)$$

$$\Leftrightarrow \quad \frac{s+\lambda-(\lambda-a_p)}{s+\lambda}\hat{y}_p(s) = \frac{k_p}{s+\lambda}\hat{r}(s)$$

$$\Leftrightarrow \quad \hat{y}_p(s) = \frac{\lambda-a_p}{s+\lambda}\hat{y}_p(s) + \frac{k_p}{s+\lambda}\hat{r}(s)$$

Let $\hat{w}^{(1)}(s) = \frac{1}{s+\lambda}\hat{r}(s)$ ; $\hat{w}^{(2)}(s) = \frac{1}{s+\lambda}\hat{y}_p(s)$. Then in the time domain,

$$y_p(t) = k_p w^{(1)}(t) + (\lambda - a_p) w^{(2)}(t).$$

Measurements at two time instants $t_1$ and $t_2$ suffice to yield

$$\begin{bmatrix} k_p \\ (\lambda - a_p) \end{bmatrix} = \begin{bmatrix} w^{(1)}(t_1) & w^{(2)}(t_1) \\ w^{(1)}(t_2) & w^{(2)}(t_2) \end{bmatrix}^{-1} \begin{bmatrix} y_p(t_1) \\ y_p(t_2) \end{bmatrix}$$

assuming the inverse exists. No derivatives are necessary!

Now the signals $w^{(1)}(t)$ and $w^{(2)}(t)$ are obtained as outputs of filters, driven by $r$ and $y_p$. Effects of initial conditions on $w^{(1)}$ and $w^{(2)}$ can be made irrelevant by making these stable filters, i.e. choosing $\lambda > 0$. Since $\lambda$ is known, $\lambda - a_p$ and $k_p$ determine the plant. It is of course desirable to use <u>history of data</u> if available, and not just isolated samples.

Let $\theta^* \triangleq (k_p \;\; \lambda - a_p)^T$ be the nominal. Let $\theta(t)$ be a running 'guess' of $\theta^*$. Then

$$e_1(t) = \theta^T(t)\, w(t) - y_p(t) = (\theta^T(t) - \theta^*)\, w(t)$$

is the <u>identification error</u>.

An identification algorithm is a rule to update the guess $\theta(t)$, and hence a differential equation for $\theta$.

The <u>gradient rule</u> for $\theta$ is

$$\dot{\theta} = -\gamma \frac{\partial}{\partial \theta}(e_1^2/2)$$

$$= -\gamma e_1 \frac{\partial e_1}{\partial \theta} \qquad \gamma > 0.$$

$$= -\gamma e_1 w$$

and is driven by the filter outputs and the identification error.

We can rewrite,

$$\dot{\theta} = -\gamma(\theta^T w - y_p) w$$

$$= -\gamma w w^T \theta + \gamma y_p w.$$

and, letting $\phi(t) \triangleq \theta(t) - \theta^* =$ <u>parameter error</u>

$$\dot{\phi} = \dot{\theta} = -\gamma \phi^T w \; w$$

$$= -\gamma w w^T \phi$$

This last equation makes clear that

$$\dot{\phi} = -\gamma \frac{\partial}{\partial \phi} \frac{1}{2}\left(\phi^T (w w^T) \phi\right)$$

a <u>degenerate descent equation</u>, since the quadratic form $\frac{1}{2}\phi^T(t)\, w(t) w^T(t)\, \phi(t)$ is of rank 1.

The above **gradient** rule makes **descent** depend on an instantaneous output error. An alternative approach is based on the integral squared error,

$$e_2(t) = \int_0^t \frac{1}{2} \left( \theta^T(t)\, w(\tau) - y_p(\tau) \right)^2 d\tau .$$

Here the argument of $\theta$ in the integrand is $t$ not $\tau$, signifying the guess at time $t$ (based possibly on all the data upto time $t$). Setting

$$\frac{\partial}{\partial \theta(t)} e_2(t) = \left( \int_0^t w(\tau) w^T(\tau)\, d\tau \right) \theta(t) - \left( \int_0^t y_p(\tau) w(\tau)\, d\tau \right)$$

$$= 0$$

we get the least squares estimate,

$$\theta_{LS}(t) = \left( \int_0^t w(\tau) w^T(\tau)\, d\tau \right)^{-1} \left( \int_0^t y_p(\tau) w(\tau)\, d\tau \right)$$

assuming that the inverse exists. To compare with the previous gradient equation we will derive a differential equation for $\theta_{LS}$. First, let,

$$P(t) \triangleq \left( \int_0^t w(\tau) w^T(\tau)\, d\tau \right)^{-1}.$$

$$\frac{d}{dt} P^{-1}(t) = w(t) w^T(t).$$

Also $$0 = \frac{d}{dt}\left(P(t)\, P^{-1}(t)\right)$$

$$= \frac{dP}{dt} P^{-1} + P \frac{d}{dt}\left(P^{-1}\right)$$

$$\Rightarrow \quad \frac{d\,P(t)}{dt} = - P(t)\; w(t)\, w^T(t)\; P(t) \,.$$

Hence $$\theta_{LS}(t) = P(t) \int_0^t w(\tau)\, y_p(\tau)\, d\tau \,.$$

satisfies

$$\dot{\theta}_{LS}(t) = \dot{P} \int_0^t \; + \; P \cdot \left(w(t)\, y_p(t)\right)$$

$$= -P(t)\; w(t)\, w^T(t) \left(P(t) \int_0^t w(\tau)\, y_p(\tau)\, d\tau\right)$$

$$+ P(t)\; w(t)\, y_p(t)$$

$$= -P(t)\; w(t)\, w^T(t)\, \theta_{LS}(t) \; + P(t)\, w(t)\, y_p(t)$$

$$= -P(t)\, w(t) \left[ w^T(t)\, \theta_{LS}(t) - y_p(t) \right]$$

$$= -P(t)\, w(t)\; e_1(t)$$

$$\Rightarrow \quad \dot{\phi}_{LS} = - P(t) \frac{\partial}{\partial \phi} \left( \frac{1}{2}\, \phi_{LS}^T(t)\, w(t)\, w^T(t)\, \phi_{LS}(t) \right)$$

again a degenerate descent equation

Note that for the (differential) equations for $\dot{P}$ and $\dot{\theta}_{LS}$ to exactly capture, ~~$\theta$~~ $P$ and $\theta_{LS}$, they have to be initialized ~~so~~ for some $t_0 > 0$ such that,

$$P(t_0) = \left( \int_0^{t_0} w(\sigma) w^T(\sigma) \, d\sigma \right)^{-1} \quad \text{exists}$$

and

$$\theta_{LS}(t_0) = P(t_0) \cdot \int_0^{t_0} w(\sigma) y_p(\sigma) \, d\sigma \, .$$

In practice one would initialize at $t=0$ with arbitrary initial conditions and verify that the effect of the initial conditions decays (rapidly) ~~as~~ as $t \to \infty$.

Summarizing the integral criterion yields.

$$\dot{\theta}(t) = -P(t) \, w(t) \left( \theta^T(t) w(t) - y_p(t) \right)$$

$$\theta(0) = \theta_0$$

$$\dot{P}(t) = -P(t) \, w(t) w^T(t) P(t)$$

$$P(0) = P_0 > 0 \, . \qquad P_0 = P_0^T$$

with solutions $P(t) = \left( P_0^{-1} + \int_0^t w(\sigma) w^T(\sigma) \, d\sigma \right)^{-1}$ and

$$\theta(t) = \left( P_0^{-1} + \int_0^t w(\sigma) w^T(\sigma) \, d\sigma \right)^{-1} \left( P_0^{-1} \theta_0 + \int_0^t w(\sigma) y_p(\sigma) \, d\sigma \right) .$$

proof of formula for $\theta(t)$:

$$\dot\theta = -P\,w\,\theta^T w + P\,w\,y_p$$

$$= -P\,w w^T \theta + P\,w\,y_p$$

$$= -P\,\dot{\overline{P^{-1}}}\,\theta + P\,w\,y_p$$

$$\Rightarrow \quad P^{-1}\dot\theta + \dot{\overline{P^{-1}}}\,\theta = +\,w\,y_p$$

$$\Rightarrow \quad \frac{d}{dt}\left(P^{-1}\theta\right) = +\,w\,y_p$$

$$\Rightarrow \quad P^{-1}(t)\,\theta(t) = P^{-1}(0)\,\theta(0) + \int_0^t w(\sigma)y(\sigma)\,d\sigma$$

$$\Rightarrow \quad \theta(t) = P(t)\left[P^{-1}(0)\,\theta(0) + \int_0^t w(\sigma)y(\sigma)\,d\sigma\right]$$

$$= \left(P_0^{-1} + \int_0^t w(\sigma)w^T(\sigma)\,d\sigma\right)^{-1}\left(P_0^{-1}\theta_0 + \int_0^t w(\sigma)y(\sigma)\,d\sigma\right)$$

QED ▨

Parameter Error

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

$$= \left(P_0^{-1} + \int_0^t w(\sigma)w^T(\sigma)\,d\sigma\right)^{-1} \cdot P_0^{-1}\underbrace{\left(\theta_0 - \theta^*\right)}_{\phi(0)}$$

$$\phi(t) \to 0 \quad \text{iff} \quad \int_0^t w(\sigma)w^T(\sigma)\,d\sigma \to \infty \quad \text{as} \quad t \to \infty$$ ▨

| INSTANTANEOUS | INTEGRAL |
|---|---|
| $\dot\theta = -\gamma\, e_1\, w$ | $\dot\theta = -P\,w\,e_1$ |
| $e_1 = \theta^T w - y_p$ | $\dot P = -P\,w w^T P$ |
| $\gamma > 0$ | $P(0) = P^T(0) \geq P_0 > 0$ |

Table of Parameter Update Laws

Generalizing Example 1 to obtain an identifier.

Let $P(s) = \dfrac{n_p(s)}{d_p(s)}$ be a scalar plant,

with $d_p(s) = s^n + \beta_n s^{n-1} + \cdots + \beta_1$

$n_p(s) = \alpha_n s^{n-1} + \cdots + \alpha_1$ ,

and $(n_p(s), d_p(s)) \equiv 1$ (co-primeness conditi-

Assume that reference inputs $r(t)$ can be chosen such that $r$ is bounded and piecewise continuous on $\mathbb{R}_+$, $\alpha_i$, $\beta_i$ all unknown, to be identified.

**Let** $\lambda(s)$ be a Hurwitz, monic polynomial of degree $n$ $= s^n + \lambda_n' s^{n-1} + \cdots + \lambda_1$. From the plant model,

$$d_p(s)\,\hat{y}_p(s) = n_p(s)\,r(s)$$

$$\frac{d_p(s)}{\lambda(s)}\,\hat{y}_p(s) = \frac{n_p(s)}{\lambda(s)}\,r(s)$$

$$\frac{\lambda(s) - (\lambda(s) - d_p(s))}{\lambda(s)}\,\hat{y}_p(s) = \frac{n_p(s)}{\lambda(s)}\,r(s)$$

$$\hat{y}_p(s) = \frac{\lambda(s) - d_p(s)}{\lambda(s)}\,\hat{y}_p(s) + \frac{n_p(s)}{\lambda(s)}\,r(s)$$

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

Here $\deg b^*(s) \leq n-1$ , $\deg a^*(s) \leq n-1$

Let $$\frac{b^*(s)}{\lambda(s)} = b^{*T}(s\mathbb{1} - \Lambda)^{-1} b_\lambda$$

$$\frac{a^*(s)}{\lambda(s)} = a^{*T}(s\mathbb{1} - \Lambda)^{-1} b_\lambda$$

where $a^* = (\alpha_1, \ldots, \alpha_n)^T$; $b^* = (\lambda_1 - \beta_1, \ldots, \lambda_n - \beta_n)^T$

$$\Lambda = \begin{pmatrix} 0 & 1 & \cdot & 0 \\ & 0 & \cdot & 1 \\ -\lambda_1 & \cdots & & -\lambda_n \end{pmatrix} \; ; \quad b_\lambda = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

We have the following controllable but possibly unobservable triples $[\Lambda, b_\lambda, b^{*T}]$, $[\Lambda, b_\lambda, a^{*T}]$, with associated state variables $w^{(2)}$ and $w^{(1)}$. Letting

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

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

we get

$$y_p(t) = a^{*T} w_p^{(1)}(t) + b^{*T} w_p^{(2)}(t)$$

$$= \theta^{*T} w_p(t)$$

where $\theta^{*T} = (a^{*T}, b^{*T}) \quad \in \mathbb{R}^{2n}$

$$w_p^T(t) = (w_p^{(1)T}, w_p^{(2)T}) \quad \in \mathbb{R}^{2n}$$

We have just written a <u>2n dimensional state</u> space realization of a degree $n$ plant $P(s)$. It thus necessarily nonminimal.

But the unobservable **modes** are those of $\lambda$ and **hence** stable.

## Identifier Structure

Define

$$(*) \quad \begin{aligned} \dot{w}^{(1)} &= \Lambda w^{(1)} + b_\lambda r \\ \dot{w}^{(2)} &= \Lambda w^{(2)} + b_\lambda y_p \end{aligned}$$

We will make 2 claims.

(a) $(*)$ is an observer for the above nonminimal realization (without knowing $\theta^*$!),

(b) for suitable hypothesis on $r(\cdot)$ and update law on $\theta$, parameter error converges.

**Claim (a)**

$$\dot{\overline{w^{(1)} - w_p^{(1)}}} = \Lambda \left(w^{(1)} - w_p^{(1)}\right)$$

$$\dot{\overline{w^{(2)} - w_p^{(2)}}} = \Lambda \left(w^{(2)} - w_p^{(2)}\right)$$

Since $\lambda(s)$ is Hurwitz $\left(w^{(i)} - w_p^{(i)}\right) \to 0$ as $t \to \infty$!

QED

Claim (b) is the main topic.

$$\text{Let} \quad \theta^T(t) = \left(a^T(t),\ b^T(t)\right)$$

$$w^T(t) = \left(w^{(1)T}(t),\ w^{(2)T}(t)\right)$$

~~$e(t) = y_p(t) - \theta^*(t)\, w(t)$~~

Notice that we can write,

$$
\begin{aligned}
y_p(t) &= \theta^{*T} w_p(t) \\
&= \theta^{*T} W(t) + \left(\theta^{*T} w_p(t) - \theta^{*T} W(t)\right) \\
&= \theta^{*T} W(t) - \varepsilon(t)
\end{aligned}
$$

transient error due to observer dynamics.

Identifier output

$$
y_i(t) = \theta^T(t)\, \underbrace{w(t)}_{\text{regressor vector}}
$$

Then identifier error

$$
\begin{aligned}
e_1(t) &\triangleq y_i(t) - y_p(t) \\
&= \theta^T(t)\, w(t) - \left(\theta^{*T} w(t) - \varepsilon(t)\right) \\
&= \left(\theta^T(t) - \theta^{*T}\right) w(t) + \varepsilon(t) \\
&= \underbrace{\phi^T(t)}_{\text{parameter error}}\, w(t) + \varepsilon(t)
\end{aligned}
$$

In certain settings it is customary to assume $\varepsilon(t) \equiv$

This is O.K. asymptotically since $\varepsilon(t) \to 0$ exponenti

Algorithms for identification (parameter update rules).

The rules considered in example 1 naturally generalize. First we can classify into instantaneous or gradient algorithms and integral or least squares algorithms. Then there can be further refined based on various normalizations and projections in parameter space.

Gradient Algorithm

$$\dot{\theta} = -\gamma \frac{\partial}{\partial \theta}\left(\tfrac{1}{2} e_1^2\right)$$

$$= -\gamma e_1 \frac{\partial e_1}{\partial \theta}$$

$$= -\gamma e_1 w \qquad \gamma > 0$$

here $\theta \in \mathbb{R}^{2n}$ the parameter space.

Least Squares Algorithm

$$\dot{\theta} = -\gamma P w e_1$$

$$\dot{P} = Q - \gamma P w w^T P \qquad \begin{matrix} Q = Q^T > 0 \\ \gamma > 0 \end{matrix}$$

here $\theta \in \mathbb{R}^{2n}$ and $P \in$ space of symmetric $2n \times 2n$ matrices.

Also $P(0) = P^T(0) > 0$, otherwise arbitrary.

Both algorithms involve the "correlation" $w(t)\, e_1(t)$ as a driving signal. By correspondence with Kalman filtering one refers to $P$ as a "covariance" matrix. The least squares algorithm is more complicated to implement but one expects it to show faster convergence (where applicable).