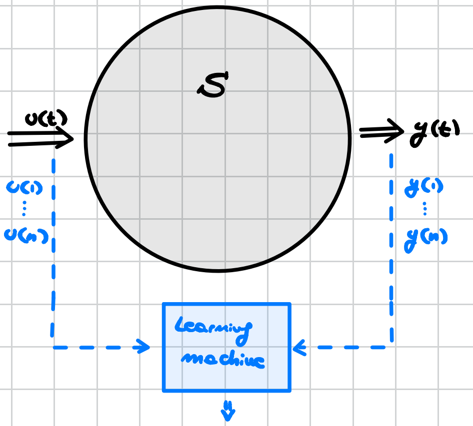
06/03 - Identification



mathematical description of S

Parametric System Identification

1) Experiment design and data collection
2) Selection of a parametric model class $\mathcal{M}(\theta) = \{M_\theta : \vartheta \in \Theta\}$
3) Choice of the identification criteria $J_N(\theta) \geq 0$
4) Minimization of $J_N(\vartheta)$ with respect of ϑ and computation of $\hat{\theta}_N$
$$\hat{\theta}_N = \arg\min_\theta J_N(\theta)$$
5) Model validation

1) Experiment design

Data are provided by someone else (no need to define an experiment), but usually you have a system where to extract. Two main choices

1) Design of $u(t)$ $\Rightarrow$ if the informatively of the resulting dataset is the thing that matter most.
2) Select the experiment length N $\Rightarrow$ confidence about the result

2) Choice of the model class

Optimize the parameters in the model class: $M_\theta = \{M_\theta : \vartheta \in \Theta\}$. We have many options: discrete / continuous time, linear / non linear, time invariant / time-varying, static / dynamic systems. This coincide with ARMA[x] models

$$y(t) = a_1 y(t-1) + \dots + b_0 u(t-d) + \dots + c_0 e(t) + \dots$$
$$\theta = [a_1 \dots a_m \; b_0 \dots b_p \; c_0 \dots c_n]^T$$

To fully specify $y(t) \sim ARMAX$ process, θ alone is not sufficient: we need also

- $e(t) \sim WN(0, \lambda^2) \rightarrow \lambda^2$
- pure delay between $u(t)$ and $y(t) \rightarrow d$

$\Big]$ not critical components

- degree of the polynomials $A(z), B(z), C(z) \rightarrow m, n, p$

λ^2 and d are not critical (we will obtain them of free), while m, n, p are critical, they need to be specified at priori

Θ is the set of all the admissible values of ϑ: e.g. $y(t) = a\, y(t-1) + e(t) \sim AR(1) \Rightarrow |a| < 1$

$$M_\theta = \{AR(1), \vartheta = a, \; |a| < 1\}$$

3) Choice of the identification criteria

Predictive approach: $J(\theta) = E[(y(t+1) - \hat{y}(t+1|t, \theta))^2]$ This is an ideal objective (we cannot compute the expectation with only 1 samples!). We define a sample based version of $J(\theta)$

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \hat{y}(t|t-k, \theta))^2$$

Why 1-step ahead (and not 2,3,...,k)? The prediction error is defined as

$$\varepsilon(t+k|t) \triangleq y(t+k) - \hat{y}(t+k|t) \quad \Rightarrow \quad VAR[\varepsilon(t+k|t)] \in [VAR[e(t)], VAR[y(t)]]$$

If the model is exact and we compute the optimal 1-step ahead prediction error

$$\varepsilon(t+1|t) = e(t) \quad \Rightarrow \quad VAR[\varepsilon(t+1|t)] = VAR[e(t)] = \lambda^2$$

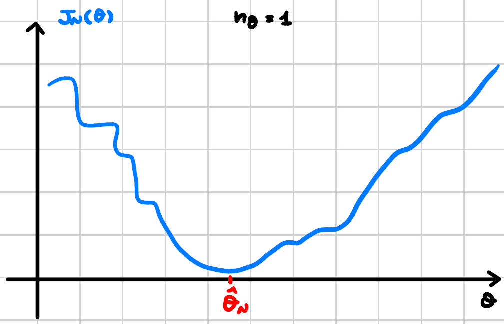
Then

$$J(\hat{\theta}) = E[e(t)] = \lambda^2$$

we obtain λ^2 for free!

optimal model parameter (real)

4) Minimization of $J_N(\vartheta)$ with respect to ϑ : $J_N(\vartheta) : \Theta \subseteq \mathbb{R}^{n_\vartheta} \to \mathbb{R}^+$



Two relevant situation:

1) $J_N(\vartheta)$ quadratic (AR, ARX)

2) $J_N(\vartheta)$ non-quadratic (ARMA, ARMAX)

5) Model validation

We made few assumption:

- S lies within the model class — We need a quality check
- m, n, p are specified at prior (already used these values at step 2)
- data have been "correctly" selected (they are informative)

Identification of ARX models

Consider the model class

no past of noise $\Rightarrow C(z)=1$

$$M_\vartheta : \; y(t) = \frac{B(z)}{A(z)} u(t-d) + \frac{\cancel{C(z)}^{\,1}}{A(z)} e(t) \qquad e(t) \sim WN(0,\lambda^2)$$

Where $A(z)$ and $B(z)$ must be written in canonical form to find the parameter vector ϑ

$$\left.\begin{aligned} A(z) &= 1 - a_1 z^{-1} - \dots - a_m z^{-m} \\ B(z) &= b_0 + b_1 z^{-1} + \dots + b_{p-1} z^{-p+1} \end{aligned}\right\} \quad \vartheta = [a_1, \dots, a_m, b_0, \dots, b_{p-1}]^T \quad \text{size of } \vartheta : \; n_\vartheta = m+p$$

In this way, our dataset is $D_N = \{ (u(1), y(1)), \dots, (u(N), y(N)) \}$

Write the model in different way...

$$\begin{aligned} M_\vartheta : \quad & A(z) y(t) = B(z) \\ & y(t) - y(t) + A(z) y(t) = B(z) u(t-d) + e(t) \\ & y(t) = (1 - A(z)) y(t) + B(z) u(t-d) + e(t) \\ & y(t) = (a_1 z^{-1} + \dots + a_m z^{-m}) y(t) + (b_0 + b_1 z^{-1} + \dots + b_{p-1} z^{-p+1}) u(t-d) + e(t) \end{aligned}$$

only quantity not available at time t

So the predictor will be written as a linear combination of $a_i y(t-i)$ and $b_{p-i}\, u(t-d-p+i)$

$$\hat{y}(t|t-1) = a_1 y(t-1) + \dots + b_0 u(t-d) + \dots + b_{p-1} u(t-d-p+1)$$

and its also the best predictor possible

optimization task: $\quad J_N(\vartheta) = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \hat{y}(t|t-1, \vartheta))^2$

Regressor (regression vector) : $\quad \varphi(t) = [\, y(t-1) \dots y(t-m) \;\; u(t-d) \dots u(t-d-p+1) \,]^T$

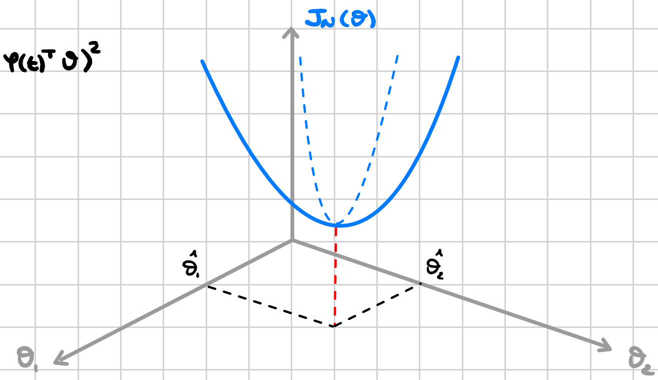
$$\hat{y}(t|t-1, \vartheta) = [a_1 \dots a_m \;\; b_0 \dots b_{p-1}] \begin{bmatrix} y(t-1) \\ \vdots \\ u(t-d-p+1) \end{bmatrix} = \vartheta^T \varphi(t)$$

compact vector form linear in ϑ

So

$$J_N(\vartheta) = \frac{1}{N} \sum (y(t) - \hat{y}(t|t-1, \vartheta))^2 = \frac{1}{N} \sum (y(t) - \varphi(t)^T \vartheta)^2$$

since $y(t)$ is linear in ϑ, $J_N(\vartheta)$ is quadratic in ϑ



Paraboloid in $\mathbb{R}^{n_\vartheta + 1}$ $\Rightarrow$ only 1 minimum

Conditions to find $\hat{\vartheta}_N$:

1) $\hat{\vartheta}_N$ is a stationary point of $J_N(\vartheta)$: $\frac{\partial}{\partial \vartheta} J_N(\vartheta)\Big|_{\vartheta=\hat{\vartheta}_N} = 0$

2) $\hat{\vartheta}_N$ is a minimum point : $\frac{\partial^2}{\partial \vartheta^2} J_N(\vartheta)\Big|_{\vartheta=\hat{\vartheta}_N} \succ 0$

1 $$\frac{\partial}{\partial \vartheta} J_N(\vartheta) = \begin{bmatrix} \frac{\partial}{\partial a_1} J_N(\vartheta) \\ \vdots \\ \frac{\partial}{\partial a_m} J_N(\vartheta) \end{bmatrix} = \frac{1}{d\vartheta}\left[J_N(\vartheta)\right] = \frac{d}{d\vartheta}\left[\tfrac{1}{N} \sum (y(t) - \varphi(t)^T\vartheta)^2\right]$$

$$= \tfrac{1}{N} \sum \frac{d}{d\vartheta}\left[y(t) - \varphi(t)^T\vartheta\right]^2$$

$$= \tfrac{1}{N} \sum 2(y(t) - \varphi(t)^T\vartheta) \frac{d}{d\vartheta}(y(t) - \varphi(t)^T\vartheta)$$

$$= \tfrac{1}{N} \sum 2(y(t) - \varphi(t)^T\vartheta)(-\varphi(t))$$

$$\frac{\partial}{\partial \vartheta} J_N(\vartheta)\Big|_{\vartheta=\hat{\vartheta}_N} = 0 \Rightarrow -\frac{2}{N} \sum \varphi(t)(y(t) - \varphi^T(t)\hat{\vartheta}_N) = 0$$

$$-\cancel{\tfrac{2}{N}} \sum \varphi(t) y(t) = -\cancel{\tfrac{2}{N}} \sum \varphi(t)\varphi(t)^T \hat{\vartheta}_N$$

$$\left[\sum \varphi(t)\varphi^T(t)\right]\hat{\vartheta}_N = \sum \varphi(t) y(t)$$ Least squares normal equation

If $\sum \varphi(t)\varphi(t)^T$ (information matrix) is non-singular $\Rightarrow$ I cannot be in a degenerate case

$$\hat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\varphi(t)^T\right]^{-1}\left(\sum_{t=1}^{N} \underbrace{\varphi(t)}_{\text{regression vector}} y(t)\right)$$ ordinary Least squares formula — This is an explicit formula

The set of optimal parameters is a function of the data only. Explicit solution for the identification problem

2 Does $\hat{\vartheta}_N$ correspond to a minimum? We have to check the second condition $\partial^2/\partial\vartheta^2 \, J_N(\vartheta)\Big|_{\vartheta=\hat{\vartheta}_N} \succ 0$

- $\frac{\partial}{\partial \vartheta} J_N(\vartheta) = -\frac{2}{N}\sum_{t=1}^{N} \varphi(t)(y(t) - \varphi^T(t)\vartheta) = -\frac{2}{T}\sum_{1}^{N} \varphi(t) y(t) + \frac{2}{N}\sum_{1}^{N} \varphi(t)\varphi(t)^T\vartheta$ We need to derive this term only

- $\frac{\partial^2}{\partial \vartheta^2} J_N(\vartheta) = \frac{d}{d\vartheta}\left[\frac{\partial}{\partial\vartheta} J_N(\vartheta)\right] = \frac{2}{N}\sum_{t=1}^{+\infty} \varphi(t)\varphi^T(t) \quad \forall\vartheta$

To be usable, we need to check that $\frac{2}{N}\sum \varphi(t)\varphi^T(t) \succ 0$ is positive definite (a quadratic matrix M is positive definite $M \succ 0$, if $\forall x \neq 0 : x^T M x > 0$). So we have to check that $x^T \frac{\partial^2}{\partial\vartheta^2} J_N(\vartheta) x > 0$

$$x^T \frac{\partial^2}{\partial\vartheta^2} J_N(\vartheta) x > 0 \Rightarrow x^T \frac{2}{N}\sum_{1}^{+\infty} \varphi(t)\varphi^T(t) x = \frac{2}{N}\underbrace{\sum_{1}^{+\infty}(\varphi^T(t)x)^2}_{\text{always positive (sum of squares)}} \geq 0$$

where $\sum(\varphi^T(t)x)^2 > 0$ is the sum of squares that is (for sure) not negative. So we can distinguish 2 cases:

1) $\frac{\partial^2}{\partial\vartheta^2} J_N(\vartheta) \succ 0$

2) $\frac{\partial^2}{\partial\vartheta^2} J_N(\vartheta) = 0$

Reasons behind the degenerative case:

1) Data are not representative of the full underlying physical process ⇒ experiment identifiability issue

2) Selected model class is too large (more than one model representing the same data) ⇒ structural identifiability issue