POLITECNICO DI MILANO





Introduction to model identification

Marco Lovera

Dipartimento di Scienze e Tecnologie Aerospaziali, Politecnico di Milano

We have a system identification problem every time that for a given system

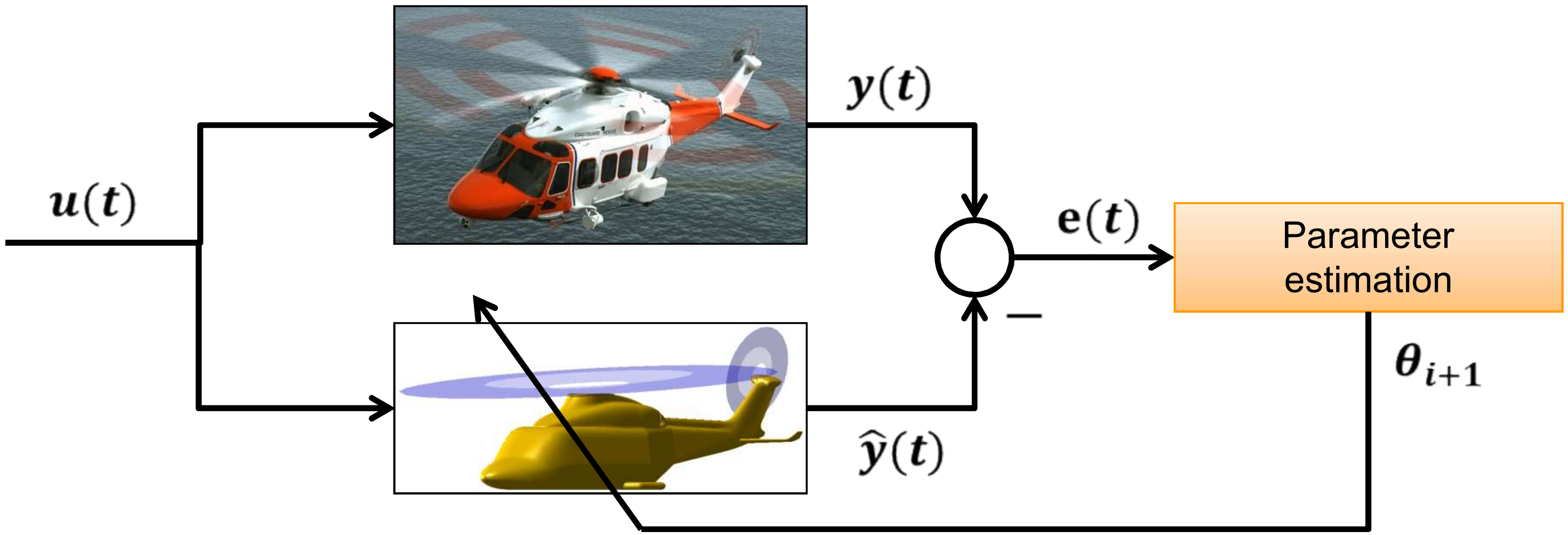
- we have a mathematical model which contains one or more parameters which are either uncertain or unknown

- It is possible to carry out experiments on the system to collect data, through which uncertainty on the parameters can be reduced.

The system identification problem

Solving a system identification problem means finding a way of «incorporating» in the model information coming from data.

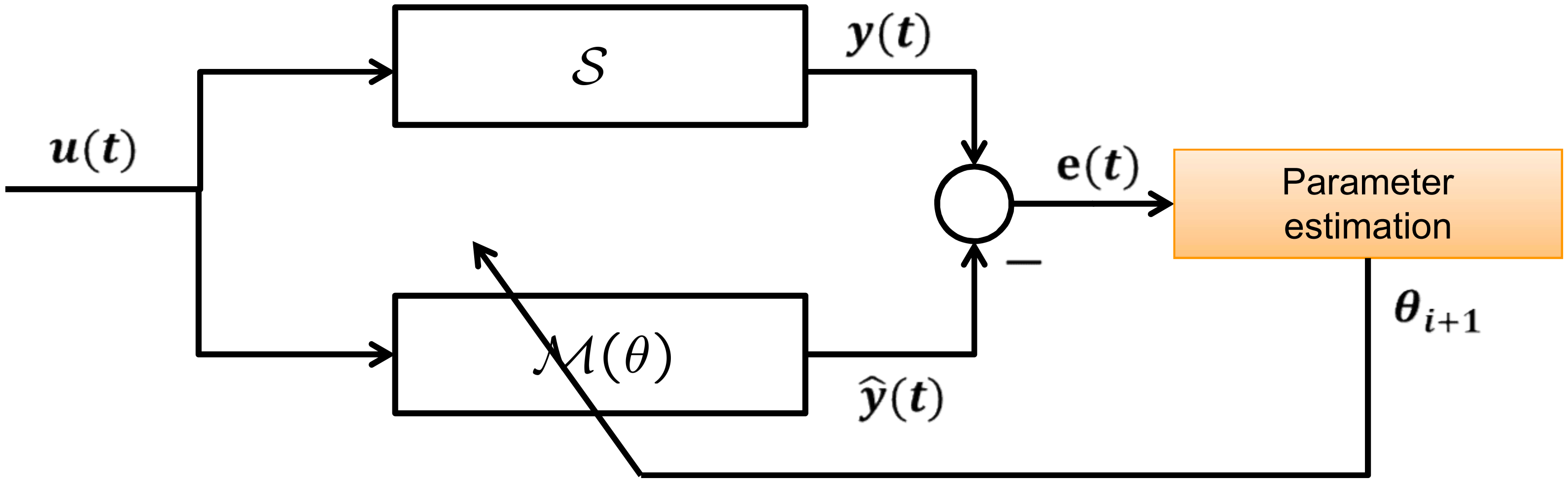
Typical example: identification of aircraft flight dynamics.



The system identification problem

In a more abstract sense, we have:

- a system $\mathcal{S}$ generating the data
- a model class $\mathcal{M}(\theta)$ depending on a parameter $\theta \in \Theta$



The system identification problem

In a realistic situation

$$\mathcal{S} \notin \mathcal{M}(\theta)$$

as the model class is defined using simplified physics to avoid undue complexity.

For example: rigid body model for flight dynamics, which neglects structural response.

Equivalently:

$$\nexists \theta^o : \mathcal{S} = \mathcal{M}(\theta^o)$$

For analysis purposes, however, things are simpler if we consider the assumption

$$\exists \theta^o : \mathcal{S} = \mathcal{M}(\theta^o)$$

or, equivalently, that

$$\mathcal{S} \in \mathcal{M}(\theta)$$

so that we know that a solution can be actually found.

Typical approach to the solution of the problem:

- Define a metric $J(\theta)$ function of
  - the data
  - the model class

- Solve the optimisation problem

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} J(\theta)$$

Note that the solution of the optimisation problem defines an estimator!

As an example, an intuitive choice for the cost function $J(\theta)$ could be the following:

$$J(\theta) = \frac{1}{N} \sum_{k=1}^{K} (y_m(k) - \hat{y}(k;\theta))^2$$

where:

- $y_m(k), \quad k = 1, \ldots, K$ samples of the measured output;
- $\hat{y}(k), \quad k = 1, \ldots, K$ samples of the model output.

In this form the problem appears trivial, that is, simply a matter of writing a model, gathering data and minimising a function.

However:

- Do we have any requirement on the data and the model class for the process to be successful?
- Can we guarantee anything about the quality of the estimates (bias, variance…)?

The model class

Example: we have a system given by

$$\dot{x}_1 = -\frac{1}{\tau_1}x_1 + \mu_1 u$$
$$\dot{x}_2 = -\frac{1}{\tau_2}x_2 + \mu_2 x_1$$
$$y = x_2$$

and we want to estimate the four parameters from measurements of u and y.

Is this possible?

The model class

Re-write the system as the cascade connection of the two blocks

$$\dot{x}_1 = -\frac{1}{\tau_1}x_1 + \mu_1 u$$
$$y_1 = x_1$$

$$\dot{x}_2 = -\frac{1}{\tau_2}x_2 + \mu_2 u_2$$
$$y = x_2$$

where the connection constraint is $u_2 = y_1$.

The transfer functions of the two blocks are

$$y_1 = G_1(s)u_1 = \frac{\mu_1}{\tau_1 s + 1} u_1$$

$$y = G_2(s)u_2 = \frac{\mu_2}{\tau_2 s + 1} u_2$$

so the cascade connection is

$$y = G_2(s)G_1(s)u = \frac{\mu_2 \mu_1}{(\tau_2 s + 1)(\tau_1 s + 1)} u$$

Now the answer to the question is clearly visible…

Structural identifiability

- A model class $\mathcal{M}(\theta)$ is locally structurally identifiable at θ^o if $\forall \theta_1, \theta_2$ in a neighborhood of θ^o we have that

$$\mathcal{M}(\theta_1) = \mathcal{M}(\theta_2) \quad \Rightarrow \quad \theta_1 = \theta_2.$$

- A model class is globally structurally identifiable if it is locally for all values of the parameter.

- Examples of structurally identifiable model structures:
  - SISO transfer functions parameterised by numerator and denominator coefficients
  - SISO state space models in canonical form.

Structural identifiability

Critical cases:

- Physically-motivated parameterisations
- Structured models (interconnected systems as in the previous case)
- MIMO models.

Data

Example: consider the (stable) system

$$y = G(s)u = \frac{\mu}{\tau s + 1} u$$

and assume the goal is to estimate the gain of the transfer function using as data the steady-state part of the step response.

This is clearly possible as

$$u(t) = U \mathrm{sca}(t) \quad \Rightarrow \quad y(t) \underset{t \to \infty}{\to} Y = \mu U$$

so the gain can be estimated as $\hat{\mu} = \dfrac{Y}{U}$

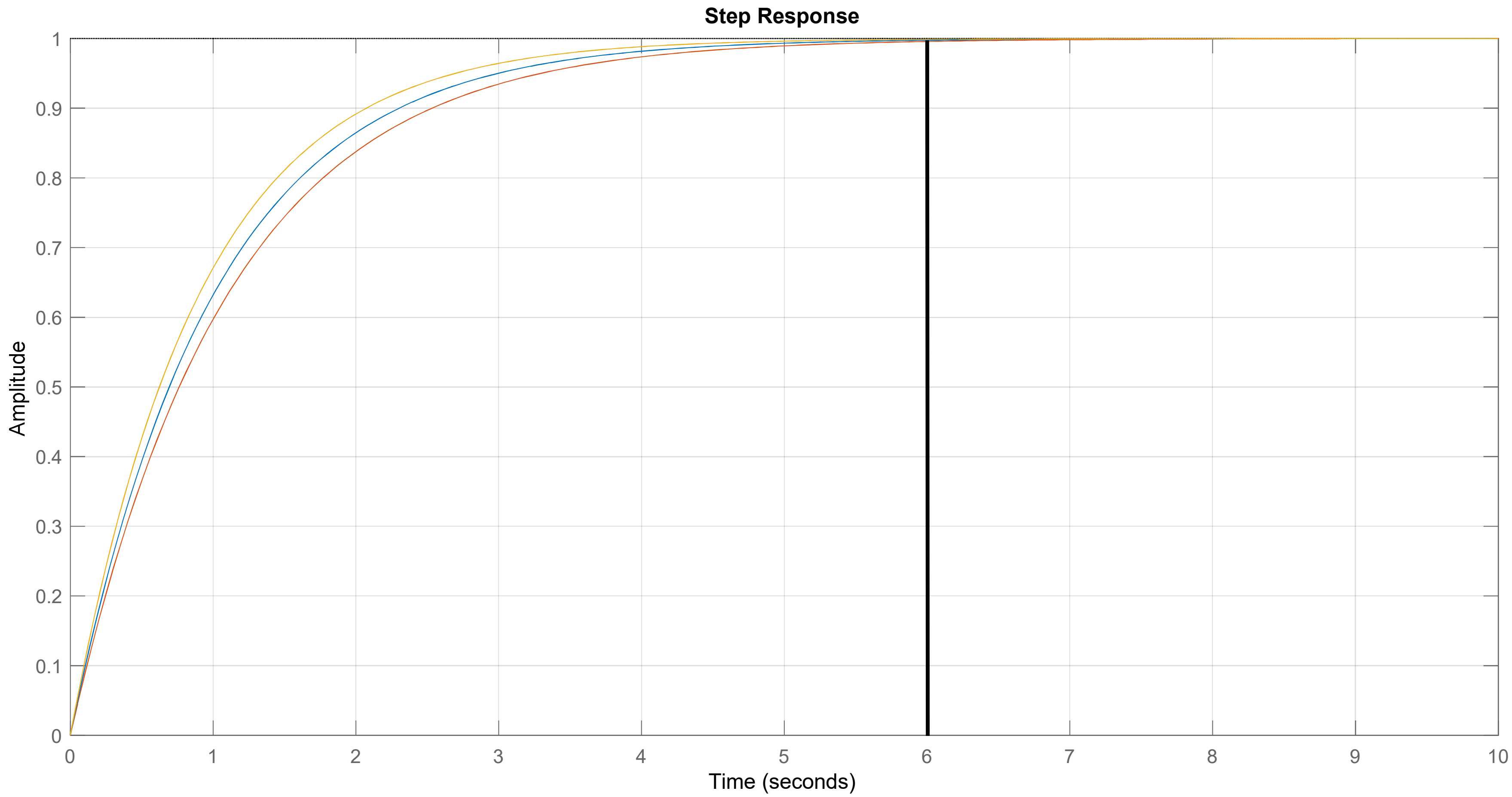
Example: consider the (stable) system

$$y = G(s)u = \frac{\mu}{\tau s + 1} u$$

and assume the goal is to estimate the gain AND the time constant of the pole using as data the steady-state part of the step response.

Is the answer the same as in the previous case?

Data


Step Response
Amplitude
Time (seconds)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
8
9
10

Example: consider the (stable) system

$$y = G(s)u = \frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} u$$

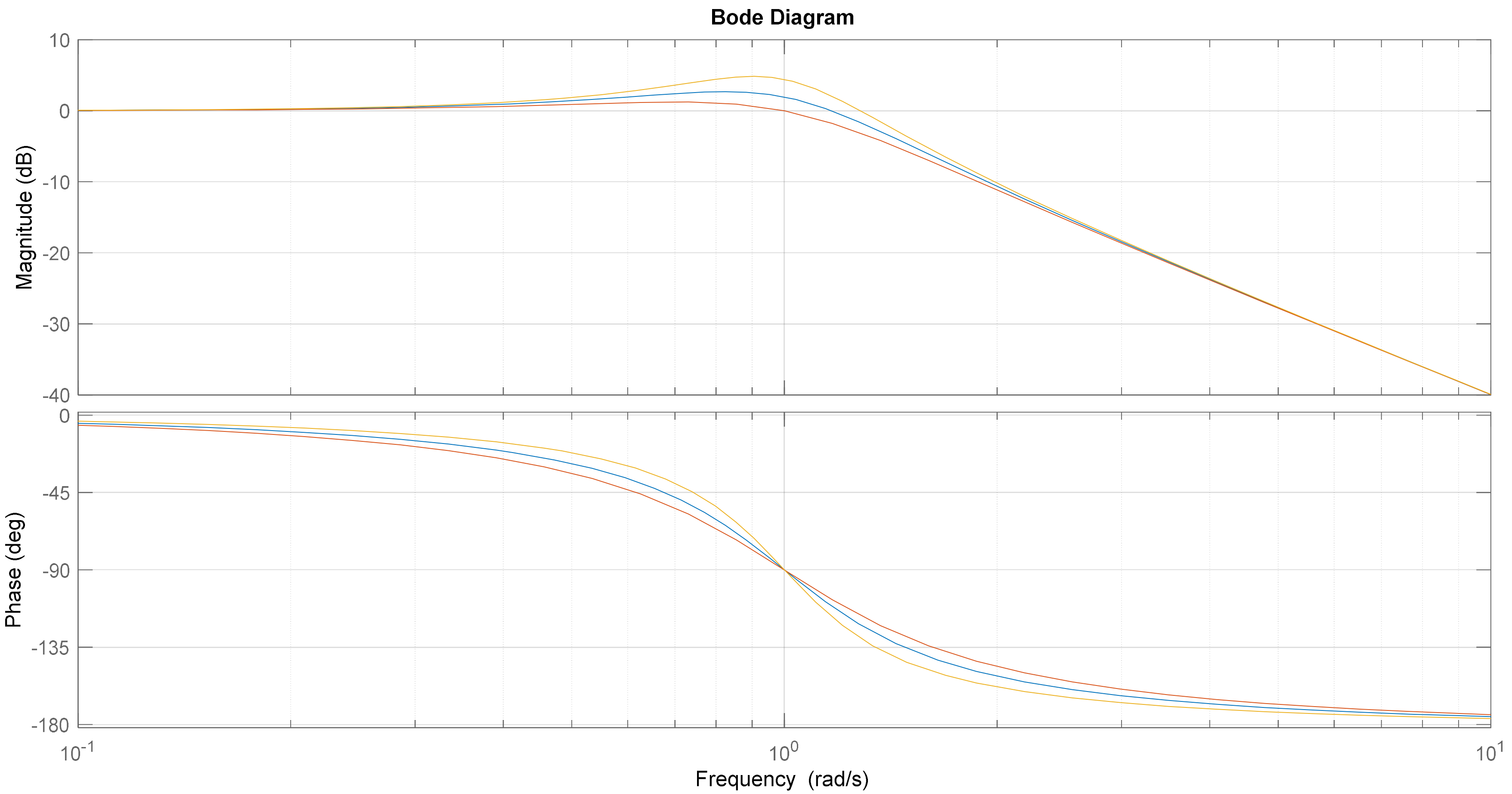
and assume the goal is to estimate the damping ratio using as data the steady-state part of response to a sinusoidal input:

$$u(t) = U \sin(\omega t) \quad \Rightarrow \quad y(t) = Y \sin(\omega t + \phi)$$

$$Y = |G(j\omega)| U \quad \phi = \arg G(j\omega).$$

By inspection of the frequency response function one sees immediately that the choice of ω is critical.

Data



Bode Diagram
10
0
-10
-20
-30
-40
Magnitude (dB)
0
-45
-90
-135
-180
Phase (deg)
10^{-1}
10^{0}
10^{1}
Frequency (rad/s)

Indeed if the excitation frequency is chosen too low or too high the sensitivity of the frequency response to changes in the damping ratio is negligible.

Note also that the steady-state response to a sinusoidal input:

$$u(t) = U \sin(\omega t) \quad \Rightarrow \quad y(t) = Y \sin(\omega t + \phi)$$

$$Y = |G(j\omega)|U \quad \phi = \arg G(j\omega).$$

provides two scalar constraints to be applied to the model to determine the value of the parameters.

Generalising, one can say that an input given by a sum of *N* sinusoids

$$u(t) = \sum_{n=1}^{N} U_n \sin(\omega_n t) \quad \Rightarrow \quad y(t) = \sum_{n=1}^{N} Y_n \sin(\omega_n t + \phi_n)$$

$$Y_n = |G(j\omega_n)| U_n \quad \phi_n = \arg G(j\omega_n), \quad n = 1, \ldots, N$$

provides 2*N* scalar constraints to be applied to the model to determine the value of the parameters.

Clearly, in view of Fourier series and Fourier transform theory this result can be significantly generalised.

Experimental identifiability

- Experimental identifiability will be define formally later.

- Informally, we can say that to achieve experimental identifiability the dataset must be «sufficiently informative» with respect to the parameters of interest.

- Qualitatively, as the steady-state response to a single sinusoid provides two constraints on the model’s frequency response function, we can anticipate that sums of an increasing number of sinusoids will lead to increasingly informative datasets.

Verification of identifiability properties

- How can one verify, at least after the fact, that a given model class and data set satisfy identifiability conditions?

- Consider the cost function, assume the minimum is at

$$\theta = \theta^{\star}$$

- and compute a local second order expansion, to get

$$J(\theta) \simeq J(\theta^{\star}) + \frac{\partial J(\theta)}{\partial \theta}|_{\theta=\theta^{\star}}(\theta - \theta^{\star}) + (\theta - \theta^{\star})^T \frac{\partial^2 J(\theta)}{\partial \theta \partial \theta^T}|_{\theta=\theta^{\star}}(\theta - \theta^{\star})$$

Verification of identifiability properties

- How can one verify, at least after the fact, that a given model class and data set satisfy identifiability conditions?

- At the minimum

$$\frac{\partial J(\theta)}{\partial \theta}|_{\theta=\theta^\star}(\theta-\theta^\star)=0$$

- Therefore

$$J(\theta)-J(\theta^\star)\simeq(\theta-\theta^\star)^T\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\theta^\star}(\theta-\theta^\star)$$

is a quadratic form in $(\theta-\theta^\star)$.

Verification of identifiability properties

- Matrix

$$\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\theta^\star}$$

is symmetric and positive semi-definite by definition, there are two possible situations.

- The matrix is positive definite: $\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\theta^\star} > 0$

- The matrix is just positive semi-definite: $\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\theta^\star} \geq 0$

- Positive definite hessian: in this case, all the eigenvalues of

$$\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}\Big|_{\theta=\theta^\star} > 0 \quad \Rightarrow \quad \lambda_i\left(\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}\Big|_{\theta=\theta^\star}\right) > 0, \quad i = 1, \ldots, n_\theta$$

are strictly positive, therefore the function is locally *increasing* along all directions in the parameter space:

- Positive semi-definite hessian: in this case, not all the eigenvalues of

$$\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\theta^\star} \geq 0 \quad \Rightarrow \quad \lambda_i\left(\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\theta^\star}\right) \geq 0, \quad i = 1, \ldots, n_\theta$$

are strictly positive, therefore the function admits, locally, directions in the parameter space along which it is non-increasing:

Verification of identifiability properties

- Therefore, if the matrix is positive definite we have a well-posed problem, as perturbing the parameter from the optimal value leads to an increased cost.
- If the matrix is just positive semi-definite, there are directions along which we can perturb the parameter without affecting the cost, so the problem does not admit a unique solution even locally.
- The latter case corresponds to lack of either structural or experimental identifiability.

Classification of model identification problems

- Besides the already defined categories of grey-box and black-box, identification problems can be classified in many different respects.

Classification of model identification problems

- Continuous-time vs discrete-time: the models we want to obtain may be formulated
  - in continuous time (differential equations)
  - in discrete-time (difference equations).

- In grey-box modelling, as the model structure is physically-motivated, continuous-time models are usually considered.

- In black-box modelling, discrete time models are the common choice, reflecting the discrete nature of sampled data.

Classification of model identification problems

- Linear vs nonlinear models: depending on the application, we might be interested in estimating:
  - linear models: for example, to capture flight dynamics near a given trim condition
  - nonlinear models: for example if dynamics involving large angles and/or fast maneuvering flight are to be captured.

The model identification process

- No matter what specific application one is facing, the process is characterised by the same steps.

- Definition of the model class: depending
  - on the intended application of the identified model (simulation, prediction, control law design…)
  - on the specific requirements of the application (flight control, aeroelastic analysis…)

  appropriate modelling choices and assumptions are made and the parameters to be estimated are defined.

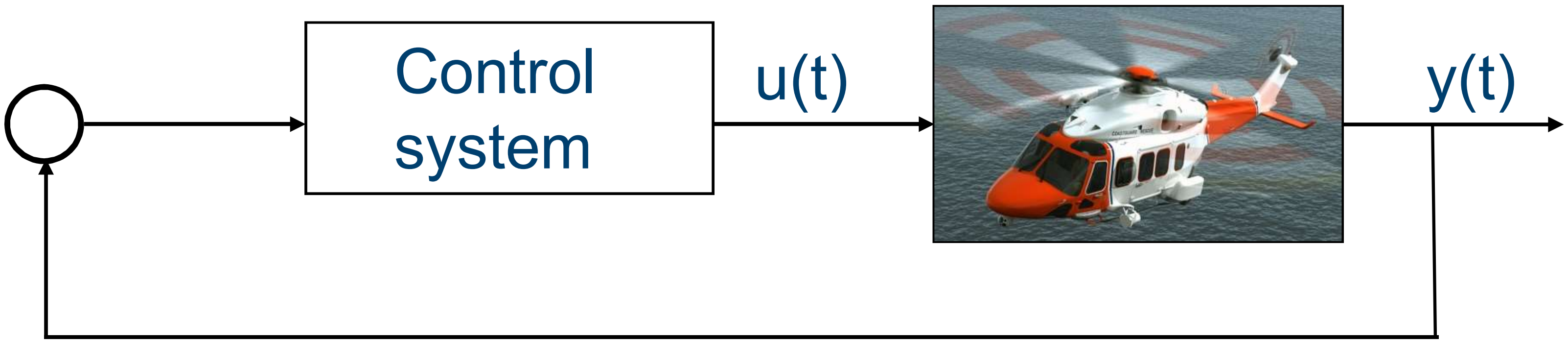
  Structural identifiability must be verified if the considered model class is «non-standard».

- Experiment design:
  - The input sequence used to excite the system must ensure experimental identifiability (so it must be «exciting» in a suitable sense)
  - On the other hand, the experiment must be safe and repeatable, therefore:
    - The response to the applied input must not lead the system into unsafe conditions
    - The shape of the input sequence must be suitable for
      - Either repeatable manual application
      - Or easy automatic implementation.

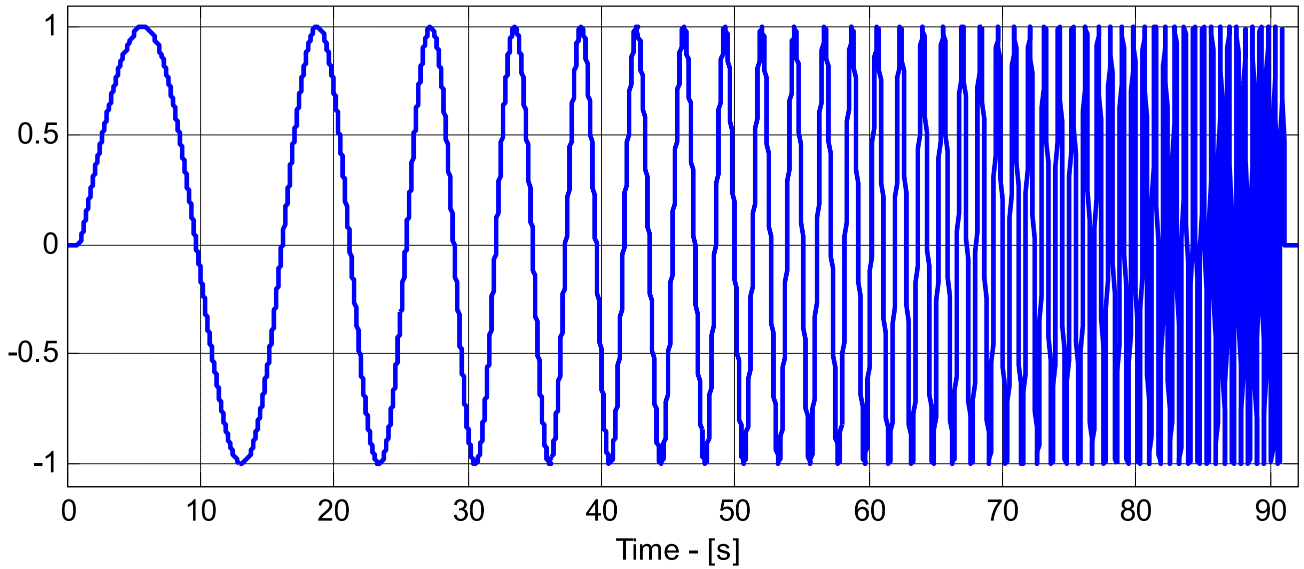
The model identification process

- Experiment design:
  - Depending on the specific application, identification experiments are carried out in open-loop:

  - Or in closed-loop, to guarantee stability during tests:

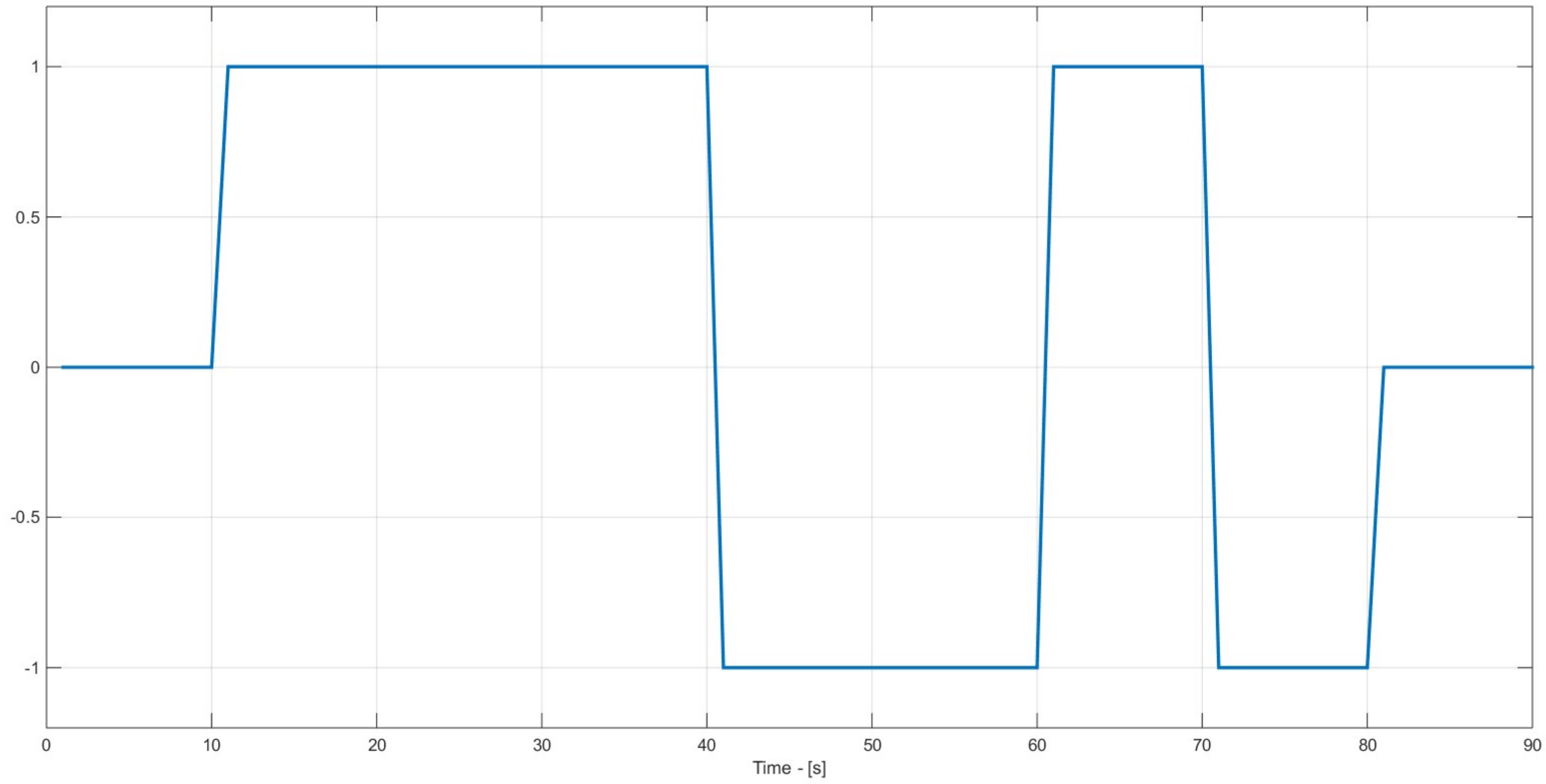


Sine sweep

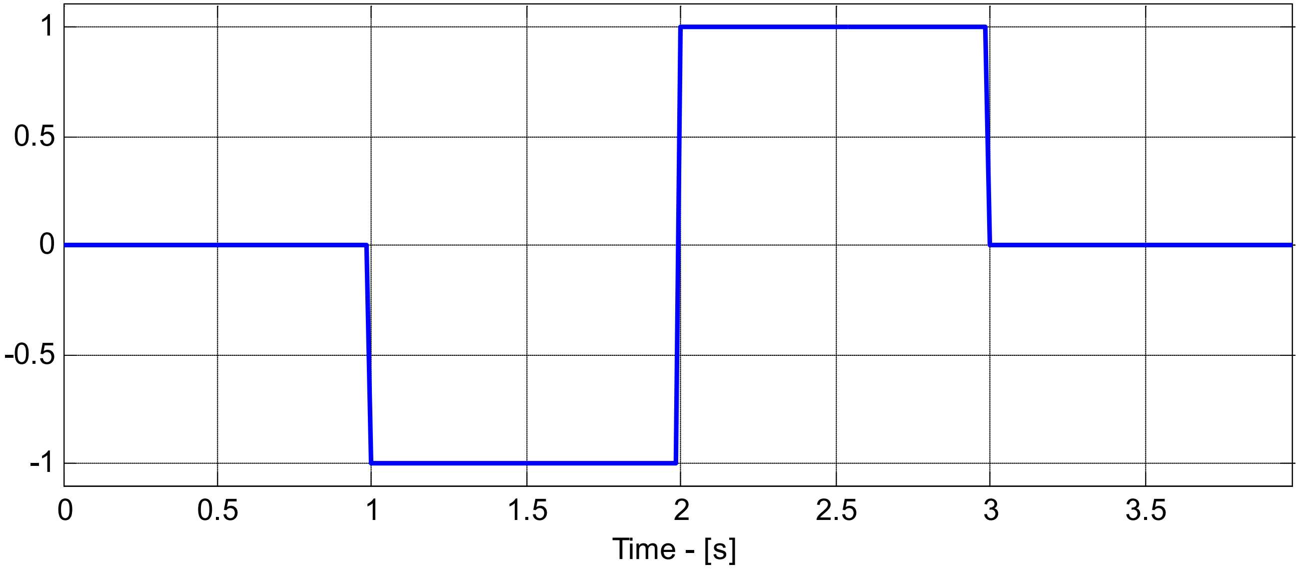


Examples of input sequences for flight testing

3211 sequence



Doublet



- Parameter estimation:
  - Assuming a suitable estimator is available for the specific problem, parameter estimates are computed.
  - As discussed previously, estimators should have small or null bias and variance as small as possible (C-R bound).

- Parameter uncertainty analysis:
  - For the computed estimates, the relevant theoretical results must be applied to check after the fact that
    - The problem was indeed well-posed.
    - Bias is indeed negligible
    - Variance is suitably small.

  - If the results are not satisfactory, a second iteration on the design and/or execution of the experiment must be carried out.

- Model (in)validation:
  - checking model characteristics to ensure they are compatible with prior knowledge (in the linear case, poles, zeros, frequency response function)
  - verifying the performance of the model using data *not employed for the estimation of the parameters*.

- Example:

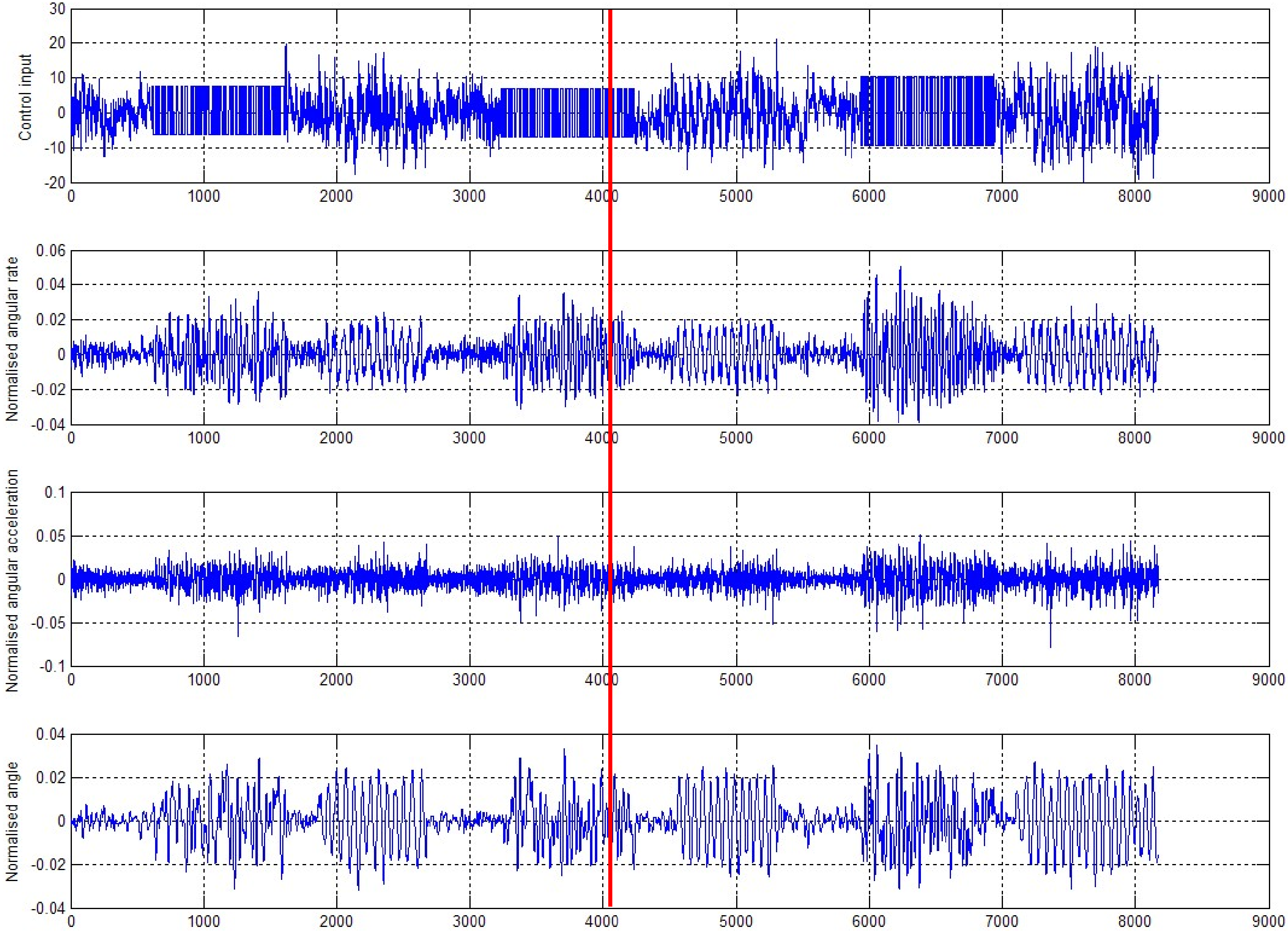
Longitudinal control

Pitch angle

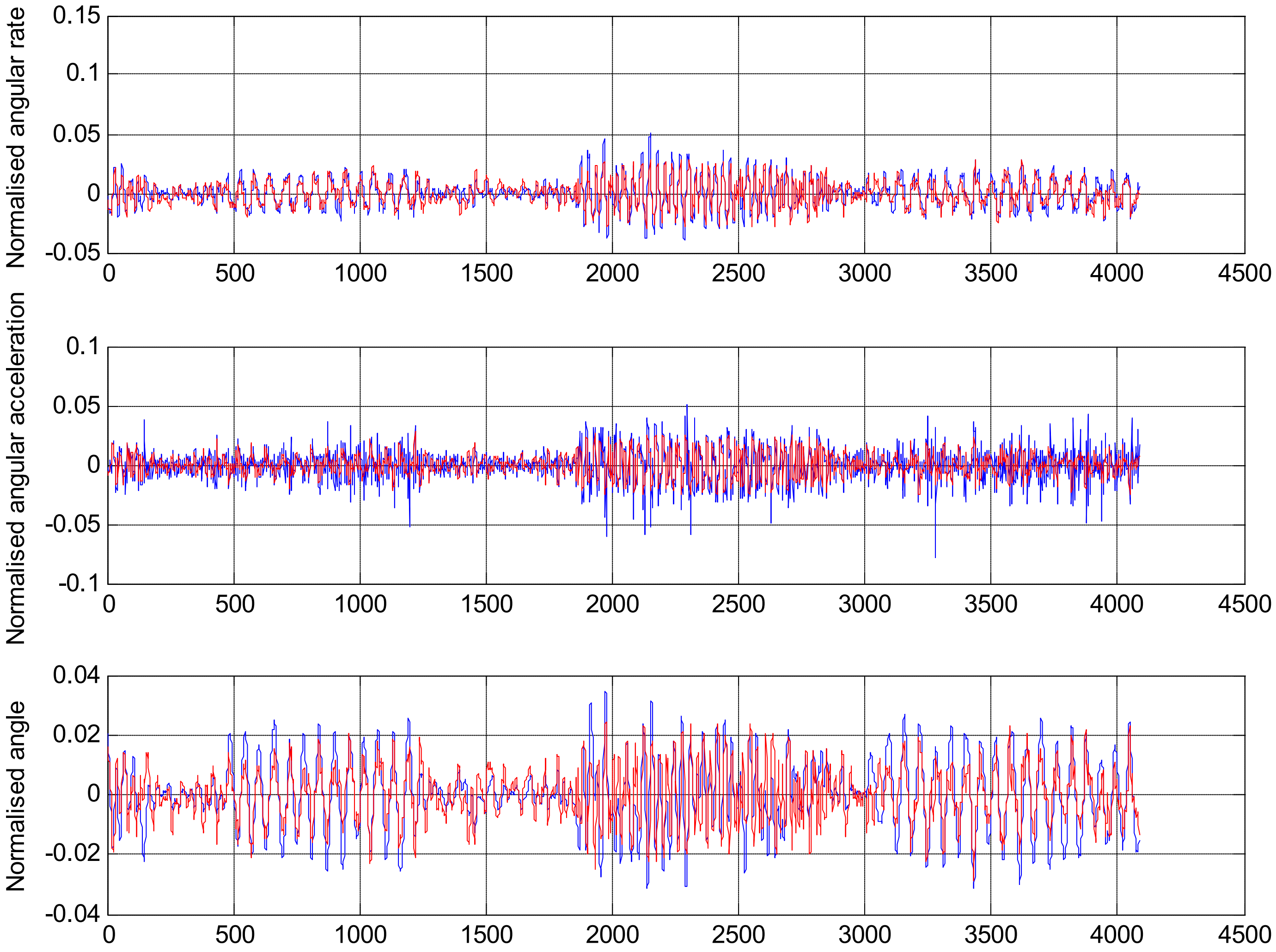
Pitch rate

Pitch acceleration

The model identification process



The model identification process



Conclusion

- Following this brief overview of the model identification process we may now turn to the question of *designing* estimators for specific problem.

- The cornestone of the methods we will develop is the so-called *maximum likelihood principle*, which will be our next topic.