1. Reminders and introduction

Consider a real-life system \( S \) operated in open loop:

\[
y(t) = G_0(z)u(t) + H_0(z)e(t) = G(z, \theta_0)u(t) + H(z, \theta_0)e(t)
\]

with \( \theta_0 \) the unknown true parameter vector

Consider also a full-order model structure \( \mathcal{M} \) for this true system \( S (S \in \mathcal{M}) \):

\[
\mathcal{M} = \{ G(z, \theta); H(z, \theta) \mid \theta \in \mathbb{R}^k \}
\]

Prediction error identification of \( S \)

An input signal \( u(t) \) \((t = 1..N)\) is applied to \( S \) and the corresponding output \( y(t) \) is collected

\[
Z^N = \{ y(t) \ u(t) \mid t = 1...N \}
\]

Based on these input-output data, prediction error identification can be used to obtain a consistent estimate \( \hat{\theta}_N \) of \( \theta_0 \):

\[
\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \left( H(z, \theta)^{-1} (y(t) - G(z, \theta)u(t)) \right)^2_{\varepsilon(t, \theta)}
\]

Choice of experimental conditions

1. the number \( N \) of data that will be collected
2. the signal \( u(t) \) that will be applied
Properties of the identified model

\( \hat{\theta}_N \) is normally distributed around \( \theta_0 \) i.e. \( \mathcal{N}(\theta_0, P_{\theta}) \)

The covariance matrix \( P_{\theta} = E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T \) of \( \hat{\theta}_N \) can be estimated from the data and \( \hat{\theta}_N \) as:

\[
P_{\theta} = \frac{\hat{\sigma}_e^2}{N} \left( \frac{1}{N} \sum_{t=1}^{N} \psi(t, \hat{\theta}_N) \psi^T(t, \hat{\theta}_N) \right)^{-1}
\]

with \( \psi(t, \theta) = \frac{\partial \hat{y}(t, \theta)}{\partial \theta} = -\frac{\partial \hat{e}(t, \theta)}{\partial \theta} \) and \( \hat{\sigma}_e^2 = \frac{1}{N} \sum_{t=1}^{N} \epsilon(t, \hat{\theta}_N)^2 \), an estimate of \( \sigma_e^2 \)

Using the distribution of \( \hat{\theta}_N \), the modeling error can be bounded at each frequency:

\[
| G(e^{j\omega}, \hat{\theta}_N) - G(e^{j\omega}, \theta_0) | < \alpha \sqrt{\text{cov}(G(e^{j\omega}, \hat{\theta}_N))}
\]

\[
< \alpha \sqrt{\Lambda_G(e^{j\omega}, \hat{\theta}_N)} \ P_{\theta} \Lambda_{G}^*(e^{j\omega}, \hat{\theta}_N)
\]

with \( \Lambda_G^T(z, \theta) = \frac{\partial G(z, \theta)}{\partial \theta} \)

The bound holds with an user-chosen probability that can be tuned via \( \alpha \in \mathbb{R} \) (e.g. the probability is 0.95 if \( \alpha = 2.45 \))

Important observations

the experimental conditions \( N \) and \( u(t) \) influence (the bound on) the modeling error

the larger \( N \), the smaller the modeling error

For a given frequency content, the most powerful \( u(t) \), the smaller the modeling error

For signals \( u(t) \) having the same power, different frequency contents could lead to different modeling errors

Example

\( S : \ y(t) = \frac{3.6z^{-1}}{1 - 0.7z^{-1}} u(t) + (1 + 0.9z^{-1}) e(t) \)

\( M : \ G(z, \theta) = \frac{b z^{-1}}{1 - f z^{-1}} \ H(z, \theta) = 1 + c z^{-1} \ \theta = \begin{pmatrix} b \\ c \\ f \end{pmatrix} \)

Let us compare the accuracy obtained with two input signals: a white noise and a cosine at \( \omega = 0.15 \)
$u_1(t)$ a white noise with variance 1.7 and $u_2(t) = 0.75 \cos(0.15t)$; both of length $N = 500$

Question:
Which signal leads to the most accurate model?

While $P_{u_1} \gg P_{u_2}$, the model $\hat{G}_2$ identified with $u_2$ has a slightly better accuracy than the model $\hat{G}_1$ identified with $u_1$

$$\sqrt{\text{cov}(\hat{G}_1(e^{j\omega}))} \geq \sqrt{\text{cov}(\hat{G}_2(\omega))} \ \forall \omega$$

A cosine at $\omega = 0.15$ seems a much better signal than a white noise to identify this true system.

How can we determine the optimal input signal for a given true system?

→ Optimal experiment design
2. Optimal experiment design

Suppose that, for the subsequent use of the model (e.g. for control), it is required that:

$$|G(e^{j\omega}, \hat{\theta}_N) - G(e^{j\omega}, \theta_0)| < r_{adm}(\omega) \quad \forall \omega$$

where \( r_{adm}(\omega) \) is a given frequency function.

The function \( r_{adm}(\omega) \) can e.g. be chosen equal to 0.1\(|G(e^{j\omega}, \hat{\theta}_N)|\) (relative error of 10\%) or can be determined using robust control considerations.

How can we choose \( u(t) \) and \( N \) for the identification experiment in order to guarantee that the model identified with this experiment meets the accuracy constraint?

We have thus to determine \( N \) and \( u(t) \) such that:

$$\alpha \sqrt{\Lambda_G(e^{j\omega}, \hat{\theta}_N) P_\theta \Lambda_G^*(e^{j\omega}, \hat{\theta}_N)} < r_{adm}(\omega) \quad \forall \omega$$

First observation: By choosing \( N \) or the power of \( u(t) \) sufficiently large, it is always possible to achieve this constraint for any given \( r_{adm}(\omega) \)

\[\implies \text{multiple solutions!}\]

Let us therefore

- fix the value of \( N \)
- and determine, for an experiment of that length \( N \), the excitation signal \( u(t) \) with the least power \( P_u = \frac{1}{N} \sum_{t=1}^{N} u^2(t) \) that nevertheless leads to an identified model with the required accuracy.
Optimization problem for the design of the excitation signal \( (N \text{ given}) \)

Determine the sequence \( u(t) (t = 1\ldots N) \) which solves:

\[
\min_{u(t) \in \{1\ldots N\}} \frac{1}{N} \sum_{t=1}^{N} u^2(t)
\]

subject to \( \alpha \sqrt{\Lambda_G(e^{j\omega}, \hat{\theta}_N)} P_{\theta} \Lambda_G^*(e^{j\omega}, \hat{\theta}_N) < r_{adm}(\omega) \forall \omega \)

**Issues with this optimization problem**

- non-linear relations between \( u(t) \) and \( P_u \) and between \( u(t) \) and \( P_{\theta} \) → convexification
- chicken-and-egg problem: dependence on \( \hat{\theta}_N \) → replace \( \hat{\theta}_N \) by an initial estimate \( \theta_{\text{init}} \) obtained e.g. via an initial identification with white noise
- infinite number of constraints i.e. at each frequency \( \omega \) → frequency grid

3. Convexification of the optimization problem

by the use of the power spectrum \( \Phi_u(\omega) \) of \( u(t) \) as alternative decision variable (instead of \( u(t) \) itself)

and by the use of asymptotic expressions for \( P_u \) and \( P_{\theta} \)

3.1 Convexification of the cost function \( P_u \)

for large \( N \),

\[
P_u = \frac{1}{N} \sum_{t=1}^{N} u^2(t) \approx \mathbb{E}u^2(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) \, d\omega
\]

\( P_u \) linear in the new decision variable \( \Phi_u(\omega) \)
3.2 Convexification of the constraint $\alpha \sqrt{\Lambda_G P_\theta \Lambda_G^*} < r_{adm}(\omega)$

for large $N$,

$$P_\theta = \frac{\sigma_2^2}{N} \left( \frac{1}{N} \sum_{t=1}^{N} \psi(t, \hat{\theta}_N) \psi^T(t, \hat{\theta}_N) \right)^{-1}$$

$$= \frac{\sigma_2^2}{N} \left( \bar{E} \psi(t, \theta_0) \psi^T(t, \theta_0) \right)^{-1}$$

$$= \frac{\sigma_2^2}{N} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} F_u F_u^* \Phi_u(\omega) + F_e F_e^* \sigma_e^2 d\omega \right)^{-1}$$

$$F_u(z, \theta_0) = \frac{1}{H_0} \frac{\partial G(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta_0} \quad \text{and} \quad F_e(z, \theta_0) = \frac{1}{H_0} \frac{\partial H(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta_0}$$

Remark. $\theta_0$ in the expression of $P_\theta$ can also be replaced by $\theta_{init}$ (see slide 18)

Let us thus rewrite the constraint as a linear function of $P_\theta^{-1}$

$$\alpha \sqrt{\Lambda_G(e^{j\omega}, \hat{\theta}_N) P_\theta \Lambda_G^*(e^{j\omega}, \hat{\theta}_N)} < r_{adm}(\omega) \quad \forall \omega$$

is equivalent (via the Schur complement) to

$$P_\theta^{-1} > R_{adm}(\omega) \quad \forall \omega$$

where $R_{adm}(\omega) = \frac{\alpha^2}{r_{adm}(\omega)} \Lambda_G^*(e^{j\omega}, \hat{\theta}_N) \Lambda_G(e^{j\omega}, \hat{\theta}_N)$

3.3. A convex formulation of experiment design

Reminder: initial optimization problem (non-convex)

Determine the sequence $u(t)$ ($t = 1 \ldots N$) which solves:

$$\min_{u(t)} \frac{1}{N} \sum_{t=1}^{N} u^2(t)$$

subject to $\alpha \sqrt{\Lambda_G(e^{j\omega}, \hat{\theta}_N) P_\theta \Lambda_G^*(e^{j\omega}, \hat{\theta}_N)} < r_{adm}(\omega) \quad \forall \omega$
Convex alternative (fixed $N$):

Determine the power spectrum $\Phi_u(\omega)$ of the excitation signal $u(t)$ which solves:

$$\min_{\Phi_u(\omega)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) \, d\omega$$

subject to $\Phi_u(\omega) \geq 0 \quad \forall \omega$ and to

$$\frac{N}{\sigma_e^2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} F_u F_u^* \Phi_u(\omega) + F_e F_e^* \sigma_e^2 \, d\omega \right) > R_{adm}(\omega) \quad \forall \omega$$

In this case, we have to restrict the structure of $\Phi_u(\omega)$ as follows:

$$\Phi_u(\omega) = |F_m(e^{j\omega})|^2 = \sum_{r=-m}^{m} c_r \, e^{j\omega \, r} \quad (\text{met } c_r = c_{-r})$$

where $c_r (r = 0...m)$ are the new decision variables of the optimization problem

The larger $m$ is chosen, the more flexible is the parametrization of the spectrum

Choosing $m = 0$ is equivalent to restrict attention to a flat spectrum (white noise)

3.4. Parametrization of the power spectrum $\Phi_u(\omega)$

The decision variable $\Phi_u(\omega)$ has an infinite dimension → a simpler and linear parametrization of $\Phi_u(\omega)$ is required

Suppose we want to be able to generate our input signal $u(t)$ of spectrum $\Phi_u(\omega)$ as the realization of a white noise filtered by a FIR filter $F_m(z)$ of order $m$:

$$u(t) = F_m(z)w(t)$$

where $w(t)$ is a white noise of variance 1 and $F_m(z) = f_0 + f_1 z^{-1} + .. + f_m z^{-m}$ an arbitrary FIR filter

4. Some alternative experiment design problems

1) Instead of $P_u$, the cost function can also be the power $P_y$ of the output signal $y(t)$ (or a combination of both)

2) Optimal experiment design can also be formulated for direct closed-loop identification

$$Z^N = \{y(t) \, u(t)| t = 1...N\} \quad \text{generated via } r(t)$$
Decision variable: $\Phi_r(\omega)$

Expression for $P_\theta$:

$$P_\theta = \frac{\sigma^2}{N} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} F_r F_r^* \Phi_r(\omega) + F_v F_v^* \sigma^2 \ d\omega \right)^{-1}$$

$$F_r(z, \theta_0) = \frac{\partial G(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta_0}$$

$$F_v(z, \theta_0) = \frac{\partial H(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta_0} - C_i d \frac{\partial G(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta_0}$$

$P_\theta^{-1}$ linear in the decision variable $\Phi_r(\omega)$

Possible formulation for direct closed-loop identification:

$$\min_{\Phi_r(\omega)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_r(\omega) \ d\omega$$

subject to $\Phi_r(\omega) \geq 0 \ \forall \omega$ and to

$$\frac{N}{\sigma^2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} F_r F_r^* \Phi_r(\omega) + F_v F_v^* \sigma^2 \ d\omega \right) > R_{adm}(\omega) \ \forall \omega$$

An other formulation will be presented in the sequel.

3) Other (dual) paradigms for (open-loop) input design:

$$\min_{\Phi_u(\omega)} \det(P_\theta)$$

subject to $\Phi_u(\omega) \geq 0 \ \forall \omega$ and to

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) \ d\omega < P_{\text{max}}$$

with $P_{\text{max}}$ the maximal power that is allowed at the input.

5. Illustration 1 (see example slides 8-12)

$$S: \ y(t) = \frac{3.6 z^{-1}}{1 - 0.7 z^{-1}} u(t) + (1 + 0.9 z^{-1}) e(t)$$

$$M: \ G(z, \theta) = \frac{b z^{-1}}{1 - f z^{-1}} \ H(z, \theta) = 1 + cz^{-1}$$

We would like to find, for an experiment of duration $N = 500$, the least powerful excitation signal $u(t)$ which leads to an identified model $G(z, \hat{\theta}_N)$ with a relative modeling error of less than $1\%$ at each $\omega$

$$|G(e^{j\omega}, \hat{\theta}_N) - G(e^{j\omega}, \theta_0)| < 0.01 \ |G(e^{j\omega}, \hat{\theta}_N)| \ \forall \omega$$

$$\Rightarrow r_{adm}(\omega)$$
Design of the optimal input signal

The required initial estimate $\theta_{\text{init}}$ is obtained via an initial experiment of duration $N = 100$ with white input ($\sigma_u^2 = 1$).

We solve then the convex optimization problem where $\theta_0$ and $\hat{\theta}_N$ have been replaced by $\theta_{\text{init}}$ e.g.

$$r_{\text{adm}}(\omega) = 0.01 \ |G(e^{j\omega}, \theta_{\text{init}})| \quad \forall \omega$$

Other choices: $\alpha = 2.45$, $m = 20$, $N = 500$

This leads to the following optimal spectrum $\Phi_u(\omega)$:

This spectrum corresponds to a power $P_{u,\text{opt}} = 0.25$

Verification

We have generated an input signal of length $N = 500$ having the optimal spectrum:

By shaping $\Phi_u(\omega)$, less power is needed to achieve the required accuracy!
The modeling error is computed/bounded using $P_\theta$ and we observe, as expected,

$$\alpha \sqrt{\Lambda_G P_\theta \Lambda_G^*} < r_{adm}(\omega) \ \forall \omega$$

Analysis of the optimal spectrum $\Phi_u(\omega)$

The optimal power spectrum concentrates the power around the frequency $\omega = 0.15 \rightarrow$ a cosine at this frequency will also be appropriate

Why concentrating the power around $\omega = 0.15$?

This is the frequency range with the pole of $G_0$ (red) and the zero of $H_0$ (blue) !!!!

6. Illustration 2: least intrusive identification experiment for control

We consider a real-life system $G_0$ (with two resonances) operated in closed loop whose objective is to reject the disturbance $v(t) = H_0(z)e(t)$:

This closed-loop delivers a product $y(t)$

Objective: (re)-identify a model of $G_0$ (and $H_0$) to update $C_{id}$ by a new robust controller designed with the model
Objective: (re)-identify a model of $G_0$ (and $H_0$) to update $C_{id}$ by a new robust controller designed with the model

→ a constraint $\alpha \sqrt{\Lambda_G P_0 \Lambda_G^*} < r_{adm}(\omega) \ \forall \omega$

Extra requirement: in order to be able to continue to produce a good $y(t)$, we would like to achieve this objective with the least intrusive identification experiment

How can we evaluate the effects of an identification experiment on the production quality?

The perturbations $y_r$ and $u_r$ are present during the experiment duration $N$

Their presence reduce the production quality

For fixed $N$, we can measure the performance degradation due to the application of a signal with power spectrum $\Phi_r(\omega)$ by:

$$J_r = \alpha_y P_{y_r} + \alpha_u P_{u_r} = \alpha_y \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{y_r}(\omega) \ d\omega \right) + \alpha_u \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{u_r}(\omega) \ d\omega \right)$$

e.g. $\alpha_u = \alpha_y = 1$

---

How can we evaluate the effects of an identification experiment on the production quality??

Normal operation vs. Identification experiment operation

$$\begin{align*}
\text{Normal operation:} & \quad \begin{cases} y(t) = S_{id}v(t) \\ u(t) = -C_{id}S_{id}v(t) \end{cases} \\
\text{Identification experiment:} & \quad \begin{cases} y(t) = y_r(t) + S_{id}r(t) + S_{id}v(t) \\ u(t) = u_r(t) + S_{id}r(t) - C_{id}S_{id}v(t) \end{cases}
\end{align*}$$

\[\text{set point} \quad \rightarrow \quad u(t) \quad \rightarrow \quad y(t)\]

\[\begin{align*}
C_d & \quad \rightarrow \quad u(t) \\
G_i & \quad \rightarrow \quad y(t)
\end{align*}\]

\[\rightarrow \text{Least intrusive identification experiment for control}\]

$$\text{arg min}_{\Phi_r(\omega)} \ J_r \ \text{subject to} \ P^{-1}_\theta \geq R_{adm}(\omega) \ \forall \omega$$

under the constraint that $P^{-1}_\theta \geq R_{adm}(\omega) \ \forall \omega$

Reminder: $P^{-1}_\theta > R_{adm}(\omega)$ is equivalent to:

$$\alpha \sqrt{\Lambda_G P_0 \Lambda_G^*} < r_{adm}(\omega)$$

Here, $r_{adm}(\omega)$ has been chosen based on robust control considerations and the duration of the experiment is fixed to $N = 500$. 

\[\text{DISC course: System identification for control} \quad 41\]

\[\text{DISC course: System identification for control} \quad 42\]

\[\text{DISC course: System identification for control} \quad 43\]
\[
\arg\min_{\Phi_r(\omega)} \mathcal{P}_u + \mathcal{P}_{y_r} \quad \text{subject to } P^{-1} \geq R_{adm}(\omega) \ \forall \omega
\]

We consider two cases for the optimization on $\Phi_r(\omega)$:

- the first one constrains $r(t)$ to be a white noise ($m = 0$)
- the second one allows $r(t)$ to have more flexible $\Phi_r(\omega)$ ($m = 10$)

Let us generate a realization of length $N = 500$ of those two spectra...

This leads to the following perturbations $y_r(t)$
\[ \Phi_r(\omega), \Phi_{ru}(\omega), \Phi_{yr}(\omega) \text{ for the white } r(t) \text{ (red) and for the flexible } r(t) \text{ (blue)} \]

\[ \Phi_r(\omega) \text{ required } J_r \text{ for } N = 500 \]

\[
\begin{array}{|l|l|}
\hline
\text{white} & 22.5 \\
\hline
\text{flexible} & 9.9 \\
\hline
\end{array}
\]

**Conclusions:**

- both experimental conditions deliver a model accurate enough for robust control (see next slide)

- \( J_r = P_{ur} + P_{yr} \) is two times larger with the RBS than with the frequency-dependent signal (i.e. 22.5 vs. 9.9)

That both the flexible and the white spectra deliver a model accurate enough for robust control can be seen by verifying in both cases that

\[ \alpha \sqrt{\Lambda G P_0 \Lambda^* G} < r_{adm}(\omega) \quad \forall \omega \]

**Explanation of the better \( J_r \) for the flexible signal:**

\[
\begin{array}{|l|}
\hline
\text{References} \\
\hline
\]