Improving the identification from relay feedback experiments ⋆

Oscar Miguel-Escrig [a], Julio-Ariel Romero-Pérez [a]

[a] *Department of System Engineering and Design. Universitat Jaume I. Castelló de la Plana. Spain.*

Abstract

This paper proposes a new method for systems identification from relay feedback test. As a difference with most of the existing methods, which are based on the describing function technique, our proposal considers a more precise approach that allows to improve the estimation results. Furthermore, the method has also the advantage of being easy to implement, which makes it to stand out among other methods that require introducing additional elements on the loop or extend the experiment duration. The validity of the proposed method has been proved using a batch of models that represent the most common dynamic behaviors of actual industrial processes. The study reveals that with the new proposal, using data from a simple relay experiment, the error in the estimation for some kind of systems can be considerably reduced with respect to previous approaches.

Key words: system identification, relay experiment

1 Introduction

Nowadays, Proportional-Integral-Derivative (PID) controllers drive the majority of control loops in industry. One of the most accepted approaches for tuning the PID is the one based on relay feedback experiments, since it is a very simple and effective way to obtain crucial information about the dynamic behavior of the system. Very popular methods, such as Ziegler-Nichols and others use the results provided by this kind of test, [1], [2]. Furthermore, many auto-tuning methods also consider variations of relay based experiments in the identification phase of the algorithms, [3], [4], [5], [6], [7], [8].

One of the firsts mentions to relay feedback experiment was done in [9], which has been further developed in [10] and significant efforts have been done to improve the initial idea [11], [12]. These works and most of the successive publications use the Describing Function (DF) method as the analysis tool, [13]. Some other investigations improve the initial idea trying to satisfy the filtering hypothesis required by DF method introducing modifications in the relay experiment, i.e., using a preload relay [14] or a saturation non-linearity [15], introducing more levels in the output of the relay [16], adding a mapping function to the relay output [17], proposing new multi-relays schemes [7], or changing the relay by a symmetric-send-on-delta (SSOD) non-linearity which allows to obtain a more sinusoidal-like signal [18].

Additionally, other lines of research have been developed to increase the estimation precision. Some interesting studies are focused on using more accurate methods, but also more time consuming. Some of these methods use the Fourier transform method [19], specifically the Fast Fourier Transform (FFT) algorithm, others use the A-locus for the parameter estimation of low order systems [20] or consider more harmonics on the response of the process [21]. Other authors propose to take into account the shape of the oscillation signal to obtain additional data [22], [23]. Another alternative, gathered in [24], propose to introduce modifications on the traditional relay experiment by considering a parasitic relay or a cascade relay, which certainly enhance the identification of the process but adds complexity to the loop structure. All these proposals, however, require to extend the experimental time, to introduce further elements on the loop or to include sophisticate calculation which could be pitfalls from the practical application point of view.

In this paper, a new method to identify points of the frequency response of the system from relay feedback

⋆ This paper was not presented at any IFAC meeting. Corresponding author O. Miguel-Escrig Tel. +34964728774. e-mail: omiguel@uji.es

Email addresses: omiguel@uji.es (Oscar Miguel-Escrig), romeroj@uji.es (Julio-Ariel Romero-Pérez).

test is presented. The proposal relies on the proper selection of the samples to get more accurate frequency response points while avoiding to overcharge the computation cost. This choice of samples leads to an improvement on the estimation of the frequency response with few points of the temporal response. The proper samples for the estimation are selected as fractions of the oscillation period and do not depend on the shape of the oscillations. Those fractions are always the same, so once the period is known, the samples used for the identification procedure can be easily selected. Unlike others proposals such as those presented in [14] and [15], where a conventional relay experiment is followed by a modified relay experiment in order to reduce the estimation error due to the Describing Function approximation, the new method does not need to extend the duration of the experiments to improve the estimation results. Furthermore, it is also demonstrated that a considerable improvement can be obtained with respect to the results of the conventional relay method proposed in [10].

The paper is organized as follows. Section 2 presents the basis for the new identification method. Section 3 develops on the ideas from the previous section and presents the guidelines to perform the identification. Section 4 compares the proposed method with the conventional relay experiment identification method and with other improved methods based on it. To perform the comparison a batch of systems are identified with all these methods, showing that the proposed method increases the accuracy of the estimation in the majority of processes. Section 5 tackles the identification of multiple points of the frequency response. Section 6 addresses the issues that appear because of the sampling and the measurement noise on the relay experiment, extending the applicability of the proposed method to noisy signals. The identification method allows a least squares formulation, which has been treated in section 7. Finally, the conclusions about this work are given in section 8.

2 Problem statement

The block diagram of a relay feedback system is shown in Figure 1. The aim of the relay based identification techniques is to extract relevant information about the unknown model $G(s)$ of the actual plant from its temporal response under the oscillatory behavior induced by the relay. In order to obtain oscillations around the operation point, in most cases the output levels of the relay are shifted, producing the generic waveform oscillation presented in Figure 2, where the error (e) and the relay output signals (u) are represented. Because of its relevance for tuning PID controllers, one of the most valuable information of $G(s)$ is the so called ultimate point, i.e. the point where the phase of $G(s)$ is $-\pi$ radians. Most of the relay identification techniques assume that the oscillations induced by the relay take place in the ultimate point. This assumption, however, is not strictly

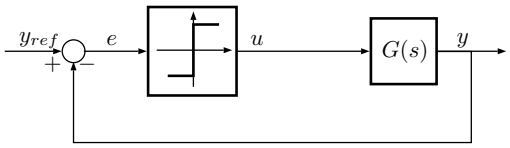

Fig. 1. Relay feedback system.

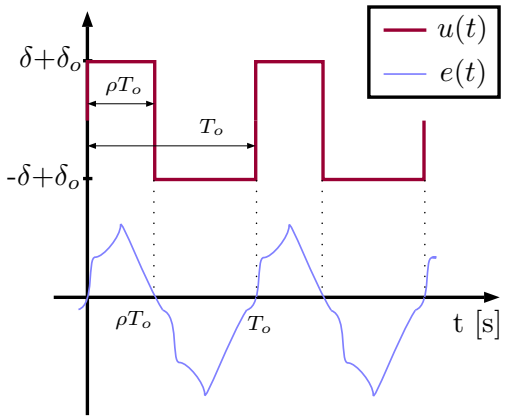

Fig. 2. General oscillation waveform for a system with relay.

true, and consequently it introduces errors in these estimation techniques.

The goal of this paper is to improve the estimation of the actual point where such oscillations occur using the data collected from a simple relay experiment, without introducing new elements on the loop neither increasing the experiment duration, since it could be important backwards for its practical application, as for example its use in systems identification phase of auto-tuning algorithms for PID. With this in mind, let us to express the error signal depicted in Figure 2 by its Fourier series expansion, which result in equation (1), (see Appendix A), where δ_o is the offset of the relay output, δ the relay height, ρ is a fraction of the oscillation period where a commutation on the relay is produced, n the harmonics and $\omega_o = 2\pi/T_o$ is the angular frequency of oscillation, with T_o being the oscillation period.

The first two terms on the right hand side of previous equation represent the magnitude of a bias as result of asymmetric oscillations. In order to induce symmetric oscillations, that is, oscillation with $\rho = 0.5$, it is necessary that:

$$y_{ref}(t) - (\delta(2\rho - 1) + \delta_o)G(0) = 0,$$

which can be achieved by setting the offset of the relay δ_o as follows:

$$\delta_o = \frac{y_{ref}(t)}{G(0)}.$$

This requires to know the system gain $G(0)$, which can be obtained by a simple step response experiment around

$$e(t) = y_{ref}(t) - (\delta(2\rho-1)+\delta_o)G(0) - \frac{4\delta}{\pi}\sum_{n}^{\infty}\frac{1}{n}\sin(n\pi\rho)\Big(\Re\{G(jn\omega_o)\}\cos(n\pi\rho - n\omega_o t) + \Im\{G(jn\omega_o)\}\sin(n\pi\rho - n\omega_o t)\Big) \tag{1}$$

the operation point or with the output of a biased relay as shown in [25]:

$$G(0) = -\frac{\int_0^{2\pi} e(t)d\omega t}{\int_0^{2\pi} u(t)d\omega t}.$$

For symmetric oscillations around y_{ref}, the bias term in equation (1) disappears and considering $\rho = 0.5$ in that equation and applying trigonometric algebra, the error signal can be rewritten as:

$$e(\tau T_o) = -\frac{4\delta}{\pi}\sum_{n=1,3,5,\dots}^{\infty}\frac{1}{n}\Big(\Re\{G(jn\omega_o)\}\sin(2\pi n\tau) + \Im\{G(jn\omega_o)\}\cos(2\pi n\tau)\Big), \tag{2}$$

where $\tau \in [0,1]$ is a dimensionless parameter defined for convenience which indicates a temporal fraction of the oscillation period. It should be noted that even values of n make the operator $\sin(n\pi\rho) = 0$, therefore, only odd values of n appear in the previous equation.

From the point of view of system identification, equation (2) can be interpreted as a simple linear equation where the real and imaginary parts of each point of the open-loop transfer function are unknown. This equation is the cornerstone of the results presented in the rest of the article. In the next sections a detailed study about the use of this equation for the estimation of frequency response points of $G(s)$ is presented.

3 Main result

As can be seen, for each harmonic considered in equation (2), a new point in the frequency response of the open-loop transfer function appears, adding its corresponding real and imaginary parts as unknowns. Hence, to identify h points of the open-loop transfer function, h number of harmonics must be taken into account (until the harmonic $n = 2h-1$) and k equations ($k = 2h$) must be considered in order to be able to resolve the system of equations, which can be expressed in matrix form as:

$$E_{k\times 1} = -\frac{4\delta}{\pi}A_{k\times 2h}B_{2h\times 2h}S_{2h\times 1}, \tag{3}$$

where each element is defined in equations (4) and (5).

Then, the points of the open-loop transfer function, whose real and imaginary parts are contained in the vector S, can be calculated as follows:

$$S = -\frac{\pi}{4\delta}(AB)^{-1}E. \tag{6}$$

The calculation of S from time response samples in E is strongly affected by the matrix AB. Fortunately, this matrix only depends on the number of harmonics and on τ_i, that is, the values of time in which the samples of $e(\tau_i T_o)$ are taken and not on the value of $e(\tau_i T_o)$ itself. Therefore, AB can be pre-calculated to guarantee the estimation of S to be as good as possible no matter the system to identify. Consequently, the main issue is to choose τ_i in order to obtain a proper value of AB.

The choice of τ_i is important because some values can lead the matrix AB in equation (6) to be singular, for example choosing $\tau_1 = 0$ and $\tau_2 = 0.5$ makes the system of equations unsolvable. Furthermore, the choice of τ_i is critical in terms of the accuracy of the solution because if they are not properly chosen, little variations on the measured data produce considerable errors in the solution. These issues are gathered numerically in the condition number of AB. Thus, to avoid these problems, τ_i have to be chosen so that the matrix AB has a proper condition number.

Lemma 1 *For any nonsingular matrix, the condition number is defined as:*

$$\kappa(\cdot) = \|\cdot\|\left\|\cdot^{-1}\right\|,$$

where the operator $\|\cdot\|$ denotes the matrix norm. Then, the condition number of the product of square matrices is related to the condition number of each matrix by:

$$\kappa(AB) \leq \kappa(A)\kappa(B).$$

Theorem 1 *Let matrices A and B be given by equation (3). If the sample time fractions τ_k are chosen such that the difference between them is:*

$$|\tau_i - \tau_j| = \frac{1}{4h}, \frac{2}{4h}, \cdots, \frac{2h-1}{4h}, \frac{2h+1}{4h}, \cdots, \frac{4h-1}{4h}, \quad \forall i,j \quad i \neq j, \tag{7}$$

where h is the number of harmonics considered, then,

$$\kappa_2(AB) \leq n, \tag{8}$$

$$E = \begin{bmatrix} e(\tau_1 T_o) \\ e(\tau_2 T_o) \\ ... \\ e(\tau_k T_o) \end{bmatrix}, \quad A = \begin{bmatrix} \sin(2\pi\tau_1) & \sin(6\pi\tau_1) & ... & \sin(2n\pi\tau_1) & \cos(2\pi\tau_1) & \cos(6\pi\tau_1) & ... & \cos(2n\pi\tau_1) \\ \sin(2\pi\tau_2) & \sin(6\pi\tau_2) & ... & \sin(2n\pi\tau_2) & \cos(2\pi\tau_2) & \cos(6\pi\tau_2) & ... & \cos(2n\pi\tau_2) \\ ... & ... & ... & ... & ... & ... \\ \sin(2\pi\tau_k) & \sin(6\pi\tau_k) & ... & \sin(2n\pi\tau_k) & \cos(2\pi\tau_k) & \cos(6\pi\tau_k) & ... & \cos(2n\pi\tau_k) \end{bmatrix}, \tag{4}$$

$$B = diag\left(1, \frac{1}{3}, ..., \frac{1}{n}, 1, \frac{1}{3}, ..., \frac{1}{n},\right) \quad \text{and} \quad S = \begin{bmatrix} \Re\{G(j\omega_o)\} \\ ... \\ \Re\{G(jn\omega_o)\} \\ \Im\{G(j\omega_o)\} \\ ... \\ \Im\{G(jn\omega_o)\} \end{bmatrix}. \tag{5}$$

where $\kappa_2(AB)$ is the euclidean condition number of AB and n is the last odd harmonic considered ($n = 2h - 1$). See proof in Appendix B, Proof (1).

Theorem 1 allows to select the times for sampling the response of relay feedback experiment in order to obtain reasonable condition numbers for matrix AB. Considering the simplest case, when $h = 1$, the possible values of τ obtained from equation (7) are:

$$|\tau_1 - \tau_2| = \frac{1}{4}, \frac{3}{4}.$$

This means that any pair of samples with these differences in time will make $\kappa_2(A) = 1$. Moreover, according to equation (8), for this particular case $\kappa_2(AB) = 1$.

According to equation (7), as more harmonics are considered, the restrictions between τ_k increase. Then, the task of obtaining τ_k, even if it is straightforward, becomes more tedious. For example, considering 2 harmonics, the restrictions between τ_k are:

$$|\tau_i - \tau_j| = \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8},$$
$$i = 1, 2, 3, 4;\ j = 1, 2, 3, 4;\ i \neq j$$

Some examples of combinations of τ_k in this case are the following:

1: $\tau_1 = 0, \quad \tau_2 = 0.125, \ \tau_3 = 0.25, \quad \tau_4 = 0.375,$

2: $\tau_1 = 0.625, \ \tau_2 = 0.75, \quad \tau_3 = 0.875, \ \tau_4 = 1,$

3: $\tau_1 = 0.125, \ \tau_2 = 0.375, \ \tau_3 = 0.75, \quad \tau_4 = 1,$

...

It is worth noticing that the solutions of equation (7) result in a finite number of combinations of $|\tau_m - \tau_n|$, covering the maximum and minimum possible differences between τ_m and τ_n. However, taking into account that $\tau_i \in [0, 1]$, there are infinite combinations of τ_i that fulfill the restrictions imposed by equation (7). In order to facilitate the selection of τ_i, the following corollary states a simple and useful criterion based on the number of harmonics considered for the estimation.

Corollary 1 *For h number of harmonics, the set of sample time fractions τ_k defined by:*

$$\tau_k = \frac{(k-1)}{4h}, \quad k = 1, 2, ..., 2h, \tag{9}$$

fulfills that $\kappa_2(AB) \leq n$, where n is the last odd harmonic considered. See proof in Appendix B, Proof (2).

Remark 1 *This corollary offers a very interesting result because for any number of harmonics h, the set of τ_k provided by equation (9) lies on the first semi-period of the limit cycle oscillations induced by the relay. Similarly, this result can be extended to the second semi-period, considering the same values of τ plus $T_o/2$, therefore, two identifications per period are possible. This is an advantage of the proposed method to be applied in auto-tuning algorithms, where the experiments must be short enough to obtain a suitable controller as fast as possible, without perturbing the processes for a long time.*

From now on, the sample time fractions τ_k will be obtained according to equation (9), which has been used to obtain the values of τ_k for some given number of harmonics, gathered in Table 1. However, it has to be kept in mind that equation (9) offers a particular solution that is extended for any number of harmonics, but there are other possible combinations of τ_k.

Considering the results exposed in the previous corollary, let us introduce the following example to show the application of the identification procedure.

Example 1 *Consider a given process to identify, which will be modeled by the transfer function G. For simplicity*

h	τ_k																			
	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	τ_{10}	τ_{11}	τ_{12}	τ_{13}	τ_{14}	τ_{15}	τ_{16}	τ_{17}	τ_{18}	τ_{19}	τ_{20}
1	0	$\frac{1}{4}$																		
2	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$																
3	0	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{5}{12}$														
4	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{6}{16}$	$\frac{7}{16}$												
5	0	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{7}{20}$	$\frac{8}{20}$	$\frac{9}{20}$										
6	0	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{5}{24}$	$\frac{6}{24}$	$\frac{7}{24}$	$\frac{8}{24}$	$\frac{9}{24}$	$\frac{10}{24}$	$\frac{11}{24}$								
7	0	$\frac{1}{28}$	$\frac{2}{28}$	$\frac{3}{28}$	$\frac{4}{28}$	$\frac{5}{28}$	$\frac{6}{28}$	$\frac{7}{28}$	$\frac{8}{28}$	$\frac{9}{28}$	$\frac{10}{28}$	$\frac{11}{28}$	$\frac{12}{28}$	$\frac{13}{28}$						
8	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{4}{32}$	$\frac{5}{32}$	$\frac{6}{32}$	$\frac{7}{32}$	$\frac{8}{32}$	$\frac{9}{32}$	$\frac{10}{32}$	$\frac{11}{32}$	$\frac{12}{32}$	$\frac{13}{32}$	$\frac{14}{32}$	$\frac{15}{32}$				
9	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{7}{36}$	$\frac{8}{36}$	$\frac{9}{36}$	$\frac{10}{36}$	$\frac{11}{36}$	$\frac{12}{36}$	$\frac{13}{36}$	$\frac{14}{36}$	$\frac{15}{36}$	$\frac{16}{36}$	$\frac{17}{36}$		
10	0	$\frac{1}{40}$	$\frac{2}{40}$	$\frac{3}{40}$	$\frac{4}{40}$	$\frac{5}{40}$	$\frac{6}{40}$	$\frac{7}{40}$	$\frac{8}{40}$	$\frac{9}{40}$	$\frac{10}{40}$	$\frac{11}{40}$	$\frac{12}{40}$	$\frac{13}{40}$	$\frac{14}{40}$	$\frac{15}{40}$	$\frac{16}{40}$	$\frac{17}{40}$	$\frac{18}{40}$	$\frac{19}{40}$

Table 1
Summary of τ_k values as a function of the number of harmonics considered.

in the calculus, we will perform the identification considering two harmonics ($h = 2$).

Then, using equation (9) the time fractions τ_k are obtained:

$$\tau_1 = 0,\ \tau_2 = \frac{1}{8},\ \tau_3 = \frac{2}{8},\ \tau_4 = \frac{3}{8}.$$

With these parameters, equation (3) results in equation (10), and then, the real and imaginary parts corresponding to the frequencies ω_o and $3\omega_o$ resulting from this system of linear equations are obtained as in equation (11), in which the error signal at the specified time fractions must be introduced to obtain the solution.

Note that the expressions presented in this system of equations are valid for any process to identify introducing the corresponding values for the error signal, i.e. no additional information about the process is required to perform the identification besides the error signal value at the required period fractions. To exemplify this let us consider two systems whose transfer functions are:

$$G_1(s) = \frac{e^{-s}}{s}, \quad G_2(s) = \frac{e^{-s}}{s+1}.$$

Figure 3 depicts the oscillation waveform for both systems and the samples to perform the identification according to corollary 1. Note that, regardless of the oscillation waveform, the samples are taken on specific fractions of the oscillation period (τ_k). This avoids searching the maximum amplitude of the signal among the samples, which is needed to perform the conventional relay identification.

Figure 4 presents the identified points obtained with both the conventional and the proposed methods. The actual placement of the oscillation points have been represented in green, the points identified with the conventional method and with the proposed method are represented in red and blue respectively. It can be seen how the new approach identifies better the oscillation points than the conventional method. This fact is shown numerically in Table 2 where the relative error of the estimation is presented. In both cases, the proposed method is more accurate than the conventional relay identification, reaching an improvement of 18.18% *and* 16.77%*. Besides, even if the accuracy with the proposed method is already higher than with the conventional one, more harmonics could be taken into account increasing even more the estimation accuracy of the oscillation point.*

4 Relation with the conventional relay method

The conventional relay method uses basically the Describing Function (DF) technique to estimate the ultimate gain of the system, [10]. The DF technique assumes that the high-order harmonics of the oscillation have a negligible effect with regard to the fundamental harmonic due to the low-pass band filtering properties of the linear part of the system. The DF method establishes that limit cycle oscillations due to a non-linearity on closed-loop systems exist if the open-loop transfer function intersects the negative inverse of the DF of the non-linearity. For the system in Figure 1 this condition results in the following equation:

$$G(j\omega_o) = -\frac{1}{\mathcal{N}}$$

$$\begin{bmatrix} e\left(0\right) \\ e\left(\frac{1}{8}T_o\right) \\ e\left(\frac{2}{8}T_o\right) \\ e\left(\frac{3}{8}T_o\right) \end{bmatrix} = -\frac{4\delta}{\pi} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0.7071 & 0.7071 & 0.7071 & -0.7071 \\ 1 & -1 & 0 & 0 \\ 0.7071 & 0.7071 & -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} \Re\{G(j\omega_o)\} \\ \Re\{G(j3\omega_o)\} \\ \Im\{G(j\omega_o)\} \\ \Im\{G(j3\omega_o)\} \end{bmatrix} \tag{10}$$

$$\begin{cases} \Re\{G(j\omega_o)\} &= -\frac{\pi}{4\delta}\Big(0.3536 \cdot e\left(\frac{1}{8}T_o\right) + 0.5 \cdot e\left(\frac{2}{8}T_o\right) + 0.3536 \cdot e\left(\frac{3}{8}T_o\right)\Big) \\ \Re\{G(j3\omega_o)\} &= -\frac{\pi}{4\delta}\Big(1.0607 \cdot e\left(\frac{1}{8}T_o\right) - 1.5 \cdot e\left(\frac{2}{8}T_o\right) + 1.0607 \cdot e\left(\frac{3}{8}T_o\right)\Big) \\ \Im\{G(j\omega_o)\} &= -\frac{\pi}{4\delta}\Big(0.5 \cdot e\left(0\right) + 0.3536 \cdot e\left(\frac{1}{8}T_o\right) - 0.3536 \cdot e\left(\frac{3}{8}T_o\right)\Big) \\ \Im\{G(j3\omega_o)\} &= -\frac{\pi}{4\delta}\Big(1.5 \cdot e\left(0\right) - 1.0607 \cdot e\left(\frac{1}{8}T_o\right) + 1.0607 \cdot e\left(\frac{3}{8}T_o\right)\Big) \end{cases} \tag{11}$$

	Classical method	**2-harm**	**Improvement**
$\|\mathbf{G_1(j\omega_o)} - \mathbf{\hat{G}_1(j\omega_o)}\|/\|\mathbf{G_1(j\omega_o)}\| \cdot \mathbf{100}$	23.4626	5.277	18.1826
$\|\mathbf{G_2(j\omega_o)} - \mathbf{\hat{G}_2(j\omega_o)}\|/\|\mathbf{G_2(j\omega_o)}\| \cdot \mathbf{100}$	18.8288	2.0513	16.7775

Table 2
Relative errors of the estimated points for $G_1(j\omega_o)$ and $G_2(j\omega_o)$.

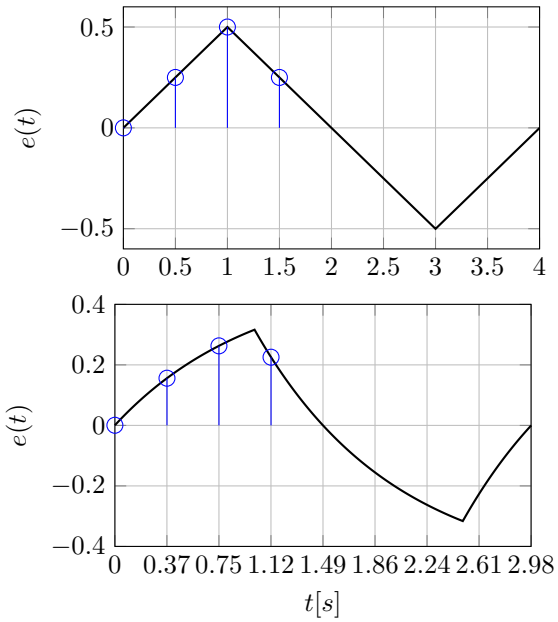

Fig. 3. Oscillation produced by the relay experiment on $G_1(s)$ (top) and $G_2(s)$ (bottom) and the samples taken for the application of the proposed method in blue.

where ω_o is the frequency of oscillation and $\mathcal{N}$ is the DF of the relay. As the equation of the DF of the relay is known ($\mathcal{N} = \frac{4\delta}{\pi A}$), the real and imaginary part of $G(j\omega_o)$ can be calculated:

$$\Re\{G(j\omega_o)\} = -\frac{\pi A}{4\delta}, \quad \Im\{G(j\omega_o)\} = 0 \tag{12}$$

where A is the amplitude of the oscillation.

Equations in (12) are a particular solution of the system in equation (6) by choosing only one harmonic. In order to prove that, let us consider the filtering hypothesis of the DF to be true, then the input to the relay will be a sinusoidal signal:

$$e(t) = A\sin(\omega_o t)$$

To include the parameter τ_k, the transformation $t = \tau_k T_o$ is done, resulting in:

$$e(\tau_k T_o) = A\sin(2\pi\tau_k).$$

As only one point of the frequency response is determined with the DF method, we will consider $h = 1$. Then, equation (3) results in equation (13).

In this case, the appropriated values of τ_k, according to corollary 1, are $\tau_1 = 0$ and $\tau_2 = 1/4$. Substituting these values in the previous expression we have:

$$\begin{bmatrix} 0 \\ A \end{bmatrix} = -\frac{4\delta}{\pi} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \Re\{G(j\omega_o)\} \\ \Im\{G(j\omega_o)\} \end{bmatrix},$$

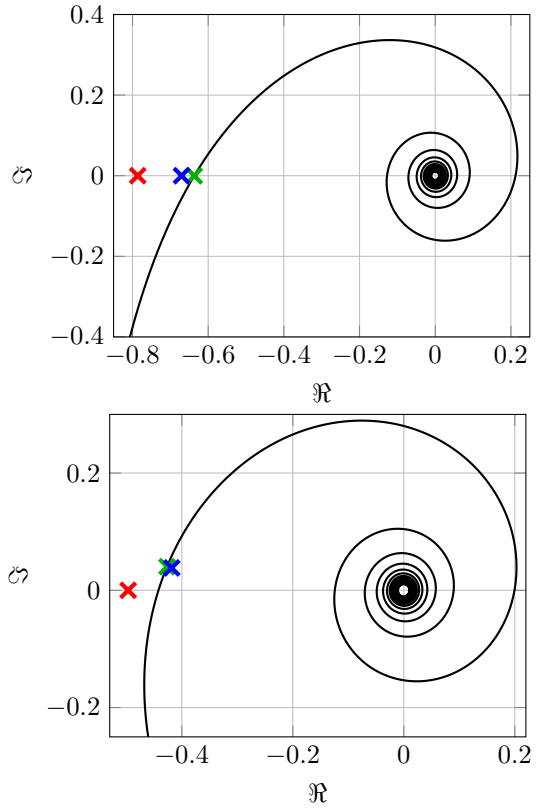

Fig. 4. Polar plots of systems $G_1(j\omega)$ (top) and $G_2(j\omega)$ (bottom) with the oscillation points (green) and the estimated oscillation points with the conventional method [10] (red) and proposed method (blue).

and solving:

$$\Re\{G(j\omega_o)\} = -\frac{\pi A}{4\delta}, \quad \Im\{G(j\omega_o)\} = 0.$$

Thus, the conventional approach of relay feedback identification, which was presented in [10], can be seen as a particular solution of the proposed method that is valid in those cases where the filtering hypothesis of the DF is fulfilled.

4.1 Extensive comparison study

In order to verify the validity of the proposed method to other systems than those considered in Example 1, it has been applied for a batch of 133 models, whose transfer functions are presented in equations (14), which describe the dynamic behavior of most of actual industrial systems, namely, models with multiple poles, complex and real poles, integrators, non-minimal phase and time delays, are included. This test batch was firstly proposed in [26], where more information about it can be found.

$$G(s) = \frac{e^{-s}}{Ts+1}, \quad T = 0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1, 1.3, 1.5, 2, 4, 6, 8, 10, 20, 50, 100, 200, 500, 1000 \tag{14a}$$

$$G(s) = \frac{e^{-s}}{(1+sT)^2}, \quad T = 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1, 1.3, 1.5, 2, 4, 6, 8, 10, 20, 50, 100, 200, 500 \tag{14b}$$

$$G(s) = \frac{e^{-s}}{(s+1)(Ts+1)^2}, \quad T = 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 2, 5, 10 \tag{14c}$$

$$G(s) = \frac{1}{(s+1)^n}, \quad n = 3, 4, 5, 6, 7, 8 \tag{14d}$$

$$G(s) = \frac{1}{(s+1)(\alpha s+1)(\alpha^2 s+1)(\alpha^3 s+1)}, \quad \alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \tag{14e}$$

$$G(s) = \frac{1}{s(1+sT_1)}e^{-sL_1}, \quad T_1 + L_1 = 1, \quad L_1 = 0.01, 0.02, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1 \tag{14f}$$

$$G(s) = \frac{Te^{-sL_1}}{(T_1 s+1)(Ts+1)}, T_1 + L_1 = 1, T = 1, 2, 5, 10, \quad L_1 = 0.01, 0.02, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1 \tag{14g}$$

$$G(s) = \frac{1-\alpha s}{(s+1)^3}, \quad \alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1 \tag{14h}$$

$$G(s) = \frac{1}{(s+1)((sT)^2+1.4sT+1)}, \quad T = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 \tag{14i}$$

The estimated points of the frequency response have been obtained with the proposed method considering 3 harmonics, for which the values of τ_k were obtained with equation (9). The estimation results are compared to the conventional relay approach (equations in (12)) as well as with two improved versions of the relay type experiments, concretely the methods known as preload-relay and saturation-relay, proposed in [14] and [15] respectively. These methods attempt to reduce the effect of neglecting the higher order harmonics by introducing slight modifications to the original relay test and calculations. The preload-relay incorporates a gain in parallel with the relay and the saturation-relay reduce the relay gain from infinite to certain value, transforming the relay into a saturation type non linearity.

The results are presented in Figures 5 and Figure 6, in which the relative error of the estimation obtained with each method and the improvements in terms of the

$$\begin{bmatrix} A\sin(2\pi\tau_1) \\ A\sin(2\pi\tau_2) \end{bmatrix} = -\frac{4\delta}{\pi} \begin{bmatrix} \sin(2\pi\tau_1) & \cos(2\pi\tau_1) \\ \sin(2\pi\tau_2) & \cos(2\pi\tau_2) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Re\{G(j\omega_o)\} \\ \Im\{G(j\omega_o)\} \end{bmatrix} \tag{13}$$

difference of relative errors are shown. It can be seen that, as expected, for the most models the relative error committed with the proposed method is lower than with the relay methods. This result is more evident in the case of the traditional relay, where reductions of about 20% in the relative error can observed in some cases. In particular, those cases where the reduction of the relative error is higher correspond to low order systems, i.e. the cases where the filtering hypothesis of the DF is hardly fulfilled.

In general, the modified relay methods improve the results obtained with the conventional relay, and its worth noting that in some few cases, these methods provide better estimations than the new proposal. This situation can be observed, for example, in models 43 to 47 for the preload-relay and models 10 to 21 for the saturation-relay. All these models present poor filtering properties because they have an isolated low frequency pole placed far from the other poles, which are placed at higher frequencies. The filtering effect of the low frequency pole does not attenuate sufficiently the higher order harmonics contribution, degrading the identification accuracy. Therefore, better results can be obtained by increasing the number of harmonics taken into account in the proposed method for these cases.

It is also important to highlight that the preload-relay and the saturation-relay methods improve their estimation results at expense of increasing the experiment duration. In fact, both methods require a two phase experiments. During the first stage a conventional relay experiment is developed in order to obtain the gain that is needed to modify the relay for the second phase. Once the gain is obtained, the second stage take place with the modified relay, at the end of which the point of frequency response where the phase of the system is $-\pi$ is estimated. Conversely, the approach proposed in this paper requires only a standard relay experiment and the estimations can be improved by increasing the number of harmonics considered in the calculations.

5 Simultaneous identification of multiple points of the frequency response

Note that in Example 1 the proposed method was used with 2 harmonics and consequently, the real and imaginary part of $G(j\omega_o)$ and $G(j3\omega_o)$ were calculated, see equation (11). In Figure 4, however, only the point corresponding to the fundamental harmonic ($G(j\omega_o)$) was represented. The truth is that the frequency response at $G(j3\omega_o)$ was calculated but with very low accuracy due to the fact that the successive harmonics have a relatively big impact on the calculation of the precedent points, i.e. harmonics $5\omega_o$, $7\omega_o$,... have an important effect on the estimation of $G(j3\omega_o)$. Therefore, in that example the estimation of $G(j3\omega_o)$ could have been improved by considering more harmonics, which implies obtaining the values of τ_k from equation (9) and solving systems of equations resultant from equation (6).

Increasing the number of harmonics considered in the calculus will improve the accuracy of the estimation of points corresponding to the frequency response of preceding harmonics. However, the points corresponding to the last harmonics considered in the calculus will have a big relative error in the estimation for the same reason that in the previous example. If an accurate estimation of these points is required, then more harmonics need to be considered. In order to illustrate the effect of high order harmonics in the estimation of other points of the frequency response of the system and in its quality, let us introduce Example 2.

Example 2 *Consider the system $G_2(s)$ from Example 1. As it has been said before, the estimation of $G_2(j3\omega_o)$ using our method with 2 harmonics is not very precise, then, in order to estimate this point with a higher accuracy let us use 3, 4 and 10 harmonics.*

The values of τ_k for $h = 3, 4, 10$ can be obtained using equation (9), or simply from Table 1. The sampling pattern for identification depending on the number of harmonics to be considered in the calculus are shown in Figure 7.

Using these samples and using equation (6), the real and imaginary parts of $G_2(j\omega_o)$ and $G_2(j3\omega_o)$ have been obtained. The estimation of these points of the frequency response for different number of harmonics considered in the calculus can be seen in Figure 8. It can be observed how by increasing the number of harmonics the estimation of $G_2(j\omega_o)$ and $G_2(j3\omega_o)$ becomes more and more accurate. This can be easily observed in Table 3, where numerical data are presented. The relative error is reduced for both $G_2(j\omega_o)$ and $G_2(j3\omega_o)$ with the increase of the number of harmonics, reaching relative errors lower than 0.09% and 0.8% for the estimation of $G_2(j\omega_o)$ and $G_2(j3\omega_o)$ respectively.

Ultimately, it is important to note that the increment in the number of harmonics also increases the computational demand of the algorithm and the data needed for identification. This can be easily seen in the size of the matrix AB which is $2h \times 2h$, where h is the number of harmonics considered. Therefore, the selection of h should be a trade-off between the complexity of the

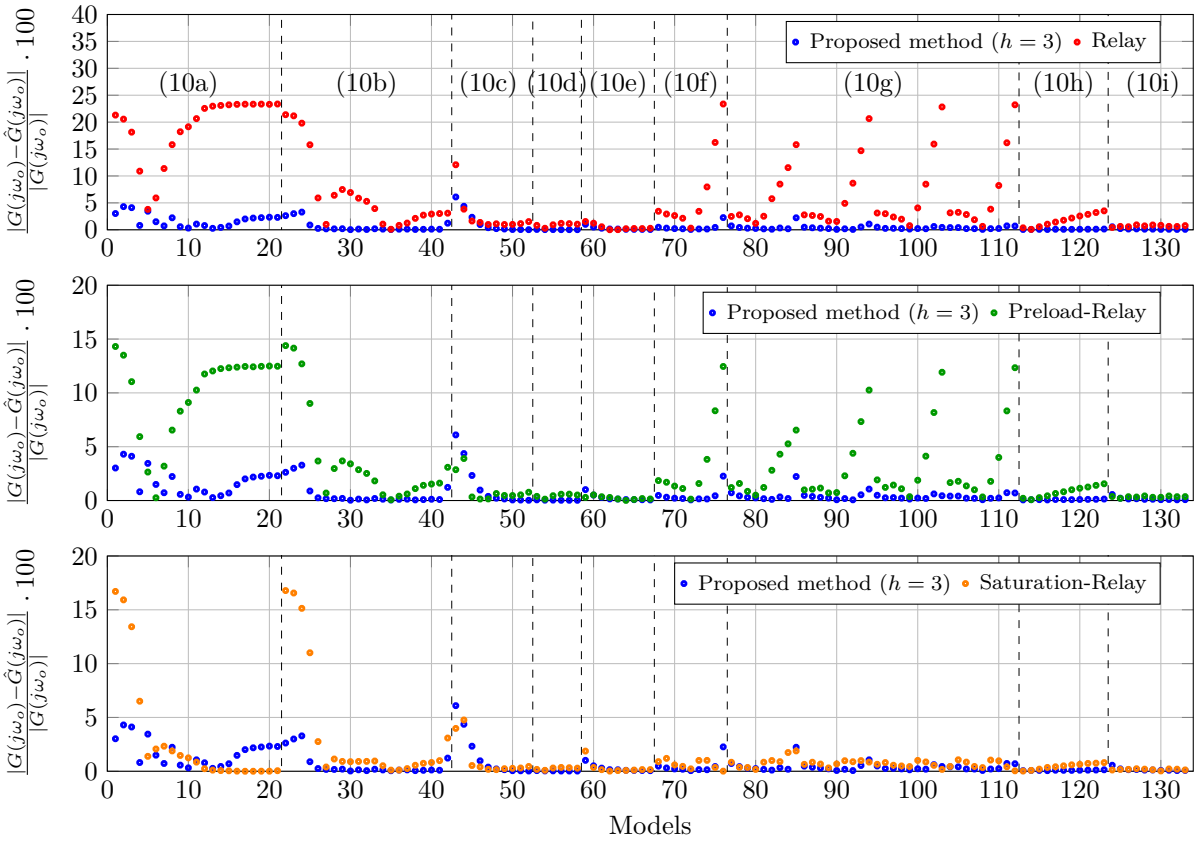

Fig. 5. Relative error of the estimated oscillation points with the proposed method with 3 harmonics and the conventional relay (top), preload-relay (middle) and saturation-relay (bottom) methods. Transfer functions of the models in the batch are indicated on the top figure.

	2-harm	**3-harm**	**4-harm**	**10-harm**
$\mathbf{\lvert G_2(j\omega_o) - \hat{G}_2(j\omega_o)\rvert / \lvert G_2(j\omega_o)\rvert \cdot 100}$	2.0513	2.2305	0.5329	0.082
$\mathbf{\lvert G_2(j3\omega_o) - \hat{G}_2(j3\omega_o)\rvert / \lvert G_2(j3\omega_o)\rvert \cdot 100}$	32.9645	20.8298	5.2931	0.7782

Table 3
Relative errors of the estimated points for $G_2(j\omega_o)$ and $G_2(j3\omega_o)$.

algorithm, the available data and the required identification accuracy.

6 Noise and sampling effects on the identification

6.1 Measurement noise effect

By considering an ideal relay in the relay feedback experiments scheme, Figure 1, it is clear that the presence of noise in the loop will affect both the measurement of the error signal and the relay output, which, in turn, will affect the oscillation waveform because of the extra switches produced near to the commutation point due to the noise. A common choice to avoid this behavior is to substitute the ideal relay for a relay with hysteresis whose value is slightly greater than the noise amplitude.

The inclusion of a hysteresis band avoids effectively the switches due to the noise but, in the conventional approach, the resulting describing function that models the non-linearity changes as it is shown in [27], and thus the inclusion of this hysteresis band has to be taken into account in the identification process. However, the inclusion of hysteresis in the relay does not affect the proposed method, which remains valid with the same equations. Note that in Appendix A the error signal equation has been obtained starting by approximating the square wave produced by the relay using a Fourier series without assuming the kind of non-linearity that produces it. Afterwards, this expression has been pushed and computed through the loop until obtaining the error signal.

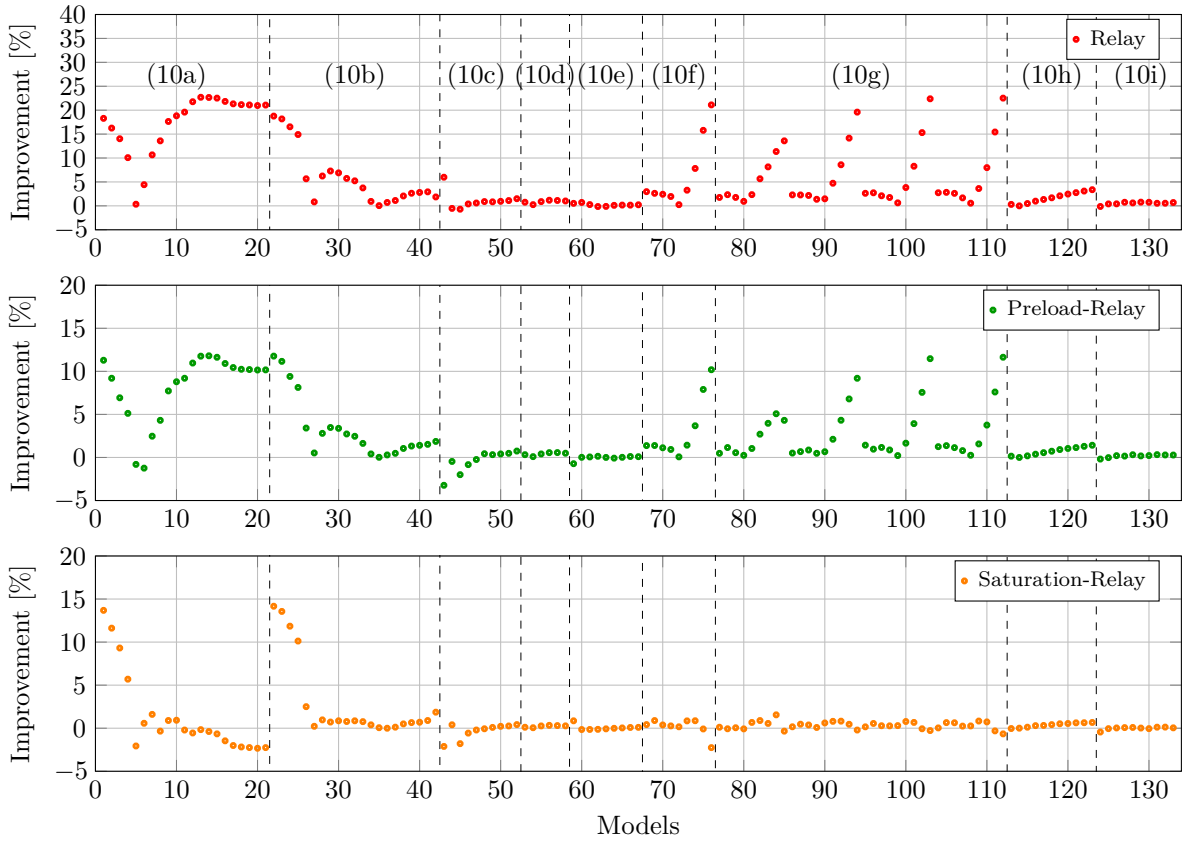

Fig. 6. Improvement (difference) achieved with the proposed method with 3 harmonics compared with the conventional relay (top), preload-relay (middle) and saturation-relay methods (bottom). Transfer functions of the models in the batch are indicated on the top figure.

Then, this equation is valid, not only for the ideal relay, but also for any non-linearity that produces a square signal with symmetric output levels. Thus, as the relay with hysteresis also produces a square wave in its output, the equation of the error signal has the same structure as with the ideal relay. In this case, however, the location of estimated points will correspond to the limit cycles induced by this kind of relay. To illustrate that the proposed method, remains valid even when a relay with hysteresis is used, let us introduce the following example:

Example 3 *Consider the system G_2 from Example 1. In this case, consider two relays with hysteresis, which have as output levels $\pm\delta$ and the hysteresis commutation thresholds are $\pm 1/2\delta$ and $\pm 3/4\delta$ respectively.*

To perform the identification we will use 4 harmonics, and as it has been done before, the values of τ_k are obtained from equation (9) or Table 1 and the real and imaginary parts of the frequency response are obtained from equation (6). The results are shown in Figure 9, where it has been also represented the results when using the ideal relay and considering the same number of harmonics in the calculus. As expected, the fact of adding a hysteresis to the relay makes the identified point to shift in frequency, which corresponds to the limit cycle induced by the new non-linearity considered in this case. Concretely, a lower frequency oscillation is obtained as the hysteresis grows. These new points, however, are accurately estimated by the proposed method without introducing any change. The relative errors in the estimation are presented in Table 4, where it can be effectively seen the validity of the method for the case of oscillations induced by a relay with hysteresis.

Even if the extra switches are avoided as a consequence of using a relay with hysteresis, the error signal could contain a non negligible noise level. A simple strategy, which has been widely used in other relay based identification methods, is to reduce the SNR (Signal to Noise Ratio) by increasing the relay output level δ. Its effectiveness can be proven as follows.

Consider $\hat{S}$ to be the estimated solution obtained with a sampled error vector with noise $\hat{E}$, which relates with the error signal E with a noise vector u: $\hat{E} = E - u$.

	Ideal relay	**hist = 1/2δ**	**hist = 3/4δ**
$\|\mathbf{G_2(j\omega_o)} - \mathbf{\hat{G}_2(j\omega_o)}\|/\|\mathbf{G_2(j\omega_o)}\| \cdot \mathbf{100}$	0.5329	0.7989	0.7184
$\|\mathbf{G_2(j3\omega_o)} - \mathbf{\hat{G}_2(j3\omega_o)}\|/\|\mathbf{G_2(j3\omega_o)}\| \cdot \mathbf{100}$	5.2931	6.8478	5.8732

Table 4
Relative errors of the estimated points for $G_2(j\omega_o)$ and $G_2(j3\omega_o)$.

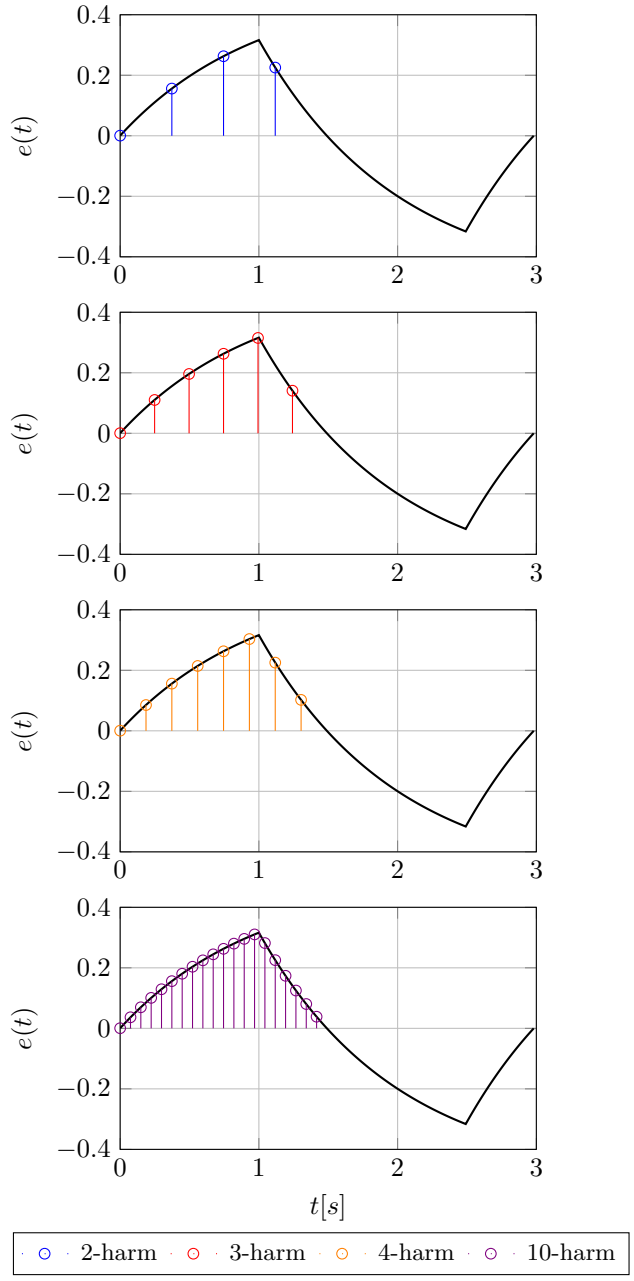

Fig. 7. Oscillation waveform resulting from relay feedback experiments for system G_2 and samples taken depending on the number of harmonics.

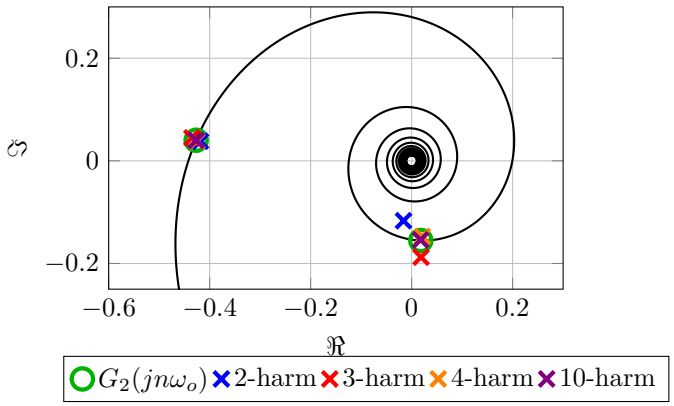

Fig. 8. Polar plot of $G_2(j\omega)$ and $G_2(j\omega_o)$ and $G_2(j3\omega_o)$ estimated with different number of harmonics.

Then the relative error on the solution defined by:

$$\frac{\left\|S - \hat{S}\right\|}{\|S\|}$$

it is related with the noise source of uncertainty and is bounded by:

$$\frac{1}{\kappa(AB)}\frac{\|u\|}{\|E\|} \leq \frac{\left\|S - \hat{S}\right\|}{\|S\|} \leq \kappa(AB)\frac{\|u\|}{\|E\|} \qquad (15)$$

This expression is proven in Appendix B, Proof (3).

Let E be the vector of samples obtained with a relay drive level δ, and $\mathcal{E}$ the vector of samples obtained with $m\delta$. Regarding to expression (2) it can be directly deduced that the samples are related by $E_i m = \mathcal{E}_i$ since the drive levels affect proportionally that expression, and therefore, $\|\mathcal{E}\| = |m|\,\|E\|$. Thus, the relative error caused by the noise, and affected by the change of the relay drive level, is bounded by:

$$\frac{1}{|m|\,\kappa(AB)}\frac{\|u\|}{\|E\|} \leq \frac{\left\|S - \hat{S}\right\|}{\|S\|} \leq \frac{\kappa(AB)}{|m|}\frac{\|u\|}{\|E\|}$$

This expression shows how the uncertainty on the solution due to noise is bounded and stresses the importance of the condition number of AB. In addition, the fact of increasing the magnitude of the drive levels of the relay

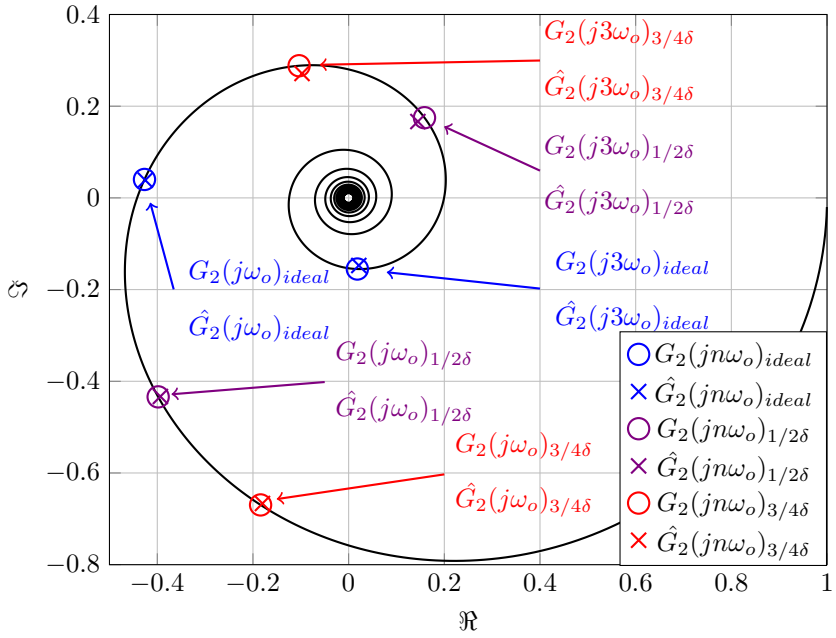

Fig. 9. Polar plot of $G_2(j\omega)$ and $G_2(j\omega_o)$ and $G_2(j3\omega_o)$ identified using an ideal relay and two relays with hysteresis.

shows to be effective for reducing both the upper and lower bounds of the uncertainty, being beneficial for the identification accuracy.

In order to illustrate this idea let us introduce the following example:

Example 4 *Consider the system G_2 from Example 1. For this case, consider additionally that we have a measurement noise whose peak-to-peak amplitude is* 0.1. *As commented before, the oscillations produced by the ideal relay cannot be used for identification in this case because its output is not a square signal. This fact can be seen in Figure 10, where the error signal $e(t)$ and the output of the relay $u(t)$ are shown for system G_2 using an ideal real in the loop. As can be seen, the noise causes a burst of switches when error signal is close to zero. In order to avoid this undesired behavior, a hysteresis band is added to the relay. We set the upper and lower switch points to* 0.05 *and* -0.05 *respectively, this will avoid the appearance of extra switches due to the noise. In addition, if we increase δ the impact of the noise is decreased. This last case has been illustrated in Figure 10, in which the above mentioned hysteresis has been added and the output levels of the relay are triplicated with regard to the original experiment.*

Once the bursts of switches due to the noise have been avoided, a square wave signal is obtained at the output of the non-linearity. Then we apply the proposed method using 3 harmonics. The values of τ_k are obtained using the equation (9) or simply from Table 1. Then, using equation (6) the real and imaginary parts of $G_2(\omega_o)$ are obtained. The results are represented in Figure 11, in

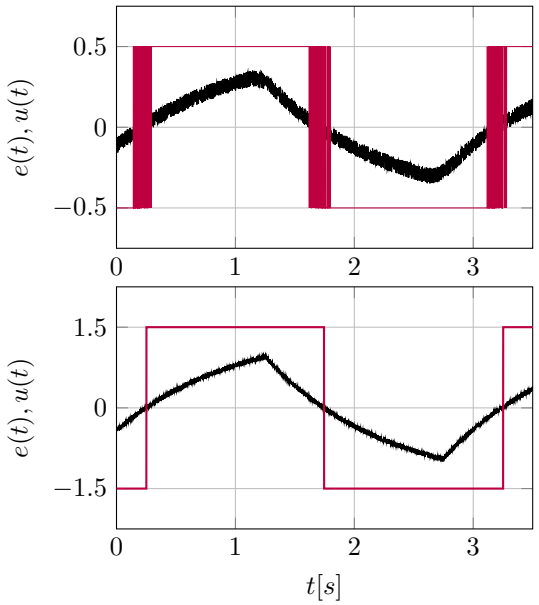

Fig. 10. Effect on the output of the relay $u(t)$ caused by the noise in the error signal $e(t)$ (top), and the adopted solution (bottom).

which it can be seen that the proposed method offers an estimation of the oscillation point which is more accurate as the output of the relay δ increases.

In Table 5, the relative errors of the estimated points are presented. The case from Example 2, where the ideal relay was used, has been included in order to compare the accuracy of both solutions. As it can be seen, the relative error decreases as δ increases, confirming that

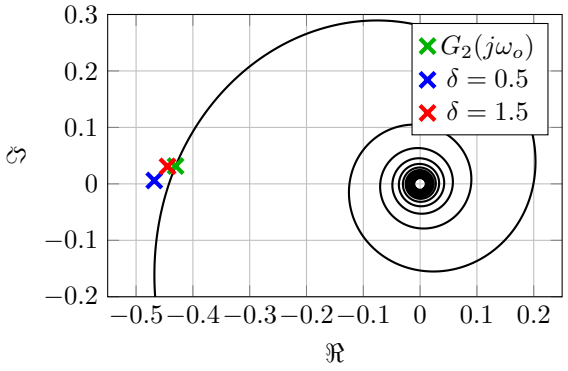

Fig. 11. Polar plot of $G_2(j\omega)$ and $G_2(j\omega_o)$ identified using a relay with hysteresis and different values for δ.

	Without noise	Noise $\delta = 0.5$	Noise $\delta = 1.5$
$\frac{\lvert\mathbf{G_2(j\omega_o)}\text{-}\hat{\mathbf{G}}_\mathbf{2}\mathbf{(j\omega_o)}\rvert}{\lvert\mathbf{G_2(j\omega_o)}\rvert} \cdot \mathbf{100}$	2.2305	10.5290	3.5952

Table 5
Relative error of the estimated point of $G_2(j\omega_o)$ using different values of δ in noisy signals compared to the case without noise.

the accuracy in noisy signals increases with δ. In fact, for $\delta = 1.5$ the accuracy is comparable to the ideal relay case without noise.

6.2 Sampling period effect

The proposed identification method relies on the proper selection of the samples to obtain a well-conditioned matrix to estimate the frequency response points accurately. However, the sampling period is commonly not precise enough to obtain the exact measure at the sample time fractions τ_i needed to perform the calculus. Strategies like interpolating the measured data can be deployed to correct that measurement mismatch, which could increase significantly the computation time needed to perform the identification. In this subsection, the effects on the accuracy of the sampling are studied when the proposed identification method is applied.

Considering that the effect of not choosing the exact samples due to a mismatch between the oscillation and sampling periods is reflected in a variation on AB and E matrices, the relative error between the expected and the obtained solution is bounded by:

$$\frac{\left\|S - \tilde{S}\right\|}{\|S\|} \leq \frac{\kappa(AB)}{1 - \kappa(AB)\,\|\Delta\| \,/\, \|AB\|} \left(\frac{\|\Delta\|}{\|AB\|} + \frac{\|\zeta\|}{\|E\|} \right) \tag{16}$$

where Δ and ζ are the additive error matrix and vector respectively on AB and E such that $\tilde{AB} = AB - \Delta$ and $\tilde{E} = E - \zeta$. This expression has been obtained in Proof (4) in Appendix B.

It can be observed that this expression is similar to the bound on the relative error produced by the noise presented in equation (15), but in that expression no variation on AB has been considered, and the source of uncertainty is different.

From equation (16) it can be seen that these uncertainties raise the upper bound of the relative error of the solution, which could induce to a high error. However, in most applications the sampling frequency is typically higher than system's dynamics, and therefore, enough samples to compute the identification method are available, which can lead to consider the small variations on AB and E negligible.

Besides, it must be kept in mind that not only the sample time fraction τ_i specified from Corollary 1 are valid, but any combination of τ_i that fulfills Theorem 1 would be appropriate for the identification, which provides multiple combinations of sample time fractions to estimate the frequency response.

7 Least Squares approach

The initial problem presented in Equation (3) can be formulated in a Least Squares (LS) fashion. This means that in order to determine some points in the frequency response of the system more equations can be considered. Equation (3) remains valid, here it has been presented with the dimension of each matrix in subindex:

$$E_{k\times 1} = -\frac{4\delta}{\pi} A_{k\times 2h} B_{2h\times 2h} S_{2h\times 1} \tag{17}$$

k being the number of equations and h the number of frequency response point to determine, which for LS method $k > 2h$.

However, despite the possibility of deploying the presented equations in a LS fashion, it does not necessarily imply an increase of precision on the obtained solution. Equations can still be linearly dependent, for example, despite having a system of linear equations (17) with $k \gg 2h$. Therefore, the question of how to chose the samples to calculate the estimated points on the frequency response remains.

Corollary 2 *The results obtained from Theorem 1 and its application derived in Corollary 1 can also be applied to the linear problem presented in equation (17) adapting the expressions for the LS formulation since matrix A is rectangular. Therefore, $\kappa(AB) \leq n$, where n is the last harmonic considered, if the sample time fractions τ_k fulfill:*

$$|\tau_i - \tau_j| = \frac{1}{2k}, \frac{2}{2k}, \cdots, \frac{k-1}{2k}, \frac{k+1}{2k}, \cdots, \frac{2k-1}{2k}, \quad \forall i, j \quad i \neq j, \tag{18}$$

where k is the number of equations. Similarly to Corollary 1, the sample time fractions τ_i can be chosen as:

$$\tau_i = \frac{(i-1)}{2k}, \quad i = 1, 2, ..., k. \tag{19}$$

to obtain a particular solution of the choices derived from expression (18). See proof in Appendix B, Proof (5).

Note that the the sample time fractions selection proposed in Corollary 2 are the same than the ones indicated by Corollary 1 in the case that the number of equations k is equal to twice the number of harmonics to calculate h. For example, if the LS approach is taken considering $k = 20$ equations, the sample time fractions obtained with equation (19) would be the same that those proposed in Table 1 for $h = 10$ harmonics.

Due to Corollary 2 the problem expressed in a Least Squares formulation allows to choose the samples so that the system is well conditioned. The following example shows the application of the LS approach and compares it with the originally proposed method.

Example 5 *Consider the process whose transfer function is defined by:*

$$G(s) = \frac{e^{-s}}{(0.5s+1)^2}$$

The identification problem has been tackled using the LS formulation presented in equation (17) considering ten equations ($k = 10$) to estimate the points of the frequency response corresponding to ω_o and $3\omega_o$, that is $h = 2$. For comparison purpose, the square matrix version of the estimation method, whose solution is given by equation (6), has been also applied using the same number of equations but assuming $h = 5$ in order to make the matrix A square.

According to Corollary 2, as both approaches consider the same number of equations, the sample time fractions would be the same, and can be read from Table 1 in the row for $h = 5$. From the stable oscillation induced by the relay that is presented in the left image of Figure 12, the corresponding samples have been taken, which have been highlighted in red. With those samples the identification with both formulation has been computed obtaining the points in the frequency response presented in the right image of Figure 12. In blue the points identified with the LS method are presented, in green the points obtained considering A a square matrix and in red the exact points.

As it can be seen, both identifications are very accurate. In the point associated with the fundamental (ω_o) and third ($3\omega_o$) harmonic the relative estimation errors are 0.082% *and* 0.7932% *respectively for both methods. The method that considers matrix A square has an accuracy of* 1.87%, 16.39% *and* 53.45% *in the next three estimated points. As expected, the relative error of higher harmonics is greater than for the lower ones, however, it is remarkable that both methods obtain the same accuracy for points $G(j\omega_o)$ and $G(j3\omega_o)$. This effect is explained in the lines below.*

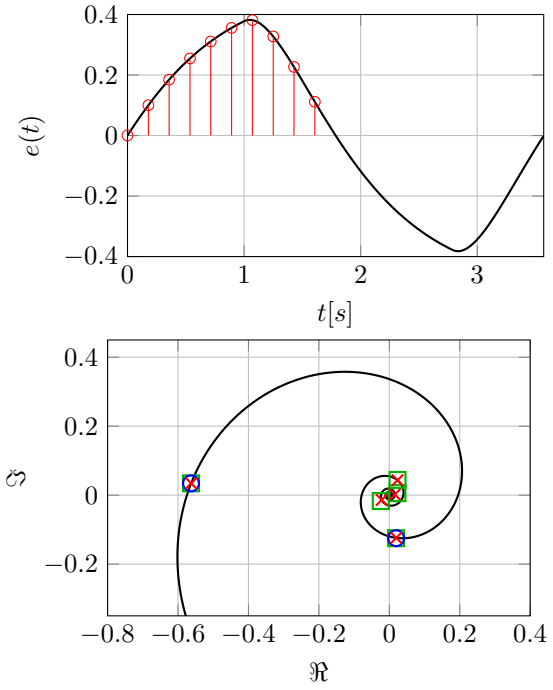

Fig. 12. Stable oscillation induced by the relay and samples taken to compute the identification (top), and frequency response estimation with both formulations (bottom), in blue with LS and in green being A square.

In fact, the frequency response points that are obtained with both approaches for ω_o and $3\omega_o$ are the same. This is due to the construction of the matrix A and the sample vector E in both formulations and the calculations performed to obtain the real and imaginary parts of these points. As explained before, the sample time fractions τ_i, and therefore, the samples $e(\tau_i T_o)$ are the same for both approaches. Besides, to obtain the frequency response points, the inversion of matrices is performed, obtaining either $B^{-1}A^{-1}$ or $B^{-1}A^{\dagger}$. These two matrices, contain some identical row vectors, which are those rows used to obtain the real or imaginary part of a given point $G(jn\omega_o)$. As these rows multiply the same sample vector E, the obtained result is the same.

Concretely in this example, for the LS formulation the matrix $B^{-1}A^{\dagger}$ is of size 4×10 and for the other formulation 10×10. The first and the second rows of $B^{-1}A^{\dagger}$ are the same that the first two rows of $B^{-1}A^{-1}$, which compute the real parts of the first and second estimated points. Similarly, the third and forth rows of $B^{-1}A^{\dagger}$ are the same that the sixth and seventh of $B^{-1}A^{-1}$, this last two rows are the ones that compute the imaginary parts of these points. Therefore, the result on the estimation

of $G(j\omega_o)$ and $G(j3\omega_o)$ with both approaches is the same when considering the same number of equations k.

Since the LS approach is equivalent to increase the number of harmonics considered in the square matrices version, comparing it with the formulation with A square for the same number of harmonics, i.e. $k = 4$ and $h = 2$, the results obtained with LS will be better because of the increased number of equations $k \gg 4$. This effect of the increase of the accuracy when more harmonics are taken into account was shown in Example 2.

The previous example shows that, the solution obtained by the LS strategy (equation (17)) with $k > 2h$ equations to estimate h points of the frequency response is the same that the first h points estimated when considering k equations to estimate $k/2$ points using equation (6). Additionally, the latter approach provide estimation results for $k/2 - h$ further points.

As it has been commented on Section 6.1, the relative uncertainty of the solution due to the presence of noise in the sampled data can be bounded by equation (15), which remains valid for the LS approach considering in its demonstration the matrix A rectangular. This expression establishes the condition number, $\kappa(AB)$, as a key factor in limiting the uncertainty. The upper bound of the condition number is determined by the matrix B, which contains the fractions $1/n$ relative to each considered harmonic n, and since the condition number of A is one, the condition number is bounded by $\kappa(AB) \leq n$. As for the LS square approach the number of harmonics considered is lower than for the square matrix version, the condition number of matrix AB is lower in LS than in the square version. This difference in $\kappa(AB)$ in each approach modifies the upper bound of the uncertainty of the solution due to the noise.

However, the arithmetic operations used to calculate the points with both approaches are the same, which means that if noise is present, it would affect in the same manner regardless of the approach. Hence, even if the upper bound of LS is lower than for the square matrix version, the LS approach does not really provide more accuracy when noise is present on the samples. The increase in the bound of the uncertainty due to noise produced by $\kappa(AB)$ in the square matrix approach is due to the fact that more points of the frequency response are calculated, accumulating their uncertainty to the bound through $\kappa(AB)$. Then, the strategy provided in section 6.1, which consists in increasing the drive level of the relay δ, is still recommendable for the LS approach to reduce the upper bound of the error due to noise.

Similarly, with regard to selection of the sample time fraction, the expression (16) provided in Section 6.2 remains valid considering matrix A rectangular. If the sampling frequency is assumed to be higher than system's dynamics, the errors produced in AB and E matrices can be neglected. Conversely, it is worth remarking that the sample time fractions proposed in equation (19) in Corollary 2 are not the only ones that make $\kappa(AB)$ minimum, which are a particular solution of equation (18).

Summarizing, the proper selection of the sample time fractions has proved to produce the same results if the same number of equations k are used regardless of the followed approach, which can be either the described in Section 3 or the LS approach described in this section. Then, if a given number of points of the frequency response are to be estimated, LS approach can be used to reduce the sizes of matrices A and B with regard to those of the square matrices approach that guarantee the same accuracy.

8 Conclusions

In this paper, a new method to obtain the frequency response points of systems from the relay experiment has been presented. Unlike previous approaches which are based on the describing function technique, this method can take into account the effect of several harmonics in the calculus, increasing the accuracy of the estimation as more harmonics are considered.

An extensive simulation study over a batch of models has been carried out using the proposed procedure and some of the most used relay-based identification methods. The proposal improves the results obtained by the conventional approaches for the majority of the models, reducing the relative error of the estimation in around 15% in some cases.

Besides, the identification of several points of the frequency response simultaneously is possible with the proposed method. Particularly, those that correspond to odd multiples of the oscillation frequency, using the same relay experiment data. Nevertheless, in order to determine its frequency response, more harmonics have to be taken into account in the calculus, increasing the computational cost of the algorithm. However, if the size of the resulting matrices is not a drawback for certain applications, considering more harmonics increases significantly the accuracy of the method.

Additionally, the effect of the noise and the sampling in the estimation is evaluated. It is shown that the method can be applied to noisy temporal responses by substituting the ideal relay by a relay with hysteresis, which constitutes one of the most common solutions. The method proves to be successful when using a relay with hysteresis and shows to increase its accuracy as the output of the relay δ increases.

The proposed method also allows a Least Squares formulation, which has been studied. The principles that

rule the estimation methodology proposed are also applicable to the LS approach, offering the same results, in terms of accuracy, while reducing the size of the matrices to consider, but at the expense of estimating less points of the frequency response.

Appendix A Fourier series expansion of $\mathbf{e(t)}$

Consider the signal $u(t)$ to be a square signal with pulse width modulated by ρ as shown in Figure 2, then the expression of $u(t)$ can be obtained through Fourier series expansion:

$$u(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \Big(a_n \cos(n\omega_o t) + b_n \sin(n\omega_o t) \Big),$$

where a_0:

$$a_0 = \frac{2}{T_o} \int_0^{\rho T_o} (\delta_o + \delta) dt + \frac{2}{T_o} \int_{\rho T_o}^{T_o} (\delta_o - \delta) dt = 2\delta(2\rho - 1) + 2\delta_o,$$

and a_n and b_n:

$$a_n = \frac{2}{T_o} \int_0^{\rho T_o} (\delta_o + \delta) \cos(n\omega_o t) dt + \frac{2}{T_o} \int_{\rho T_o}^{T_o} (\delta_o - \delta) \cos(n\omega_o t) dt,$$

$$b_n = \frac{2}{T_o} \int_0^{\rho T_o} (\delta_o + \delta) \sin(n\omega_o t) dt + \frac{2}{T_o} \int_{\rho T_o}^{T_o} (\delta_o - \delta) \sin(n\omega_o t) dt.$$

After some calculations:

$$u(t) = \delta(2\rho - 1) + \delta_o + \frac{2\delta}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \Big(\sin(2\pi n\rho) \cos(n\omega_o t) + (1 - \cos(2\pi n\rho)) \sin(n\omega_o t) \Big),$$

and arranging:

$$u(t) = \delta(2\rho - 1) + \delta_o + \frac{4\delta}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin(\pi n\rho) \cos(\pi n\rho - n\omega_o t).$$

Then, $y(t)$ is obtained by the relation $y = G \cdot u$, where G is the open-loop transfer function of the system, which results in expression (20), and then as $e(t) = y_{ref}(t) - y(t)$, the error signal $e(t)$ is obtained, resultin in equation (21).

Appendix B Mathematical proofs

Proof (1) Let us consider the euclidean norm ($\|\cdot\|_2$). The euclidean condition number of B is the ratio between the maximum and minimum eigenvalues:

$$\kappa_2(B) = \frac{\lambda_{max}(B)}{\lambda_{min}(B)} = \frac{1}{1/n} = n.$$

Then, applying the results on Lemma 1 about the condition number of the product of matrices, the condition number of A must be forced to 1. This can be attained by forcing A to be orthogonal. By definition, an orthogonal matrix Q follows that $QQ^T = Q^TQ = I$, which can be reformulated as:

$$\langle q_i, q_j \rangle = \begin{cases} 0, & \forall i, j, \quad i \neq j \\ 1, & \forall i, j, \quad i = j \end{cases}$$

where q_i, q_j are the ith and jth row vectors of Q.

Defining α_k as the kth row vector of A, and re-writing $\theta_k = 2\pi\tau_k$ for the sake of brevity, the terms of the product AA^T are:

$$\begin{aligned} \langle \alpha_i, \alpha_j \rangle &= \sum_{p=1}^{h} \sin((2p-1)\theta_i) \sin((2p-1)\theta_j) \\ &\quad + \cos((2p-1)\theta_i) \cos((2p-1)\theta_j) \\ \langle \alpha_i, \alpha_j \rangle &= \sum_{p=1}^{h} \cos((2p-1)(\theta_i - \theta_j)) \end{aligned} \tag{22}$$

Considering $i \neq j$ in equation (22), i.e. the terms outside the main diagonal, we can solve $\langle \alpha_i, \alpha_j \rangle = 0$, which has the following general solution in terms of τ and h:

$$|\tau_i - \tau_j| = \frac{1}{4h}, \frac{2}{4h}, \cdots, \frac{2h-1}{4h}, \frac{2h+1}{4h}, \cdots, \frac{4h-1}{4h}.$$

Considering $i = j$ in equation (22) result in $\langle \alpha_i, \alpha_j \rangle = h$ regardless of the values θ_k, so the matrix A will be similar to an orthogonal matrix, more specifically:

$$AA^T = A^TA = hI. \tag{23}$$

From the definition of the condition number and the euclidean norm:

$$\begin{aligned} \kappa_2(A) &= \|A\|_2 \left\|A^{-1}\right\|_2 \\ &= \sqrt{\lambda_{max}(A^TA)} \sqrt{\lambda_{max}((A^{-1})^TA^{-1})}, \end{aligned}$$

where $\lambda_{max}(\cdot)$ denotes the largest eigenvalue of the matrix. Applying some matrix algebra and the result on equation (23):

$$\begin{aligned} \kappa_2(A) &= \sqrt{\lambda_{max}(A^TA)\lambda_{max}((AA^T)^{-1})} \\ &= \sqrt{\lambda_{max}(hI)\,\lambda_{max}\left(\frac{1}{h}I\right)} = \sqrt{h\frac{1}{h}} = 1. \end{aligned}$$

□

$$y(t) = (\delta(2\rho-1)+\delta_o)G(0) + \frac{4\delta}{\pi}\sum_n^\infty \frac{1}{n}\sin(n\pi\rho)\Big(\Re\{G(jn\omega_o)\}\cos(n\pi\rho - n\omega_o t) + \Im\{G(jn\omega_o)\}\sin(n\pi\rho - n\omega_o t)\Big) \quad (20)$$

$$e(t) = y_{ref}(t) - (\delta(2\rho-1)+\delta_o)G(0) - \frac{4\delta}{\pi}\sum_n^\infty \frac{1}{n}\sin(n\pi\rho)\Big(\Re\{G(jn\omega_o)\}\cos(n\pi\rho - n\omega_o t) + \Im\{G(jn\omega_o)\}\sin(n\pi\rho - n\omega_o t)\Big) \quad (21)$$

Proof (2) For a given number of harmonics h, consider in equation (7) the differences between the consecutive sample time fractions to be:

$$\tau_k - \tau_{k-1} = \frac{1}{4h}, \quad k = 2, ..., 2h$$

and choosing $\tau_1 = 0$, the set of τ_k is:

$$\tau_1 = 0,\ \tau_2 = \frac{1}{4h},\ \tau_3 = \frac{2}{4h},\ \dots,\tau_{2h} = \frac{2h-1}{4h}.$$

□

Proof (3) For the sake of brevity consider $\mathcal{A} = AB$. Then, since $\|E\| = \|\mathcal{A}S\| \leq \|\mathcal{A}\|\,\|S\|$ and $S - \hat{S} = \mathcal{A}^{-1}u$:

$$\frac{\left\|S-\hat{S}\right\|}{\|S\|} = \frac{\left\|\mathcal{A}^{-1}u\right\|}{\|S\|} \leq \frac{\|\mathcal{A}\|\left\|\mathcal{A}^{-1}\right\|\|u\|}{\|E\|} = \kappa\frac{\|u\|}{\|E\|}$$

being κ the condition number of $\mathcal{A}$. Furthermore, $\|u\| = \left\|\mathcal{A}(S-\hat{S})\right\| \leq \|\mathcal{A}\|\left\|S-\hat{S}\right\|$ and $\|S\| \leq \left\|\mathcal{A}^{-1}\right\|\|E\|$ imply that:

$$\frac{\left\|S-\hat{S}\right\|}{\|S\|} \geq \frac{\|u\|}{\|\mathcal{A}\|\,\|S\|} \geq \frac{\|u\|}{\|\mathcal{A}\|\,\|\mathcal{A}^{-1}\|\,\|E\|} \geq \frac{1}{\kappa}\frac{\|u\|}{\|E\|}$$

□

Proof (4) Let $Ax = b$ be the exact system of equations and $(A-E)\tilde{x} = b - e$ the approximation due to uncertainties in A and b. Let $B = A^{-1}E$ such that $\|B\| < 1$. Then, considering also that $(A-E) = A(I-B)$:

$$\begin{aligned} x - \tilde{x} &= A^{-1}b - (A-E)^{-1}(b-e) \\ &= A^{-1}b - (A(I-B))^{-1}(b-e) \\ &= A^{-1}b - (I-B)^{-1}A^{-1}(b-e) \\ x - \tilde{x} &= (I-(I-B)^{-1})A^{-1}b + (I-B)^{-1}A^{-1}e \end{aligned}$$

Therefore, by considering $1/\|x\| \leq \|A\|/\|b\|$:

$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq \left\|I-(I-B)^{-1}\right\| + \frac{\left\|(I-B)^{-1}\right\|\left\|A^{-1}\right\|\|e\|\,\|A\|}{\|b\|}$$
$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq \left\|I-(I-B)^{-1}\right\| + \left\|(I-B)^{-1}\right\|\kappa(A)\frac{\|e\|}{\|b\|}$$

where $\kappa(A)$ is the condition number of A. Since $\|B\| < 1$, $(I-B)^{-1}$ can be developed by Neumann series:

$$(I-B)^{-1} = \sum_{k=0}^\infty B^k$$

Then:

$$\left\|(I-B)^{-1}\right\| \leq \sum_{k=0}^\infty \|B\|^k = \frac{1}{1-\|B\|}$$

and similarly:

$$\left\|I-(I-B)^{-1}\right\| \leq \frac{\|B\|}{1-\|B\|}$$

thus, considering also that $\|B\| \leq \kappa(A)\,\|E\|/\|A\|$:

$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq \frac{\kappa(A)}{1-\kappa(A)\,\|E\|/\|A\|}\left(\frac{\|E\|}{\|A\|} + \frac{\|e\|}{\|b\|}\right)$$

□

Proof (5) From the LS formulation, A is a rectangular matrix and B a square matrix. The condition number of the product is:

$$\kappa(AB) = \|AB\|\left\|(AB)^\dagger\right\|$$

where $M^\dagger$ denotes the Moore-Penrose pseudoinverse, which can be expressed as $M^\dagger = (M^TM)^{-1}M^T$. Developing the pseudoinverse for the product:

$$\begin{aligned}(AB)^\dagger &= ((AB)^TAB)^{-1}(AB)^T = (B^TA^TAB)^{-1}B^TA^T \\ &= B^{-1}(A^TA)^{-1}(B^T)^{-1}B^TA^T\end{aligned}$$

since $B^T = B$:

$$(AB)^\dagger = B^{-1}(A^TA)^{-1}A^T = B^{-1}A^\dagger$$

Therefore as $\|AB\| \leq \|A\|\,\|B\|$ and $\left\|(AB)^\dagger\right\| \leq \left\|B^{-1}\right\|\left\|A^\dagger\right\|$:

$$\kappa(AB) \leq \|A\|\,\|B\|\left\|B^{-1}\right\|\left\|A^\dagger\right\| = \kappa(A)\kappa(B)$$

As in Theorem 1 B is a square matrix with the maximum eigenvalue being 1 and minimum $1/n$ therefore $\kappa(B) = n$. Then, $\kappa(A) = \|A\|\left\|A^\dagger\right\|$ and $\|A\| = \sqrt{\lambda_{max}(A^TA)}$. Since the construction of the matrix A is the same than in Theorem 1:

$$\langle \alpha_i, \alpha_j \rangle = \begin{cases} 0, & \forall i,j, \quad i \neq j \\ h, & \forall i,j, \quad i = j \end{cases}$$

where α_k is the k-th row vector of A^T. Therefore, $\|A\| = \sqrt{h}$. And for $\left\|A^\dagger\right\| = \sqrt{\lambda_{max}((A^\dagger)^TA^\dagger)}$:

$$(A^\dagger)^T = \left[(A^TA)^{-1}A^T\right]^T = A((A^TA)^{-1})^T = A(A^TA)^{-1}$$

and then:

$$(A^\dagger)^TA^\dagger = A(A^TA)^{-1}(A^TA)^{-1}A^T = A(A^TAA^TA)^{-1}A^T$$

since $A^TA = hI$,

$$(A^\dagger)^TA^\dagger = A(h^2I)^{-1}A^T = \frac{1}{h^2}AA^T$$

and as $\lambda_{max}(AA^T) = \lambda_{max}(A^TA)$, $\lambda_{max}((A^\dagger)^TA^\dagger) = 1/h$ obtaining:

$$\kappa(A) = \sqrt{h\frac{1}{h}} = 1$$

which makes $\kappa(AB) \leq n$. □

References

[1] J. G. Ziegler and N. B. Nichols, "Optimum settings for automatic controllers," *trans. ASME*, vol. 64, no. 11, 1942.

[2] B. D. Tyreus and W. L. Luyben, "Tuning pi controllers for integrator/dead time processes," *Industrial & Engineering Chemistry Research*, vol. 31, no. 11, pp. 2625–2628, 1992.

[3] K. J. Åström and T. Hägglund, *Automatic tuning of PID controllers*. Instrument Society of America (ISA), 1988.

[4] K. J. Åström and T. Hägglund, "A new auto-tuning design," *IFAC Proceedings Volumes*, vol. 21, no. 7, pp. 141–146, 1988.

[5] K. J. Åström, T. Hägglund, and A. Wallenborg, "Automatic tuning of digital controllers with applications to hvac plants," *Automatica*, vol. 29, no. 5, pp. 1333–1343, 1993.

[6] M. Friman and K. V. Waller, "A two-channel relay for autotuning," *Industrial & engineering chemistry research*, vol. 36, no. 7, pp. 2662–2671, 1997.

[7] M. T. da Silva and P. R. Barros, "A robust relay feedback structure for processes under disturbances: Analysis and applications," *Journal of Control, Automation and Electrical Systems*, vol. 30, no. 6, pp. 850–863, 2019.

[8] C. Lorenzini, L. F. A. Pereira, and A. S. Bazanella, "A generalized forced oscillation method for tuning proportional-resonant controllers," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 3, pp. 1108–1115, 2019.

[9] R. Kochenburger, "Analyzing contactor servomechanisms by frequency-response methods," *Electrical Engineering*, vol. 69, no. 8, pp. 687–692, 1950.

[10] K. J. Åström and T. Hägglund, "Automatic tuning of simple regulators with specifications on phase and amplitude margins.," *Automatica*, vol. 20, no. 5, pp. 645–651, 1984.

[11] L. Ljung, *System identification: theory for the user*. Prentice-hall, 1987.

[12] T. Liu and F. Gao, *Industrial process identification and control design: step-test and relay-experiment-based methods*. Springer Science & Business Media, 2011.

[13] N. M. Krylov and N. N. Bogoliubov, *Introduction to non-linear mechanics*. Princeton university press, 1949.

[14] K. K. Tan, T. H. Lee, S. Huang, K. Y. Chua, and R. Ferdous, "Improved critical point estimation using a preload relay," *Journal of Process Control*, vol. 16, no. 5, pp. 445–455, 2006.

[15] C.-C. Yu, *Autotuning of PID Controllers. A Relay Feedback Approach*. Springer-Verlag London Limited, second ed.

[16] S. W. Sung, J. H. Park, and I.-B. Lee, "Modified relay feedback method," *Industrial & engineering chemistry research*, vol. 34, no. 11, pp. 4133–4135, 1995.

[17] T. H. Lee, Q.-G. Wang, and K. K. Tan, "A modified relay-based technique for improved critical point estimation in process control," *IEEE Transactions on Control Systems Technology*, vol. 3, no. 3, pp. 330–337, 1995.

[18] J. Sánchez, M. Guinaldo, A. Visioli, and S. Dormido, "Identification of process transfer function parameters in event-based pi control loops," *ISA transactions*, vol. 75, pp. 157–171, 2018.

[19] Q.-G. Wang, C.-C. Hang, and Q. Bi, "Process frequency response estimation from relay feedback," *Control Engineering Practice*, vol. 5, no. 9, pp. 1293–1302, 1997.

[20] I. Kaya and D. P. Atherton, "Parameter estimation from relay autotuning with asymmetric limit cycle data," *Journal of Process Control*, vol. 11, no. 4, pp. 429–439, 2001.

[21] J. Lee, S. W. Sung, and T. F. Edgar, "Integrals of relay feedback responses for extracting process information," *AIChE journal*, vol. 53, no. 9, pp. 2329–2338, 2007.

[22] W. L. Luyben, "Getting more information from relay-feedback tests," *Industrial & engineering chemistry research*, vol. 40, no. 20, pp. 4391–4402, 2001.

[23] T. Thyagarajan and C.-C. Yu, "Improved autotuning using the shape factor from relay feedback," *Industrial & engineering chemistry research*, vol. 42, no. 20, pp. 4425–4440, 2003.

[24] Q.-G. Wang, T. H. Lee, and L. Chong, *Relay feedback: analysis, identification and control*. Springer Science & Business Media, 2012.

[25] S.-H. Shen, J.-S. Wu, and C.-C. Yu, "Use of biased-relay feedback for system identification," *AIChE Journal*, vol. 42, no. 4, pp. 1174–1180, 1996.

[26] K. Åström and T. Hägglund, *Advanced PID Control*. ISA-The Instrumentation, Systems, and Automation Society, 2006.

[27] M. A. Johnson and M. H. Moradi, *PID control. New Identification and Design Methods*. Springer, 2005.