Transfer Function Parameter Identification by Modified Relay Feedback

Soltesz, Kristian; Hägglund, Tore; Åström, Karl Johan

Published in:
2010 American Control Conference

DOI:
10.1109/ACC.2010.5531322

2010

Document Version:
Peer reviewed version (aka post-print)

Transfer Function Parameter Identification by Modified Relay Feedback

Kristian Soltesz
Dept. of Automatic Control
Lund University
Lund, Sweden
Email: kristian.soltesz@control.lth.se

Tore Hägglund
Dept. of Automatic Control
Lund University
Lund, Sweden
Email: tore.hagglund@control.lth.se

Karl Johan Åström
Dept. of Automatic Control
Lund University
Lund, Sweden
Email: karl_johan.astrom@control.lth.se

***Abstract*— This paper proposes a method of finding low order models of a SISO transfer function based on relay feedback. Parameter identification is posed as a (non-convex) squared output error minimization problem, numerically solved utilizing Newton-Raphson iteration with back tracking line search. Focus lies on computing the cost function gradient and Hessian with respect to the parameter vector and on finding a feasible starting point. The method is demonstrated for FOTD model identification. A modified relay method is used to ensure good excitation around a predefined phase angle fo the system. The method requires no a priori system information. The identification method is evaluated on a batch of common process industry processes. Finally, conclusions and suggestions on future work are provided.**

I. Introduction

The use of relay feedback [1] as process identification method has been around for a long time. Its main application has been in automatic tuning of PID controllers in process industry, where it is still broadly used, due to its simplicity and reliability.

The original method yields the point on the Nyquist curve corresponding to the phase crossover frequency. The method has been augmented with various modifications of the relay non-linearity, [2] being one of the more elegant, resulting in the possibility to identify a point on the Nyquist curve other than that corresponding to the phase crossover frequency.

Several alternative data analysis methods have been proposed. Mats Lilja utilized least square regression to identify low order time delayed transfer function models from frequency domain data (i.e. several points on the Nyquist curve) [3].

Here an optimization method, yielding a transfer function description of the process to be identified, is presented. A discrete time counterpart of the method is outlined in [4]. The method is based on Newton-Raphson iteration over a cost function of the transfer function parameters. Cost derivatives (Jacobian and approximation of Hessian) are obtained through simulation of an augmented system. Due to non-convexity of the cost function in the transfer function parameters, a close-to optimal initial parameter guess is desirable. Such initial guess has here been obtained by gridding the normalized time delay of the model, evaluating the cost for each grid point, and choosing the parameters corresponding to the minimum as starting point for the optimization.

Input signals generated through a modified relay feedback are considered, since it allows for signal energy concentration around a frequency corresponding to a pre-defined phase lag of the system to be identified, without a priori system information. For PI(D) tuning applications, a frequency corresponding to a point in the third quadrant of the Nyquist curve is preferable. Since PI provides a phase lag, the obtained model needs not be accurate for phase lags larger than $145°$, whereas accuracy up to the phase crossover frequency can be of interest when considering PID control, due to the phase lead of the controller [5].

In order to verify generality of the method, it has been tested on the AMIGO[1] batch, consisting of nine classes of processes, cf. [5]. Per design, the process models of the batch are representative for process control industry, which is also the main target application field of the material which follows.

II. Optimization Method for Identification

Here the proposed identification method is presented. Time is assumed to be continuous.

A. Objective

Our aim is to identify parameters $\boldsymbol{\theta} = [\boldsymbol{b}\ \boldsymbol{a}\ L]^T$ ($\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^n, L \in \mathbb{R}_+$) of the time delayed strictly proper continuous time transfer function process model

$$P(s) = \frac{B(s)}{A(s)}e^{-Ls} = \frac{\sum_{j=1}^{n} b_j s^{n-j}}{s^n + \sum_{i=1}^{n} a_i s^{n-i}}e^{-Ls}. \qquad (1)$$

If the number of zeros is believed to be $m < n-1$, we assign $b_1 = \cdots = b_{n-m-1} = 0$. Given input sequence $u(t)$ and corresponding output sequence $y(t)$, we formulate the objective as to minimize the mean squared output error

$$J(\hat{\boldsymbol{\theta}}) = \frac{1}{2}\int_{t_0}^{t_f} (\hat{y}(t) - y(t))^2 dt, \qquad (2)$$

where

$$\hat{y}(t) \triangleq \mathscr{L}^{-1}(\hat{P}(s)) \cdot U(s)). \qquad (3)$$

[1]AMIGO stands for Approximate M-constrained Integral Gain Optimization. The AMIGO test batch was originally used to obtain guidelines for a Ziegler-Nichols type tuning scheme.

The problem is convex in $\hat{\boldsymbol{b}}$ and $\hat{L}$. However, it is non-convex in $\hat{\boldsymbol{a}}$. For example, letting the model be defined through $\hat{\boldsymbol{\theta}} = [\hat{a}_1\ \hat{a}_2\ \hat{a}_3\ 0\ 0\ \hat{b}_3\ \hat{L}]^T$ and freezing all parameters except $\hat{a}_3$ according to $\hat{\boldsymbol{\theta}} = [1\ 0\ \hat{a}_3\ 0\ 0\ 1\ 0]^T$ yields

$$J(\hat{\boldsymbol{\theta}}) = \frac{1}{2}\int_0^1 \mathscr{L}^{-1}(\hat{P}(s)^2) = \frac{1}{2}\int_0^1 \sin(\hat{a}_3 t)^2 dt$$

which is clearly not convex in $\hat{a}_3$.

B. Newton-Raphson Method

Due to the general non-convexity of (2) there exists no known method, guaranteeing convergence to the global minimum. A candidate method, which has proved successful for the problem instances we have analyzed, has been the Newton-Raphson approach, involving the computation of $\nabla J(\hat{\boldsymbol{\theta}})$ and $\nabla^2 J(\hat{\boldsymbol{\theta}})$ in each iteration.

C. Evaluation of Gradient

The gradient is given by

$$\nabla J(\hat{\boldsymbol{\theta}}) = \int_{t_0}^{t_f} \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \frac{1}{2}(\hat{y} - y)^2 dt = \int_{t_0}^{t_f} \frac{\partial \hat{y}}{\partial \hat{\boldsymbol{\theta}}} dt. \tag{4}$$

Introducing the canonical controllable state space form of $\hat{P}(s)$ yields

$$\frac{\partial \hat{\boldsymbol{x}}}{\partial t} = \hat{\boldsymbol{A}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{B}}u \tag{5}$$

$$\hat{y} = \hat{\boldsymbol{C}}\hat{\boldsymbol{x}}, \tag{6}$$

where

$$\frac{\partial \hat{x}_1}{\partial t} = -\hat{\boldsymbol{a}}^T\hat{\boldsymbol{x}} + u \tag{7}$$

$$\frac{\partial \hat{x}_k}{\partial t} = \hat{x}_{k-1},\ 2 \le k \le n \tag{8}$$

$$\hat{y} = \hat{\boldsymbol{b}}^T\hat{\boldsymbol{x}}. \tag{9}$$

In order to calculate $\nabla J(\hat{\boldsymbol{\theta}})$, we need to evaluate

$$\frac{\partial \hat{y}}{\partial \hat{\boldsymbol{\theta}}} = \boldsymbol{C}\frac{\partial \hat{\boldsymbol{x}}}{\partial \hat{\boldsymbol{\theta}}}. \tag{10}$$

From (9) we obtain

$$\frac{\partial \hat{y}}{\partial \hat{b}_k} = \hat{x}_k,\ 1 \le k \le n. \tag{11}$$

Finding partial derivatives of $\hat{y}$ w.r.t. the components of $\hat{\boldsymbol{a}}$ is somewhat more involving. From (3) we obtain

$$\hat{Y}(s) = \frac{\hat{B}(s)}{\hat{A}(s)}e^{-\hat{L}s}U(s) \Rightarrow \tag{12}$$

$$\Rightarrow \frac{\partial \hat{Y}(s)}{\partial \hat{a}_k} = -\frac{s^{n-k}}{\hat{A}(s)}\hat{Y}(s),\ 1 \le k \le n \tag{13}$$

The dynamics of (13) can be incorporated in the state space description (5), (6) by augmenting n states $\hat{\boldsymbol{z}}$ to the state vector $\hat{\boldsymbol{x}}$, forming $\hat{\boldsymbol{x}}_{\boldsymbol{e}} = [\hat{\boldsymbol{x}}^T\ \ \hat{\boldsymbol{z}}^T]^T$. Letting the augmented states take on the roles

$$\hat{z}_k = -\frac{\partial \hat{y}}{\partial \hat{a}_k},\ 1 \le k \le n \tag{14}$$

we utilize (13) to obtain the augmented state dynamics

$$\frac{\partial \hat{z}_1}{\partial t} = \hat{y} - \hat{\boldsymbol{a}}^T\hat{\boldsymbol{z}} = \hat{\boldsymbol{b}}^T\hat{\boldsymbol{x}} - \hat{\boldsymbol{a}}^T\hat{\boldsymbol{z}} \tag{15}$$

$$\frac{\partial \hat{z}_k}{\partial t} = \hat{z}_{k-1},\ 2 \le k \le n. \tag{16}$$

The augmented system in $\hat{\boldsymbol{x}}_{\boldsymbol{e}}$ provides the desired parameter derivatives

$$\hat{y} = \hat{\boldsymbol{b}}^T\hat{\boldsymbol{x}} \tag{17}$$

$$\frac{\partial \hat{y}}{\partial \hat{\boldsymbol{b}}} = \boldsymbol{I}_n\hat{\boldsymbol{x}} \tag{18}$$

$$\frac{\partial \hat{y}}{\partial \hat{\boldsymbol{a}}} = -\boldsymbol{I}_n\hat{\boldsymbol{z}}. \tag{19}$$

Finally, from (1), we obtain

$$\frac{\partial \hat{Y}(s)}{\partial \hat{L}} = -s\frac{\hat{B}(s)}{\hat{A}(s)}e^{-\hat{L}s}U(s). \tag{20}$$

Using (7)-(9) the parameter derivative can be written

$$\frac{\partial \hat{y}}{\partial \hat{L}} = \hat{a}_n\hat{b}_1\hat{x}_n - \hat{b}_1 u + \sum_{j=1}^{n-1}(\hat{a}_j\hat{b}_1 - \hat{b}_{j+1})\hat{x}_j. \tag{21}$$

D. Hessian Approximation

The Hessian of (2) is given by

$$\nabla^2 J(\hat{\boldsymbol{\theta}}) = \int_{t_0}^{t_f} \left(\frac{\partial \hat{y}}{\partial \hat{\boldsymbol{\theta}}}\right)^2 + (\hat{y} - y)\frac{\partial^2 \hat{y}}{\partial \hat{\boldsymbol{\theta}}^2} dt. \tag{22}$$

The first term in (22) is quadratic, i.e. ≥ 0. Under the realistic assumption that the output error $\hat{y} - y$ is uncorrelated with its derivatives in the components of $\hat{\boldsymbol{\theta}}$, the time average of the second term is small. Thus it can be neglected, motivating the Hessian approximation

$$\nabla^2 J(\hat{\boldsymbol{\theta}}) \approx \int_{t_0}^{t_f} \left(\frac{\partial \hat{y}}{\partial \hat{\boldsymbol{\theta}}}\right)^2 dt. \tag{23}$$

III. FOTD Model Identification

In this section we utilize the proposed optimization method to obtain FOTD models, parametrized as

$$\hat{P}(s) = \frac{\hat{b}}{s + \hat{a}}e^{-\hat{L}s}, \tag{24}$$

i.e. corresponding to parameter vector $\hat{\boldsymbol{\theta}} = [\hat{b}\ \hat{a}\ \hat{L}]^T$. A motivation for choosing a modified relay feedback as the source of input signal is followed by the proposal of a method for finding initial parameters $\hat{\boldsymbol{\theta}}_0$ for the optimization. Finally, attention is given to some practical implementation related issues.

A. Input Signal

Existing PID tuning methods such as Ziegler-Nichols [6], λ [7], (A)MIGO [8] as well as a promising MIGO extension, presented by Garpinger [9] rely on accurate LF process models. Of particular interest is the phase region $[-\pi, -\frac{\pi}{2}]$ rad, determining the sensitivity properties of the system. Additionally, the λ and MIGO methods utilize a static gain estimate.

Describing function analysis indicates that negative feedback connection of a proper, possibly time delayed, monotone LTI system P and a relay non-linearity results in limit cycle oscillations. The fundamental harmonic of the oscillation occurs at the phase crossover frequency of P. These observations are the basis of the identification method proposed by Åström and Hägglund in 1984 [1]. Replacing the relay with the two channel (TC) relay non-linearity shown in figure 1 allows for an energy concentration at a frequency corresponding to an arbitrary third quadrant phase angle of P, as described by Friman and Waller in [2].

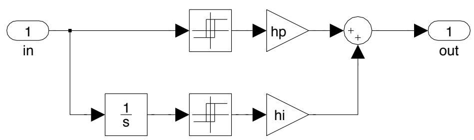

Fig. 1. Two channel relay.

The describing function of the TC relay is given by

$$N(a) = \frac{4h_p}{\pi a} - \frac{4h_i}{\pi a} i. \tag{25}$$

The corresponding phase angle is thus

$$\varphi_{TC} = \arctan\left(\frac{h_i}{h_p}\right). \tag{26}$$

By choosing h_p, h_i the phase of (25) can be chosen arbitrarily in the range $[0, \frac{\pi}{2}]$ rad, i.e. the fundamental limit cycle will occur at angular frequency ω_φ corresponding to phase $\varphi = -\pi + \varphi_{TC} \in [-\pi, -\frac{\pi}{2}]$ rad of P.

The Fourier series expansion of the symmetric T-periodic square wave $u(t)$ with amplitude A_u is given by

$$u(t) = \sum_{k=1}^{\infty} \frac{4A_u}{\pi k} \sin\left(\frac{2\pi k t}{T}\right). \tag{27}$$

Hence, the input signal energy content at the phase crossover frequency is

$$\frac{\int_T \left(\frac{4A_u}{\pi} \sin\left(\frac{2\pi t}{T}\right)\right)^2 dt}{\int_T u^2(t) dt} = \frac{8}{\pi^2} \approx 0.8, \tag{28}$$

i.e. 80 %, under relay feedback (disregerading the initial convergence phase). Remaining energy lies at integer multiples of the phase crossover frequency.

For the two-channel relay, the above analysis will additionally depend on the LTI system, but the key observations still hold:

- Most input signal energy is issued at the fundamental frequency of the limit cycle oscillation.
- Remaining energy is issued at integer multiples of the fundamental frequency.

If little energy is supplied in the overtones, or if these are heavily attenuated by P, effectively all identification data originates from the single frequency ω_φ. Since $\hat{\boldsymbol{\theta}} = [\hat{b}\ \hat{a}\ \hat{L}]^T$ has three components, this results in an under-determined problem. Generally, if one requires good model fit for a *range* of phase angles, a broader spectrum input is needed. One way to achieve this, is to alter φ_{TC} (by means of h_p, h_i in (25)) part way through the experiment, and hence obtain frequency data corresponding to at least two separate phase angles φ_1, φ_2 within the third quadrant. Subsequently, the cost function terms J_k and its derivatives ∇J, $\nabla^2 J$ corresponding to φ_k can be weighted together, with weights w_k being functions of corresponding signal energies E_{y_k}, in order to distribute model error over φ in a desired manner.

It is clear, from the above reasoning, that static gain information from obtained models is unreliable. If the aim of identification is to utilize a tuning method explicitly requiring a static gain estimate, e.g. λ or AMIGO, this can be obtained by augmenting the experiment with a step response.

B. Initial Parameter Values

Since the cost function (2) is non-convex in $\hat{\boldsymbol{\theta}}$, a starting point $\hat{\boldsymbol{\theta}}_0$ close to the global minimum is essential in order to avoid convergence of the Newton-Raphson iteration to a local minimum far from the global one. Assuming that the process dynamics to be identified are de facto (approximately) FOTD, the following paragraphs suggest a methodology for choosing $\hat{\boldsymbol{\theta}}_0$.

The FOTD system (24) can be re-parametrized in normalized time delay $\hat{\tau} = \frac{\hat{L}}{\hat{L}+1/\hat{a}}$, average residence time $\hat{T}_{ar} = 1/\hat{a} + \hat{L}$ and static gain $\hat{K} = \hat{b}/\hat{a}$. Of these parameters τ is the most difficult to estimate since it requires a separation between delay and lag, while $\hat{T}_{ar}$ is typically easy to estimate. The following, heuristic, grid-based method aims at yielding a feasible starting point $\boldsymbol{\theta}_0$ for the Newton-Raphson iteration, by first estimating τ.

Assume that the input–output data set $\{u(t),\ y(t)\}$, $t \in [t_0, t_f]$ is the outcome of a TC relay feedback experiment. Truncating the data set, only to include the last N periods of converged limit cycle oscillation yields the new data set $\{u(t),\ y(t)\}, t \in [t_N, t_f]$. Let A_u and A_y be the amplitudes of the first harmonics in $u(t)$ and $y(t)$, $t \in [t_N, t_f]$, respectively. These are readily given by the Fourier transform as

$$A_u = \left| \frac{2}{t_f - t_N} \int_{t_N}^{t_f} u(t) e^{-i \frac{2\pi N}{t_f - t_N} t} dt \right|, \tag{29}$$

$$A_y = \left| \frac{2}{t_f - t_N} \int_{t_N}^{t_f} y(t) e^{-i \frac{2\pi N}{t_f - t_N} t} dt \right|. \tag{30}$$

The phase- and magnitude of $\hat{P}(i\omega_\varphi)$, are given by

$$\angle \hat{P}(i\omega_\varphi) = -\hat{L}\omega_\varphi - \tan^{-1}(\frac{1}{\hat{a}}\omega_\varphi) = \varphi = -\pi + \varphi_{TC} \tag{31}$$

$$|\hat{P}(i\omega_\varphi)| = A_y \frac{\hat{b}/\hat{a}}{\sqrt{1 + \omega_\varphi^2 (1/\hat{a})^2}} A_u, \tag{32}$$

where φ_{TC} is the TC relay phase from (26). For a given normalized time delay $\hat{\tau}$ we can insert $\hat{L} = \frac{\hat{\tau}}{1-\hat{\tau}} \frac{1}{\hat{a}}$ into (31) and solve the resulting convex equation in $\hat{a}$ numerically. The obtained $\hat{a}$ can now be inserted into (32), yielding $\hat{b}$. By griding $\hat{\tau}$-space we obtain a family of models $\hat{P}_{\hat{\tau}_i}(s)$. The cost (2) is evaluated for all $\hat{P}_{\hat{\tau}_i}(s)$. Subsequently, $\hat{\boldsymbol{\theta}}_0$ is chosen to be the parameters of the model corresponding to the smallest cost function value.

The outcome of this procedure is illustrated in figure 2 for the FOTD processes $\boldsymbol{\theta} = [5/4\ 5/4\ 1/5]^T \Leftrightarrow \tau = 0.2$ (solid), $\boldsymbol{\theta} = [2\ 2\ 1/2]^T \Leftrightarrow \tau = 0.5$ (dashed) and $\boldsymbol{\theta} = [5\ 5\ 4/5]^T \Leftrightarrow \tau = 0.8$ (dotted), all with average residence time $T_{ar} = 1.0$ and steady state gain $K = 1.0$. Introducing

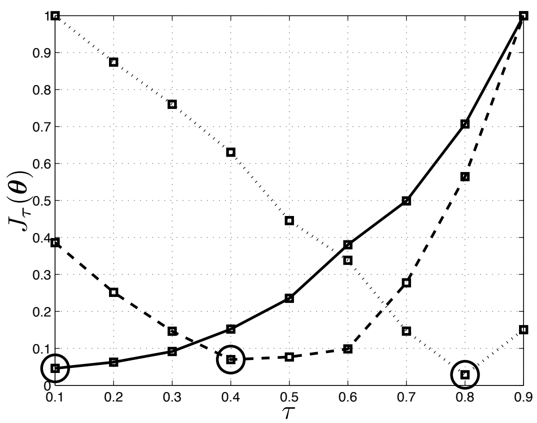

Fig. 2. Normalized cost $J_\tau \boldsymbol{\theta})/\max J(\boldsymbol{\theta})$ as function of normalized time delay τ for FOTD processes with $\tau = 0.2$ (solid), $\tau = 0.5$ (dashed) and $\tau = 0.8$ (dotted).

the grid $\tau_i \in \{0.1i,\ i = 1..9\}$, the method yields either the correct τ or its grid neighbors.

C. Model Order Validation

When identifying processes where the order of P exceeds that of $\hat{P}$, an inherent model reduction takes place. The cancellation of one or several poles is compensated for by a change in delay estimate $\hat{L}$. If the input u has a narrow spectrum, the obtained model $\hat{P}$ can still be accurate around the frequency corresponding to the spectral peak. However, accuracy local to one point might not be enough for feasible controller synthesis.

Therefore, a test for checking the validity of a FOTD model is desirable. An instructive such test is provided in increasing the model order to SOTD and identifying the parameters $\boldsymbol{\theta}_{0,+}$ of the new model $\hat{P}_+$. If $|\hat{L} - \hat{L}_+|$ is large compared to $\hat{L}$, it is motivated to de facto increase model order to SOTD.

IV. Experimental Procedure

In this section we outline the experimental procedure. Data was generated in MATLAB/Simulink using the TC relay feedback connection shown in Figure 3.

A. Data Generation

Parameters h_p, h_i in (26) corresponding to $\varphi_{TC} = 0.4\pi$ rad, i.e. $\approx 75°$ were chosen. Other $\varphi_{TC} \in [0, \frac{\pi}{2}]$ rad

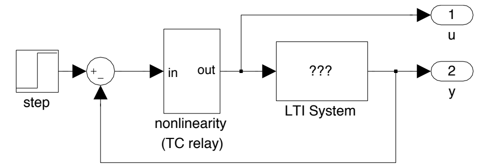

Fig. 3. Simulink model generating test data for the proposed identification method. The contents of the non-linearity block are shown in Figure 1. (The step is used to initialize a limit cycle oscillation.)

would shift the phase dependence on model accuracy. However, the identification methodology would remain unaltered.

Each data generating simulation lasted 11 zero crossings of LTI input $u(t)$. Identification data was generated for all 133 batch processes.

B. Identification

In this first paper, we consider the ideal measurement noise and load disturbance free case. The only modifications applied to the above presented theory has been those of discretization (i.e. exchanging integrals for sums, the Fourier transform for the FFT, etc.).

Fundamental frequency amplitudes of in- and outputs were found by applying the FFT versions of (29), (30) on truncated versions of $u(t), y(t)$, corresponding to the two last oscillation periods. (As a comment it should be mentioned that the chosen number of relay switches was found heuristically, so that the last two relay periods could be considered converged limit cycle.)

Subsequently, an initial parameter vector $\hat{\boldsymbol{\theta}}_0$ was determined by means of (31), (32) and the described $\hat{\tau}$-grid method with grid size $\hat{\tau}_i \in \{0.1i,\ i = 1..9\}$.

The Newton-Raphson optimization was applied over 7 iterations, which was found to be adequate, considering cost convergence for the different batch processes.

Back tracking line search, cf. [10], was added to increase convergence rate. The method is illustrated below, with δ being the step length, while $\alpha = 0.25, \beta = 0.5$ are user-defined parameters.

$$\texttt{while } J(\hat{\boldsymbol{\theta}} + \delta\Delta\hat{\boldsymbol{\theta}}) > J(\hat{\boldsymbol{\theta}}) + \alpha\delta\nabla J(\hat{\boldsymbol{\theta}})^T \Delta\hat{\boldsymbol{\theta}}, \delta := \beta\delta$$

Finally, bounds on time delay estimate $\hat{L}$ were introduced, forcing it to be strictly non-negative and less than a half period of the fundamental frequency component in $u(t)$.

V. Results

Results from the identification of one particular transfer function are presented in detail, exploiting key features of the proposed method. This is followed by a compilation of the model errors obtained by running the method on a batch [5].

A. Instance Study

Here, results from identifying $P(s) = \frac{1}{(s+1)^2}e^{-s}$ are presented. The choice of process is motivated by the fact that process order is higher than model order. This has two fundamental implications:

- There exists no FOTD model with 'good' fit for all frequencies. However, the proposed method is expected to yield one with good fit around the phase φ in the third quadrant.
- The initial guess provided by $\hat{\tau}$-gridding is sub-optimal, since the model structure assumption is invalid, demonstrating the benefit of the Newton-Raphson optimization.

Figure 4 shows identification input $u(t)$, generated by the TC relay feedback, together with corresponding process output and converging model output.

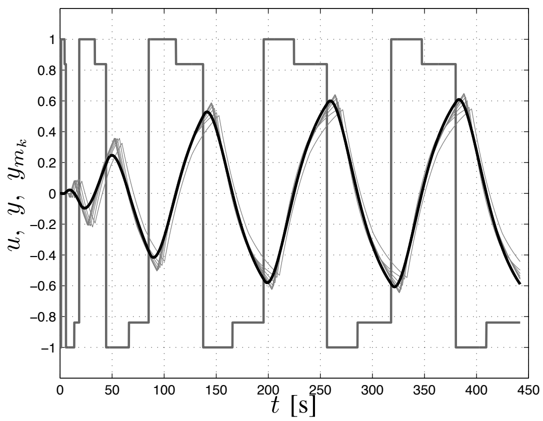

Fig. 4. TC relay output $u(t)$ (grey), process output $y(t)$ (solid, black) and converging model outputs $y_{m_k}(t),\ k \in \{1, \ldots, 7\}$ (grey, thin).

Figure 5 shows the Nyquist curve of P together with those of the obtained FOTD model $\hat{P}$ and the corresponding initial model $\hat{P}_0$ provided by the τ-gridding. Not unexpectedly, $\hat{P}_0$ provides a better all-over fit, whereas $\hat{P}$ shows a better fit in the third quadrant (which is achieved at expense of a worse fourth quadrant fit). Both models provide good fits at the phase angle $\varphi = -\pi + \varphi_{TC}$, corresponding to the fundamental harmonic of the process input $u(t)$.

The observations presented above generally hold for the AMIGO batch.

A complementary representation of performance is given by the step response. Figure 6 shows the step responses of P, $\hat{P}$ and $\hat{P}_0$ in figure 5.

As expected, the final model $\hat{P}$ has a worse static gain estimate than the initial model $\hat{P}_0$.

Note the over-estimation of L, shown in the lower plot of figure 6, being a consequence of lower model than process order. A second order model (provided a feasible $\boldsymbol{\theta}_{+,0}$ is given by

$$\hat{P}_+(s) = \frac{0.001s + 1.06}{(s + 1.26)(s + 0.84)} e^{-1.01s}. \tag{33}$$

Figure 7 shows the initial part of the step responses of P, $\hat{P}$ and $\hat{P}_+$.

The model order test of section III yields

$$\frac{|\hat{L} - \hat{L}_+|}{\hat{L}} = 0.26. \tag{34}$$

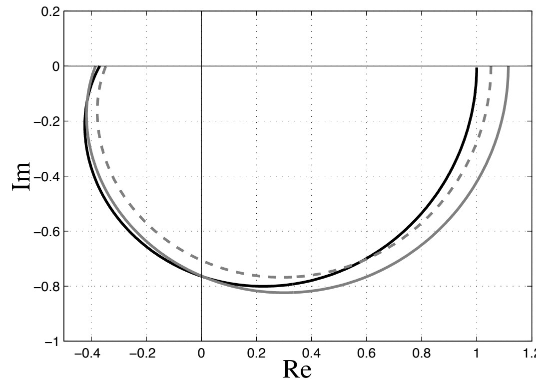

Fig. 5. Nyquist curve of $P(s) = \frac{1}{(s+1)^2} e^{-s}$ (black), $\hat{P}_0(s) = \frac{0.52}{s+0.49} e^{-1.35s}$ (grey, dashed) and $\hat{P}(s) = \frac{0.57}{s+0.51} e^{-1.37s}$ (grey, solid).

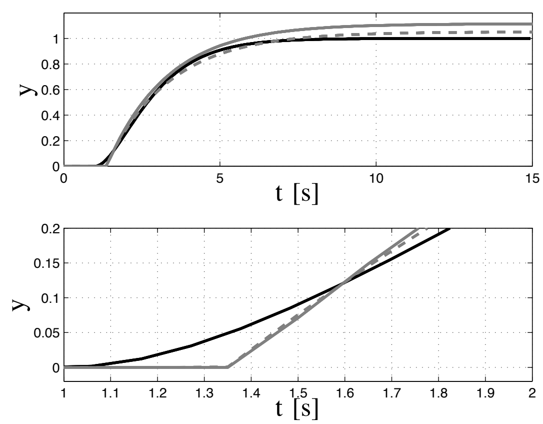

Fig. 6. Step response of $P(s) = \frac{1}{(s+1)^2} e^{-s}$ (black), $\hat{P}_0(s) = \frac{0.52}{s+0.49} e^{-1.35s}$ (grey, dashed) and $\hat{P}(s) = \frac{0.57}{s+0.51} e^{-1.37s}$ (grey, solid). The lower plot is a magnification of the bottom left part of the upper plot.

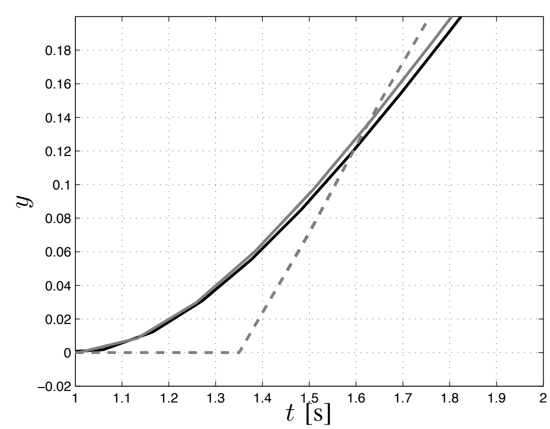

Fig. 7. Step responses of $P(s) = \frac{1}{(s+1)^2} e^{-s}$, $\hat{P}(s) = \frac{0.57}{s+0.51} e^{-1.37s}$ and $\hat{P}_+(s) = \frac{0.001s+1.06}{(s+1.26)(s+0.84)} e^{-1.01s}$

Another interesting observation is that $\hat{T}_{ar} = 3.33$ for the FOTD model and $\hat{T}_{+,ar} = 2.98$ for the SOTD model, which are both good estimates, given $T_{ar} = 3.0$ for the process. However, assume all input energy was issued at the frequency ω_φ, i.e. $u(t) = \sin(\omega_\varphi t)$. Asymptotically the cost would be minimized (to $J = 0$) when $|P(i\omega_\varphi)| = |\hat{P}(i\omega_\varphi)|$ and $\angle P(i\omega_\varphi) = \angle \hat{P}(i\omega_\varphi)$, where the left hand sides are constants and the right hand sides are given by (32) and (31), respectively. This is an under-determined system in $\hat{\boldsymbol{\theta}}$, with unique solution $\forall \hat{L} \in \mathbb{R}_+$, as indicated in section III.

B. Batch Study

Figure 8 shows a compilation of gain errors $|P| - |\hat{P}|$ plotted against process phase, for the processes of the test batch.

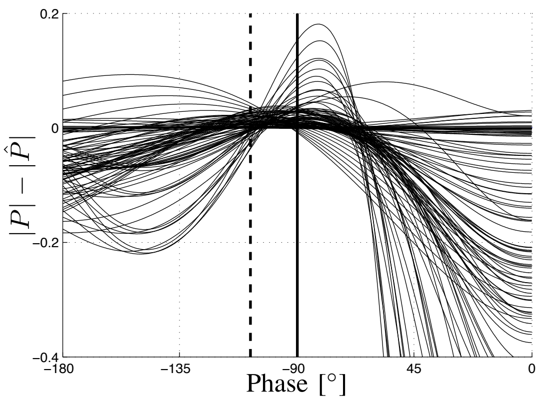

Fig. 8. Gain error $|P| - |\hat{P}|$ as function of process phase for the processes of the AMIGO test batch.

As expected, the method yields best fit close to the phase $-115°$ corresponding to the first harmonic of the input signal $u(t)$ (marked by a dashed line in figure 8).

For larger negative phase values within the third quadrant, the errors are negative for most processes, corresponding to conservative models, concerning sensitivity.

VI. Conclusions

A method for computing partial derivatives of the output error in model transfer function parameters has served as basis for a gradient search (Newton-Raphson) approach to system identification. The method is applicable to all proper, possibly time delayed, transfer functions.

The following, highly interrelated, items need to be decided, prior to applying the method: cost function, model order (choice and verification), input signal, initial parameters and halting criterion. Particular attention has to be given to the input signal, ensuring spectral content at frequencies for which model validity is crucial.

This paper was mainly confined to the case of FOTD model structure, utilizing a quadratic cost function and TC relay feedback for input generation.

Initial parameters were obtained by means of a heuristic gridding strategy and no explicit attention was given to halting criteria for the optimization.

A method for model order validation was suggested.

The approach proved successful for a large number of common process types and instances thereof.

VII. Future Work

There are several directions for potential future work related to the proposed identification method.

One obvious continuation would be to combine the identification method with one or several PID-tuning methods and evaluate the obtained closed loop performance. A related issue is the investigation of how process- and measurement noise affect the identification and ultimately the closed loop performance.

Another interesting direction is that of MIMO control. Especially TITO systems are common in process industry. Hence an extension of the method to the identification of TITO dynamics would be of high interest.

It would also be interesting to evaluate performance of the method using higher order models, possibly with modifications regarding cost function and input signal. SOTD models (with one zero) are of particular interest, covering essentially all modeling needs for PID design.

References

[1] K. J. Åström and T. Hägglund, "Automatic tuning of simple regulators with specifications on phase and amplitude margins," *Automatica*, vol. 20, pp. 645–651, 1984.
[2] M. Friman and K. V. Waller, "A two-channel relay for autotuning," *Industrial and Engineering Chemistry Research*, vol. 36, pp. 2662–2671, 1997.
[3] M. Lilja, "Least squares fitting to a rational transfer function with time delay," in *Preprints Control '88*. Oxford: IEEE, Jan. 1988.
[4] K. J. Åström and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records," in *Proc. IFAC Conference on Self-Adaptive Control Systems*, Teddington, UK, Jan. 1965.
[5] K. J. Åström and T. Hägglund, *Advanced PID Control*. ISA, 2006.
[6] J. G. Ziegler and N. B. Nichols, "Optimum settings for automatic controllers," *Transaction fo the ASME*, pp. 759–768, 1942.
[7] E. B. Dahlin, "Designing and tuning digital ccontrollers," *Instruments and Control Systems*, vol. 42, pp. 77–83, 1968.
[8] T. Hägglund and K. J. Åström, "Revisiting the Ziegler-Nichols tuning rules for PI control," *Asian Journal of Control*, vol. 4, no. 4, pp. 364–380, Dec. 2002.
[9] O. Garpinger and T. Hägglund, "A software tool for robust PID design," in *Proc. 17th IFAC World Congress, Seoul, Korea*, July 2008.
[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, 2004.