



HAL
open science

Useful redundancy in parameter and time delay estimation for continuous-time models

Huong S Ha, James S Welsh, Mazen Alamir

► **To cite this version:**

Huong S Ha, James S Welsh, Mazen Alamir. Useful redundancy in parameter and time delay estimation for continuous-time models. *Automatica*, 2018, 99 (sept.), pp.455-462. 10.1016/j.automatica.2018.06.023 . hal-01500012

HAL Id: hal-01500012

<https://hal.science/hal-01500012>

Submitted on 4 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Useful redundancy in parameter and time delay estimation for continuous-time models

Huong Ha^a, James S. Welsh^a, Mazen Alamir^b,

^a*School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*

^b*Gipsa-lab, Control Systems Department. University of Grenoble, France*

Abstract

This paper proposes an algorithm to estimate the parameters, including time delay, of continuous time systems based on instrumental variable identification methods. To overcome the multiple local minima of the cost function associated with the estimation of a time delay system, we utilise the useful redundancy technique. Specifically, the cost function is filtered through a set of low-pass filters to improve convexity with the useful redundancy technique exploited to achieve convergence to the global minimum of the optimization problem. Numerical examples are presented to demonstrate the effectiveness of the proposed algorithm.

Key words: continuous time identification; time delay, useful redundancy, filtering.

1 Introduction

The goal of system identification is to estimate the parameters of a model in order to analyse, simulate and/or control a system. In the time domain, there are two typical approaches to identify a system, i.e. discrete-time (DT) identification and continuous-time (CT) identification. For several decades, DT identification has been dominant due to the strong development of the digital computer. More recently, estimation using continuous-time identification methods has received much attention due to advantages such as providing insights to the physical system and being independent of the sampling time [15] [17] [25]. For example, with irregular sampling time, the DT model becomes time-varying; hence the DT system identification problem becomes more difficult while for CT identification, the system is still time-invariant. In reality, irregular sampling occurs in many cases, e.g. when the sampling is event-triggered, when the measurement is manual and also in the case of missing data [5].

In some CT identification problems, one wants to estimate the system parameters and any unknown time delay, as there exist many practical examples including, chemical processes, economic systems, biological systems, that possess time delays. It is important to estimate the delay accurately since a poor estimate can result in poor model order selection and inaccurate estimates of the system parameters.

There are many approaches to estimate a system time delay [9]. A simple approach is to consider the impulse response data, e.g. estimate the time delay by finding where the impulse response becomes nonzero [10] or by noting the delay where the correlation between input and output is maximum [10] [11]. Another approach is to model the delay by a rational polynomial transfer function using a Padé or similar approximation and then estimate the time delay as part of the system parameters [1] [2] [18]. In [27] [6] [7], the time delay and system parameters of a Multiple Input Single Output CT system are estimated in a separable way using an iterative global nonlinear least-squares or instrumental variable method.

Recently, a method of estimating the parameters and time delay of CT systems has been suggested in [12] which is based on a gradient technique. The parameters and the time delay are estimated separately, i.e. when one is estimated, the other is fixed which is then repeated in an iterative manner. In this approach, the Simplified Refined Instrumental Variable (SRIVC) method is used to estimate the parameters whilst the time delay is estimated using the Gauss-Newton method. In addition, due to the effects of multiple minima in the cost function to be minimized for the time delay [20] [23] [8], a low-pass filter is employed to increase convexity. As shown in [8] [14] [13], a suitable low-pass filtering operation on the estimation data can help to extend the global convergence region of the cost function, hence improve the accuracy of

tion and inaccurate estimates of the system parameters.

* The material in this paper was not presented at any conference.

Email addresses: huong.ha@uon.edu.au (Huong Ha), james.welsh@newcastle.edu.au (James S. Welsh), mazen.alamir@grenoble-inp.fr (Mazen Alamir).

the time delay estimate.

In this paper, we adopt the idea of using a low-pass filter where instead of using only one filter, we suggest to use multiple low-pass filters and incorporate the useful redundancy technique [3] [4]. Useful redundancy is a technique that was developed to avoid local minima when solving a nonlinear inverse problem. The concept is to generate a family of cost functions that have different local minima, but share the same global minimum with the original cost function of the optimization problem. Whenever the algorithm is trapped in a minimum, the solver path is switched to another solver path in a way such that the minimum found using the new solver path corresponds to a decrease in the original cost function. This allows the algorithm to cross local minima and converge to the global minimum, hence improving the accuracy of the estimated parameters. For the algorithm described in this paper, the multiple cost functions are generated by filtering the original time delay cost function through a number of low-pass filters with different cut-off frequencies that span the system bandwidth.

The paper is organized as follows. Section 2 describes the model setting and Section 3 recalls both the SRIVC algorithm and the SRIVC-based time delay estimation using a low-pass filter. Section 4 formulates the algorithm of the new method and provides analysis on the effectiveness of the proposed method. Section 5 presents numerical experiments and results for both regular and irregular sampling schemes. Finally, the conclusion will be drawn in Section 6.

2 Model setting

Consider a continuous-time linear, time invariant, single input single output system,

$$x(t) = G_0(p)u(t - \tau_0) = \frac{B(p)}{A(p)}u(t - \tau_0), \quad (1)$$

with

$$\begin{aligned} B(p) &= b_0p^m + b_1p^{m-1} + \dots + b_m, \\ A(p) &= p^n + a_1p^{n-1} + \dots + a_n, \quad n \geq m \end{aligned}$$

where τ_0 is the time delay, $u(t)$, $x(t)$ are the input and deterministic output of the system respectively and p is the differential operator, i.e. $p^{(i)}x(t) = d^i x(t)/dt^i$. In addition, the following assumptions are made:

Assumption 2.1 *Polynomials $B(p)$ and $A(p)$ are coprime.*

Assumption 2.2 *The system is asymptotically stable.*

Assumption 2.3 *The high frequency gain of $G_0(p)$ is 0, i.e. $G_0(p)$ is strictly proper.*

The deterministic output $x(t_k)$ is measured as $y(t_k)$ in the presence of noise, i.e.

$$y(t_k) = x(t_k) + e(t_k). \quad (2)$$

Furthermore, we consider the sampling time of the input $u(t_k)$ and output data $y(t_k)$ as either regular or irregular. The time-varying sampling interval is denoted as,

$$h_k = t_{k+1} - t_k, \quad k = 1, 2, \dots, N-1, \quad (3)$$

where N is the length of the data.

The objective of a CT system identification problem is to estimate the time delay, τ_0 , and the parameters $a_1, a_2, \dots, a_n, b_0, b_1, \dots, b_m$ of the CT model in (1), using the measured input and output data, $u(t_k)$ and $y(t_k)_{k=1}^N$ respectively.

3 Parameter and time delay identification of Continuous-time models with SRIVC and filtering

There exists a large number of continuous time identification methods see, for example, [17] [28]. In this paper, we consider the Simplified Refined Instrumental Variable Continuous Time method (SRIVC), which is developed in the literature by Young and Garnier [16][17][28]. We begin with a basic description of the SRIVC method.

3.1 Traditional SRIVC method

To simplify the description we first assume the time delay τ_0 is known. The SRIVC method can then be summarized as follows.

From (1) and (2),

$$\frac{A(p)}{A(p)}y(t_k) = \frac{B(p)}{A(p)}u(t_k - \tau_0) + e(t_k), \quad (4)$$

which becomes,

$$A(p)y_A(t_k) = B(p)u_A(t_k - \tau_0) + e(t_k), \quad (5)$$

with

$$y_A(t_k) = \frac{1}{A(p)}y(t_k), \quad u_A(t_k - \tau_0) = \frac{1}{A(p)}u(t_k - \tau_0). \quad (6)$$

From (5), a linear regression model can be formed as

$$y_A^{(n)}(t_k) = \Phi_N(t_k, \tau_0)\theta + e(t_k) \quad (7)$$

where

$$\begin{aligned} \Phi_N(t_k) &= [-y_A^{(n-1)}(t_k), -y_A^{(n-2)}(t_k), \dots, -y_A(t_k), \\ &\quad u_A^{(m)}(t_k - \tau_0), u_A^{(m-1)}(t_k - \tau_0) \dots u_A(t_k - \tau_0)], \\ \theta &= [a_1, a_2, \dots, a_n, b_0, b_1, \dots, b_m]^T, \\ y_A^{(i)}(t_k) &= \frac{d^i y_A(t_k)}{dt^i}, \quad u_A^{(i)}(t_k - \tau_0) = \frac{d^i u_A(t_k - \tau_0)}{dt^i}. \end{aligned} \quad (8)$$

To solve the linear regression problem in (7), the SRIVC algorithm uses an instrumental variable (IV) method. The instruments in the SRIVC method are chosen as the estimated noise free outputs, i.e. the regressor becomes,

$$\hat{\Phi}_N(t_k) = [-\hat{x}_A^{(n-1)}(t_k), -\hat{x}_A^{(n-2)}(t_k), \dots, -\hat{x}_A(t_k), u_A^{(m)}(t_k - \tau_0), u_A^{(m-1)}(t_k - \tau_0) \dots u_A(t_k - \tau_0)] \quad (9)$$

where

$$\hat{x}_A(t_k) = \frac{1}{\hat{A}(p)} \hat{x}(t_k); \hat{x}(t_k) = \frac{\hat{B}(p)}{\hat{A}(p)} u(t_k - \tau_0), \quad (10)$$

with $\hat{B}(p)$ and $\hat{A}(p)$ estimates of $B(p)$ and $A(p)$.

3.2 Implementation of the CT filtering operation for irregular sampled data and arbitrary time delay

The SRIVC method can be used to estimate system parameters from regular sampled data as well as from irregular sampled data. However, one difficulty in implementing the SRIVC method for irregular sampled data in the presence of an arbitrary time delay is the CT filtering operation, e.g.,

$$u_A(t_k - \tau) = \frac{1}{A(p)} u(t_k - \tau), \quad (11)$$

when τ is an arbitrary time delay.

There are two reasons for this difficulty:

- (1) $u(t_k - \tau)$ is not available from the measured data,
- (2) The digital simulation of $u(t_k - \tau)$ is generally performed in state-space form hence, an equivalent discrete-time state space representation of the CT state space model needs to be computed. For irregular data, as the sampling interval $h(t_k)$ is time-varying, the computational load using standard methods, e.g. *expm* in Matlab, to compute the transformation matrices will be large.

As suggested in [12], the two problems mentioned above can be solved as follows,

- (1) $u(t_k - \tau)$ can be constructed from the neighbouring data based on the inter-sample behaviour, e.g. zero-order-hold (ZOH), first-order-hold (FOH), etc.
- (2) To reduce the computational load, the time-varying sampling interval is divided into two intervals: the first interval is a multiple of a constant sampling period, the second interval is the residual. The discretization matrix of the first interval is pre-computed using a standard method and stored in an array. The discretization matrix for the second interval is computed by a fast approximation, e.g. the 4th order Runge-Kutta (RK4) method. The final discretization matrix of the sampling interval h_k is a product of the two matrices.

Specifically, convert the transfer function (TF) model in (11) to the equivalent CT state-space model,

$$\begin{cases} \dot{\mathbf{z}}(t - \tau) = \mathbf{F}_c \mathbf{z}(t - \tau) + \mathbf{G}_c u(t - \tau) \\ u_A(t - \tau) = \mathbf{H}_c \mathbf{z}(t - \tau). \end{cases} \quad (12)$$

The equivalent DT state-space model will be,

$$\begin{cases} \mathbf{z}(t_{k+1} - \tau) = \mathbf{F}(h_k) \mathbf{z}(t_k - \tau) + \mathbf{M}(t_k) \\ u_A(t_k - \tau) = \mathbf{H}_c \mathbf{z}(t_k - \tau) \end{cases} \quad (13)$$

where $\mathbf{F}(h_k)$, $\mathbf{M}(t_k)$ are computed as follows,

- (a) Denote $t_{k,1}, t_{k,n_k}$ ($n_k \geq 2$) as $t_k - \tau, t_{k+1} - \tau$ respectively; $t_{k,i}, i = 2, \dots, n_k - 1$ as the transition time-instants of $u(t)$ between $t_k - \tau$ and $t_{k+1} - \tau$; and $h_{k,i} = t_{k,i+1} - t_{k,i}$.
- (b) Let Δ be a constant sampling period; $m_{k,i}$ be a positive integer; $\delta_{k,i} \geq 0$ such that:
$$h_{k,i} = m_{k,i} \Delta + \delta_{k,i} < (m_{k,i} + 1) \Delta.$$
- (c) Compute $\mathbf{F}(m\Delta)$, $\mathbf{G}(m\Delta)$.
- (d) Compute $\mathbf{F}(\delta_{k,i})$, $\mathbf{G}(\delta_{k,i})$ using, e.g. RK4.
- (e) Finally, compute $\mathbf{F}(h_k)$, $\mathbf{M}(t_k)$ as follows,

$$\begin{aligned} \mathbf{F}(h_k) &= e^{\mathbf{F}_c h_k}, \\ \mathbf{M}(t_k) &= \sum_{i=1}^{n_k-1} \prod_{j=i+1}^{n_k-1} \mathbf{F}(h_{k,j}) \mathbf{G}(h_{k,j}) u(t_{k,i}), \end{aligned}$$

with

$$\begin{aligned} \mathbf{F}(h_{k,i}) &= \mathbf{F}(m_{k,i} \Delta) \mathbf{F}(\delta_{k,i}), \\ \mathbf{G}(h_{k,i}) &= \mathbf{F}(m_{k,i} \Delta) \mathbf{G}(\delta_{k,i}) + \mathbf{G}(m_{k,i} \Delta). \end{aligned}$$

Now that we can generate filtered signals of the irregular sampled data, we describe the traditional SRIVC algorithm [16][17][28] in Algorithm I.

ALGORITHM I

Step 1. Initialization

- (1) Create the stable state variable filter (SVF),

$$F(p) = \frac{1}{(p + \omega_c^{SVF})^n}, \quad (14)$$

where ω_c^{SVF} is chosen to be larger than or equal to the bandwidth of the system.

- (2) Filter $y(t_k)$ and $u(t_k - \tau_0)$ via the SVF to generate derivatives of the signals, i.e. $y_A^{(n)}(t_k), y_A^{(n-1)}(t_k), \dots, y_A(t_k)$ and $u_A^{(m)}(t_k - \tau_0), u_A^{(m-1)}(t_k - \tau_0), \dots, u_A(t_k - \tau_0)$ by using the method described in [16].
- (3) Use the least squares method to estimate the initial parameters,

$$\hat{\theta}_0 = [\Phi_N \Phi_N^T]^{-1} \Phi_N Y_N. \quad (15)$$

with

$$\begin{aligned} \Phi_N &= [\varphi_N^T(t_1) \dots \varphi_N^T(t_N)], \\ Y_N &= [y_A^{(n)}(t_1) y_A^{(n)}(t_2) \dots y_A^{(n)}(t_N)]^T, \end{aligned} \quad (16)$$

where φ_N is defined in (8).

Step 2. Iterative estimation

for $j = 1$:convergence^a

- (1) If the estimated TF model is unstable, reflect the right half plane zeros of $\hat{A}(p, \hat{\theta}^{j-1})$ into the left half plane and construct the new estimate.
- (2) Generate the instrumental variable,

$$\hat{x}(t_k) = \frac{\hat{B}(p, \hat{\theta}^{j-1})}{\hat{A}(p, \hat{\theta}^{j-1})} u(t_k - \tau_0). \quad (17)$$

- (3) Filter the input $u(t_k - \tau_0)$, output $y(t_k)$ and the instrumental variable $\hat{x}(t_k)$ by the continuous time filter

$$F_c(p) = \frac{1}{\hat{A}(p, \hat{\theta}^{j-1})}, \quad (18)$$

to generate the derivatives of these signals.

- (4) Using the prefiltered data, generate an estimate $\hat{\theta}^j$ using the IV method,

$$\hat{\theta}^j = [\Psi_N \Phi_N^T]^{-1} \Psi_N Y_N. \quad (19)$$

where Ψ_N is the IV matrix generated by the instrumental variables $\hat{x}(t_k)$ and Φ_N is the regression matrix,

$$\begin{aligned} \Phi_N &= [\varphi_N^T(t_1) \dots \varphi_N^T(t_N)], \\ \Psi_N &= [\hat{\varphi}_N^T(t_1) \dots \hat{\varphi}_N^T(t_N)], \end{aligned} \quad (20)$$

where $\hat{\varphi}_N$ is defined in (9).

end

^a Convergence requires the relative error between the estimated parameters of two consecutive iterations to be less than an $\varepsilon > 0$ where ε is a small value chosen to achieve the desired accuracy.

3.3 SRIVC-based time delay estimation with filtering

A recently developed method [12] to estimate both the time delay and parameters for the CT model (2) considers it as a separable nonlinear least squares problem. The SRIVC algorithm and the Gauss-Newton method are used in [12] to estimate the system parameters and the time delay respec-

tively. In this problem the cost function for the time delay estimation has multiple minima [12], hence a low-pass filter is utilized to extend the global convergence region [12] [13] [14].

Let $L(p)$ be a CT low-pass filter with cut off frequency ω_c^L . Applying the filter $L(p)$ to the linear regression (7), we obtain

$$\bar{y}_A^{(n)}(\rho) = \bar{\varphi}_N(\rho)\theta + \bar{\varepsilon}(\rho), \quad (21)$$

where $\rho = [\theta^T, \tau]$ and $\bar{\cdot}$ represents the signal y, u or e filtered by $L(p)$.

When the estimated parameter $\hat{\theta}$ is a function of τ , i.e.

$$\hat{\theta}(\tau) = [\bar{\Psi}_N(\rho) \bar{\Phi}_N^T(\rho)]^{-1} \bar{\Psi}_N(\rho) \bar{Y}_N, \quad (22)$$

then the time delay τ can be estimated as [12][21],

$$\hat{\tau} = \arg \min_{\tau} \bar{J}_N(\tau) = \arg \min_{\tau} \bar{J}_N(\rho)|_{\theta=\hat{\theta}(\tau)}, \quad (23)$$

with

$$\bar{J}_N(\rho) = \frac{1}{2(N-s+1)} \bar{\varepsilon}^T(\rho) \bar{\varepsilon}(\rho), \quad (24)$$

where s is chosen to guarantee that at sample s , $t_s \geq \tau$.

By using the Gauss-Newton method, the system parameters and the time delay can be iteratively estimated,

$$\begin{aligned} \hat{\tau}^{j+1} &= \hat{\tau}^j - \mu^j [\nabla^2 \bar{J}_N(\hat{\tau}^j)]^{-1} \nabla \bar{J}_N(\hat{\tau}^j) \\ \hat{\theta}^{j+1} &= [\bar{\Psi}_N(\hat{\theta}^j, \hat{\tau}^{j+1}) \bar{\Phi}_N^T(\hat{\theta}^j, \hat{\tau}^{j+1})]^{-1} \bar{\Psi}_N(\hat{\theta}^j, \hat{\tau}^{j+1}) \bar{Y}_N(\hat{\theta}^j), \end{aligned} \quad (25)$$

where μ^j is the step size and

$$\begin{aligned} \nabla \bar{J}_N(\hat{\tau}^j) &= \bar{\varepsilon}_\tau^T \bar{\varepsilon}(\rho), \\ \nabla^2 \bar{J}_N(\hat{\tau}^j) &= \bar{\varepsilon}_\tau^T \bar{\varepsilon}_\tau - \bar{\varepsilon}_\tau^T \bar{\varepsilon}_\theta (\bar{\varepsilon}_\theta^T \bar{\varepsilon}_\theta)^{-1} \bar{\varepsilon}_\theta^T \bar{\varepsilon}_\tau, \\ \varepsilon_\tau &= \frac{\partial \varepsilon(\rho)}{\partial \tau} \Big|_{\rho=\hat{\rho}^j} = pG(p, \hat{\theta}^j) u(\hat{\tau}^j), \\ \varepsilon_\theta &= \frac{\partial \varepsilon(\rho)}{\partial \theta} \Big|_{\rho=\hat{\rho}^j} = -\bar{\Psi}_N^T(\hat{\theta}^j, \hat{\tau}^j). \end{aligned} \quad (26)$$

The SRIVC-based time delay estimation with filtering [12] is summarized in Algorithm II.

ALGORITHM II

Step 1. Initialization

- (1) Set the initial value $\hat{\tau}^0$, the boundaries^a τ_{min}, τ_{max} , $\Delta\tau_{min}$ ^b, the cut-off frequency ω_c^L of $L(p)$, the cut-off frequency of $F(p)$, ω_c^{SVF} , and a small positive ε to indicate convergence.
- (2) Based on the initial value $\hat{\tau}^0$, use the SRIVC method to compute $\hat{\theta}^0$.

(3) Compute $\nabla^2 \tilde{J}_N(\hat{\tau}^0)$ and $\nabla \tilde{J}_N(\hat{\tau}^0)$ from (26).

Step 2. Iterative estimation

for $j = 1$:convergence

(1) Compute $\Delta \hat{\tau}^j$ using the filtered input/output data, equation (26) and

$$\Delta \hat{\tau}^j = -[\nabla^2 \tilde{J}_N(\hat{\tau}^j)]^{-1} \nabla \tilde{J}_N(\hat{\tau}^j). \quad (27)$$

- (2) (a) Compute $\hat{\tau}^{j+1} = \hat{\tau}^j + \Delta \hat{\tau}^j$
 If $\hat{\tau}^{j+1} \notin [\tau_{min}, \tau_{max}]$, let $\Delta \hat{\tau}^j = \Delta \hat{\tau}^j / 2$ and repeat this step.
 If $|\Delta \hat{\tau}^j| \leq \Delta \tau_{min}$, break.
 (b) Estimate $\hat{\theta}^{j+1}$ using the SRIVC method and time delay $\hat{\tau}^{j+1}$ from the filtered input/output data.
 (c) Compute $\tilde{J}_N(\hat{\tau}^{j+1})$. If $\tilde{J}_N(\hat{\tau}^{j+1}) \geq \tilde{J}_N(\hat{\tau}^j)$, let $\Delta \hat{\tau}^j = \Delta \hat{\tau}^j / 2$ and return to (a).
 (3) If $\tilde{J}_N(\hat{\tau}^j) - \tilde{J}_N(\hat{\tau}^{j+1}) \geq \varepsilon$, go to Step 1, else break.

end

Step 3. Refined parameter estimation

Repeat Step 2 with low-pass filter $L(p) = 1$.

^a τ_{min} and τ_{max} are boundaries for the delay, which are known a priori, i.e. $\tau_{min} \leq \tau_0 \leq \tau_{max}$

^b $\Delta \tau_{min}$ is the limit value of the step size of the time delay estimate using Gauss-Newton method.

Remark 3.1 The smaller the cut-off frequency ω_c^L , the larger the global convergence region. However, when ω_c^L is chosen too small, information from the data is lost, which affects the accuracy of the parameter estimation [12]. The suggestion in [12] is,

$$\omega_c^L \geq \frac{1}{10} bw, \quad (28)$$

where bw is the bandwidth of the system $G_0(p)$ (rad/sec).

4 SRIVC-based time delay estimation with multiple filtering and redundancy method

In this section, we propose a method that improves convergence to the global minimum of the time delay optimization problem using the useful redundancy technique [3][4]. The useful redundancy technique was developed to avoid local minima when solving a non-linear inverse problem. As stated in the previous section, the main problem in time delay estimation is that the cost function $J_N(\rho)$ possesses many local minima [12]. The filtering operation [14] described in Section 3.3 can be employed in order to increase the global convergence region of the cost function associated with the

time delay estimation. However, when the data is very noisy or the initial value, $\hat{\tau}^0$, is located far from the global minimum, using only one filter does not guarantee that the solution of the optimization problem converges to the global minimum.

To demonstrate how useful redundancy can be utilized in our estimator, we first describe the useful redundancy technique.

4.1 The useful redundancy technique

We define the useful redundancy technique by quoting directly from [4].

Definition 1 [4]. Consider an optimization problem,

$$\min_{\rho} J_0(\rho).$$

Then it is called *M-safely redundant* if and only if the following conditions hold:

- (1) There exists a finite M cost functions J_i sharing the same global minimum $\rho^* \in \mathbb{R}^n$.
- (2) There exists a solver (or an iterative scheme) ℓ and a finite number of iterations $r^* \in \mathbb{N}$ such that for some $\gamma \in [0, 1)$ and all $\rho \in \mathbb{P}$ the following inequality holds,

$$\Delta_N^\gamma(\rho) = \min_{i \in \{0, \dots, M\}} [J_0(\ell^{(r^*)}(\rho, J_i)) - \gamma J_0(\rho)] \leq 0 \quad (29)$$

where $\ell^{(r^*)}(\rho, J_i)$ is the candidate solution obtained after r^* iterations of ℓ using the cost function J_i and starting from the initial guess ρ . \square

The solver path (ρ, J_i) is defined as the sequence of iterates $\ell^{(j)}(\rho, J_i)$ for the solver ℓ when the cost function J_i starts from an initial guess ρ . Condition 2 means that for any initial ρ , there always exists a solver path ℓ that corresponds to a decrease in the original cost J_0 after at most r^* iterations. It is proven in [4] that if an optimization problem is *M-safely redundant* following from Definition 1, that convergence to a global minimum is guaranteed.

An algorithm that describes the useful redundancy technique is given in Algorithm III.

ALGORITHM III

Step 1. Initialization Choose an initial value $\rho^{(0)}$.

Step 2. Iterative estimation

for $k=0$:converge

- (1) Set $i = 0$ and $failure = true$
- (2) While $failure$ do
 - (a) Use the solver path (ρ, J_i) starting from $\rho^{(k)}$, find the corresponding minima $\xi^{(k,i)}$.

(b) Compute
 $failure = \left(J_0(\xi^{(k,i)}) > (1 - \varepsilon)J_0(\rho^{(k)}) \right)^a$
(c) If $failure$ then $i = (i + 1) \bmod M$,
Else set $k = k + 1, \rho^{(k)} = \xi^{(k,i)}$.
End while

end

^a ε is a predefined small value that is chosen based on the desired accuracy.

4.2 Theoretical results related to the time delay cost function

To construct an M -safely redundant optimization problem for the estimation of the time delay, we need a cost function J_0 and multiple solver paths that satisfy the two conditions in Definition 1. Here, the solver paths are generated by filtering the time delay estimation error using a set of low-pass filters with different cut-off frequencies. The cost function J_0 is formulated from these filtered cost functions such that there always exists a solver path that corresponds to a decrease in J_0 after a finite number of iterations. Next we describe the set of filters and the cost function J_0 required to satisfy the conditions in Definition 1.

Consider a continuous-time, linear, time-invariant SISO system described by

$$y(t) = G_0(p)u(t - \tau_0) + e(t). \quad (30)$$

We make a further assumption on the noise $e(t)$.

Assumption 4.1 $e(t)$ is white random process uncorrelated with u having intensity λ .

For an estimate $G(p, \theta)$ and τ , the estimation error $\varepsilon(t, \theta, \tau)$ can be computed as,

$$\varepsilon(t, \theta, \tau) = y(t) - G(p, \theta)u(t - \tau). \quad (31)$$

As mentioned in Section 3, the delay can be estimated by minimizing the cost function $J(\theta, \tau)$, where $J(\theta, \tau) = \int_{-\infty}^{\infty} \varepsilon(t, \theta, \tau)^2 dt$. If the estimation error is filtered by the low-pass filter $L(p)$, then an estimate of θ and τ can be computed by,

$$(\hat{\theta}, \hat{\tau}) = \arg \min J(\theta, \tau), \quad (32)$$

where

$$\begin{aligned} \bar{J}(\theta, \tau) &= \int_{-\infty}^{\infty} \bar{\varepsilon}(t, \theta, \tau)^2 dt = \int_{-\infty}^{\infty} \left\{ L(p)\varepsilon(t, \theta, \tau) \right\}^2 dt \\ &= \int_{-\infty}^{\infty} \left\{ L(p)[y(t) - G(p, \theta)u(t - \tau)] \right\}^2 dt \end{aligned} \quad (33)$$

which by Parseval's theorem is equivalent to,

$$\bar{J}(\theta, \tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[|G_0(j\omega)e^{-j\omega\tau_0} - G(j\omega, \theta)e^{-j\omega\tau}|^2 \times \Psi_u(\omega) + \Psi_v(\omega) \right] |L(j\omega)|^2 d\omega, \quad (34)$$

with $\Psi_u(\omega)$ the spectral density of $u(t)$ and the spectral density $\Psi_v(\omega) = \lambda$ follows from Assumption 4.1. If the transfer function $G_0(p)$ is known; the input signal white noise, i.e. $\Psi_u(\omega) = 1$; and $\delta\tau = \tau - \tau_0$, the cost function $\bar{J}(\theta, \tau)$ can be written as,

$$\begin{aligned} \bar{J}(\delta\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[|1 - e^{-j\omega\delta\tau}|^2 |G_0(j\omega)e^{-j\omega\tau_0}|^2 \right. \\ &\quad \left. \times \Psi_u(\omega) + \lambda \right] |L(j\omega)|^2 d\omega \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left[(1 - \cos(\omega\delta\tau)) |G_0(j\omega)|^2 + \frac{\lambda}{2} \right] \\ &\quad \times |L(j\omega)|^2 d\omega. \end{aligned} \quad (35)$$

Recall that we are concerned with how to choose a set of filters and the cost function J_0 such that the two conditions of an M -safely redundant problem are satisfied. To satisfy the first condition we need to ensure all the cost functions share the same global minimum. The second condition requires the filter set to be constructed such that for any initial value of $\delta\tau$, there always exists a cost function whose solver path corresponds to a decrease in the cost function J_0 after a finite number of iterations. Note that these conditions can be checked if we know the locations of the global minimum and the extrema, i.e. the minima and maxima of $\bar{J}(\delta\tau)$.

First we establish a result for the location of the global minimum of the (non)filtered delay cost function.

Theorem 4.1 Consider the system $G_0(p)$ as described in (30). When $\lambda = 0, \forall \delta\tau \in \mathbb{R}, \forall$ low-pass filters $L(p) \neq 0$, such that $\bar{J}(\delta\tau) \geq \bar{J}(0)$, the equality $\bar{J}(\delta\tau) = \bar{J}(0)$ occurs if and only if $\delta\tau = 0$.

Proof. The proof of Theorem 4.1 is provided in Appendix A.1.

Theorem 4.2 provides a result for the locations of the extrema of the filtered cost function $\bar{J}(\delta\tau)$.

Theorem 4.2 If $L(p)$ is selected such that $L(p)G_0(p)$ is an ideal low-pass filter with cut-off frequency ω_c rad/sec, then the locations of the i^{th} positive extrema of the time delay cost function $\bar{J}(\delta\tau)$ can be approximated by,

$$\tilde{\delta\tau}_i \simeq \frac{2i+1}{4}T_c - \frac{T_c}{(2i+1)\pi^2}, \quad (36)$$

where $T_c = 2\pi/\omega_c$ and $i \geq 1$. When i is even, the corresponding extremum is a minimum and when i is odd, it is a maximum.

Proof. The proof of Theorem 4.2 is provided in Appendix A.2.

Remark 4.1 From Theorem 4.2, the locations of the extrema for the filtered cost function are known, hence the global convergence region is,

$$|\delta\tau| < \frac{3\pi}{2\omega_c} - \frac{2}{3\pi\omega_c}, \quad (37)$$

corresponding to the distance from the maxima closest to the global minimum.

The result of Remark 4.1 can be seen in Fig. 1, which provides the plot of $\bar{J}(\delta\tau)$ when $\omega_c = 2\pi$ (the constant λ of the integral in (35) is set to 0 here for simplicity).

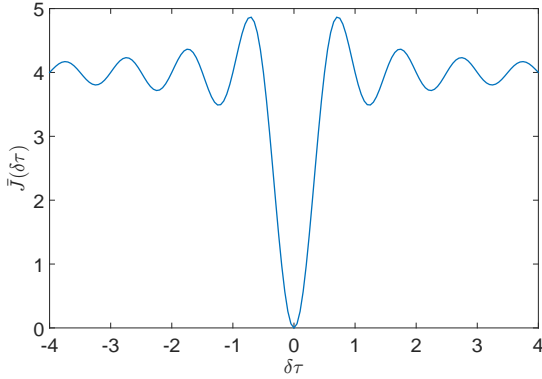


Fig. 1. Plot of $\bar{J}(\delta\tau)$ when $\omega_c = 2\pi$

4.3 Time delay estimation with the useful redundancy technique

Now we can specify a filter set that satisfies the two conditions in Definition 1.

Condition 1: From Theorem 4.1, the first condition is satisfied for any set of low-pass filters as the global minimum of all the filtered time delay cost functions occurs when $\delta\tau = 0$.

Condition 2: The second condition depends critically on the choice of filters and cost function J_0 . The filter set needs to be chosen such that for any initial value of $\delta\tau$, there always exists a cost function whose solver path corresponds to a decrease in the cost function J_0 after a finite number of iterations.

To verify that condition 2 is satisfied, consider the necessary and sufficient conditions. To satisfy the necessary condition, we need to show that for any initial value $\delta\tau_0$, there always exists a path that has a minima location, $\delta\tau_1$, closer to the global minimum w.r.t. $\delta\tau_0$. For the sufficient condition, we need to establish that $J_0(\delta\tau_1) < J_0(\delta\tau_0)$.

First, we check the necessary condition. A simple choice is a filter set $L_k(p)$, $k = \overline{1, n_f}$ where the cut-off frequencies, $\omega_{c,k}$, of $L_k(p)G_0(p)$ span from $1/\beta$ to 1 of the system bandwidth ($10 \geq \beta > 1$)¹. The cut-off frequencies are chosen based on the corresponding periods, $T_{c,k}$, being linearly spaced, i.e.,

$$T_{c,k} = \beta T_{bw} + (k-1) \frac{(1-\beta)T_{bw}}{n_f-1}, \quad k = \overline{1, n_f}, \quad (38)$$

where $T_{bw} = 2\pi/bw$, bw is the bandwidth of the system $G_0(p)$ (rad/sec). This is due to the local minima of the filtered cost function being linearly dependant on the period of the cut-off frequency (Theorem 4.2).

Finally we need to prove the existence of such a filter set that satisfies the necessary condition of Condition 2 in Definition 1.

Theorem 4.3 There exists a filter set $L_k(p)$, $k = \overline{1, n_f}$ chosen such that, $\forall \beta \geq \frac{5/4 - 1/(5\pi^2)}{3/4 - 1/(3\pi^2)}$, where $L_k(p)G_0(p)$ are ideal low-pass filters chosen with linearly spaced periods such that the cut-off frequencies, $\omega_{c,k}$, spanning from $1/\beta$ to 1 of the system bandwidth, i.e.

$$T_{c,k} = \beta T_{bw} - (k-1) \frac{(\beta-1)T_{bw}}{n_f-1}, \quad k = \overline{1, n_f}, \quad (39)$$

where $T_{bw} = 2\pi/bw$ and $T_{c,k} = 2\pi/\omega_{c,k}$. Then,

$$\forall \delta\tau_0 \neq 0, \exists L_q(p) : |\xi(L_q(p), \delta\tau_0)| < |\delta\tau_0|, \quad (40)$$

where $\xi(L_q(p), \delta\tau_0)$ is the corresponding minimum found using the filtered time delay cost function generated by $L_q(p)$ with the initial delay $\delta\tau_0$.

Proof. The proof of Theorem 4.3 is provided in Appendix A.3.

Remark 4.2 From Theorem 4.3, we see that with the choice of,

$$\beta \geq \frac{5/4 - 1/(5\pi^2)}{3/4 - 1/(3\pi^2)}, \quad (41)$$

there always exists an n_f , that can be computed by (A.8) in Appendix A.3, that ensures the filter set defined in Theorem 4.3 satisfies the necessary part of Definition 1, Condition 2.

Remark 4.3 Note that (A.8) provides a loose lower bound. It is possible to have a smaller value of n_f and still obtain a filter set that satisfies the necessary condition.

Remark 4.4 If $G_0(p)$ does not have resonant peaks then it doesn't matter if $G_0(p)$ is known or unknown, as any ideal

¹ From Remark 3.1, β is chosen ≤ 10 .

low pass filter $L_k(p)$ with bandwidth smaller than the bandwidth of $G_0(p)$ will allow $L_k(p)G_0(p)$ to approximate the desired low pass filter behaviour. If $G_0(p)$ has resonant peaks and the bandwidth of $L_k(p)$ is chosen significantly smaller than that of $G_0(p)$ then $L_k(p)G_0(p)$ will approximate the desired low pass filter behaviour.

An empirical method can also be utilized to determine the number of filters. For example, to check if a set of n_f filters with the ratio β satisfies the necessary condition, we can plot n_f cost functions that follow from (A.2) in Appendix A.2 and check the minima locations using (36) to see that for any initial delay, $\delta\tau_0$, there always exists a cost function where the corresponding minima $\delta\tau_i$ is found closer to the global minimum with respect to $\delta\tau_0$. Fig. 2 shows a set of six cost functions with the ratio $\beta = 2$ for a system bandwidth of 2π rad/sec, i.e. the cut-off frequencies span from $1/2$ to 1 (Hz). From the figure, we can see that for any initial value of the time delay less than 15 sec, there always exists a cost function where the minima is found closer to the global minimum with respect to the initial delay. This can be confirmed by computing the minima of all the cost functions using (36) and observing the path to the global minimum.

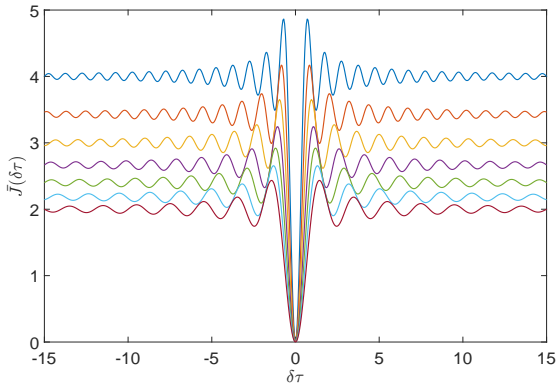


Fig. 2. Plots of $J(\delta\tau)$ set of filters with linearly spaced periods

Note that in practice, as $G_0(p)$ is not known a priori, and with the difficulty of implementing an ideal low-pass filter with irregular data, the filters $L_k(p)$ are implemented as Butterworth filters with cut-off frequencies $\omega_{c,k}$. Due to this, the extrema locations are not exactly those specified in (36). We discuss the effect of this difference by first citing the following theorem from [14], as a lemma.

Lemma 4.1 [14] Consider a system with transfer function $F(s) = L(s)G_0(s)e^{-s\tau_0}$, let u be the input signal and w the corresponding output signal of the system. Then we have the following relationship,

$$\bar{J}(\delta\tau) = 2[\gamma_w(0) - \gamma_w(\delta\tau)] = 2\gamma_w(0)[1 - \rho_w(\delta\tau)], \quad (42)$$

where $\gamma_w(t)$ is the autocorrelation function of $w(t)$ and $\rho_w(\delta\tau) = \gamma_w(\delta\tau)/\gamma_w(0)$.

Proof. See proof in [14].

With respect to the model setting in (30), $\gamma_w(\delta\tau)$ in Lemma 4.1 is the autocorrelation of the output of the system. Hence from Lemma 4.1, for an input signal of Gaussian distributed white noise, the filtered delay cost function $\bar{J}(\delta\tau)$ will have the form,

$$\bar{J}(\delta\tau) = \sum_{k=1}^{n_1} a_k e^{-b_k \delta\tau} + \sum_{k=1}^{n_2} h_k e^{-c_k \delta\tau} \cos(d_k \delta\tau + \psi_k), \quad (43)$$

where b_k is the absolute value of the k^{th} real pole of $L(s)G_0(s)$; c_k and d_k are the absolute values of the real and imaginary parts of the k^{th} complex pole of $L(s)G_0(s)$; and weighting coefficients h_k are related to the magnitude of the frequency response at the corresponding frequency. Augmenting a low-pass filter $L(s)$ to $G_0(s)$ obviously introduces slow poles into the system and by choosing $L(s)$ to be a Butterworth filter, it inherently introduces complex poles to the system. Hence, from (43), the filtered delay cost function will exhibit a slightly underdamped response. It can then be seen that the locations of the extrema of the cost function are associated with the underdamped response of the Butterworth filter. Hence, the augmentation of $L(s)$ with $G_0(s)$ when $L(s)$ is implemented as a Butterworth filter provides a cost function similar in nature to Fig. 1.

Next, to establish the sufficient condition, we need to derive a cost function J_0 that shares the same global minimum with the filtered cost functions where the minimum path corresponds to a decrease in J_0 .

To analyse this condition, we take into consideration the model error that is generated in the iterations when estimating the rational part of the system due to an inaccurate estimate of the delay. Consider we have an estimate for the delay, $\hat{\tau}$, then when estimating the rational component of (30), we have the following relationship in the Laplace domain,

$$\hat{G}(s) = G_0(s)e^{-(\tau_0 - \hat{\tau})s}, \quad (44)$$

where $\hat{G}(s)$ is the estimate for the rational part of the system. It is obvious that when $\hat{\tau} \neq \tau_0$, $\hat{G}(s) \neq G_0(s)$. Hence there always exists model error in estimating the rational part of the system if $\hat{\tau} \neq \tau_0$.

We now quantify the model error to understand the impact of it on the cost function. Here, a model reduction technique based on an identification method [19] is used to find a rational transfer function $G_r(s)$. The idea is to generate a noise-free data set from the system $G_0(s)e^{-(\tau_0 - \hat{\tau})s}$, then use the SRIVC method to obtain a rational system $\hat{G}_r(s)$ that has the same model order as the true system and satisfies (44). For example, Fig. 3 shows the graphs of ten filtered delay cost functions after including the model error for the system $G(s) = 2e^{-8s}/(0.25s^2 + 0.7s + 1)$.

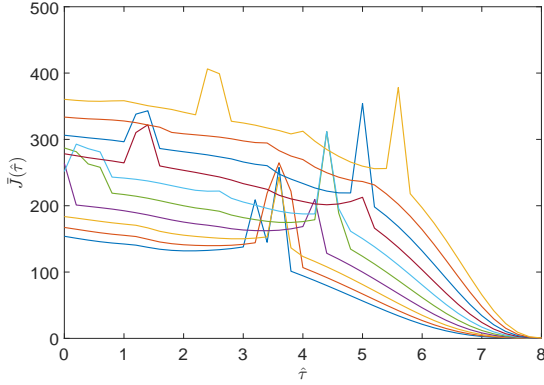


Fig. 3. Plot of $\bar{J}(\hat{\tau})$ for set of filters $L_k(p)$ defined as in Theorem 3.

We now define the cost function J_0 as,

$$J_0(\delta\tau) = \frac{1}{n_f} \sum_{k=1}^{n_f} \bar{J}_k(\delta\tau), \quad (45)$$

where $\bar{J}_k(\delta\tau) = 100 \frac{\|y(t) - \hat{x}_k(t)\|_2}{\|y(t) - \mathbb{E}\{y(t)\}\|_2}$, $\hat{x}_k(t)$ is the estimated system output when the filter $L_k(s)$ is used. Note that $\bar{J}_k(\delta\tau)$ is the normalized version of the filtered cost function, $\bar{J}_k(\delta\tau)$.

The averaging is used to obtain a smooth decreasing cost function J_0 as there are always errors in the estimation due to noise and model error. Fig. 4 shows the average normalized mean square error for the ten filtered delay cost functions shown in Fig. 3. It can be seen that J_0 has the same global minimum as the ten filtered cost functions. Also shown in Fig. 4 is the minimum path through the filtered cost function that corresponds to a decrease in J_0 . Note that the ‘*’ in Fig. 4 corresponds to a switch between cost functions $J_k(\delta\tau)$.

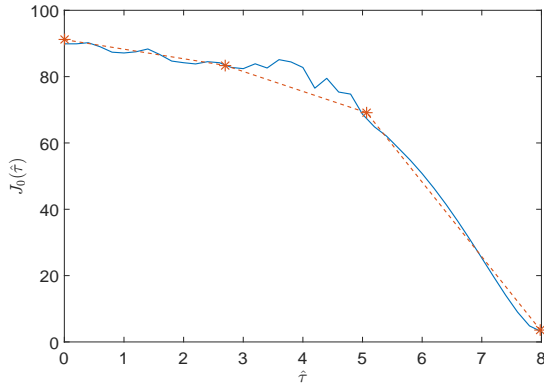


Fig. 4. Plot of the average normalized mean square error of the filtered cost delay function (*: switching point, solid line: J_0).

Therefore, in summary, the *M-safely redundant* optimization problem for the time delay estimation is defined as,

$$\min_{\tau} J_0(\tau), \quad (46)$$

with $J_0(\tau)$ defined as in (45). The filter set consists of n_f Butterworth filters with cut-off frequencies, chosen such that the corresponding periods are linearly spaced, that span from $1/\beta$ to 1 of the system bandwidth where $\beta \geq \frac{5/4 - 1/(5\pi^2)}{3/4 - 1/(3\pi^2)}$ and n_f is chosen using the methodology discussed in Section 4.3.

4.4 Algorithm

The proposed algorithm utilizing useful redundancy with multiple filters is described in Algorithm IV.

ALGORITHM IV

Step 1. Initialization

- (1) Select a set of low-pass filters as suggested in Section 4.3.
- (2) Set the boundaries $\tau_{min}, \tau_{max}, \Delta\tau_{min}, \Delta\tau_{max}$ ^a, and the SVF cut-off frequency ω_c^{SVF} .
- (3) Choose an initial value $\hat{\tau}_0$ of the time delay.

Step 2. Iterative estimation

for $i=1$:converge

- (1) for $k=1:n_f$
 - (a) Choose the low-pass filter $L_k(p)$.
 - (b) Set the initial delay $\hat{\tau}_{i,k}^0$ as $\hat{\tau}_{i-1}$
 - (c) Use the SRIVC algorithm and Gauss-Newton method as described in Algorithm II (without Step 3) to estimate the time delay $\hat{\tau}_{i,k}$.
 - (d) Compute $J_0(\hat{\tau}_{i,k})$ (follows (45)).

end

- (2) Choose $\hat{\tau}_i = \underset{\hat{\tau}_{i,k}}{\operatorname{argmin}} J_0(\hat{\tau}_{i,k})$.

- (3) If $|(\hat{\tau}_i - \hat{\tau}_{i-1})/\hat{\tau}_i| \geq \epsilon$, go to Step 1, else break.^b

end

Step 3. Refine parameter estimation

Repeat Step 2 with the low-pass filter $L(p) = 1$.

^a Note that $\Delta\tau_{max}$ is used to constrain the increment $\Delta\hat{\tau}$ in the time delay estimation, i.e. when $\Delta\hat{\tau}^j > \Delta\tau_{max}$, set $\Delta\hat{\tau}^j = \Delta\tau_{max}$. This is to ensure the Wolfe conditions are satisfied at each iteration in the Gauss-Newton technique [26].

^b ϵ is a small value selected to obtain the desired accuracy.

5 Numerical Examples

In this section, we demonstrate the effectiveness of the algorithm through numerical examples. These examples are commonly used in the literature [12] [16] [24] and act as defacto benchmarks to compare the performance between different CT system identification algorithms.

5.1 Case 1

Case 1 considers the second order system used in [12], i.e.

$$G(s) = \frac{2e^{-8s}}{0.25s^2 + 0.7s + 1}. \quad (47)$$

In this example, we consider two sampling schemes for the input and output data.

- (1) Regular (uniform) sampling: The excitation input signal is a PRBS (pseudo-random binary sequence) of maximum length. The sampling time is 50ms, the number of stages in the shift register is 10 and the number of samples, $N = 4000$.
- (2) Irregular (nonuniform) sampling: The input excitation signal is a PRBS of maximum length. The number of stages in the shift register is 10 and the clock period is 0.5s. The input and output data are sampled at an irregular time instant t_k , with a sampling interval h_k uniformly distributed as, $h_k \sim U[0.01, 0.09](s)$. The realization of h_k is randomized for each run. The number of samples, $N = 4000$.

The additive output disturbance is Gaussian distributed white noise with zero mean designed to give Signal to Noise Ratios (SNR) of 5dB and 15dB. The SNR is defined as,

$$SNR = 10 \log \frac{P_x}{P_e},$$

with P_x and P_e the average power of the noise-free output $x(t_k)$ and the disturbance noise $e(t_k)$ respectively.

The system order is assumed known a priori (order 2). The frequency ω_c^{SVF} is chosen as 1 (rad/sec) which is same as the value used in [12]. To evaluate the algorithm, the time delay is initialized from different values $\hat{\tau}^0 = 0, 1, 3, 5, 7, 9$ (similar to the experiment in [12]) and also from a random value drawn from a uniform distribution $U[0, 9](s)$.

For the two sampling schemes, 100 different data sets are generated for each noise level. The system parameters and time delay are estimated using both the algorithm described in [12] and the algorithm proposed in this paper:

- First we consider the algorithm developed in [12]: The cut off frequency of the filter is chosen as 1/10 of the system bandwidth as suggested in [12]. For regular data, the command '*tdsrvc*' from the CONTSID toolbox [22] is used to estimate the time delay and the system parameters.

- For the multiple filtering algorithm utilizing the useful redundancy technique as proposed in this paper, the number of filters is set to 10 and the cut off frequencies, chosen based on linearly spaced periods, of the filters span from 1/10 to 1 of the system bandwidth. The order of the Butterworth filter is set to 10.

For both algorithms, the time delay boundaries τ_{min}, τ_{max} are set to 0 and 15 seconds respectively. The maximum number of iterations for the Gauss-Newton method (Step 2 in Algorithm II and IV) is set to 10. The threshold, ϵ , to determine convergence is set to 10^{-3} .

We consider the estimated time delay to be the global minimum when the relative error, ϵ_r , is less than 1%, i.e.

$$\epsilon_r = \frac{|\hat{\tau} - \tau_0|}{\tau_0} \times 100\%, \quad (48)$$

where $\hat{\tau}$ and τ_0 are the estimated delay and the true system delay respectively.

Fig. 5 contains plots of the 10 cost functions corresponding to the filters used in the useful redundancy method. It can be seen that they all share the same global minimum, however they all possess different local minima.

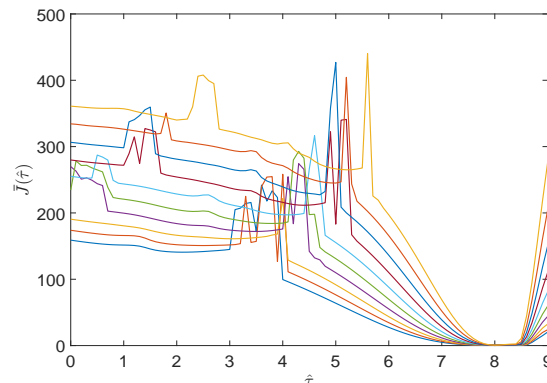


Fig. 5. Case 1: Cost functions of the 10 filters.

Fig. 6 presents a plot of J_0 , for a single dataset, showing the trajectory of the time delay estimate when the sampled data is regular and the SNR is 15dB. The algorithm is started with the initial value of the time delay, $\hat{\tau}^0 = 0$. The '*' in Fig. 6 represents the algorithm switching to another solver path. The algorithm stops when the condition in Algorithm IV is satisfied for the threshold $\epsilon = 10^{-3}$, i.e. when the relative error in the time delay between consecutive iterations is 10^{-3} .

Next we demonstrate the effectiveness of the proposed algorithm by showing the global convergence of 100 datasets with different initial values for the delay. The results are provided in Tables 1 and 2 where they are also compared to results from the existing method [12]. Note that the results presented here for the existing method are different to those

Table 1
Global convergence as a percentage of 100 Monte Carlo simulations for case 1

Sampling scheme	Initial delay	Existing method [12]						Proposed method					
		0s	1s	3s	5s	7s	9s	0s	1s	3s	5s	7s	9s
Regular sampled data	SNR = 5dB	65%	100%	100%	97%	100%	100%	100%	100%	100%	100%	100%	100%
	SNR = 15dB	69%	89%	92%	99%	100%	100%	100%	100%	100%	100%	100%	100%
Irregular sampled data	SNR = 5dB	1%	1%	0%	17%	100%	90%	100%	100%	100%	100%	100%	100%
	SNR = 15dB	1%	0%	7%	18%	100%	89%	100%	100%	100%	100%	100%	100%

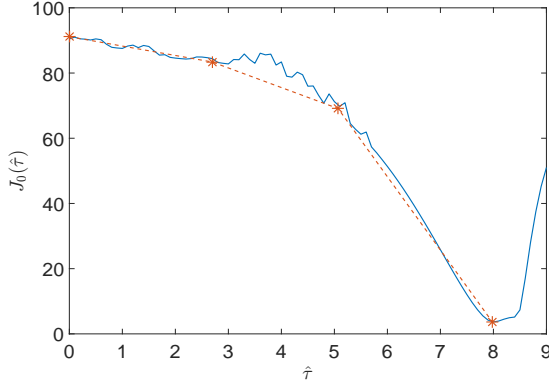


Fig. 6. Case 1: The original cost function and time delay estimate trajectory for a single dataset (*: switching point, solid line: J_0).

achieved in [12] as we chose a larger bound on the delay. In [12], τ_{max} was set to 10 seconds while here, it is set to 15 seconds.

Table 2
Global convergence as a percentage of 100 Monte Carlo simulations for case 1 with a randomly selected initial delay

Sampling scheme	SNR	Existing method [12]	Proposed method
Regular sampled data	5dB	90%	100%
	15dB	97%	100%
Irregular sampled data	5dB	40%	100%
	15dB	41%	100%

The results presented in Tables 1 and 2 show that the proposed method, which utilizes the useful redundancy technique, performs much better than the existing method [12] irrespective of the initial delay. When the initial delays are close to the true system delay, the global convergence percentage is high, i.e. approximately 100% for both algorithms. However, when the initial delays are poorly selected, the global convergence percentage is low for the existing method, e.g. with the initial delay = 0, the percentage convergence for the existing method [12] is less than 1% while the proposed method achieves 100% global percentage convergence for both SNRs and the two sampling schemes.

5.2 Case 2: Rao-Garnier test system

In this section, we consider a system based on the Rao-Garnier continuous time benchmark [16],

$$G(s) = \frac{(-6400s + 1600)e^{-8s}}{s^4 + 5s^3 + 408s^2 + 416s + 1600}$$

This system is linear, non-minimum phase with complex poles and a time delay. As with case 1, the experiment is conducted using both regular and irregular sampling schemes. For each sampling scheme, the input excitation signal is a PRBS of maximum-length. In the regular sampling case, the sample time is 10ms, the number of stages in the shift register is 10, the number of samples, $N = 8000$. For the irregular sampling, the number of stages in the shift register is 10 and the clock period is 0.5s. The input and output data are sampled at an irregular time instant t_k , where the sampling interval h_k is uniformly distributed as, $h_k \sim U[0.01, 0.05](s)$. The number of samples, $N = 4500$.

We consider two SNR levels: 5dB and 15dB. The frequency ω_c^{SVF} is chosen as 25 rad/sec, which is approximately the bandwidth of the system. Again, the initial values for the time delay, $\hat{\tau}^0$, are selected as 0, 1, 3, 5, 7, 9 as well as a random value from the uniform distribution $U[0, 9](s)$. The delay boundaries τ_{min}, τ_{max} are set to 0 and 15 sec respectively.

For the existing method, the cut off frequency of the filter is chosen as 1/10 of the system bandwidth. For the proposed method, the cut off frequencies of the filters, chosen based on linearly spaced periods, span from 1/10 to 1 of the system bandwidth and the number of filters used is 15. The order of the Butterworth filters is set to 10. The maximum number of iterations for the Gauss-Newton method (Step 2 in Algorithm II and IV) is 10. The threshold, ϵ , for convergence is 10^{-3} .

For comparison, 100 different data sets are generated for each noise level of each sampling scheme.

A graph showing the trajectory of a time delay estimate is plotted in Fig. 7 for a single data set. Switching between cost functions is indicated by '*'. Note that, if only one filter were to be used, there is a high chance of convergence

Table 3
Global convergence as a percentage of 100 Monte Carlo simulations for Case 2

Sampling scheme	Initial delay	Existing method [12]						Proposed method					
		0s	1s	3s	5s	7s	9s	0s	1s	3s	5s	7s	9s
Regular sampled data	SNR = 5dB	1%	52%	22%	43%	76%	68%	99%	99%	99%	99%	99%	99%
	SNR = 15dB	0%	60%	6%	28%	70%	69%	100%	100%	100%	100%	100%	100%
Irregular sampled data	SNR = 5dB	8%	16%	23%	26%	66%	48%	100%	100%	100%	100%	100%	100%
	SNR = 15dB	9%	10%	19%	27%	55%	50%	100%	100%	100%	100%	100%	100%

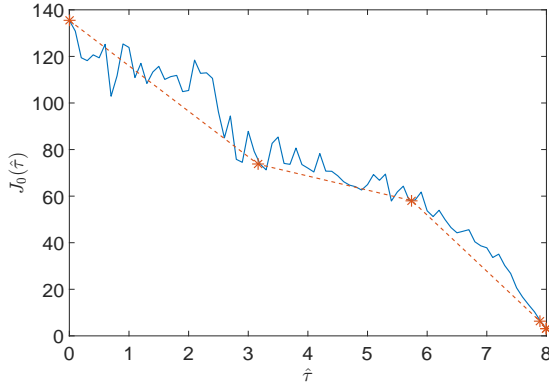


Fig. 7. Case 2: The original cost function and time delay estimate trajectory for a single data set (*: switching point, solid line: J_0)

to a local minima. However, by using the useful redundancy technique, the proposed algorithm can avoid the local minima and converge to the global minimum.

Table 4
Global convergence as a percentage of 100 Monte Carlo simulations for case 2 with a randomly selected initial delay

Sampling scheme	SNR	Existing method [12]	Proposed method
Regular sampled data	5dB	52%	99%
	15dB	46%	100%
Irregular sampled data	5dB	48%	100%
	15dB	21%	100%

The numerical results of the experiments are provided in Table 3 and 4. We can see that for both sampling schemes, the proposed method based on the useful redundancy technique performs much better as compared to the existing method. For all the initial values of delay used in this experiment, the existing method [12] never convergences 100% to the global minimum. However the proposed method utilizing useful redundancy still achieves a very high global convergence percentage, i.e. mostly 100% with any initial delay for both SNRs and sampling schemes.

6 Conclusion

The paper presents a new algorithm to estimate the parameters and time delay of a continuous-time system from regularly and irregularly sampled data. The idea is based on Instrumental Variable methods and employing the useful redundancy technique to enhance the global convergence by generating multiple cost functions by filtering the data several times. The paper also develops theoretical results related to the minima locations of the filtered delay cost function and the choice of filters to ensure the algorithm can converge to a global minimum. Numerical results show a significant improvement in the global convergence rate of the time delay estimation as compared to existing methods irrespective of the SNR.

References

- [1] M. Agarwal and C. Canudas. On-line estimation of time-delay and continuous-time process parameters. *International journal of control*, 46(1):295–311, 1987.
- [2] S. Ahmed, B. Huang, and S.L. Shah. Parameter and delay estimation of continuous-time models using a linear filter. *Journal of Process Control*, 16(4):323 – 331, 2006.
- [3] M. Alamir. On useful redundancy in dynamic inverse problems related optimization. 41(2):976 – 981, 2008.
- [4] M. Alamir, J.S. Welsh, and G.C. Goodwin. Redundancy versus multiple starting points in nonlinear system related inverse problems. *Automatica*, 45(4):1052 – 1057, 2009.
- [5] K.J. Åström and B. Bernhardsson. Systems with Lebesgue sampling. In A. Rantzer and C.I. Byrnes, editors, *Directions in Mathematical Systems Theory and Optimization*, pages 1–13. 2003.
- [6] A. Baysse, F.J. Carrillo, and A. Habbadi. Time domain identification of continuous-time systems with time delay using output error method from sampled data. *Proceeding of the 18th IFAC World Congress*, 44(1):9041–9046, 2011.
- [7] A. Baysse, F.J. Carrillo, and A. Habbadi. Least squares and output error identification algorithms for continuous time systems with unknown time delay operating in open or closed loop. *Proceeding of the 16th IFAC Symposium on System Identification*, 45(16):149–154, 2012.

[8] M. Björklund, M. Nihtilä, and T. Söderström. Algorithms for time delay estimation using a low complexity exhaustive search. *Proceedings of the 9th IFAC Symposium on Identification and System Parameter Estimation*, pages 1254–1259, 1991.

[9] S. Björklund. *A survey and comparison of time delay estimation methods in linear systems*. PhD thesis, Lund Institute of Technology, 2003.

[10] C. Carlemalm, S. Halvarsson, T. Wigren, and B. Wahlberg. Algorithms for time delay estimation using a low complexity exhaustive search. *IEEE Transactions on Automatic Control*, 44(5):1031–1037, 1999.

[11] G.C. Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, 75(2):236–255, 1987.

[12] F. Chen, H. Garnier, and M. Gilson. Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. *Journal of Process Control*, 25:19–27, 2015.

[13] D. Eckhard, A.S. Bazanella, C.R. Rojas, and H. Hjalmarsson. Cost function shaping of the output error criterion. Manuscript provisionally accepted for publication in *Automatica*, 2016.

[14] G. Ferretti, C. Maffezzoni, and R. Scattolini. On the identifiability of the time delay with least-squares methods. *Automatica*, 32(3):449 – 453, 1996.

[15] H. Garnier. Direct continuous-time approaches to system identification. overview and benefits for practical applications. *European Journal of Control*, 24:50–62, 2015.

[16] H. Garnier and L. Wang. *Identification of Continuous-time Models from Sampled Data*. Springer-Verlag, London, 2008.

[17] H. Garnier and P.C. Young. The advantages of directly identifying continuous-time transfer function models in practical applications. *International Journal of Control*, 87(7):1319–1338, 2013.

[18] P.J. Gawthrop and M.T. Nihtilä. Identification of time delays using a polynomial identification method. *Systems & Control Letters*, 5(4):267 – 271, 1985.

[19] C. Gu. *Model Order Reduction of Nonlinear Dynamical Systems*. PhD thesis, University of California, Berkeley, 2011.

[20] V. Kaminskas. Algorithms for time delay estimation using a low complexity exhaustive search. *Proceedings of the 5th IFAC Symposium on Identification and System Parameter Estimation*, pages 669–677, 1979.

[21] L.S.H. Ngia. Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using kautz and laguerre filters. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 48(6):562–579, 2001.

[22] A. Padilla, H. Garnier, and M. Gilson. Version 7.0 of the CONTSID toolbox. *Proceedings of the 17th IFAC Symposium on System Identification*, 48(28):757 – 762, 2015.

[23] R. Pupeikis. Recursive estimation of the parameters of linear systems with time delay. *Proceedings of the 7th IFAC Symposium on Identification and System Parameter Estimation*, pages 787–792, 1989.

[24] G.P. Rao and H. Garnier. Numerical illustrations of the relevance of direct continuous-time model identification. *Proceedings of the 15th IFAC World Congress*, 2002.

[25] H. Unbehauen and G.P. Rao. Continuous-time approaches to system identification-A survey. *Automatica*, 26(1):23–35, 1990.

[26] P. Wolfe. Convergence conditions for ascent methods. *SIAM Review*, 11(2):226–235, 1969.

[27] Z. Yang, H. Iemura, S. Kanae, and K. Wada. Identification of continuous-time systems with multiple unknown time delays by global nonlinear least-squares and instrumental variable methods. *Automatica*, 43(7):1257 – 1264, 2007.

[28] P.C. Young. Parameter estimation for continuous-time models - a survey. *Automatica*, 17(1):23–39, 1981.

A Appendix

A.1 Proof of Theorem 1

Proof A.1 From (35), when $\lambda = 0$, $\bar{J}(\delta\tau)$ becomes,

$$\bar{J}(\delta\tau) = \frac{1}{\pi} \int_{-\infty}^{\infty} (1 - \cos(\omega\delta\tau)) |G_0(j\omega)L(j\omega)|^2 d\omega. \quad (\text{A.1})$$

$\forall \delta\tau \in \mathbb{R}$, $\omega \geq 0$, we have

$$\begin{aligned} 1 - \cos(\omega\delta\tau) &\geq 0, \quad \forall \delta\tau, \\ |G_0(j\omega)L(j\omega)|^2 &\geq 0, \quad \forall \omega. \end{aligned}$$

Therefore, $\bar{J}(\delta\tau) \geq 0$, $\forall \delta\tau \in \mathbb{R}$, $\omega \geq 0$. Note that $\bar{J}(0) = 0$, hence we have $\bar{J}(\delta\tau) \geq \bar{J}(0)$, $\forall \delta\tau \in \mathbb{R}$, $\omega \geq 0$. The equality occurs when $L(j\omega) = 0$, $\forall \omega$ or $G_0(j\omega) = 0$, $\forall \omega$ or $\cos(\omega\delta\tau) = 1$, $\forall \omega$. As the system $G_0(p)$ and the filter $L(p)$ are not 0 (condition of Theorem 1), the equality only occurs when $\cos(\omega\delta\tau) = 1$, $\forall \omega$, hence $\delta\tau = 0$. \square

A.2 Proof of Theorem 2

Proof A.2 If $L(s)$ is selected such that $L(s)G_0(s)$ is an ideal low-pass filter with cut-off frequency ω_c , then $\bar{J}(\delta\tau)$ becomes,

$$\begin{aligned} \bar{J}(\delta\tau) &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} |1 - e^{-j\omega\delta\tau}|^2 d\omega + C \\ &= \frac{2\omega_c}{\pi} - \frac{2 \sin(\omega_c \delta\tau)}{\pi \delta\tau} + C, \end{aligned} \quad (\text{A.2})$$

where $C = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda |L(j\omega)|^2 d\omega$.

The minima and maxima of $\bar{J}(\delta\tau)$ occur at the roots of,

$$\frac{d\bar{J}(\delta\tau)}{d\delta\tau} = -\frac{2}{\pi} \frac{\omega_c \cos(\omega_c \delta\tau) \delta\tau - \sin(\omega_c \delta\tau)}{\delta\tau^2}. \quad (\text{A.3})$$

From (A.3), we can see that the extrema of $\bar{J}(\delta\tau)$ are also the extrema of the function $\text{sinc}(\omega_c \delta\tau)$. However, the minima of $\bar{J}(\delta\tau)$ will be the maxima of $\text{sinc}(\omega_c \delta\tau)$ and the maxima of $\bar{J}(\delta\tau)$ will be the minima of $\text{sinc}(\omega_c \delta\tau)$. Note that for the function $\text{sinc}(\omega_c \delta\tau)$, the locations of the i^{th} ($i \geq 1$) positive extremum $\tilde{\delta\tau}$ can be approximated by,

$$\begin{aligned}\tilde{\delta\tau}_i &\simeq (i + \frac{1}{2}) \frac{\pi}{\omega_c} - \frac{1}{(i + \frac{1}{2})\pi\omega_c} \\ &\simeq \frac{2i+1}{4} T_c - \frac{T_c}{(2i+1)\pi^2}.\end{aligned}\quad (\text{A.4})$$

where T_c is the corresponding period of the cut-off frequency ω_c . The minima occurs when i is odd and the maxima occurs when i is even.

Therefore, for $\bar{J}(\delta\tau)$, the positive extrema can also be approximated using (A.4), the only difference is that the minima of $\bar{J}(\delta\tau)$ occurs when i is even and the maxima occurs when i is odd. \square

A.3 Proof of Theorem 3

Proof A.3 Here we use the positive extrema locations of the time delay cost function in (36). Since the filtered time delay cost function $\bar{J}(\delta\tau)$ is an even function, we only need to prove the theorem for $\delta\tau_0 > 0$.

Denote $\tilde{\delta\tau}_{i,k}^{\min}$ as the i^{th} positive minimum of the filtered time delay cost function $\bar{J}(\delta\tau)$ generated by the filter $L_k(p)$; $\tilde{\delta\tau}_{i,k}^{\max}$ as the i^{th} positive maximum of the filtered time delay cost function $\bar{J}(\delta\tau)$ generated by the filter $L_k(p)$. From Theorem 4.2, they can be computed by,

$$\begin{aligned}\tilde{\delta\tau}_{i,k+1}^{\min} &= \frac{4i+1}{4} T_{c,k+1} - \frac{T_{c,k+1}}{(4i+1)\pi^2}, \\ &= \left(\frac{4i+1}{4} - \frac{1}{(4i+1)\pi^2} \right) \left(\beta - k \frac{\beta-1}{n_f-1} \right) T_{bw} \\ \tilde{\delta\tau}_{i,k+1}^{\max} &= \frac{4i-1}{4} T_{c,k+1} - \frac{T_{c,k+1}}{(4i-1)\pi^2}, \\ &= \left(\frac{4i-1}{4} - \frac{1}{(4i-1)\pi^2} \right) \left(\beta - k \frac{\beta-1}{n_f-1} \right) T_{bw}.\end{aligned}\quad (\text{A.5})$$

From (A.5), for any filter $L_k(p)$, we have the following property,

$$\tilde{\delta\tau}_{i,k}^{\max} < \tilde{\delta\tau}_{i,k}^{\min} < \tilde{\delta\tau}_{i+1,k}^{\max}, \quad \forall i \in \mathbb{Z}^+. \quad (\text{A.6})$$

Consider the two following cases:

Case 1: $\delta\tau_0 \leq \tilde{\delta\tau}_{1,1}^{\max}$, then from (A.6), we have, $\xi(L_1(p), \delta\tau_0) = 0$, which is smaller than $\delta\tau_0$ as $\delta\tau_0 \neq 0$. Therefore, $\xi(L_1(p), \delta\tau_0) < \delta\tau_0$. \square

Case 2: $\delta\tau_0 > \tilde{\delta\tau}_{1,1}^{\max}$. Denote i_0 such that,

$$\tilde{\delta\tau}_{i_0-1,1}^{\max} < \delta\tau_0 \leq \tilde{\delta\tau}_{i_0,1}^{\max}, \quad i_0 \in \mathbb{N}, i_0 \geq 2. \quad (\text{A.7})$$

Now consider a filter set $L_k(p)$, $k = \overline{1, n_f}$ defined as in Theorem 4.3, with,

$$n_f \geq (1/M + \beta - 2)/(1/M - 1), \quad n_f \in \mathbb{N}, \quad (\text{A.8})$$

where

$$M = \frac{4i_0 - 3 - 4/((4i_0 - 3)\pi^2)}{4i_0 - 1 - 4/((4i_0 - 1)\pi^2)}. \quad (\text{A.9})$$

We need to prove the filter set satisfies the condition in (40). First, we prove that the filter set $L_k(p)$, $k = \overline{1, n_f}$ has the following property,

$$\tilde{\delta\tau}_{i+1,k+1}^{\max} \geq \tilde{\delta\tau}_{i,k}^{\min},$$

where $i = \overline{1, i_0 - 1}$, $k = \overline{1, n_f - 1}$.

From (A.8),

$$\begin{aligned}n_f &\geq \frac{1/M + \beta - 2}{1/M - 1} \\ \Rightarrow M &\leq 1 - \frac{(\beta - 1)/(n_f - 1)}{\beta - (n_f - 2)(\beta - 1)/(n_f - 1)}.\end{aligned}\quad (\text{A.10})$$

Note that, as $\beta > 1$,

$$\begin{aligned}1 - \frac{(\beta - 1)/(n_f - 1)}{\beta} &> 1 - \frac{(\beta - 1)/(n_f - 1)}{\beta - (\beta - 1)/(n_f - 1)} \\ &> 1 - \frac{(\beta - 1)/(n_f - 1)}{\beta - (n_f - 2)(\beta - 1)/(n_f - 1)}.\end{aligned}\quad (\text{A.11})$$

From (A.10) and (A.11),

$$1 - \frac{(\beta - 1)/(n_f - 1)}{\beta - (k - 1)(\beta - 1)/(n_f - 1)} \geq M, \quad (\text{A.12})$$

where $k = \overline{1, n_f - 1}$, hence, combining (A.9) and (A.12),

$$\frac{\beta - k(\beta - 1)/(n_f - 1)}{\beta - (k - 1)(\beta - 1)/(n_f - 1)} \geq \frac{4i_0 - 3 - 4/((4i_0 - 3)\pi^2)}{4i_0 - 1 - 4/((4i_0 - 1)\pi^2)}, \quad (\text{A.13})$$

where $k = \overline{1, n_f - 1}$.

Next, consider the function

$$f(x) = \frac{4x + 1 - 4/((4x + 1)\pi^2)}{4x + 3 - 4/((4x + 3)\pi^2)}. \quad (\text{A.14})$$

Taking the derivative w.r.t. x of (A.14),

$$\begin{aligned} \frac{\partial f(x)}{\partial x} &= \frac{8 + \frac{16}{(4x+1)\pi^2} + \frac{16(4x+3)}{(4x+1)^2\pi^2} + \frac{64/\pi^4}{(4x+1)(4x+3)^2}}{\left((4x+3) - 4/((4x+3)\pi^2)\right)^2} \\ &\quad - \frac{\frac{16}{(4x+3)\pi^2} + \frac{16(4x+1)}{(4x+3)^2\pi^2} + \frac{64/\pi^4}{(4x+1)^2(4x+3)}}{\left((4x+3) - 4/((4x+3)\pi^2)\right)^2} \end{aligned}$$

Now $\forall x \geq 1$,

$$\begin{aligned} \frac{\partial f(x)}{\partial x} &> \frac{8 - \frac{16}{(4x+3)\pi^2} - \frac{16(4x+1)}{(4x+3)^2\pi^2} - \frac{64/\pi^4}{(4x+1)^2(4x+3)}}{\left((4x+3) - 4/((4x+3)\pi^2)\right)^2} \\ &> \frac{8 - \frac{16}{(4x+3)\pi^2} - \frac{16}{(4x+3)\pi^2} - \frac{64/\pi^4}{(4x+1)^2(4x+3)}}{\left((4x+3) - 4/((4x+3)\pi^2)\right)^2} \\ &> \frac{8 - 2\frac{16}{7\pi^2} - \frac{64}{175\pi^4}}{\left((4x+3) - 4/((4x+3)\pi^2)\right)^2} > 0. \end{aligned}$$

Hence, $f(x)$ is an increasing function for $x \geq 1$, so we have, $f(i_0 - 1) \geq f(i)$, $i = \overline{1, i_0 - 1}$, or,

$$\begin{aligned} \frac{4i_0 - 3 - 4/((4i_0 - 3)\pi^2)}{4i_0 - 1 - 4/((4i_0 - 1)\pi^2)} &\geq \frac{4i + 1 - 4/((4i + 1)\pi^2)}{4i + 3 - 4/((4i + 3)\pi^2)}, \quad i = \overline{1, i_0 - 1}. \end{aligned} \quad (\text{A.15})$$

From (A.13) and (A.15),

$$\frac{\beta - k(\beta - 1)/(n_f - 1)}{\beta - (k - 1)(\beta - 1)/(n_f - 1)} \geq \frac{4i + 1 - 4/((4i + 1)\pi^2)}{4i + 3 - 4/((4i + 3)\pi^2)},$$

where $i = \overline{1, i_0 - 1}$, $k = \overline{1, n_f - 1}$, which can be rewritten as,

$$\begin{aligned} \left(\frac{4i+3}{4} - \frac{1}{(4i+3)\pi^2}\right) \left(\beta - k \frac{\beta-1}{n_f-1}\right) &\geq \left(\frac{4i+1}{4} - \frac{1}{(4i+1)\pi^2}\right) \left(\beta - (k-1) \frac{\beta-1}{n_f-1}\right), \end{aligned} \quad (\text{A.16})$$

for $i = \overline{1, i_0 - 1}$, $k = \overline{1, n_f - 1}$.

From (A.16) and (A.5), we conclude,

$$\tilde{\delta}\tau_{i+1, k+1}^{\max} \geq \tilde{\delta}\tau_{i, k}^{\min}, \quad i = \overline{1, i_0 - 1}, k = \overline{1, n_f - 1}. \quad (\text{A.17})$$

Now, we will prove that $\tilde{\delta}\tau_{i,1}^{\max} \geq \tilde{\delta}\tau_{i,n_f}^{\min}$, $i = \overline{1, i_0}$. Similar to the previous analysis, it's trivial to show that,

$$\frac{5 - 4/(5\pi^2)}{3 - 4/(3\pi^2)} \geq \frac{4i + 1 - 4/((4i + 1)\pi^2)}{4i - 1 - 4/((4i - 1)\pi^2)}, \quad \forall i \in \mathbb{Z}^+$$

As $\beta \geq \frac{5 - 4/(5\pi^2)}{3 - 4/(3\pi^2)}$, therefore, for $\forall i \in \mathbb{Z}^+$,

$$\left(4i - 1 - \frac{4}{(4i - 1)\pi^2}\right) \beta \geq \left(4i + 1 - \frac{4}{(4i + 1)\pi^2}\right).$$

Then by multiplying with T_{bw} , we have,

$$\tilde{\delta}\tau_{i,1}^{\max} \geq \tilde{\delta}\tau_{i,n_f}^{\min}, \quad \forall i \in \mathbb{Z}^+. \quad (\text{A.18})$$

Lastly, we show that for any filter set $L_k(p)$, $k = \overline{1, n_f}$ defined as in Theorem 4.3 with $n_f \geq (1/M + \beta - 2)/(1/M - 1)$, $n_f \in \mathbb{N}$,

$$\forall \delta\tau_0 \neq 0, \exists L_q(p) : \xi(L_q(p), \delta\tau_0) < \delta\tau_0.$$

Case 2-A:

$$\tilde{\delta}\tau_{i_0-1,1}^{\min} < \delta\tau_0 \leq \tilde{\delta}\tau_{i_0,1}^{\max} \quad (\text{A.19})$$

From (A.6) and (A.19), it is obvious that $\xi(L_1(p), \delta\tau_0) = \tilde{\delta}\tau_{i_0-1,1}^{\min}$, and $\tilde{\delta}\tau_{i_0-1,1}^{\min} < \delta\tau_0$, hence $\xi(L_1(p), \delta\tau_0) < \delta\tau_0$. \square

Case 2-B:

$$\tilde{\delta}\tau_{i_0-1,1}^{\max} < \delta\tau_0 \leq \tilde{\delta}\tau_{i_0-1,1}^{\min}. \quad (\text{A.20})$$

From (A.5) we see that, $\forall i \in \mathbb{Z}^+$,

$$\tilde{\delta}\tau_{i,n_f}^{\min} < \tilde{\delta}\tau_{i,n_f-1}^{\min} < \dots < \tilde{\delta}\tau_{i,2}^{\min} < \tilde{\delta}\tau_{i,1}^{\min}, \quad (\text{A.21})$$

hence,

$$\tilde{\delta}\tau_{i_0-1,n_f}^{\min} < \tilde{\delta}\tau_{i_0-1,n_f-1}^{\min} < \dots < \tilde{\delta}\tau_{i_0-1,1}^{\min}. \quad (\text{A.22})$$

Now combining (A.18) and (A.20),

$$\begin{aligned} \tilde{\delta}\tau_{i_0-1,n_f}^{\min} &\leq \tilde{\delta}\tau_{i_0-1,1}^{\max}, \\ \tilde{\delta}\tau_{i_0-1,1}^{\max} &< \delta\tau_0 \leq \tilde{\delta}\tau_{i_0-1,1}^{\min}, \end{aligned}$$

we have,

$$\tilde{\delta}\tau_{i_0-1,n_f}^{\min} < \delta\tau_0 < \tilde{\delta}\tau_{i_0-1,1}^{\min}. \quad (\text{A.23})$$

Considering (A.22) and (A.23), there always exists a value $q \geq 2$ such that $\tilde{\delta}\tau_{i_0-1,q}^{\min} < \delta\tau_0 < \tilde{\delta}\tau_{i_0-1,q-1}^{\min}$. Following from (A.17), we have $\tilde{\delta}\tau_{i_0-1,q-1}^{\min} \leq \tilde{\delta}\tau_{i_0,q}^{\max}$. Therefore,

$$\tilde{\delta}\tau_{i_0-1,q}^{\min} < \delta\tau_0 < \tilde{\delta}\tau_{i_0,q}^{\max}. \quad (\text{A.24})$$

Combining (A.6) and (A.24),

$$\xi(L_q(p), \delta\tau_0) = \tilde{\delta}\tau_{i_0-1,q}^{\min}, \quad (\text{A.25})$$

therefore, $\xi(L_q(p), \delta\tau_0) < \delta\tau_0$. \square