ON-LINE INPUT IDENTIFICATION AND ACTIVE NOISE CANCELLATION: AN OVERVIEW OF RECENT RESULTS

Rapporto interno N. 3, gennaio 2010

Politecnico di Torino
Corso Duca degli Abruzzi, 24-10129 Torino-Italia
On-line input identification and Active Noise Cancellation: an overview of recent results

L. Pandolfi*
Politecnico di Torino, Dipartimento di Matematica,
Corso Duca degli Abruzzi 24, 10129 Torino — Italy,
luciano.pandolfi@polito.it
January 26, 2010

Abstract

We give an overview of a recent method we proposed for the solution of on-line deconvolution problems, which can be applied to Active Noise Cancellation (ANC). The method has been studied both for lumped parameter systems and for distributed systems in input-output and in state space form. Finally, we show that the method can be extended to identify inputs in case of non-fickian diffusion, which is modelled by a partial differential equation with infinite memory. This problem is of interest for example in biology, when a solute—a drug for example—has to be detected, but this same equation is encountered in different applications, for example thermodynamics and viscoelasticity. The analysis of this example will give a feeling of the method of proof used in this kind of study.

1 Introduction

Input identification is an ubiquitous problem in science since it amounts to identify the input to a system, on the basis of measures taken on the output. This problem is called “deconvolution” in the special case that the system is linear and shift invariant since in this case we have to solve for the unknown \( u \) the convolution equation

\[
y(t) = \int d\eta(s)u(t - s)
\]

In general, the integral is a multiple integral and \( \eta \) is a suitable measure. This problem is at the core of image reconstruction (in astronomy, biomedical

*This paper fits into the research program of the GNAMPA-INDAM.
applications etc.) and it has a different degree of difficulty, depending on the properties of the measure $\eta$.

We are here interested in causal systems, and $t$ represents time. In this case, the convolution equation takes the form

$$y(t) = \int_0^t d\eta(s)u(t - s)$$

(1)

If it happens that

$$d\eta(s) = \delta(s) + K(s), \quad K(s) \in L_{\text{loc}}^1(0, +\infty)$$

($\delta$ is Dirac’s delta) then the problem is well posed\(^1\) and easily solved: its solution is equivalent to the construction of the inverse system which turns out to be causal. Unfortunately, in most of the applications,

$$d\eta(s) = K(s) \in L_{\text{loc}}^1(0, +\infty)$$

and now the solution of the Volterra integral equation is an ill posed problem in the sense of Hadamard: the solution $u$ might not exist and, more important, even if it exists it does not depend continuously on the datum $y$ in any reasonable sense. This is an important point since available data are always corrupted by errors, and in general are sampled at a finite number of time instants.

The fact that the problem is ill posed corresponds to the fact that the inverse system cannot be causal or proper. For the solution of this ill posed problem, several “regularization” or “penalization” techniques have been proposed. These techniques can be “off-line” or “on-line”. Off-line techniques accumulate all the information during a certain interval of time $[0, T]$, which are then elaborated in order to find an approximant $^\wedge u(\cdot)$ of the unknown input $u(\cdot)$. These methods cannot be used for control applications. Instead, on-line deconvolution at a certain time $t$ produces a value $^\wedge u(t)$ of an estimate of $u(t)$, based solely on information obtained at previous times. Actually, it is possible that the signal constructed by the algorithm at time $t$ is an approximation of $u(t - \tau)$ where $\tau$ is a delay which conceivably can be made as small as we wish. Techniques of this class can be used for example in Active Noise Cancellation (ANC) paradigm, which is as follows: at time $t$ an estimate $^\wedge u(t)$ of the unknown noise $u(t)$ is fed to the system “in opposition of phase”, so to reduce the effect on the output of the unwanted “noise” $u(t)$. See a description of these ideas in [18].

Our goal is:

\(^1\)in the sense of Hadamard. This means that the problem is uniquely solvable and the solution depends continuously on the data when the problem is formulated in suitable normed spaces.
to describe an algorithm for online deconvolution in the simple framework of linear lumped systems described in state space form;

to describe an input-output version which can be applied also to classes of distributed systems;

Our final goal is to present a new result on input identification for a class of systems with memory which is encountered in several applications, for example in the study of diffusion when Fick law does not hold, as in the case of diffusion in polymers: we wish to detect the rate of immission of a solute (for example a drug) in a nonfickian fluid. The analysis of this specific problem will give a feeling of the techniques used in this approach to input identification.

2 Linear lumped systems in state space form

We consider a linear time invariant finite dimensional system described by

\[ \dot{x} = Ax + Bu, \quad y = Cx, \quad x(0) = 0 \]  

(i.e., we assume that the initial condition is known; hence \( x(0) = 0 \) without restriction). Here \( x \in \mathbb{R}^n \) and the matrices are constant. In order to sketch the ideas in the simplest form, we shall assume that \( u \) and \( y \) are scalar, but [10] consider a general multivariable system.

The idea to identify \( u \) is as follows: we construct a “model for the system”

\[ \dot{w} = Aw + Bv, \quad z = Cw, \quad w(0) = 0. \]  

In practice, system (3) will be a numerical simulation of (2). The idea is to compare \( y(t) \) and \( z(t) \) and to force \( z(t) \) to track \( y(t) \) as precisely as possible, using a suitable input \( v(t) = \hat{u}(t) \). Hopefully, under suitable assumptions, this input \( \hat{u}(t) \) will track \( u(t) \). For this, we fix a time delay \( \tau > 0 \) which, conceivably, can be as small as we wish. This is discussed below. We sample the output at times \( \tau_k = k\tau \). So, available data are

\[ \xi_k = y(\tau_k) + \theta_k, \quad k > 0 \]

where \( \theta_k \) is an unknown error in the measures. We assume that the tolerance of the error is known:

\[ |\theta_k| < h. \]

At time \( t = \tau_k \) we compute the estimate \( \hat{u}(t), \ t \in [\tau_k, \tau_{k+1}) \), on the basis of measures taken at previous times. We proceed recursively as follows:

step 0 to be performed at \( \tau_0 = 0 \): no previous measures. The best we can do is to choose \( \hat{u}(t) = 0 \), for \( 0 \leq t < \tau_1 \).
**step 1** at time $\tau_1$ we have one measure $\xi_1$. Using this measure we perform the algorithm describe below and we get $\hat{u}(t)$ for $t \in [\tau_1, \tau_2)$. This is a candidate estimate of $u_{[0, \tau_1]}$.

**step k** we have the previous observations $\xi_0 = 0$, $\xi_1$, $\ldots$, $\xi_k$. It is our decision to consider solely the last measure $\xi_k$. Using solely this piece of information $\xi_k$ we perform the algorithm describe below and we get $\hat{u}(t)$ for $t \in [\tau_k, \tau_{k+1})$. This is a candidate estimate of $u_{[\tau_{k-1}, \tau_k]}$.

The algorithm is as follows: we first fix a *penalization parameter* $\alpha > 0$ (conceivably small, $\alpha \in (0, 1)$).

At time $\tau_k$, $k \geq 1$, the function $\hat{u}(t)$ is known for $t < \tau_k$. So also $w(t)$ is known for $t < \tau_k$. We consider square integrable functions $v(t)$, $t \in [\tau_k, \tau_{k+1})$ and we define

$$\hat{u}_{|t \in [\tau_k, \tau_{k+1})} (t) = \arg \min_v \left\{ \|z(\tau_{k+1}) - \xi_k\| + \alpha \int_{\tau_k}^{\tau_{k+1}} v^2(s) \, ds \right\}.$$  

Note that

$$z(\tau_{k+1}) = C \left[ w(\tau_k) + \int_{\tau_k}^{\tau_{k+1}} e^{A(\tau_{k+1} - s)B}v(s) \, ds \right]$$

does depend on the still unspecified function $v(t)$.

It is easily computed that

$$\hat{u}_{|t \in [\tau_k, \tau_{k+1})} (t) = - \left[ \alpha I + \Lambda_{(k)}^* \Lambda_{(k)} \right]^{-1} \Lambda_{(k)}^* \left[ C e^{A\tau} w(\tau_k) - \xi_k \right]$$  \hspace{1cm} (4)

where

$$\Lambda_{(k)} v = \int_{\tau_k}^{\tau_{k+1}} C e^{A(\tau_{k+1} - s)B}v(s) \, ds.$$  

Note that while $x(\tau_k)$ is not known, instead $w(\tau_k)$ is known since it is the state of our “model system”. We denote $\hat{u}_{\alpha,h}(t)$ the function so constructed (this function depends on the noise $\theta_k$ and not directly on it’s tolerance $h$ but the notation above is standard).

**Remark 1**

- The number $\alpha$ is a “penalization parameter”. It’s effect is to reduce sensitivity to the errors $\theta_k$ and to the sampling errors.

- We decided to take into account only one last measure in the previous algorithm. We expect that the algorithm can be modified so to take into account several previous measures (i.e. we can perform more observations then updating of $\hat{u}(t)$). It is to be expected that this will have a positive effect on the robustness of the algorithm, but it will produce a less prompt reconstruction.
Now we state a result concerning “consistency of the algorithm”. I.e. we give conditions under which $\hat{u}(t)$ approximates the unknown input $u(t)$ (and so also $w(t)$ converges to $x(t)$). We state the result for the case of scalar system, $u \in \mathbb{R}$, $y \in \mathbb{R}$ and in the case of systems of relative degree 1. I.e. when $CB \neq 0$. The multivariable case and the case $\ker CB = 0$ are more delicate and are treated using Morse quasicanonical form, see [10]. We have (from [10, 12]):

**Theorem 2** Let the unknown real valued input $u(t)$ be locally square integrable.

Let us assume that the system is of relative degree 1.

Let $\tau$ be the step of the observation, $h$ the tolerance of the error and $\alpha$ the penalization parameter. Let $\tau \to 0$, $h \to 0$ and $\alpha \to 0$ while respecting the following conditions:

$$\frac{\sqrt{\tau}}{\alpha} \to 0, \quad \frac{h}{\alpha} \to 0;$$

Then,

$$\hat{u}_{\alpha,h} \to u \quad \text{in } L^2(0,T) \text{ for every } T > 0.$$

If $u$ is continuous on an interval $[a, b] \subseteq (0, T]$ then the convergence is uniform on $[a, b]$ and if $u$ is continuous on an interval $[0, b] \subseteq [0, T]$ and furthermore $u(0) = 0$, then the convergence is uniform on $[0, b]$.

Note the special role of $t = 0$, which is due to the fact that at the initial time $t = 0$ we don’t have previous measures, so that the best we can do is to impose $\hat{u}(0) = 0$.

If we have additional a priori information on the regularity of the unknown input $u$, then it is possible to give explicit convergence estimates, see [10, 12]. The estimates are important since the result above is a “consistency results”: it gives no hint on the way the penalization parameter $\alpha$ has to be chosen. Let us now look at the the following explicit estimate, from [12]:

**Theorem 3** Let us assume that the unknown input $u$ is of bounded variation and that a bound $M$ for its variation is known:

$$u(t) = \int_0^t d\mu(s), \quad V(u) = \int_0^T d|\mu|(s) < M < +\infty.$$

Then we have

$$\|v - u\|_{L^2(0,T)} \leq C \left\{ \frac{h + \sqrt{\tau}}{\alpha} + \alpha V(u) + \sqrt{\alpha} \right\}. \quad (5)$$

Here $C$ is a suitable constant, which can be explicitly computed.
The previous estimate shows that in order to have a small reconstruction error, the penalization parameter $\alpha$ has to be “large” as compared with the step $\tau$ and the error tolerance $h$, which are often the parameters which cannot be changed.

**Remark 4** The previous estimate requires an a priori information on the tolerance of the errors in the measures. This can be realistic for some applications but not in every case. Moreover, when using estimates as (5) in order to choose $\alpha$, this is a worst case choice; but, it is conceivable that the worst possible level of noise is never attained. So, the choice of $\alpha$ suggested by (5) is conservative and the following query makes sense: is it possible to update the values of the penalization parameter $\alpha$ at every step of the algorithm, using information on the measures $\xi$?

We shall discuss this point in Section 3.3.

Let us discuss the previous procedure by looking at simple simulations. These simulations concern the problem of numerical differentiation, which is the simplest version of an input identification problem.

### 2.1 Numerical differentiation and simple simulations

**Numerical differentiation** is the following problem. We know samples of a function $y(t)$ and we want to compute numerically its derivative $y'(t)$. Note that the values $y(t)$ are not known. Only the values $y(\tau_k)$ (corrupted by noise) are known.

Numerical differentiation is a special instance of the problem we have seen, since it amounts to reconstruct the input $u$ of the system

\[ \dot{x} = u, \quad y = x, \quad x(0) = 0. \]  

(6)

A well known fact is that numerical differentiation is an ill-posed problem: i.e., even if solvable, the solution of the problem does not depends continuously on the data.

Numerical computation of the derivative is possibly the simplest example of an ill posed problem and its solution depends on suitable regularization algorithms. The next example shows that the choice of the algorithm is more delicate then we might expect.

**Example 5** We consider system (6). There is a unique input $u$ which produces the given output and the goal is the determination of $u$ on the basis of measures taken on $y(t)$ at discrete times. Likely, the algorithm that first comes to the mind is as follows: we consider a “model system”

\[ \dot{w} = v \]
on every time interval $[\tau_k, \tau_{k+1}]$ and, at the time $\tau_k$ we choose

$$w(\tau_k) = \xi_k.$$  

On the interval $[\tau_k, \tau_{k+1})$ we take the input $v$ with minimal norm among those inputs driving the output of the model from $\xi_k$ to $\xi_{k+1}$.

We apply this algorithm and we see that we shall find meaningless results. Let the he input, unknown to us, be $u(t) \equiv 0$. The observation interval is $[0, 1]$ and let us consider a sequence of measures and reconstructions. When the number of steps is $n$ then we consider the case that the error tolerance is $h = 1/n$.

We observe at times $\tau_k = k/n$ and we get the observations $\xi_k = x(\tau_k) + \theta_k$ where $\theta_k$ is a disturbance unknown to us, but such that $||\theta_k|| \leq 1/n$.

The “control” $v$ of minimal norm which transfers $\xi_k$ to $\xi_{k+1}$ is easily computed:

$$v_{n|[\tau_k, \tau_{k+1})}(t) = n[\theta_{k+1} - \theta_k].$$

Now we compute $v_n$ explicitly in the following cases: the case that $\theta_k = 1/n$ for each $k$ or $\theta_k = (-1)^k/n$. In the first case we have

$$v_n(t) \equiv 0.$$  

This is exactly the sought for input $u$ so that it seems that the proposed identification procedure is very efficient. However, in the second case we obtain

$$v_n(t) = \begin{cases} 
-2 & \text{on } [\tau_{2k}, \tau_{2k+1}) \\
+2 & \text{on } [\tau_{2k+1}, \tau_{2k+2}) 
\end{cases}$$

and this sequence of functions does not converge neither pointwise nor in $L^2$-norm to $u(t) \equiv 0$. In fact it converges to $u(t) \equiv 0$, but only weakly in $L^2(0, 1)$.

Instead, reconstruction of the derivative can be achieved using the algorithm in the previous section, thanks to Theorems 2 and 3. This is illustrated in the following simulations. The input function in Fig. 1 is

$$u(t) = \begin{cases} 
1 & \text{if } 2 \leq t \leq 3 \\
0 & \text{if } t < 2, \ 3 < t < 10 
\end{cases}$$

while Fig. 2 presents the case of periodic input functions (the functions sin $t$ and cos $t$).

The rationale we can deduce from these and similar simulations:

- the penalization parameter $\alpha$ is used to transform an ill-posed problem to a well posed one: the $\alpha$-problem is always solvable and the solution
The numerical computation of the derivative is quite easy. The case
\[ \dot{x} = x_1, \quad \dot{x}_1 = u, \quad y = x \]
is much more delicate and it is difficult to improve the reconstruction of the signal beyond a certain limit. This is related to the fact that the system is not minimum phase.

2.2 Notes on the references

The idea we used in the reconstruction algorithm has to be traced back to the technique introduced in control theory by Krasovskii school, and described

---

2 in accordance with the fact that the problem with \( \alpha = 0 \) is not well posed.
Figure 2: left: $u(t) = \sin t$, right $u(t) = \cos t$, $N = 1000$, $h = .1$, $\alpha = .5$

$N =$steps , $h =$error tolerance, $\alpha =$penalization

in the book [31].

Possibly, the very first papers in which the method was explicitly applied to input identification problems are [?, 33, 39, 46]. This idea has then been used in several papers and analyzed in the books [47, 38]. The book [47] deals with finite dimensional system while [38] considers distributed systems. Both these books assume from the outset that the observation is $y = x$, the full state of the system, then adapted to special cases in which only part of the components of the state is observed. The first paper in which a general state-space (finite dimensional) system has been studied is [10]. The algorithm presented in Section 2 is taken from this paper, and even when applied to the special case $y = x$, it is not the same as the algorithm used in the Russian literature, although it is inspired by that one.

3 Active Noise Cancellation

The algorithm described above can be used for noise cancelation. Most of the algorithms used for this introduce high gain feedbacks and this is so also for the algorithm to be presented here. These algorithms often change the input-output response of the system even if the noise happens to be absent. Instead, we construct an algorithm of noise cancellation which is active only in the presence of noise.

Let us illustrate the method in the following case:

$$\dot{x} = Ax + D(v - u) + Df_0(t), \quad y = Cx \quad (7)$$
where $f_0(t)$ is the \textit{known input} which regulates the system, $v$ is an unwanted noise, and $u$ is the control that has to be constructed so to reduce this noise.

The “model system” is now
\begin{equation}
\dot{x}' = Ax + D(\omega - u^{(M)}) + Df_0, \quad \dot{y}(t) = C\dot{x}(t)
\end{equation}
where $\omega$ is an estimate of the unknown noise $v$ and $u^{(M)}$ is the control applied to the “model system” (it will be $u^{(M)}(t) = u(t - \tau)$). Let us introduce the following notation:

- $u^{(M)}$ (as above) for the control which acts on the model and $u^{(S)} = u$ for the control which acts on the system.
- an index $(k)$, as in $z_{(k)}$, denotes a function defined on the interval $[\tau_k, \tau_{k+1})$.

The core of the compensation method is a procedure which, performed at each time step, constructs a function $\hat{v}_{(k-1)}$, defined on $[\tau_{k-1}, \tau_k)$, an estimate of $v$ on this interval. This procedure is an adaptation of the algorithm described in Section 2 to the case of system (7) and its model (8). In the interval $[\tau_k, \tau_{k+1})$, we shall use the functions $\hat{v}_{(r)}$ $(r < k)$ in order to construct: 1) $\omega_{(k)}$, which should act on the model system so to mimic the disturbance $v$; 2) the controls $u^{(S)}_{(k)}$ and $u^{(M)}_{(k)}$ which should reduce the action of the disturbance from the outputs, respectively, of the system and of its model.

The functions $\omega$, $u^{(S)}$ and $u^{(M)}$ are defined by
\begin{equation}
\begin{cases}
\omega_{(k)}(t) = \hat{v}_{(k-1)}(t - \tau) \\
u^{(S)}_{(k)}(t) = \hat{v}_{(k-1)}(t - \tau) \\
u^{(M)}_{(k)}(t) = \hat{v}_{(k-2)}(t - 2\tau) .
\end{cases}
\end{equation}

Hence, we have
\begin{equation}
u^{(M)}(t) = u^{(S)}(t - \tau) = \omega(t - \tau).
\end{equation}

The interconnections of the system and of its model are described by the block diagram, where $1/z$ denotes the delay of one step.

This algorithm amounts to the construction of a feedback loop which is inactive, and does not change the dynamic of the system, in the ideal case that disturbances are not present. See [13] for details. This paper treats also the more realistic case that an upper bound for the duration (in time) of the process is not a priori given.
3.1 A simulation

The problem we consider here is described in [30], see also [44].

The system is a robot harm, which has to move a certain load. The path to be tracked is imposed by an external operator, using the exterior signal $F_0$. To achieve path tracking, a nominal value $M_0$ of the load to be transported is identified first and a "controller" is constructed, so to track the prescribed path when transporting the nominal load. The system so constructed tracks the prescribed path if there is no disturbance, and asymptotically tracks the path under disturbances which fades away. The resulting system is described by the following block diagram, taken from [30].

The frequency domain description of the system is

$$X(\lambda) - F_0(\lambda) = \frac{1}{M_0} \frac{1}{\lambda^2 + K_1 \lambda + K_2} D(\lambda)$$

using a common notation, the reconstructed signal is denoted $\hat{u}$. So, we cannot use $^\sim$ to denote Laplace transform. Here, Laplace transform is denoted by an upper case letter. Later on, we need a matrix notation for which capital letters have to be used. So, we shall change the notations as specified below.
where $D$ is (the Laplace transform of) a disturbance and $M_0$, $K_1$ and $K_2$ are constant. The values of the constants $K_1$ and $K_2$ are so chosen to ensure asymptotic stability and tracking the path described by $F_0$.

The controller has been designed using the value of the nominal load to be transported, so that it can’t insure satisfactory path tracking if the load does not have the nominal value. As stated in [30] “in the case of robot manipulators, it is almost impossible to compute the disturbances one by one. Moreover, parameter variations such as mass variation cannot be known. Therefore, it is smart to compute the disturbance from the input/output relationship of the motor”. This estimate is then used to force the harm to track the prescribed path, even under variations of the parameters.

In the approach we have described we follow this same circle of ideas. In order to see a simulation, we use the output

$$Y(\lambda) = \lambda X(\lambda)$$

(so that the resulting system has relative degree one). When the disturbances, in particular the mass variation, are compensated using the deconvolution filter in Section 3, we get the following block diagram:

Now we see the following simulation, taken from [13]. We plot the desired track $F_0(t)$ (dotted: the extreme case that the desired path is a sinusoid) and the errors committed with and without the application of the deconvolution filter, when the load to be transported is different from the nominal one (it weights three times more). For clarity, we consider the case that there is no external disturbance: $D = 0$. The simulation gives the plot in Fig. 3 (the values of the parameters are specified in the caption of the figure). This figure plots the desired path and the errors committed with and without the deconvolution filter and shows precise tracking, after a short transient, when the deconvolution filter is used.
In principle, when the harm transports the nominal weight then the compensator we described in Section 3 does not produce any signal and does not affect the free dynamic of the system. In fact, a “small” signal is produced anyhow since, due to the measures being taken at discrete times, the observer detects a discrepancy between the observation on the system and on its model.

3.2 Limitation of the method

The performance of the algorithm is unacceptable in the presence of periodic inputs of high frequencies (applications to robotics often are not affected by such nasty signals).

This is the first limitation of the algorithm, which is common to every ANC algorithm, see [34, 62]. A second and more important limitation, already noted, is that the penalization parameter $\alpha$ has to be tuned on the basis of preliminary analysis of the expected noise. So, we need this preliminary piece of information on the nature of the disturbance affecting the system. And then, $\alpha$ is chosen on the basis of the worst noise, which might never affect the system. We noted that a too small value of $\alpha$ amplifies the effect of the errors in the measures. So, we now present an algorithm for adaptive determination of $\alpha$. 

Figure 3: $\tau = 0.01$, $\alpha = 1/85$, $h = 0.1$. 

$F_0$ and $X$ without (left) and with compensator (right).
3.3 Adaptive determination of $\alpha$

The point with the value of $\alpha$ is that if $\alpha$ is too large then reconstruction is slow, and of poor quality; but, if it is too small, the estimate of the noise may present wild oscillations and this is the phenomenon we would like to avoid. So, we wonder whether it is possible to adaptively change the value of $\alpha$ from step to step, on the basis of the available measures. This is in fact possible for the case of systems of relative degree 1. This limitation is discussed below.

The idea of the algorithm is as follows (we describe the algorithm in the case $y = x$. See [51] for the general case.):

We fix the sampling time $\tau$ and a coefficient $\mu \geq 1$. Furthermore we fix two numbers $\epsilon$ and $\gamma$, with $0 < \epsilon < \gamma < 1$. We use the algorithm described above for the construction of $\hat{v}$ at each time step, but now we consider the value of $\alpha$ as a free parameter at each step. So, at each step we compute a certain function $v_\alpha$ defined on $[\tau_k, \tau_{k+1})$, which is a candidate approximant of the disturbance $v$ on the interval $[\tau_k, \tau_{k+1})$. Using this function $v_\alpha$ we can compute $w(\tau_{k+1})$ which is compared with $\xi_{k+1}$ in order to chose the value of $\alpha$. Two cases are possible:

- we have $\|w(\tau_{k+1}) - \xi_{k+1}\| \leq \mu h^{\gamma}$. For every $\alpha$.

  In this case we choose
  \[ \alpha = \alpha_k = \mu h^{\gamma - \epsilon}. \]

- otherwise, it is possible to prove the existence of a unique value of $\alpha$ such that
  \[ \|w(\tau_{k+1}) - \xi_{k+1}\| = \mu h^{\gamma}. \]  \hspace{1cm} (11)

This unique value $\alpha = \alpha_k$ we choose on this interval, and the estimate $\hat{v}(k)$ of the disturbance $v(t)$ on $[\tau_k, \tau_{k+1})$ is $v_\alpha = v_{\alpha_k}$.

Note that this process is easily performed since an explicit form of $\hat{v}$ as a function of $\alpha$ is known, see (4).

The meaning of $\mu$ and the rationale of the choice: if $\mu = 1$ then we choose $\alpha$ according to the pure discrepancy principle: that value of $\alpha$ which makes the discrepancy $\|w(\tau_{k+1}) - \xi_{k+1}\|$ equal to the measurement error: to reduce this difference less then the measurement error is meaningless. Now, the value of $\mu \geq 1$ is introduced since also exact equality may be too stringent.

The effect of this algorithm is to prevent $\alpha$ from becoming too small so that the system possibly will not give such good attenuation of the noise, but surely will not undergo wild oscillations.

This algorithm is described in [51] where the case $y \neq x$ and additional discussions are presented (also concerning the more realistic case that an
upper bound for the time is not given). We take the following simulation from this paper. The problem has been described in [64] and correspond to a problem of interest in flight control. A certain function (a control) has to be constant but, due to the failure of a component, its values suddenly changes, as described by the graph in Fig. 4 (topo left). The desired output of the system is the top line in Fig. 4, top right while the line going down is the output without compensation. The figures on the second line describe the desired and real outputs when the compensator is active (left) and the values of $\alpha$ automatically chosen by the algorithm at each iteration (right).

Finally we comment on the reason why this method can be applied only to systems of relative degree 1. The simplest example of a system of relative degree 2 is the system

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = u, \quad y = x_1.$$  \hfill (12)

The proposed algorithm identifies $x_2$ as an “artificial input” to the system $\dot{x}_1 = x_2$. Of course, $x_2$ is estimated with a certain error.
Once $x_2$ is known, this information can be used to estimate the input $u$. But, in order to identify $u$ with an acceptable tolerance then we need a “very precise” estimate of $x_2$ while the quality of the reconstruction of $x_2$ is limited by condition (11).

We conjecture that an algorithm which might be used so to have an adaptive algorithm for ANC in systems with higher relative degree, is the algorithm described in [14].

4 Distributed systems

The algorithm for input identification and deconvolution has been extended to a class of systems with memory. Actually, the adaptive algorithm has not been extended yet, and the algorithm has not been tested for ANC in these cases.

We describe briefly the classes of systems that have been treated up to now, and then we shall see few simulations.

We consider the problem

$$y(t) = \int_0^t K(t, s) u(s) \, ds \quad (13)$$

where now $K(t, s)$ is a square matrix.

The idea for the reconstruction is as follows: in the noiseless case the output $y(t)$ being given by (13). We try to impose the equality

$$y(t) = \int_0^t K(t, s) u(s) \, ds = \int_0^t K(t, s) v(s) \, ds$$

where $v$ is unknown. This being an ill posed problem, we introduce a positive “regularization parameter” $\alpha$ and we try to solve instead

$$\alpha v(t) + \int_0^t K(t, s) v(s) \, ds = y(t) = \int_0^t K(t, s) u(s) \, ds \quad (14)$$

Hopefully, when $\alpha \to 0$, under suitable assumptions, the solution $v_\alpha$ to this problem will converge to the unknown function $u$.

This method is related to the so-called Lavr'entev regularization method, see [56] and references therein.

Of course, we have to take into account noises in the observation and the fact that observations are often taken at discrete times in order to transform this idea into a practical algorithm. This we are going to show.

\footnote{From now on, there will be no need to distinguish the control $u$ from the noise. Moreover, upper case letter will be used to denote kernels of Volterra integral operators, possibly matrix valued, so that Laplace transform is denoted with $\hat{\cdot}$. Consequently, this $\hat{\cdot}$ notation can’t be used for the estimate of $u$, which will be denoted $v$.}
We first consider the deconvolution problem, $K(t, s) = K(t - s)$ and $K(t)$ possibly with a weak singularity at $t = 0$. For example, $K(t)$ may be an Abel kernel, $K(t) = t^{-\gamma}, \gamma < 1$.

Convolution equations with Abel kernels are encountered in different application, see [20, 21, 22, 43] and they gained popularity also in control theory, see [8, 40, 42, 48, 58, 60] and references therein.

4.1 Systems with weakly singular kernels

Let the kernel in (13) be scalar, $K(t, s) = K(t - s)$, and satisfy the following conditions, see [11]:

- the inequalities

$$0 < \frac{M_1}{|\lambda|^{\gamma_1}} \leq |\hat{K}(\lambda)| \leq \frac{M_2}{|\lambda|^{\gamma_2}}, \quad \gamma_1 > 0, \quad \gamma_2 > 0;$$

- the following Sector Condition:\footnote{which extends real positive kernels.} There is a set

$$S_{r, \theta} = \{\lambda \in C, |\lambda| < r, |\arg \lambda| > \theta\}, \quad 0 < \theta < \pi$$

and a positive number $\nu_S$ (larger than the abscissa of convergence of the Laplace transform $\hat{K}(\lambda)$) such that

$$\Re \lambda > \nu_S \implies \hat{K}(\lambda) \notin S_{r, \theta}.$$

The set $S_{r, \theta}$ is the part of a sector centered on the negative real axis, outside a disk centered at 0. So, we call this the “sector condition”. This condition is satisfied in particular by the Abel kernels

$$K(t) = \frac{1}{\Gamma(\gamma)}, \quad \gamma < 1$$

since in this case (with $\Gamma(z)$ the Euler Gamma function)

$$\hat{K}(\lambda) = \Gamma(1 - \gamma) \frac{1}{\lambda^{1-\gamma}}, \quad \gamma < 1.$$ 

Under these conditions, it has been proved that the function $\hat{v}(t)$ given by $\hat{v}(t) = v_k$ for $t \in [k\tau, (k + 1)\tau)$ and

$$v_k = \frac{[y(\tau_k) + \xi_k] - w(\tau_k)}{\alpha}$$ (15)
$(\xi_k)$ is the noise in the measures and $\alpha$ is the regularization parameter) approximates the unknown input $u(t)$. Here $w(\tau_k)$ is given by

$$w(\tau_{k+1}) = w(\tau_k) + \int_{\tau_k}^{\tau_{k+1}} K(\tau_{k+1} - s)v(s)\,ds$$

$$+ \int_{0}^{\tau_k} F(\tau_k - s)v(s)\,ds, \quad F(t) = K(t + \tau) - K(t) \quad (16)$$

$(\tau)$ is the step in the observation and $\tau_k = k\tau$.) So, the reconstruction algorithm at every step uses only data measured at previous times.

The main result in [11] is:

**Theorem 6** Let $u \in L^2_{\text{loc}}(0, +\infty)$ and let $T > 0$ be fixed. For every $\epsilon > 0$ there exist $\tau_\epsilon$, $h_\epsilon$ and $\alpha_\epsilon$ such that if $\hat{v}(t)$ is given by (15) with $\alpha = \alpha_\epsilon$ then for all $h \in (0, h_\epsilon)$, $\tau \in (0, \tau_\epsilon)$ we have

$$\|u - \hat{v}\|_{L^2(0, T)} < \epsilon.$$ 

Furthermore, let the input $u$ be differentiable with square integrable derivative. Under the stated conditions, for every $T > 0$ there exist positive number $M$, $\delta_1$ and $\delta_2$ such that the reconstruction error satisfies

$$\|u - v\|_{L^2(0, T)} \leq M \left[ \alpha^{1/(1+\gamma_2)} + \frac{\sqrt{\tau}}{\alpha} + \frac{h}{\alpha} \right].$$

### 4.2 Numerical evaluation of Caputo fractional derivatives

As we noted, the prototype of the deconvolution problem is the problem of numerical computation of the derivative, which amounts to the identification of the “input” $u(t)$ from (6), i.e.

$$y(t) = \int_{0}^{t} u(s) \, ds.$$ 

Fractional order derivatives are important in several applications and very popular in physics now, see [20, 21] and moreover are the prototype of the identification problems in this section. In fact, it turns out that the algorithm referred to in Section 4.1 can be applied to the numerical computation of Caputo fractional derivative of order up to 2. It has to be suitably modified for the evaluation of Caputo fractional derivatives of higher order, as we see now.

By definition, the derivative $D^\gamma_t$, fractional derivative of order $\gamma \in (0, 1)$, in the sense of Caputo, is the operator

$$f(\cdot) \longrightarrow \frac{1}{\Gamma(1-\gamma)} \int_{0}^{t} (t - s)^{-\gamma} f'(s) \, ds = (D^\gamma_t f)(t) = (J^{1-\gamma}Df)(t)$$
where \( J^{1-\gamma} \) denotes Riemann-Liouville fractional integral (\( \Gamma \) is the Euler Gamma function).

Fractional derivatives of order \( n + \gamma \), \( n \geq 1 \) and \( \gamma \in (0, 1) \) are defined as

\[
(D^{n+\gamma}_tf)(t) = \frac{1}{\Gamma(1-\gamma)} \int_0^t \frac{1}{(t-s)^\gamma} (D^{n+1}_sf)(s) \, ds .
\]

So, the computation of the fractional derivative can be interpreted as the following inverse problem. Let \( u \) be the function to be computed, i.e.

\[
u = D^{n+\gamma}_tf = J^{1-\gamma}f^{(n+1)}.
\]

We first apply \( J^{\gamma} \) to both sides. Using the semigroup property of fractional integration,

\[
J^{\gamma}J^{\sigma} = J^{(\gamma+\sigma)},
\]

we get

\[
J^{\gamma}u = (Jf^{(n+1)})(t) = f^{(n)}(t) - f^{(n)}(0).
\]

If \( n > 0 \), we apply \( J^n \) to both the sides and we get the following equation for \( u \):

\[
(J^{n+\gamma}u)(t) = f(t) - \sum_{j=0}^{n} f^{(j)}(0) \frac{t^j}{j!} = g(t) \quad (17)
\]

where

\[
(J^{n+\gamma}u)(t) = (K \ast u)(t), \quad K(t) = \frac{1}{\Gamma(n+\gamma)} t^{n+\gamma-1} \quad (18)
\]

(we use \( \ast \) to denote convolution.)

Note that \( u \) is the unknown function to be approximated and that equality (17) holds on the interval \( 0 \leq t \leq T \) where the measures are taken. If we consider \( T' > T \) and then we extend the unknown function \( u \) to \( t \in (T, T') \) then we get an extension of the function \( g(t) \). We stress that this does not affect the values of \( g(t) \) for \( t \leq T \); i.e., the transformation from \( u \) to \( g \) is causal.

As in [26, 43], we assume that the numbers \( f^{(j)}(0) \) are known for physical reasons, for example because they have to be zero. So, we assume that the function \( g(t) \) in (17) is (approximately) known and we are reduced to solve a Volterra integral equation of the first kind with Abel kernel in the unknown \( u \).

The Laplace transform of \( K(t) \) is

\[
\hat{K}(\lambda) = \frac{1}{\lambda^{n+\gamma}} .
\]

Sector condition holds if and only if \( n + \gamma < 2 \).
Let us consider instead the case \( n + \gamma \geq 2 \). In this case the algorithm has to be modified in an interesting way. We consider again Eq. (14), which in this case has to be replaced with

\[
\alpha v + (J^{n+\gamma}v)(t) = g = (J^{n+\gamma}u)(t).
\]  

Let \( v_\alpha \) be its solution. It is easily seen that \( v_\alpha \) does not converge to \( u \) when \( n + \gamma \geq 2 \). So, the idea now is to adapt the order of the penalization parameter to the order of the derivative, but of course this can’t be obtained by simply replacing \( \alpha \) with \( \alpha^\epsilon \), for a suitable exponent \( \epsilon \). Instead, we note that there exists a natural number \( k \) such that

\[
\frac{n + \gamma}{k} \in (0, 2)
\]

so that

\[
K_{1/k}(t) = \frac{t^{(n+\gamma-1)}}{\Gamma((n+\gamma)/k)}, \quad \hat{K}_{1/k}(\lambda) = \frac{1}{\lambda^{(n+\gamma)/k}}
\]

now satisfies the sector condition.

It is known that

\[
\left[ J^{(n+\gamma)/k} \right]^j \phi = (K_{1/k}^* \phi)_j = J^{(n+\gamma)/k} \phi.
\]

In particular,

\[
\left[ J^{(n+\gamma)/k} \right]^k \phi = (K_{1/k}^* \phi) = J^{(n+\gamma)} \phi.
\]

This equality suggests that we replace Eq. (19) with the problem:

\[
(\alpha \delta + J^{(n+\gamma)/k})^k \ast v = (\alpha \delta + K_{1/k})^* \ast v = g
\]  

where \( \delta \) is the Dirac delta function; i.e., we penalize inside the \( k \)-th power. \( L^2 \)-convergence of \( v = v_\alpha \), the solution of (21), to the unknown function \( u \) is proved in [52] (where also the effect of disturbances and of discretization is studied.)

**Remark 7** Problem (21) is a singular perturbation problem, of an order which is adapted to the singularity of the kernel. This idea of using singular perturbations of higher order in the context of Lavrent’ev method has been exploited in [61, 29, 14]. In particular, the paper [14] considers the multivariable case, i.e. the case that \( u \) and \( y \) are vector of the same dimensions, and the matrix kernel \( K(t) \) does not satisfy the condition that \( K(0) \) is invertible. Instead, \( K(t) \) has a zero for \( t = 0 \), with a certain “structure” (see [49] for the definitions.) We conjecture that this result can be used in order to remove the condition of “relative degree 1” in ANC problems.

\[\text{the exponent } ^* \text{ denotes } k\text{-th order convolution, see [24].}\]
4.2.1 Few simulations

In this section we examine few simulations on the numerical evaluation of Caputo fractional derivatives, taken from [52]. In every practical application the signal \( f(t) \) will be bounded, and corrupted by an error whose tolerance \( h \) denotes relative error (alternatively, the tolerance may denote the variance of a random noise with zero mean, but we assumed that the error has uniform distribution on \([-h, h]\)). In order to present few simulations we work with functions whose fractional derivatives can be easily computed in closed form. The simplest examples correspond to unbounded functions \( f(t) \) and this case is of course more demanding and forces us to consider \( h \) as an absolute tolerance, otherwise the error will increase without bounds.

The derivative we compute is

\[
D^{2+1/4}\alpha f.
\]

The tolerance \( h \) must be tight since we are computing derivatives of order higher then 2.

Figure 5 presents the exact graphs of \( D^{2+1/4}\alpha f \) as computed from its exact expression in closed form and, superimposed, the graph computed with the algorithm in [52] using the values of the parameters specified in the captions of the figures. These are the time interval, \([0, T]\) (\(T\) is specified), the number \( N \) of the steps, i.e. \( \tau = [T/N] \), the error tolerance \( h \) and the penalization parameter \( \alpha \).

The graph in Fig. 5, left, is that of \( D^{2+1/4}\alpha f \) when the function \( f(t) \) is

\[
f(t) = \begin{cases} 
\frac{t^3}{6} & \text{if } 0 < t < 1 \\
\frac{t^2}{2} - \frac{t}{2} + \frac{1}{6} & \text{if } t \geq 1.
\end{cases}
\]

We have \( f^{(3)}(t) = 1 \) on \((0, 1)\), and equal to 0 for \( t > 0 \) so that

\[
f^{(0/4)}(t) = \begin{cases} 
\frac{4}{3\Gamma(3/4)} t^{3/4} & \text{if } 0 < t < 1 \\
\frac{3\Gamma(3/4)}{4} \left\{ t^{3/4} - (t - 1)^{3/4} \right\} & \text{if } t \geq 1.
\end{cases}
\]

Fig. 5, right, shows the case that \( D^{0/4}\alpha f = \text{const} \), which is obtained when \( f(t) = 12t^{3/4}/45 \) so to have \( (D^{0/4}\alpha f)(t) = \Gamma(1/4) \).

The last graphs, presented in Fig. 6, concern the function \( f(t) = \sin t \), \( t \in (0, 50) \). We don’t use the exact expression for the fractional derivative given in [45] in terms of hypergeometric functions now. Instead the fractional integral of \( f^{(3)}(t) = -\sin t \) is computed numerically and compared with the estimate given by the algorithm in Section 4.1. Thanks to the fact that the function \( f(t) \) is now bounded, we compare the graph obtained when \( h = 0.01 \) and \( h = 0.05 \).
4.3 A class of nonlinear problems

The algorithms we presented has been extended to a class of nonlinear systems in [56]. We describe the results. We consider the nonlinear input-output relation

\[ y(t) = \int_0^t K(t,s)u(s)ds + \int_0^t F(t,s,u(s))ds, \quad (22) \]

Here \( u \) and \( y \) are vectors of the same dimension, say \( y, u \) belong to \( \mathbb{R}^n \).

Let \( \Delta = \{(t,s) : 0 \leq s \leq t \leq T\} \). We assume that

\[ K(t,s) : \Delta \rightarrow \mathbb{R}^n, \quad F(t,s,u) : \Delta \times \mathbb{R}^n \rightarrow \mathbb{R}^n \]

and:

- on the kernel \( K(t,s) \):
  - it is continuous for \( 0 \leq s \leq t \leq T \);
  - it satisfies
    \[ K(t,t) = I, \quad t \in [0,T]. \quad (23) \]
    - the derivative \( K_t(t,s) \) exists a.e. and \( \sup_{t \in [0,T]} \int_0^t \|K_t(t,s)\|^2ds \leq B \).

- on the nonlinear term \( F(t,s,u) \):
  - it is continuous and the partial derivative \( F_t(t,s,u) \) exists for a.e. \( (t,s) \in \Delta \) and for all \( u \in \mathbb{R}^n \).
Figure 6: Parameters: $T = 50$, $N = 500$, $a = 0.3$ and $h = 0.01$ (left) $h = 0.05$ (right).

\[ \text{for each } u, v \in \mathbb{R}^n \text{ and a.e. } (t, s) \in \Delta \text{ we have} \]
\[ \| F(t, s, v) - F(t, s, u) \| \leq N(t, s) \| v - u \|, \]
\[ \| F_i(t, s, v) - F_i(t, s, u) \| \leq N(t, s) \| v - u \| \]
\[ \sup_{t \in [0, T]} \int_0^t N^2(t, s) ds \leq N. \]

- for every $t \in [0, T]$, the function
\[ u \rightarrow F(t, t, u) : \mathbb{R}^n \rightarrow \mathbb{R}^n \]

is monotone in the sense of operator theory (see [6]), i.e.
\[ \langle x - y, f(x) - f(y) \rangle \geq 0, \quad \forall x, y \quad (24) \]

(the crochet denotes inner product).

The crucial conditions are condition (23) and the monotonicity condition (24). We study an algorithm for the on-line determination of the unknown input $u(t)$, under the assumption that $u$ is piecewise $W^{1,2}(0, T)$; i.e. there exist finitely many points $t_1 < \cdots < t_n$ in $(0, T)$, such that the restriction of $u$ to each interval $(t_i, t_{i+1})$ belongs to $W^{1,2}(t_i, t_{i+1})$.

Let the output $y(t)$ given in (22) be read with an error error $\xi(t)$ of known tolerance $h$:
\[ \| \xi(\cdot) \|_{L^2(0, T)} < h. \]

Then we define a candidate approximant $v_\alpha$ of $u$ as the solution of the following Volterra integral equation:
\[
\alpha v(t) + \int_0^t [K(t, s)v(s) + F(t, s, v(s))]ds = z(t)
\]

where
\[
z(t) = \frac{1}{\alpha} \int_0^t e^{-\frac{1}{\alpha}(t-s)}[y(s) + \xi(s)]ds.
\]

Then we have the consistency result

**Theorem 8** Let \( \alpha = \alpha(h) \) be a function such that
\[
\lim_{h \to 0^+} \alpha(h) = 0, \quad \lim_{h \to 0^+} \frac{h}{\alpha(h)} = 0
\]

and let \( v_{\alpha(h)}(t) \) solve (25). Then we have:
\[
v_{\alpha(h)}(t) \longrightarrow u(t)
\]
in \( L^2(0, T; \mathbb{R}^n) \).

Results on error estimates and uniform convergence on intervals where \( u(t) \) is smooth are similar to those in Theorem 6 and are not reported here. See [35] for a different approach.

## 5 Non Fickian diffusion equation

We now extend the identification algorithm to a class of distributed systems with delays, so to have a feeling of the techniques used to justify the results presented up to now. This same class of systems appears in many different applications, which are briefly summarized below. We concentrate on one of these application, which models non fickian diffusion. For example, this is the case of diffusion in polymers.

### 5.1 Derivation of the model

Let a solute diffuse in a solvent. For simplicity we assume that the solvent stays in a one dimensional channel, say on the interval \([0, \pi]\). Let \( \theta(x, t) \) be the concentration of the solute at time \( t \) and position \( x \) and let \( J(x, t) \) be the flux at time \( t \) through the unit area at position \( x \).

The diffusion equation is obtained as a combination of:

- Mass conservation:
in a given interval \([a, b] \subseteq [0, \pi]\), the variation of the mass in unit time is equal to the amount of the solute that diffuses throughout the boundary of the interval \([a, b]\), plus the contribution of external sources:

\[
\frac{d}{dt} \int_a^b \theta(x, t) \, dx + \int_a^b F(x, t) \, dt = -J(b, x) + J(a, x) = -\int_a^b \frac{\partial}{\partial x} J(x, t) \, dx + \int_a^b F(x, t) \, dx.
\]

Here \(F(x, t)\) represents the contribution of external sources distributed along the segment \([0, \pi]\). We divide both sides with \(b - a\) and we pass to the limit for \(b \to x, a \to x\). We find

\[
\theta_t(x, t) = -J_x(x, t) + F(x, t).
\]

- A constitutive law:

  this law states how the velocity of diffusion is related to the concentration. The classical law is

  - **Fick law**

    The flux is proportional to the gradient of the concentration: \(J = -c \theta_x(x, t)\). Combining these equations together we find

    \[
    \theta_t(x, t) = \frac{\partial}{\partial x} [c(x) \theta_x(x, t)] + F(x, t),
    \]

  an equation which coincides with the classical heat equation.

  It is found experimentally that Fick law holds at low concentration and it does not hold at all in certain materials, like colloidal solutions or polymers. The shortcoming of Fick law is that it postulates that the molecules of the solute move freely in the solvent, and can react immediately to a gradient of the concentration. In particular, the concentration of the solute reaches its value at the boundary of \(\Omega\) in time 0, after an abrupt change of the concentration within \(\Omega\). As a consequence, there cannot be a physical surface which separates the region reached by the solute from the one which has not yet been reached. Experiments instead reveal such sharp separation in materials with complex molecular structure, see \([9, 27]\).

  For this reasons, a different constitutive equation has been proposed:

  - **Non-fickian diffusion law:**
The relation of $J$ and $\theta$ takes into account past memory of the concentration:

$$J(x, t) = - \int_0^t N(t - s) c(x) \theta_x(x, s) \, ds + \Phi(x, t).$$

Note that now $\Phi(x, t)$ includes in particular the contribution of the integral for $t < 0$. In fact, the integral should be

$$\int_{-\infty}^t = \int_0^0 + \int_0^t.$$

The contribution of the first integral on the right side is inserted in $\Phi(x, t)$.

Combining mass concentration and this non Fickian diffusion law we get the equation

$$\theta_t(x, t) = \frac{\partial}{\partial x} \int_0^t N(t - s) c(x) \theta_x(x, s) \, ds + F(x, t).$$

(27)

**Remark 9** A first instance of this relation was proposed in the following form:

$$\epsilon J_t + J = -c \theta_x \text{ so that } N(t) = \frac{1}{\epsilon} e^{-t/\epsilon}.$$

It was then realized that only one parameter $\epsilon$ is not sufficient to catch the complexity of the diffusion process, and the general form was proposed.

Equation (28) appears in different applications, for example as a model of viscoelastic materials. It was also proposed as a heat equation with finite diffusion velocity (see [5, 25]). After that it has been studied in a great number of papers. Among the oldest ones, see for example [3, 16, 23, 41]. In more recent times, this equation has been studied from the point of view of the stability theory and controllability, see for example [2, 4, 7, 36, 50, 63, 17]. See [50] for the existence and properties of the solutions. A recent approach based on moment theory is in [53, 54]. The results in these papers are reported in Section 5.3.

We shall study equation (27) with $c(x) \equiv 1$, i.e. we consider the equation

$$\theta_t = \int_0^t N(t - s) \theta_{xx}(s) \, ds + F(x, t).$$

(28)

This equation has to be provided with suitable initial and boundary conditions. We shall consider the case of homogeneous Dirichlet boundary condition

$$\theta(0, t) = 0, \quad \theta(\pi, t) = 0$$

and initial condition

$$\theta(x, 0) = \xi(x).$$
5.2 Input identification with accessible sources

We now consider the following problem: we assume \( \mu(t; x) = 0 \) for \( t < 0 \) so that \( F(x, t) \) represents the effects of distributed external source. We assume that

\[
F(x, t) = B(x)u(t)
\]

where \( u(t) \) is a scalar function. So, there is an external source of solute. The operator \( B \in L^2(0, \pi) \) is assume to be known; it might be the characteristic function of the set from which the solute enters the system; and we assume that this region is accessible to measures. The goal is the identification of the input function \( u(t) \); for example the goal could be to detect abrupt changes in this function.

An abstract form of Eq. (28) is

\[
\theta(t) = \int_0^t N(t-s)A\theta(s) \, ds + Bu(t)
\]

in \( L^2(0, \pi) \), where \( A \) is the operator

\[
\text{dom } A = H^1_0(0, \pi) \cap H^2(0, \pi), \quad (A\phi)(x) = \phi''(x).
\]

Hence, \( A = A^* \leq 0 \). The operator \( B \) acts from \( \mathbb{R} \) to \( L^2(0, \pi) \) and it is the multiplication with the function \( B(x) \):

\[
(Bu)(x) = B(x)u \quad \forall u \in \mathbb{R}.
\]

The observation being located at the source, we can assume

\[
y(t) = C\theta(t) = B^*\theta(t) = \int_0^\pi B(x)\theta(x, t) \, dx + \xi(t)
\]

\textit{(colocated input and output.)} The function \( \xi(t) \) represents the error in the measure and we assume that the tolerance \( h \) is known: \( |\xi(t)| < h \) for every \( t \). So, the measure is an averaged concentration of the solute at the source which is only indirectly related to the function \( u(t) \) since \( \mu(t) \) also depends on past activity.

A formula for the impulse response of Eq. (28) is unknown even in the simple case of one dimensional space variable, so that we can’t proceed as in Sections 4. Instead, we mimic the finite dimensional procedure in Sect. 2. We disregard the fact that measures have to be taken or elaborated at discrete times and we assume that the measures are taken at every time instant.

We assume that the kernel \( N(t) \) is of class \( C^3 \), with \( N(0) = 1 \) and moreover we assume that \( N(t) \) decays exponentially and it is a \textit{positive real function}: the Laplace transform \( \hat{N}(\lambda) \) is defined in a right half plane \( \Re \lambda > -\sigma, \sigma > 0 \); and, if \( \Re \lambda > 0 \) then \( \Re \hat{N}(\lambda) > 0 \).
Under solely these assumptions identification of $u$ might be impossible: it is impossible for example when $B = 0$. So we assume that the system is minimum phase:

$$\Re \lambda > 0 \implies \ker B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B = 0.$$  \hspace{1cm} (30)

Note that this in particular implies

$$\ker B \neq 0 \text{ so that } B \neq 0. \hspace{1cm} (31)$$

We first study the case $h = 0$, i.e. the ideal case of exact measures. We associate a model to our system,

$$w_t = \int_0^t N(t - s)\Delta w(s) \, ds + Bv(t), \quad z(t) = B^*w(t)$$

where, following the suggestions we have seen before, we propose

$$v(t) = -\frac{1}{\alpha} [z(t) - y(t)].$$

This quantity can be computed at each time $t$ and it has the following expression in terms of $w(t)$ and the unknown quantity $\theta(t)$:

$$v(t) = -\frac{1}{\alpha} B^*[w(t) - \theta(t)]. \hspace{1cm} (32)$$

Let us introduce the error

$$e(t) = w(t) - \theta(t), \quad \text{so that} \quad v(t) = -\frac{1}{\alpha} B^*e(t).$$

Note that $e(t)$ is not known. The function $e(t)$ solves the following equation:

$$e_t = \int_0^t N(t - s)\Delta e(s) \, ds - \frac{1}{\alpha} BB^*e(t) - Bu(t)$$

with zero initial and boundary conditions (of Dirichlet type).

We can’t use this formula in order to compute $e(t)$, since the right hand side depends on the unknown signal $u(t)$. But, we can elaborate on this formula in order to get needed information.

The elaboration we do is to study the Laplace transform. Laplace transform require that $u(t)$ be defined on $[0, +\infty)$, which is not an assumption. May be the system evolves only on a finite time interval of time. But, the formula for the estimate $v(t)$ is (32), which is causal: the value $v(t)$ at a certain time $t$ does not depend on the future. Hence we can introduce an arbitrary extension of $u$, for example $u = 0$, and this will not affect the results.

28
We get:
\[
\dot{e}(\lambda) = -[\lambda I - \tilde{N}(\lambda)A + \frac{1}{\alpha}BB^*]^{-1} B\dot{u}(\lambda),
\]
\[
\dot{v}(\lambda) = \frac{1}{\alpha} B^*[\lambda I - \tilde{N}(\lambda)A + \frac{1}{\alpha}BB^*]^{-1} B\dot{u}(\lambda),
\]
\[
\dot{B}\dot{u}(\lambda) - B\dot{u}(\lambda) = -
\left[
I + \frac{1}{\alpha} BB^*(\lambda I - \tilde{N}(\lambda)A)^{-1}
\right]^{-1} B\dot{u}(\lambda)
\]
\[
= -\alpha R \left[
\alpha I + B^* \left(\lambda I - \tilde{N}(\lambda)A\right)^{-1} B
\right]^{-1} \dot{u}(\lambda). \tag{33}
\]

The last row has been obtained thanks to the equality
\[
[I + BB^*]^{-1} B = B[I + B^* Z B]^{-1}.
\]

We note:

**Theorem 10** The inverse \((\lambda I - \tilde{N}(\lambda)A)^{-1}\) exists in \(\Re \lambda > 0\). The set of imaginary points in which the inverse does not exist is denumerable.

**Proof.** Let \(\Re \lambda > 0\) and let \(\theta_0\) satisfy
\[
(\lambda I - \tilde{N}(\lambda)A)\theta_0 = 0.
\]

Then we have also
\[
0 = \langle (\lambda I - \tilde{N}(\lambda)A)\theta_0, \theta_0 \rangle + \langle \theta_0, (\lambda I - \tilde{N}(\lambda)A)\theta_0 \rangle
\]
\[
= 2 \langle \Re \lambda \rangle \|\theta_0\|^2 + 2 \langle \Re \tilde{N}(\lambda) \rangle \|(-A)^{1/2} \theta_0\|.
\]

If \(\Re \lambda > 0\) then we must have \(\theta_0 = 0\) since \(\Re \tilde{N}(\lambda) > 0\).

The inverse might not exist on the imaginary axis but in this case we must have
\[
\frac{i\omega}{\tilde{N}(i\omega)} = -n^2
\]
(since \([-n^2]\) is the spectrum of \(A\)). The set of the numbers \(\omega\) which satisfy this equation is denumerable, since \(\lambda \rightarrow \lambda/\tilde{N}(\lambda)\) is an analytic function.

We now prove our main result here:

**Theorem 11** For every \(T > 0\), \(v_\alpha \rightarrow u\) in \(L^2(0, T)\) for \(\alpha \rightarrow 0^+\).

**Proof.** The idea is as follows: with \(T\) fixed, we extended \(u(t)\) to a square integrable function on \((0, +\infty)\). We shall use Parseval formula in order to compute the \(L^2(0, +\infty)\)-norm of \(Bu_\alpha(t) - Bu(t)\) as follows:
\[
\int_0^{+\infty} |Bu_\alpha(t) - Bu(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\dot{B}v(i\omega) - B\dot{u}(i\omega)|^2 \, d\omega
\]
\[
= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \alpha B \left[\alpha I + B^* \left(i\omega I - \tilde{N}(i\omega)A\right)^{-1} B\right]^{-1} \dot{u}(i\omega) \, d\omega. \tag{34}
\]
Of course, we must prove that the integral on the right side (equivalently, on the last line) makes sense. Furthermore, we prove that $|\hat{v}(i\omega) - \hat{u}(i\omega)|^2$ converges to 0 for every $\omega$ in such a way that we can use dominated convergence theorem. So, we get

$$\lim_{\alpha \to 0+} \int_{0}^{+\infty} |Bv_\alpha(t) - Bu(t)|^2 \, dt = 0;$$

i.e., $v_\alpha$ approximates the unknown input $u$, as wanted.

We need two steps in order to fulfill this program:

**Step 1** we prove

$$\alpha B \left[ \alpha I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1} \in H^\infty$$

and there exists $M > 0$ such that the following inequality holds for every $\alpha > 0$ and $\lambda$ with $\Re \lambda > 0$:

$$\left| \alpha B \left[ \alpha I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1} \right| < M$$

In this way we see that the integral at the right side of (34) makes sense since $\hat{u}(\lambda)$ is square integrable.

**Step 2** we prove that for every fixed $\lambda$ in $\Re \lambda \geq 0$ and such that $\left( \lambda I - \hat{N}(\lambda)A \right)^{-1}$ exists, we have

$$\lim_{\alpha \to 0+} \alpha B \left[ \alpha I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1} = 0.$$ 

As $\ker B = 0$ (see (31)), this is equivalent to

$$\lim_{\alpha \to 0+} \alpha \left[ \alpha I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1} = 0.$$ 

This limit combined with boundedness noted above, and $|\hat{u}(i\omega)|^2$ being integrable, justifies the use of dominated convergence theorem, as wanted.

So, we first prove boundedness, as required in **Step 1**. We need two properties: we first need the existence of

$$\left[ \alpha I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1}.$$
i.e. we fix $\alpha > 0$ and $\lambda$ with $\Re \lambda > 0$ and we must prove

$$\begin{bmatrix} \alpha I + B^* \left( \lambda I - \tilde{N}(\lambda)A \right)^{-1} B \end{bmatrix} u = 0 \implies u = 0.$$  \hspace{1cm} (35)

Once the existence of the inverse is known, the required boundedness property boils down to the following property: there should not exist a bounded sequence $\{\alpha_n\}$ and sequences $\{\lambda_n\}$ in $\Re \lambda_n > 0, ||u_n|| = 1$ such that

$$\frac{1}{\alpha_n} \begin{bmatrix} \alpha_n I + B^* \left( \lambda_n I - \tilde{N}(\lambda_n)A \right)^{-1} B \end{bmatrix} u_n = \begin{bmatrix} I + \frac{1}{\alpha_n} B^* \left( \lambda_n I - \tilde{N}(\lambda_n)A \right)^{-1} B \end{bmatrix} u_n \to 0.$$ \hspace{1cm} (36)

Note that condition (36) with $\lambda_n = \lambda, u_n = u$ and $\alpha_n = \alpha$ (hence not convergent to 0) reduces to (35). So, the two properties can be proved in one shot, by proving that condition (36) is impossible. For this, it is sufficient to prove that the limit of the real part cannot be 0.

As $\{u_n\}$ is in a finite dimensional space, there exists a subsequence (still denoted $\{u_n\}$) such that $u_n \to u_0$, and $||u_0|| = 1$. Let $a_n$ be the real part of (36), so that

$$a_n = \left\langle \begin{bmatrix} I + \frac{1}{\alpha_n} B^* \left( \lambda_n I - \tilde{N}(\lambda_n)A \right)^{-1} B \end{bmatrix} u_n, u_n \right\rangle + \left\langle u_n, \begin{bmatrix} I + \frac{1}{\alpha_n} B^* \left( \lambda_n I - \tilde{N}(\lambda_n)A \right)^{-1} B \end{bmatrix} u_n \right\rangle .$$

We prove that the sequence $\{a_n\}$ cannot converge to 0. And, in fact,

$$a_n = ||u_n||^2 + \frac{1}{\alpha_n} \left\langle B^* \left( \lambda_n I - \tilde{N}(\lambda_n)A \right)^{-1} B u_n, u_n \right\rangle + ||u_n||^2 + \frac{1}{\alpha_n} \left\langle B^* \left( \tilde{N}(\lambda_n)A \right)^{-1} Bu_n, u_n \right\rangle = 2||u_n||^2 + 2 \frac{\Re \lambda_n}{\alpha_n} \left\langle \left( \lambda_n I - \tilde{N}(\lambda_n)A \right)^{-1} Bu_n, u_n \right\rangle - 2 \frac{\Re \tilde{N}(\lambda_n)}{\alpha_n} \left\langle \left( \tilde{N}(\lambda_n)A \right)^{-1} Bu_n, A \left( \lambda_n I - A \right)^{-1} Bu_n \right\rangle .$$

Both the second and third terms are nonnegative for $\Re \lambda_n > 0$ so that

$$a_n \geq 2 ||u_n||^2 \geq 2,$$

as wanted (note that dissipativity of $A$ has been used).
Now we prove the statement in Step 2. We fix $\lambda$ in $\Re \lambda \geq 0$, such that the inverse $(\lambda I - \hat{N}(\lambda)A)^{-1}$ exists. This value of $\lambda$ is kept fixed. We prove the equivalent property
\[
\lim_{\alpha \to 0^+} \alpha \left[ \alpha I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1} = 0.
\]
If not, then there exist a sequence $\alpha_n \to 0^+$ and a sequence $\{u_n\}$ of real numbers such that $|u_n| = 1$, for which
\[
v_n = \alpha_n \left[ \alpha_n I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right]^{-1} u_n
\]
does not converge to 0.

Note that $\{v_n\}$ is bounded, thanks to Step 1. and it must be
\[
\lim \left[ \alpha_n I + B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B \right] v_n = \lim \alpha_n u_n = 0.
\]
This implies the contradiction $v_n \to 0$. In fact, $\{v_n\}$ being real valued and bounded, it has a convergent subsequence $v_{n_k} \to v_0$. As $\alpha_n \to 0^+$, we see that
\[
B^* \left( \lambda I - \hat{N}(\lambda)A \right)^{-1} B v_0 = 0.
\]
Assumption (30) implies $v_0 = 0$. So, every subsequence of the bounded sequence $\{v_n\}$ converges to 0, i.e. we get the contradictory statement that $v_n \to 0$.

This completely justifies the use of Parseval inequality and the exchange of limit and integral and we see that
\[
\lim_{\alpha \to 0^+} \hat{u}_\alpha = u \quad \text{in } L^2(0,T).
\]

The proof is now complete in the ideal case $h = 0$. If $h \neq 0$ then we have an additional term of the type $h/\alpha$. In this case $\hat{u} = \hat{u}_{h,\alpha}$ and we have the following consistency result:

**Theorem 12** Let us assume that $\alpha \to 0^+$ and $h \to 0$ while respecting the condition $h/\alpha \to 0$. Then,
\[
\hat{u}_{\alpha,h} \to u \quad \text{in } L^2(0,T).
\]

**Remark 13** The arguments above have been modeled on those presented in [15], where a parabolic equation is studied, i.e. an equation with infinite diffusion speed. In contrast with this, the case studied here is more similar to a hyperbolic equation, since it has finite propagation speed. In fact, it is (the integrated form of) the wave equation when $N(t) \equiv 1$. 32
5.3 A moment theory approach to Eq. (28)

In order to complete our picture of Eq. (28), we present recent results on the representation of the solutions. These results are connected with a study of controllability under boundary controls, and have been used to solve a source identification problem. These results can be found in [53, 54] while the source identification problem has been studied in [55].

Let us consider the sequence \( \{\sin nx\} \), which is the sequence of the eigenfunctions of the operator \( A \) in (29). These eigenfunctions are not normalized but have constant norm \( \sqrt{\pi/2} \) and the corresponding eigenvalue is \( -n^2 \).

We proceed formally: we expand \( \theta(\cdot, t) \) in series of the eigenfunctions \( \phi_n(x) \):

\[
\theta(x, t) = \sum_{n=1}^{+\infty} \theta_n(t) \phi_n(x).
\]

Due to the linearity of the problem, we can distinguish the contribution of \( \xi \) and the contribution of \( F(x, t) \) and it turns out that the representation of \( \theta(x, t) \) has the following form:

\[
\begin{align*}
\theta(x, t) &= \sum_{n=1}^{+\infty} z_n(t) \xi_n \phi_n(x) + \sum_{n=1}^{+\infty} \phi_n(x) \left[ \int_0^t z_n(t-s) F_n(s) \, ds \right], \\
F_n(t) &= \int_0^\pi \phi_n(x) F(x, t) \, dx.
\end{align*}
\]

Here,

\[
\begin{align*}
z_n'(t) &= -n^2 \int_0^t N(t-s) z_n(s) \, ds.
\end{align*}
\]

This is a formal representation. In order to prove that \( \theta(x, t) \) is well defined by this formula, we must prove that the sequence

\[
\{z_n(t)\}
\]

is a “base” in \( L^2(0, T) \) for a suitable time \( T \). In order to study system (27) with boundary control in the Dirichlet condition,

\[
\theta(0, t) = u(t) \in L^2(0, T),
\]

we need to know that the sequence

\[
\left\{ n \int_0^t N(t-s) z_n(s) \, ds \right\}
\]

is a “base” in \( L^2(0, T) \) for a suitable time \( T \).

There are many different definitions of “bases” in Hilbert spaces. We are most lucky when we can prove that a given sequence is a Riesz sequence
(called an \(L\)-sequence in Russian literature.) We recall the definition: let \(H\) be a separable Hilbert space and let \(\{z_n\}\) be a sequence of elements of \(H\). This set of elements is called a Riesz sequence if there exists an orthonormal base \(\{e_n\}\) (possibly on a different Hilbert space \(H_1\)) and a bounded boundedly invertible transformation \(T\) (between \(H_1\) and \(H\)), such that
\[
T e_r = z_r .
\]
If the sequence \(\{z_n\}\) is complete in \(H\) then it is called a Riesz base, see [1, 65].

Using Bari theorem (see [19]), it is possible to prove that both the sequence (39) and (40) are Riesz sequences provided that \(N(t) \in C^3(0, +\infty)\) and \(N(0) = 1\).

These facts can be used in order to prove both results on the existence and properties of the solutions and controllability/observability results. Concerning existence and properties of the solutions we have in particular:

**Theorem 14** Both the sequences in (37) converge and define a function \(\theta \in C(0,T;L^2(0,\pi))\) for every \(T > 0\), which furthermore depends continuously on \(\xi \in L^2(0,\pi)\) and \(F \in L^2(0,T;L^2(0,\pi))\).

If the forcing terms (both boundary and distributed) are zero and the initial condition \(\xi\) belongs to \(H^1(0,\pi)\), then the output
\[
y(t) = \int_0^t \left[ \frac{\partial}{\partial x} \theta(x,s) \right]_{x=0} ds
\]
belongs to \(L^2(0,T)\) for every \(T > 0\) and is continuous from \(H^1(0,\pi)\) to \(L^2(0,T)\).

Concerning controllability and observability we have:

**Theorem 15** There exists a number \(T_0\) such that:

- if the forcing terms (both boundary and distributed) are zero and if \(y(t) = 0\) on \([0,T_0]\) then \(\xi = 0\). Furthermore, if \(\xi \neq 0\), the inverse transformation: \(y(\cdot) \rightarrow \xi: L^2(0,T_0) \rightarrow L^2(0,\pi)\) is continuous, i.e. observability holds.

- Let a target \(\eta \in L^2(0,\pi)\) (as well as the initial conditions \(\xi \in L^2(0,\pi)\)) be given. There exists a control \(u(t) \in L^2(0,T_0)\) such that the solution \(\theta(t)\) of problem (28) (now with boundary control \(\theta(0, t) = u(t)\)) satisfies \(\theta(T_0) = \eta\), i.e. controllability holds.

These results have been applied in [55] in order to solve a source identification problem: the function \(F(x,t)\) has the form \(F(x,t) = B(x)g(t)\) and \(g(t)\) is known. It is required to identify \(B(x)\) on the basis of \(g(t)\) and observations taken on the flux at the boundary.
Remark 16 See [28] for a further discussion of controllability concepts for Eq. (28).

6 Conclusions

In this paper we have given an overview of recent results concerning input identification to linear finite dimensional systems and their extensions to distributed systems and in particular to a class of distributed systems with infinite memory. This class of systems, encountered for example in thermodynamics, chemistry and viscoelasticity, has the noticeable property of being close in spirit to hyperbolic systems since it has finite propagation speed.

A simple example of systems with a finite “signal propagation speed” are of course systems with finitely many lags. Input identification for this class of systems is not so often studied as it would deserve.

The input identification algorithm has been applied to Active Noise Cancellation in regions of \( \mathbb{R}^n \), \( n \geq 1 \). An approach often used is either to discretize the system first, and to solve the problem for the resulting finite dimensional approximation, or to use frequency domain techniques. Alternatively, the problem can be studied in its original state space formulation first. Discretization is then used at the moment of realizing the feedback loop, see for example [37, 57, 59]. The approach presented here seems to be applicable to this problem too without the need of a preliminary finite dimensional approximation.

References


37


