Approximating distributional behaviour of linear systems using Gaussian function and its derivatives

Athanasios A. Pantelous[1] and Nicos Karcanias[2]

[1]Department of Mathematical Sciences, University of Liverpool, UK
(e-mail: A.Pantelous@liverpool.ac.uk)
[2]School of Engineering and Mathematical Sciences, City University, UK
(; e-mail: N.Karcanias@ city.ac.uk)

Abstract: The paper is concerned with defining families of smooth functions that can be used for the approximation of impulsive types of solutions for linear systems. We review the different types of approximations of distributions in terms of smooth functions and explains their significance in the characterization of system properties where impulses were used for their characterisation. For controllable systems, we establish an interesting relation between the time t and sigma (volatility) in the approximation of distributional solutions. An algorithm is then proposed for the calculation of the coefficients of the input required to minimize the distance of our desired target state before and after approximation is proposed. The optimal choice of sigma is derived for a pre-determined time t for the state transition.

Keywords: Linear systems; approximating distributional behaviour; Gaussian function and its derivatives

1. INTRODUCTION

The use of distributions in the study of linear system problems is a well-established subject going back to Gupta and Hasdorff (1963), Zadeh and Desoer (1963), Verghese (1979), Verghese and Kailath (1979), Karcanias and Kouvaritakis (1979), Campbell (1980, 1982), Willems (1981), Jaffe and Karcanias (1981), Cobb (1982, 1983), Karcanias and Hayton (1982), Karcanias and Kalogeropoulos (1989), Willems (1991), and references there in. The work so far has dealt with the characterisation of basic system properties such as infinite poles and zeros Verghese (1979), Verghese and Kailath (1979) for regular and singular (implicit) systems, as well as the study of fundamental control problems where the solution is expressed in terms of distributions. Typical problems are those dealing with the notions of Almost (A, B)-invariance and almost controllability subspaces Willems (1981), Jaffe and Karcanias (1981).

In particular, the study of distributional solutions plays a key role in many areas in systems and control such as:

(i) Controllability, Observability.

(ii) Infinite zero characteristic behaviour.

(iii) Almost invariant subspaces, almost controllability spaces.

(iv) Dynamics of singular systems etc.

The distributional characterization is also linked to solution of a number of control problems. The solution of such problems have theoretical significance, given that distributions cannot be constructed and only smooth functions can be constructed and implemented. The idea of approximating distributional inputs with smooth functions that achieve a similar control objective was first introduced by Gupta and Hasdorff (1963), Gupta (1966). In this paper, we review different possible approximations of Dirac distributions when infinite, or finite time domain is used that for their approximation and give some necessary and sufficient conditions for such approximation. We also consider the problem of approximation of the sum of Dirac distributions and its derivatives and their use as inputs to systems.

We assume that the system is controllable, and under this assumption we establish an interesting connection between time t and sigma (volatility) parameter of the approximating, cumulative Gaussian density function. Surprisingly, the fraction t/sigma is more or less constant and a probabilistic criterion is given describing the first (minimum - stopping) time that the desired state is being achieved. Finally, a new algorithm is proposed for the calculation of the coefficients of our smooth input signal that approximate the distributional input that carries out the transfer of the origin to a desirable point in the space. In this algorithm, we want to minimize the distance of the desired state to be transferred before and after approximation. The optimal choice of sigma is derived for a pre-determined time t. The results provide a rigorous statement of the early idea presented by Gupta (1966).

2. PROBLEM DEFINITION

We consider the linear time invariant (LTI) system

$$\underline{x}'(t) = A\underline{x}(t) + \underline{b}u_o(t), \qquad (2.1)$$

where $\underline{x}(t) \in \mathcal{C}^{\infty}\left(\mathbb{F}, \mathcal{M}(n\times 1;\mathbb{F})\right)$ (smooth function over the field $\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$, which elements belong to the algebra $\mathcal{M}(n\times 1;\mathbb{F})$), and $u_o(t)\in\mathbb{F}$ are the state, input vectors, respectively and $A\in\mathcal{M}(n\times n;\mathbb{R})$, $\underline{b}\in\mathcal{M}(n\times 1;\mathbb{R})$. For the simplicity, we can assume, that A is simple and expressed as

$$A = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}, \tag{2.2}$$

where $\lambda_i \neq 0$ for every $i \in \underset{\sim}{n}$ ($\underset{\sim}{n} \triangleq \{1,2,...,n\}$).

The paper deal with the following *key* question: *"Can we develop an approximation to impulsive behaviour with a respective approximation of the related system and control properties?"*

The answer to this question underpins, the development of a smooth approximation of impulsive trajectories and thus also of the related system and control properties. A number of control problems involving distributional solutions relate to the adjustment of initial conditions with distributional inputs, resulting to distributional state trajectories; these imply changing the given state of a linear system to a desired state in minimum time. The important questions that arise are:

(i) How can we approximate distributions and their derivatives by different families of smooth functions and their derivatives?

(ii) What are the different types of approximation?

(iii) What is the impact of the approximation on the properties of the associated control problem and on the nature of the resulting transition, when smooth functions are used?

It is assumed that the input to the LTI is a linear combination of Dirac δ-function and its first $n-1$ derivatives, ie

$$u_o(t) = \sum_{k=0}^{n} a_k \delta^{(k)}(t). \tag{2.3}$$

which is a linear combination of Dirac δ-function and its first $n-1$ derivatives, where $\delta^{(k)}$ or $\frac{d^k\delta}{dt^k}$ is the k^{th} derivative of the Dirac δ-function, and a_k for $i \in \underset{\sim}{n}_o$ ($\underset{\sim}{n}_o \triangleq \{0,1,2,...,n-1\}$) are the magnitudes of the delta function and its derivatives. We shall denote the state of the system at time $t=0^-$ as $\underline{x}(0^-)$ and at time $t \geq 0^+$ achieves as $\underline{x}(0^+)$. We assume that $\underline{x}(0^-) = [0 \quad 0 \quad \dots \quad 0]^T$ at $t=0^-$ and $\underline{x}(0^+) = [x_1 \quad x_2 \quad \dots \quad x_n]^T$ at $t \geq 0^+$. Furthermore, we assume that the system is *controllable* and thus we can transfer the state in any desired point of the state space.

Furthermore, we assume that our system is *controllable*, i.e. we can transfer the state in any desired point. Let the state of the system at time $t=0^-$ is $\underline{x}(0^-)=\underline{0}$ and at time $t=0^+$ achieves $\underline{x}(0^+)$. The existence of an input that transfers the state of the system (2.1) from $\underline{x}(0^-)=\underline{0}$ to $\underline{x}(0^+)$ requires that the vector $\underline{x}(0^+)$ belongs to the controllable subspace of the pair. The necessary and sufficient condition for the state of a system (1) to be transferred from $\underline{x}(0^-)=\underline{0}$ at time $t=0^-$ to some $\underline{x}(0^+)\in\{A|\underline{b}\}$ at $t=0^+$ by the action of control input of type (2.3) is that the resulting trajectory $\underline{x}(t)$ is expressed as $\underline{x}(t)=\sum_{k=0}^{n-1}\beta_k\delta^{(k)}(t)$ where the coefficients β_k for $i \in \underset{\sim}{n}_o$ are chosen to be the components of $\underline{x}(0^+)$ along the subspace $\underline{b}, A\underline{b}, A^2\underline{b}, \dots, A^{n-1}\underline{b}$, respectively according to some projections law.

In the next section, we consider some background results on the approximation of Dirac delta function are presented.

3. APPROXIMATIONS OF DIRAC DELTA FUNCTION

Defining smooth functions, which can approximate in some sense distributions, has been a problem that has been considered in the literature. In this section, we review these results and we suggest a systematic and very rigorous procedure for generating sequences of its derivatives. If one of the different approximations of Dirac δ-function is being followed, (see Gupta and Hasdorff, (1963), Gupta, (1966), Zemanian, (1987), Cohen and Kirchner, (1991), Estrada and Kanwal, (2000), Kanwal, (2004) etc) the change of the state in some minimum practical time depends mainly upon how well the approximations have being generated. The relationship between the approximation used and the resulting time for the transition is an important issue that is considered subsequently.

The Dirac δ-function can be viewed as the limit of sequence function

$$\delta(t) = \lim_{a\to 0}\delta_a(t), \tag{3.1}$$

where $\delta_a(t)$ is called a *nascent* delta function. This limits is in the sense that

$$\lim_{a\to 0}\int_{-\infty}^{+\infty}\delta_a(t)f(t)dt = f(0). \tag{3.2}$$

These properties can often be simulated by using a smooth, finite approximation of the Dirac distribution. Such approximations have additional advantages. In fact, approximating the Dirac distribution by a smooth function may actually be a better representation of the solution sought in the particular problem, especially if the width of the approximation function can be coupled to the physics of the problem. Following the ideas of Cohen and Kirschner (1991), a suitable approximating function, which is convenient for computations, should satisfy the following important properties everywhere on the domain under consideration:

1. Its limit with some defining parameter is the Dirac distribution (see eq. (3.1)).

2. It is positive, decreases monotonically from a finite maximum at the source point, and tends to zero at the domain extremes.

3. Its derivative exists and is continuous function.

4. It is symmetric about the source point, for instance 0 (see eq. (3.1) and (3.2)).

5. It can be represented by a reasonably simple Fourier integral (for infinite domains) or Fourier series (for finite domains).

Next, we discuss the appropriate approximation of Dirac function based on the finiteness or infiniteness of the time domain.

3.1. Infinite Time Domain

We first point out that the best nascent delta function depends on the particular application. Some well known and very useful in applications nascent delta functions are the Gaussian and Cauchy distributions, the rectangular function, the derivative of the sigmoid (or Fermi-Dirac) function, the Airy function etc; see for instance Gupta (1966), Zemanian (1987), Estrada and Kanwal (2000), Kanwal (2004) et al. and recently the use of a finite difference formula which is converted into an appropriate sequence; see Boykin (2003). A short review of such approximations is given next.

A very useful in applications nascent delta function are:

● The Cauchy distributions,

$$\delta_a(t) = \frac{1}{\pi}\frac{a}{a^2+t^2} = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{ikt-|ak|}dk\,,$$

● The rectangular function,

$$\delta_a(t) = \frac{rect(t/a)}{a} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \sin c\left(\frac{ak}{2\pi}\right)e^{ikt}dk\,,$$

● The derivative of the sigmoid (or Fermi-Dirac) function,

$$\delta_a(t) = \partial_t \frac{1}{1+e^{-t/a}} = -\partial_t \frac{1}{1+e^{t/a}},$$

● The Airy function

$$\delta_a(t) = \frac{1}{a}A_i\left(\frac{t}{a}\right).$$

Following Boykin (2003), the finite difference formula may be easily converted into a sequence that approaches a derivative of the Dirac delta function in one dimension.

Thus, we obtain

$$\delta_a(t) = \begin{cases} \frac{1}{a}, & -\frac{a}{2} < t < \frac{a}{2} \\ 0, & |t| > \frac{a}{2} \end{cases}, \qquad (3.3)$$

which approaches $\delta(t)$ as $a \to 0$. Moreover, an expression for the derivatives of the Dirac delta can be given by the following equation,

$$\frac{d^k}{dx^k}\delta(x) = \lim_{\substack{a\to 0\\ k\to 0}}\left[\left(\frac{1}{h}\right)^k \sum_{j=0}^{k} a_j \delta_a\left(x + b_j h\right)\right], \qquad (3.4)$$

where $x = t_o - t$ and we use

$$\frac{d^k}{du^k}\delta(u)|_{t-t_o} = (-1)^k \frac{d^k}{du^k}\delta(u)|_x\,.$$

The expression (3.4) is exactly what we might obtain by simply making the substitution $f(t) \to \delta_a(t)$ in the following finite difference approximation for the k^{th} derivative of a test function f evaluated at t_o which can be represented as

$$\frac{d^k}{dt^k} f(t)|_{t=t_o} \approx \left(\frac{1}{h}\right)^k \sum_{j=0}^{k} a_j f\left(t_o + b_j h\right). \qquad (3.5)$$

Note that a_j and b_j are suitable chosen constants and (3.5) becomes exactly in the limit $h \to 0$. Furthermore, due to the fact that f is sampled at discrete points, we can write

$$\frac{d^k}{dt^k} f(t)|_{t=t_o} = \lim_{h\to 0}\left\{\left(\frac{1}{h}\right)^k \sum_{j=0}^{k} a_j \int_{-\infty}^{+\infty} \delta\left(t-\left(t_o + b_j h\right)\right) f(t)\,dt\right\} \qquad (3.6)$$

3.2. Finite Time Domain

Unfortunately, the Gaussian distribution is not a good approximation of the Dirac distribution on a finite domain, namely that the first derivative (which is important in this paper) can be discontinuous at a special point. Thus, recently, a different approximation has been proposed by Cohen and Kirschner (1991), which satisfies all the properties (**1**) through (**5**). This is the β-distribution of the classical probability theory. This distribution has the expression

$$\beta_\pi(\theta) = \begin{cases} \frac{(\pi+\theta)^{a-1}(\pi-\theta)^{b-1}}{(2\pi)^{2a-1}B(a,b)}, \forall\theta \in \mathcal{J} \\ 0 \qquad\qquad\qquad\qquad , otherwise \end{cases} \qquad (3.7)$$

where $\mathcal{J}$ is a finite interval and

$$B(a,b) \triangleq \int_{\mathcal{J}} (\pi+\theta)^{a-1}(\pi-\theta)^{b-1}\,d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

where also $\Gamma(x)$ is the well-known Gamma distribution. Since, in the next few lines of the present paper, the infinite time domain is used, the interesting reader may consult Cohen and Kirschner (1991) for further details.

3.3 Why a sum of Dirac Delta Functions?

However, in our approach, our time domain is *infinite* and the classical Gaussian distribution, i.e.

$$\delta(t)=\lim_{\sigma\to 0}\frac{1}{\sigma\sqrt{2\pi}}e^{-t^2/2\sigma^2}=\lim_{\sigma\to 0}\frac{1}{\sigma}\Phi\left(\frac{t}{\sigma}\right), \qquad (3.8)$$

where $\Phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is being used.

Consequently, the approximate expression for the controller (2.3) is given by

$$u_\sigma(t)=\sum_{k=0}^{n-1}a_k\frac{1}{\sigma^{k+1}}\Phi^{(k)}\left(\frac{t}{\sigma}\right), \qquad (3.9)$$

where $\Phi^{(i)}\left(\frac{t}{\sigma}\right)=\left(\frac{d^i}{dt^i}\left(\frac{t}{\sigma}\right)^i\right)\Phi\left(\frac{t}{\sigma}\right)$.

Then, we take the limit

$$u_o(t)=\lim_{\sigma\to 0}u_\sigma(t). \qquad (3.10)$$

Moreover, at the end of this section, we are answering to another significant question: "*why a sum of Dirac delta functions*?"

Considering the results of 2nd section and the whole discussion till that part of the 3nd section, generally speaking, we should point out that the input for the linear differential system (2.1) should be given by a *single-layer distribution;* see Zemanian (1987), Estrada and Kanwal (2000) and Kanwal (2004). This kind of distributions has a huge importance in many applications.

Lemma 3.1 If $\mathcal{U}$ is a bounded closed set in $\mathbb{F}$ and $\mathcal{Y}$ is a neighbourhood of $\mathcal{U}$, then there exists a function such that $n=1$ on $\mathcal{U}$, $n=0$ outside $\mathcal{Y}$, and $0\le n\le 1$ over $\mathbb{F}$. □

Definition 3.1 Let $\mathcal{S}$ be a piecewise regular curve in $\mathbb{F}$ and σ is a locally integrable function defined on $\mathcal{S}$. The linear continuous functional $\sigma\delta_S$ on the space $\mathcal{D}$ of infinitely differentiable complex-valued functions on $\mathbb{F}$ with compact support is defined as

$$\langle\sigma\delta_S,\varphi\rangle=\int_S\varphi(\xi)\sigma(\xi)\delta S$$

$\forall\ \varphi\in\mathcal{D}$ and is called *single (or simple) layer on* $\mathcal{S}$ *with density* σ. □

Note that $\sigma\delta_S(x)=\int_S\delta(x-\xi)\sigma(\xi)\delta S_\xi$.

Definition 3.2 Let $\mathcal{S}$ be a piecewise regular curve in $\mathbb{F}$ and $\mu\delta_S$. The linear continuous functional $-d/dt(\mu\delta_S)$ on the space $\mathcal{D}$ of infinitely differentiable complex-valued functions on $\mathbb{F}$ with bounded support is defined as

$$\langle -d/dt(\sigma\delta_S),\varphi\rangle=\int_S\sigma(\xi)\frac{d\varphi(x-\xi)}{dt}\delta S\quad\forall\ \varphi\in\mathcal{D}.\ \square$$

Consequently, it can be easily shown that every distribution $\sigma\delta_S(x)$ that has compact support is of finite order, see Zemanian (1987) Estrada and Kanwal (2000). Thus, it is deduced that every distribution $\sigma\delta_S(x)$ whose support is the point $x=\tau$ has the form $\sum_{k=0}^{n-1}c_k\delta^{(k)}(t-\tau)$, i.e. a linear independent combination of Dirac δ-function and its first $n-1$ derivatives. Consequently, since we are interesting to transfer the state of system (2.1) at time $t=0^-$ from the initial point $\underline{x}(0^-)$ and at time $t=0^+$ to achieve $\underline{x}(0^+)$, (2.3) is appropriate, when the support point is $\tau=0$.

4. MAIN RESULTS

In this section, we will try to answer to the following questions: "*if we wish to achieve state* $\underline{x}(0^+)$ *at time* $t=0^+$ *what are the necessary coefficients* a_k *for* $k\in\underset{\sim}{n}$ *and what is the optimal choice of volatility* σ *that it takes the state there at time* $t\ge 0^+$ *?*" In this direction, the following known results are significant.

Lemma 4.1 The solution of system (2.1) is given by

$$\underline{x}(t)=e^{At}\int_{-\infty}^{t}e^{-A\tau}\underline{b}u_o(\tau)d\tau, \qquad (4.1)$$

where A is diagonal and $u_o(\tau)$ is given by combining (3.9) and (3.10). □

Remark 4.1 In the general case, the matrix A is not always simple. However, the problem described above can be solved similarly. In this case, we should generate n linearly independent vectors $v_1,v_2,\ldots,v_n$ and a $n\times n$ similarity transformation $Q=[v_1,v_2,\ldots,v_n]$ that takes A into the Jordan canonical form. In the next lines, we present briefly the more essential part. Further details are omitted, since they are far beyond the scopes of the present version of the paper.

Thus, there exists an invertible matrix $Q\in\mathcal{M}(n\times n;\mathbb{F})$ such as $J=Q^{-1}AQ$, where $J\in\mathcal{M}(n\times n;\mathbb{F})$ is the Jordan canonical form of matrix A. Analytically,

$$J=block\ diag\{J_o,J_{q+1},J_{q+2},\ldots,J_k\}$$

• The block diagonal matrix $J_o=block\ diag\{J_1,J_2,\ldots,J_q\}$, where

$$J_i=\begin{bmatrix}\lambda_i & 1 & & 0\\ & \lambda_i & 1 & \\ & & \ddots & 1\\ 0 & & & \lambda_i\end{bmatrix}\in\mathcal{M}(\tau_i\times\tau_i;\mathbb{F})$$

is also a diagonal matrix with diagonal elements the eigenvalue λ_i, for $i=\underset{\sim}{q}$. Consequently, the dimension of J_o is $s\times s$, $s\triangleq\sum_{i=1}^{q}\tau_i$.

• Also, each block matrix $J_j = block\ diag\left\{J_{j,1}, J_{j,2}, \ldots, J_{j,d_j}\right\}$,

$$J_{j,z_j} = \begin{bmatrix} \lambda_j & 1 & & & \\ & \lambda_j & 1 & & \\ & & \lambda_j & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_j \end{bmatrix} \in \mathcal{M}\left(z_j \times z_j; \mathbb{F}\right)$$

for $j = q+1, q+2, \ldots, k$, and $z_j = \underset{\sim}{d}_j$. □

However, only for the simplicity of calculations, we have already assumed that the matrix A is in diagonal form. Consequently, the solution (4.1) is transposed into

$$\underline{x}(t) = \lim_{\sigma \to 0} \left\{ e^{At} \int_{-\infty}^{t} e^{-A\tau} \underline{b} u_\sigma(\tau) d\tau \right\},$$

or equivalently,

$$\underline{x}(t) = e^{At} \lim_{\sigma \to 0} \left[\int_{-\infty}^{t} e^{-A\tau} \underline{b} \sum_{k=0}^{n-1} a_k \frac{1}{\sigma^{k+1}} \Phi^{(k)}\left(\frac{\tau}{\sigma}\right) d\tau \right].$$

Remark 4.2 In order to make our calculations affordable due to the long number of terms that get involved, we consider the fact that $\Phi(t/\sigma) = 1/\sqrt{2\pi}\ e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2}$ and its derivatives tend to zero very *strongly* with $t/\sigma \to \infty$ (note that $\sigma \to 0$). Thus, by letting $t/\sigma = K(t,\sigma)$, where $K(t,\sigma)$ is chosen large enough (i.e. $K(t,\sigma) \to \infty$) that the assumption as stated above is valid, i.e.

$$\Phi(t/\sigma) \triangleq \Phi\left(K(t,\sigma)\right) \overset{K(t,\sigma)\to\infty}{\to} 0,$$

and its derivatives

$$\Phi^{(k)}(t/\sigma) \triangleq \Phi^{(k)}\left(K(t,\sigma)\right) \overset{K(t,\sigma)\to\infty}{\to} 0, \text{ for } k \in \underset{\sim}{n}_o. \square$$

Actually, the choice of $K(t,\sigma)$ depends on the choice of time t and the volatility σ (note that $\sigma \to 0$). In practice, the time t can be fixed, since we can pre-define the time in order to change the initial state of the system in (almost) zero time, for instance it can be $t \propto 10^{-6}$ seconds. So, as we will see analytically in the next paragraphs, the problem can be transferred into a distance-minimization problem, since we want to determine

$$\sigma^* = \inf\{\sigma \in \mathbb{R}_+ : \textit{for a fixed time } t^*, \textit{ such that } \Phi^{(k)}\left(K(t^*,\sigma)\right) \to 0, k \in \underset{\sim}{n}_o\}.$$

For the optimal choice of σ^*, we have to minimize the distance $\underline{x}(t^*) - \underline{\hat{x}}(t^*)$ using the Euclidian norm, i.e.

$$\left\| \underline{x}(t^*) - \underline{\hat{x}}(t^*) \right\|_2 \to 0, \tag{4.2}$$

where $\underline{x}(t^*)$ is the desired state and $\underline{\hat{x}}(t^*)$ is given by the approximation procedure, see equation (4.1).

The following lemma is required for the subsequent developments. Our objective is to re-write the equation (4.1).

Lemma 4.2 The approximated expression (4.3) holds,

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} u_\sigma(\tau) d\tau \approx \Phi^{-1}\left(K(t,\sigma) + \lambda_i \sigma\right) e^{\frac{1}{2}\lambda_i^2 \sigma^2} \sum_{k=0}^{n-1} a_k \lambda_i^k, \tag{4.3}$$

where

$$\Phi^{-1}(x) \triangleq \int_{-\infty}^{x} \Phi(y) dy = \sqrt{2}\,\mathrm{erf}^{-1}(2x-1), \text{ when } x \in (0,1).$$

Proof: Substituting the expression (3.9) into the integral $\int_{-\infty}^{t} e^{-\lambda_i \tau} u_\sigma(\tau) d\tau$, and we obtain

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} \sum_{k=0}^{n-1} \frac{a_k}{\sigma^{k+1}} \Phi^{(k)}(\tau/\sigma) d\tau = \sum_{k=0}^{n-1} a_k \int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi^{(k)}(\tau/\sigma)}{\sigma^{k+1}} d\tau.$$

Let us start with

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi(\tau/\sigma)}{\sigma} d\tau = \int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\tau^2}{2\sigma^2}} d\tau$$
$$= e^{\frac{1}{2}\lambda_i^2\sigma^2} \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\tau}{\sigma} + \lambda_i \sigma\right)^2} d\tau = e^{\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(\frac{t}{\sigma} + \lambda_i \sigma\right)$$

Now, we will calculate

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi'(\tau/\sigma)}{\sigma^2} d\tau = e^{-\lambda_i \tau} \frac{1}{\sigma} \Phi(\tau/\sigma)\Big|_{-\infty}^{t} + \lambda_i \int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi(\tau/\sigma)}{\sigma} d\tau$$
$$= e^{-\lambda_i t} \frac{1}{\sigma} \Phi(t/\sigma) + \lambda_i e^{\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(\frac{t}{\sigma} + \lambda_i \sigma\right),$$

and

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi''(\tau/\sigma)}{\sigma^3} d\tau$$
$$= e^{-\lambda_i \tau} \frac{1}{\sigma^2} \Phi'(\tau/\sigma)\Big|_{-\infty}^{t} + \lambda_i \int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi'(\tau/\sigma)}{\sigma^2} d\tau$$
$$= e^{-\lambda_i t}\left[\lambda_i \frac{1}{\sigma} \Phi(t/\sigma) + \frac{1}{\sigma^2} \Phi'(t/\sigma)\right] + \lambda_i^2 e^{\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(\frac{t}{\sigma} + \lambda_i \sigma\right).$$

Similarly, we can prove that

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi^{(k)}(\tau/\sigma)}{\sigma^{k+1}} d\tau$$
$$= e^{-\lambda_i t} \sum_{m=1}^{k} \lambda_i^{m-k+1} \frac{1}{\sigma^{k-m+1}} \Phi^{(k-m)}\left(\frac{t}{\sigma}\right) + \lambda_i^k e^{\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(\frac{t}{\sigma} + \lambda_i \sigma\right).$$

Now, we choose $t/\sigma = K(t,\sigma)$, (note that $\sigma \to 0$) where $K(t,\sigma)$ is chosen large enough (i.e. $K(t,\sigma) \to \infty$) such as

$$\Phi(t/\sigma) \triangleq \Phi\left(K(t,\sigma)\right) \overset{K(t,\sigma)\to\infty}{\to} 0,$$

and its derivatives

$$\Phi^{(k)}(t/\sigma) \triangleq \Phi^{(k)}\left(K(t,\sigma)\right) \overset{K(t,\sigma)\to\infty}{\to} 0, \text{ for } k \in \underset{\sim}{n}_o.$$

Consequently, we have

$$\int_{-\infty}^{t} e^{-\lambda_i \tau} \frac{\Phi^{(k)}(\tau/\sigma)}{\sigma^{k+1}} d\tau \approx \lambda_i^k e^{\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(K(t,\sigma)+\lambda_i\sigma\right),$$

and (4.3) is proven. □

Furthermore, combining expressions (4.1) and (4.3), we take

$$\underline{x}_i\left(K(t,\sigma)\sigma\right) \approx e^{\lambda_i K(t,\sigma)\sigma+\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(K(t,\sigma)+\lambda_i\sigma\right) b_i \sum_{k=0}^{n-1} a_k \lambda_i^k, \quad \text{for } i=1,2,\ldots,n. \quad (4.4)$$

Note that since we have assumed that $\frac{t}{\sigma} = K(t,\sigma) \to \infty$ then, we have to consider the solution of (2.1) at time $t^* = K(t,\sigma)\sigma$ (which can be $\propto 10^{-6}$). □

Consequently, the following lemma derives.

Lemma 4.3 For the diagonal matrix A we obtain the system

$$\underline{a} = V^{-1}\underline{\hat{x}}(t^*), \quad (4.5)$$

where $t^* = K(t,\sigma)\sigma$, $V = V(\lambda_1, \lambda_2, ..., \lambda_n)$ is the Vandermonde matrix,

$$\underline{\hat{x}}_i(t^*) \triangleq \frac{\underline{x}_i\left(K(t,\sigma)\sigma\right)}{e^{\lambda_i K(t,\sigma)\sigma+\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(K(t,\sigma)+\lambda_i\sigma\right) b_i}, \text{ for } i \in \underset{\sim}{n} \quad (4.6)$$

and $\underline{a} = \left[a_o \quad a_1 \quad \cdots \quad a_{n-1}\right]^T$.

Proof: The expression (4.4) can be re-written as follows

$$\underline{\hat{x}}_i(t^*) \triangleq \frac{\underline{x}_i\left(K(t,\sigma)\sigma\right)}{e^{\lambda_i K(t,\sigma)\sigma+\frac{1}{2}\lambda_i^2\sigma^2} \Phi^{-1}\left(K(t,\sigma)+\lambda_i\sigma\right) b_i} \approx \sum_{k=0}^{n-1} a_k \lambda_i^k,$$

for $i \in \underset{\sim}{n}$. Now by making some simple algebra, we have

$$\underline{\hat{x}}(t^*) = \begin{bmatrix} \underline{\hat{x}}_1(t^*) \\ \underline{\hat{x}}_2(t^*) \\ \vdots \\ \underline{\hat{x}}_n(t^*) \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{n-1} a_k \lambda_1^k \\ \sum_{k=0}^{n-1} a_k \lambda_2^k \\ \vdots \\ \sum_{k=0}^{n-1} a_k \lambda_n^k \end{bmatrix},$$

or equivalently

$$\underline{\hat{x}}(t^*) = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ & & \vdots & & \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix} \underline{\alpha},$$

since the eigenvalues λ_i's are distinct, the Vandermonde matrix exists, from which we can obtain the expression (4.6). □

Now, we will return to the expression (4.2), i.e.

$$\left\|\underline{x}(t^*) - \underline{\hat{x}}(t^*)\right\|_2 = \sqrt{\sum_{i=1}^{n}\left[\underline{x}_i\left(K(t,\sigma)\sigma\right) - \underline{\hat{x}}_i\left(K(t,\sigma)\sigma\right)\right]^2} \to 0$$

if and only if we determine $K(t,\sigma)$ such that

$$\Phi^{-1}\left(K(t,\sigma)+\lambda_i\sigma\right) = e^{-\frac{1}{2}\lambda_i^2\sigma^2}. \quad (4.7)$$

The expression (4.2) is very elegant because it transfers a pure system and control theory problem into a standard statistical problem. Theoretically speaking, we have already assumed that $\sigma \to 0$, so expression (4.7) gives

$$\Phi^{-1}\left(K(t,\sigma)\right) = 1 \Leftrightarrow \int_{-\infty}^{K(t,\sigma)} \Phi(x)dx = 1. \quad (4.8)$$

In probability theory and statistics, the normal distribution or Gaussian distribution $\Phi(x)$ is a continuous probability distribution that often gives a good description of data that cluster around the mean. The graph of the associated probability density function is bell-shaped, with a peak at the mean, and is known as the Gaussian function or bell curve.

Actually, in our case we are interested for

$$\int_{-\infty}^{K(t,\sigma)} \Phi(x)dx,$$

which is the **cumulative distribution function** (cdf) of a random variable X evaluated at a number $K(t,\sigma)$ (with other words, it is the probability of the event that *X* is less than or equal to $K(t,\sigma)$). Fortunately, since we want the above expression to be equal to 1, i.e. (4.8) holds, we have only to look a standard cumulative normal distribution table, and we can straightforwardly determine the value of $K(t,\sigma)$, which can be given by the expression

$$\Pr\left[K \le K(t,\sigma)\right] = 1,$$

assuming that *K* follows Gaussian (Normal) distribution. In practice, we can accept the value of $K(t,\sigma)$ to be equal or greater to 3.90. Consequently, we have obtained an analytic formula for the best choice of volatility σ, which is given by

$$\sigma^* = \frac{t^*}{K(t,\sigma)} = K(t,\sigma)^{-1} t^* \approx 0.256 \cdot t^*. \quad (4.9)$$

Remark 4.3 It is clear from (4.9) that the choice of the optimal σ^* depends on the desired t^* and vice versa. □

Now we are ready to propose the main result of this paper, which can be concluded into the following algorithm.

Algorithm CIZT

(**C**hange **I**n **Z**ero **T**ime)

1st Step: Define the desired

$$\underline{x}\left(t^*\right)=\left[\begin{array}{cccc} x_1\left(t^*\right) & x_2\left(t^*\right) & \cdots & x_n\left(t^*\right)\end{array}\right]^T$$

for the state transmission.

2nd Step: Pre-determined the required time t^* for the state transmission, then using the expression (4.9), the optimal volatility $\sigma^*=0.256\cdot t^*$ (since $K\left(t^*,\sigma^*\right)=3.9$) is given.

3rd Step: Finally, the coefficients $\underline{a}=\left[\begin{array}{cccc} a_o & a_1 & \cdots & a_{n-1}\end{array}\right]^T$ are calculated by (4.5), i.e. $\underline{a}=V^{-1}\hat{\underline{x}}\left(t^*\right)$, where

$$V=V\left(\lambda_1,\lambda_2,...,\lambda_n\right)=\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{n-1} \\ & & \vdots & & \\ 1 & \lambda_n & \lambda_n^2 & \ldots & \lambda_n^{n-1}\end{bmatrix}$$

is the Vandermonde matrix, and

$$\hat{\underline{x}}_i\left(t^*\right)\approx\frac{\underline{x}_i\left(K\left(t^*,\sigma^*\right)\sigma^*\right)}{e^{\lambda_i K\left(t^*,\sigma^*\right)\sigma^*}b_i},\ \text{for}\ i\in\underset{\sim}{n}.$$

Remark 4.4 From the control viewpoint and the type of the application (that the change of the initial conditions required), it is significant to choose an appropriate time for the initial state transition. As we can see in the following example, considering Remark 4.3, the time can take any desired value, but the volatility σ^* should always satisfy the expression (4.9) in order to have an excellent approximation, see also Remark 4.2.

Example 4.1 (See Gupta, 1966) Consider the system

$$\begin{bmatrix} x_1'(t) \\ x_2'(t)\end{bmatrix}=\begin{bmatrix} -2 & 0 \\ 0 & -3\end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t)\end{bmatrix}+\begin{bmatrix} 1 \\ 2\end{bmatrix}u_o(t)$$

where $\underline{x}(t)\in\mathcal{C}^\infty\left(\mathbb{R},\mathcal{M}(2\times1;\mathbb{R})\right)$ and $u_o(t)\in\mathbb{R}$ are the state vector and the input, respectively. The square diagonal matrix $A=\begin{bmatrix} -2 & 0 \\ 0 & -3\end{bmatrix}\in\mathcal{M}(2\times2;\mathbb{R})$, and the input vector $\underline{b}=\begin{bmatrix} 1 \\ 2\end{bmatrix}\in\mathcal{M}(2\times1;\mathbb{R})$ are derived.

For this system, it is desired to change the state from $\underline{x}\left(0^-\right)=\begin{bmatrix} 0 \\ 0\end{bmatrix}$ at time 0^- to $\underline{x}\left(t^*\right)=\begin{bmatrix} 3 \\ 4\end{bmatrix}$ at time $t^*=10^{-6}$ seconds (or 1 microsecond). For this task, we want to design an input (2.3) to achieve this in 1 microsecond.

Here the step of our **CIZT** algorithm should be run.

1st Step: The desired state is $\underline{x}\left(t^*\right)=\begin{bmatrix} 3 \\ 4\end{bmatrix}$.

2nd Step: We have pre-determined the required time $t^*=10^{-6}$, so the optimal volatility $\sigma^*=2.56\cdot10^{-7}$ (since $K\left(t^*,\sigma^*\right)=3.9$).

3rd Step: Then,

$$\hat{\underline{x}}_1\left(10^{-6}\right)\approx3e^{2\cdot10^{-6}}\ \text{and}\ \hat{\underline{x}}_2\left(10^{-6}\right)\approx2e^{3\cdot10^{-6}},$$

The inverse of the Vandermonde matrix is

$$V^{-1}=V^{-1}(-2,-3)=\begin{bmatrix} 1 & -2 \\ 1 & -3\end{bmatrix}^{-1}=\begin{bmatrix} 3 & -2 \\ 1 & -1\end{bmatrix}.$$

Finally, the coefficients $\underline{a}=\begin{bmatrix} a_o \\ a_1\end{bmatrix}$ are calculated, i.e.

$$\underline{a}=\begin{bmatrix} a_o \\ a_1\end{bmatrix}=\begin{bmatrix} 3\hat{\underline{x}}_1\left(10^{-6}\right)-2\hat{\underline{x}}_2\left(10^{-6}\right) \\ \hat{\underline{x}}_1\left(10^{-6}\right)-\hat{\underline{x}}_2\left(10^{-6}\right)\end{bmatrix}$$

$$=\begin{bmatrix} 5.000006 \\ 1.000001\end{bmatrix}\approx\begin{bmatrix} 5 \\ 1\end{bmatrix}.$$

5. CONCLUSIONS

In this paper, a methodology has been proposed for approximating the distributional trajectory that transfers the state of a linear differential system in (almost) zero time by using the impulse-function and its derivatives. Actually, the input vector has to be made as a linear combination of the δ-function of Dirac and its derivatives. However, the approximation is based on the Gaussian (Normal) function. The work has involved the following three distinct problems:

(i) We have started with the impulsive trajectory that transfers the origin to a point in the state space and used this as the central point motivating the need to approximate distributions by smooth functions.

(ii) After that, we have examined the family of Gaussian functions, which may be used to approximate distributions and we have defined an appropriate Euclidean metric to measure how good the approximation is and investigates the link of the σ parameter of Gauss functions to the time and inevitably to the distance from the desirable initial state.

(iii) We have pre-determined the minimal time required for achieving a solution to the above standard controllability problem in terms of approximations to the distributional solutions, by using Gaussian families for the approximation. Finally, the CIZT algorithm has been proposed for the calculation of the coefficients of our input function.

As further research, of special interest is the link of approximation to the energy and time requirements for the

transfer of the origin to a point within the R-sphere when the approximations to the distributional solutions is carried out. Such problems can be examined under restrictions on the energy of the input signal and we can qualify the links of the approximation on the energy and time requirements for the control signal. Clearly similar problems can be defined for the dual problem of reconstructibility.

REFERENCES

Bowen, J.M. (1994) Delta function terms arising from classical point source fields, *Am. J. Phys.* Vol. 62, pp. 511-515.

Boykin, T.B. (2003) Derivatives of the Dirac delta function by explicit construction of sequences, *Am. J. Phys.* Vol. 72, pp. 462-468.

Campbell, S.L. (1980) *Singular systems of differential equations*. I. Research Notes in Mathematics 40. Pitman, Boston.

Campbell, S.L. (1982) *Singular systems of differential equations*. II. Research Notes in Mathematics 61. Pitman, Boston.

Cobb, J.D. (1982) On the Solutions of Linear Differential Equations with Singular Coefficients, *Journal of Differential Equations*, Vol. 46, No. 3, pp. 310-323.

Cobb, J.D. (1983) Descriptor Variable Systems and Optimal State Regulation, *IEEE Transactions on Automatic Control*, Vol. 28, No. 5, pp. 601-611.

Estrada, R. and Kanwal, R.P. (2000) *Singular integral equations*, Birkhauser, Boston.

Gupta, S.C. and Hasdorff, L. (1963), Changing the state of a linear system by use of normal function and its derivatives, *International Journal of Electronics* Vol. 14, pp. 351-359.

Gupta, S.C. (1966) *Transform and state variable methods in linear systems*, Wiley New York, U.S.A.

Jaffe, S. and Karcanias, N. (1981), Matrix pencil characterization of almost (A, B)-invariant subspaces: a classification of geometric concepts. *International Journal of Control* Vol. 33, pp. 51-93.

Kanwal, R.P. (2004) *Generalized Functions: Theory and applications*, Birkäuser, 3rd edition, U.S.A..

Karcanias, N. and Kouvaritakis, B. (1979), Zero time adjustment of initial conditions and its relationship to controllability subspaces, *International Journal of Control* Vol. 29 (5), pp. 749-765.

Karcanias, N. and Hayton, G.E. (1982), Generalised autonomous dynamical systems, algebraic duality and geometric theory. *Proceedings of 8th IFAC World Congress*, Kyoto, Japan, Pergamon Press, pp. 289-294.

Karcanias, N. and Kalogeropoulos, G.I. (1989), Geometric Theory and Feedback Invariants of Generalized Linear Systems: A Matrix Pencil Approach, *Circuits, Systems and Signal Processing*, Special issue on Singular Systems Vol. 8 (3), pp. 395-397.

Verghese, G.C. (1979), *Infinite Frequency Behaviour of Generalised Dynamical Systems*, PhD Thesis, Stanford University, 1979.

Verghese, G.C. and Kailath, T. (1979), Impulsive Behaviour and Dynamical Systems, *Proceedings of 4th Symposium on Mathematical Theory of Networks and Systems*, Delft, The Netherlands, pp162-168.

Willems, J.C. (1981). Almost Invariant spaces: An Approach to High Gain Feedback-Part I: Almost controlled invariant subspaces, *IEEE Transactions Automatic Control*, Vol. AC-26, pp. 235-252.

Willems, J.C. (1991), Paradigms and Puzzles in the Theory of Dynamical Systems, *IEEE Transactions on Automatic Control*, Vol. 36, pp 259-294.

Zadeh, Z.A. and Desoer, C.A. (1963), Linear system theory, the state space approach, McGraw-Hill, New York, USA.

Zemanian, A.H. (1987), Distribution theory and transform analysis: An introduction to generalized functions with applications, Dover Publications, Inc, New York, USA.