BU-1398-M

Integral Transform Estimation and Compartmental Models

running title: **Transform Estimation and Compartmental Models**

MARTHA CONTRERAS

Biometrics Unit, Cornell University, Ithaca, NY 14853, USA. Research supported by New York State hatch project NYC-151413. Tech. Rep. BU-1398-M. Author to whom correspondence should be addressed. e-mail: mpc14@cornell.edu

CHANGMEI LIU

Biometrics Unit, Cornell University, Ithaca, NY 14853, USA. Research supported by NSF grant DEB-925370 (Presidential Faculty Fellowship Award) to Carlos Castillo-Chavez in Cornell University. e-mail: cl103@cornell.edu.

GEORGE CASELLA

Biometrics Unit, Cornell University, Ithaca, NY 14853, USA. Research supported by NSF No. DMS-9625440. e-mail: gc15@cornell.edu.

LOUISE M. RYAN

Harvard School of Public Health, Department of Biostatistics and Dana Farber Cancer Institute, Boston, MA 02115, USA. e-mail: ryan@jimmy.harvard.edu

AND

CHARLES VAN LOAN

Department of Computer Science, Cornell University, Ithaca, NY 14853, USA. e-mail: cv@cs.cornell.edu.

Abstract

As an alternative to standard estimation methodology, such as ordinary least squares or weighted least squares, we investigate tractable estimation methodology involving the use of integral transforms, particularly the use of the Laplace transform. To this, we establish theoretical results in transform estimation, and point out the numerically and analytically desirable properties of transform estimation, which are not directly available in standard least squares estimation, when dealing with implicitly defined functions such as those arising from a convolution or as the solution to a system of differential equations. Moreover, we address the choice of the variable of the transformation or the kernel of the transform in theory and make some recommendations for the practical selection of such variable. As an example of how transform estimation could be efficiently and meaningfully implemented, we propose convergent algorithms for estimating the flow rates or the parameters, from a discrete set of measurements, in linear time invariant compartmental models.

1 Introduction

Integral transform estimation has been widely used in the engineering and statistical literature, see for instance [1, 2, 3, 4, 5, 6, 7] and more recently [8] employed the use of the Laplace transform for estimation when dealing with a linear time invariant compartmental system. Most of these authors used the Laplace transform to obtain a closed form expressions to the convolution operator or to the solution of a differential equation. However, most authors also acknowledged the arbitrariness in estimation due to the infinite freedom of choosing the variable in the kernel of the Laplace transform.

Because it motivates our approach to integral transform estimation and that of [8], we provide some detail of the work of [1, pp. 86-89]. They dealt with the inverse problem of estimating the coefficients of a second order differential equation from a discrete set of observations. They proposed the used of the Laplace transform to

obtain a closed form expression of the implicitly defined model function and in their numerical simulation, they obtained coefficient estimates which were numerically the same as if they had not applied their transform methodology, but rather had approached the problem via least squares. However, it appears that [8] are the first in employing the use of the Laplace transform for estimation when dealing with a linear time invariant compartmental system.

This paper consists of two main sections. Since they serve as an excellent source for the practicality of transform estimation, Section 2 gives a brief review of general compartmental models and the Laplace transform. However, we first review standard estimation methodology involving compartmental models and its computational demands, then we introduce transform estimation as a general methodology for implicitly defined functions. To this, we obtain results in the theory of transform estimation and established convergence for a proposed algorithm involving the Laplace transform in compartmental models. Section 3 focuses on the numerical implementation of transform estimation in compartmental models. More specifically, we make some practical recommendations for the appropriate choice for the variable of the transformation in the Laplace kernel and establish general convergence results of our proposed algorithms. Lastly, we conclude with remarks about the numerical and analytical utility of Laplace transform estimation and point out other applications of transform estimation.

2 Compartmental Models

Compartmental systems have a wide range of applicability particularly in biomedical engineering where they are used to model the kinetics of distribution of materials through the organism, pharmacology, chemical kinetics, and in diagnosis and therapy such as clinical pharmacology and pharmacotherapy. Thus, the problem of estimating the flow rates and subsequently the "concentration" curve in any one of the compartments from a discrete set of data, is of considerable importance.

Formally, by a compartmental system we mean a biological or a physical system which is made up of a finite number subsystems, called compartments or states, each of which is homogeneous and well-mixed, and the compartments interact by exchanging materials. There may be inputs from the environment into one or more compartments, and there may be outputs or excretions from one or more compartments into the environment, see for instance [9]. Nonetheless, the dynamics of a general compartmental system are modeled by the following

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad t \geq 0 \\ x(0) &= 0 \\ y(t) &= Cx(t), \end{aligned} \tag{2.1}$$

where the ith equation is given by

$$\dot{x}_i = \sum_{j=1}^{n} a_{ij}x_j + \sum_{k=1}^{r} b_{ik}u_k(t), \quad i = 1, ..., n.$$

In the terminology of control theory, x, u, and y are referred to as the state/response function, input, and output vectors, respectively. However, for our purposes, we will refer to the system given by (2.1) as a compartmental system where $A = [a_{ij}]$ is the $n \times n$ compartmental matrix representing the interaction between compartments whose entries are typically unknown and must be determined from a discrete set of measurements. The entries represent linear combinations of the flow/transfer rates from compartment j to compartment i, where compartment j is denoted by x_j, the jth component of the vector $x(t)$, and where $\dot{x}(t)$ is the rate of change of the response function with respect to time. In general, the entries of A may depend on the state $x(t)$, on time t, and on a vector of unknown scalar parameters. However, for the remainder of this paper we will assume that A consists solely of unknown scalars so that (2.1) is then a *linear time invariant compartmental system.*

To determine these entries or parameters, an experiment is conducted in which r inputs excite the compartments thus causing them to interact with one another. The r inputs are regarded as the transposed column vector, $u(t) = (u_1(t), u_2(t), ..., u_r(t))^{\mathrm{T}}$,

where $u(t)$ is the input or forcing function. The paths by which the r inputs enter the n compartments is represented by a $n \times r$ matrix $B = [b_{ik}]$, called the input distribution matrix where entry b_{ik} is positive if input $u_k(t)$ enters compartment i, and zero otherwise. Since it might not be possible or practical to observe each individual state by itself, we define a $p \times n$ matrix C, known as the output connection matrix, representing the paths from compartments to sampling devices where entry c_{ij} is positive if compartment j influences output function component y_i; otherwise $c_{ij} = 0$. Hence, we see that $y(t) = (y_1(t), y_2(t), ..., y_p(t))^T$.

Due to conservation of mass or from the mass balance equations, it can be established that the compartmental matrix, A, has non-negative off diagonal elements, non-positive diagonal elements, and its column sums are non-positive, see [10]. This implies that the matrix is diagonally column dominant and that the eigenvalues of A have a non-positive real part and that none are purely imaginary. The former follows from Gerschgorin's theorem which can be found in [10, 11], for instance. Furthermore, these matrices need not be symmetric or normal since the entries correspond to physical flow rates from one state to another. This property of A will play a critical role in the estimation process, nonetheless. (These matrices, actually their transpose negated, belong to the more general class of matrices known as M-Matrices, [12]).

Through the use of integrating factors, the solution to (2.1), or the input-output relation, can be deduced to be

$$\begin{aligned} y(t) = Cx(t) &= C \int_0^t e^{(t-\tau)A} Bu(\tau) d\tau \\ &:= Ce^{tA} * Bu(t), \end{aligned} \tag{2.2}$$

where $*$ denotes the convolution operator.

Prior to stating our algorithm and because it will motivate our approach to estimation, we review the standard method of estimation involving compartmental models.

2.1 Standard Estimation

There is a large literature pertaining to the estimation and identifiability of a compartmental system from what is referred to as input-output experiments, see [10, 13] for example. However, estimation of the flow rates is not a trivial matter numerically since it involves the specification and coding of a highly nonlinear model function. In [14], they attempt to address this problem by proposing a method of estimation which does not involve the encoding of partial derivatives in their optimization scheme. However, the success of their methodology is still dependent on the accurate approximation of the matrix exponential. In what follows, we briefly review standard estimation methodology and its computational demands.

Suppose the outputs are measured at discrete times $t_1 < t_2 < ... < t_m$. Then we introduce the $p \times 1$ vectors y^i, $i = 1, ..., m$, where y^i is a set of measurements gathered in all of the p compartments at time t_i, that is an approximation to the true outputs that satisfies

$$y^i = y(t_i) + \epsilon_i, \tag{2.3}$$

where it is usual to assume that the statistical expected value of the $p \times 1$ error vector ϵ is zero, given the true value of the parameters at each fixed sampling time t_i, and that the errors have finite statistical variance as well. However, we will make no further mention of the statistical properties of the estimator(s) in this paper.

We suppose that the response function, $y(t)$, is approximated by these discrete set of measurements y^i at times t_i. Then under the assumption that the entries of A are uniquely, either locally or globally, least squares identifiable, they are estimated from the following model

$$y^i = C \int_0^{t_i} e^{A(t_i - \tau)} Bu(\tau) d\tau + \epsilon_i, \quad i = 1, ..., m.$$

That is, we seek the local or global minimizer of the constrained nonlinear least squares problem,

$$\begin{aligned} &\min_{a_{kj}} \sum_{i=1}^{m} ||y(t_i) - y^i||^2; \quad k, j = 1, ..., n, \\ &\text{subject to} \\ &a_{kj} \geq 0, \quad k \neq j, \text{ and} \\ &a_{kk} \leq -\sum_{j=1}^{n} a_{jk}, \quad k \neq j, \end{aligned} \tag{2.4}$$

where $||F(\cdot)||^2 = F(\cdot)^T F(\cdot)$ and F denotes the $p \times 1$ nonlinear vector-valued function.

Some facts about (2.4) are in order. The constraints are needed to insure a meaningful solution to (2.4); that is, it can be shown, see [15], that the constraints imply that the solution to the compartmental system is nonnegative and stable for all time, t. It can also be shown that a sufficient condition for the local identifiability of (2.4) is that the Jacobian of the objective function is of full rank [16]. The proof relies on showing that under the proper restrictions on the compartmental matrix, forcing function, and input matrix, the sequence of iterates from Newton's method is bounded. This sequence will be a uniformly bounded sequence in a separable Hilbert space from which it follows, see [17], that we can extract a weakly convergent subsequence which converges to a solution of (2.4). However, if the Hilbert space is finite dimensional, as it is in this case, then the convergence will be strong so that we have a strong solution to (2.4).

However, estimating the transfer rates or parameters of a compartmental system from data is an inverse problem which is typically ill-posed. That is, identifiability and continuous dependence on the data are concerns. For these reasons, to insure continuous dependence on the data, (2.4) is usually regularized in the sense of Tychonoff, see for instance [18]. Regularization is then accomplished in two ways: either (2.4) is linearized and then regularized, as is the case if one employs the Levenberg-Marquardt algorithm, or the objective function in (2.4) is regularized and then linearized as is the case when one employs a penalized least squares method, see [18].

Regardless of which regularization procedure one employs, the accurate solvability of (2.4) is dependent on the accuracy of the numerical approximation of the matrix exponential. The numerical approximation of the matrix exponential could be a sensitive issue since compartmental matrices are usually not symmetric or normal, thus their spectral decomposition need not exist. Furthermore, even if the spectral decomposition exists, the eigenvectors could be numerically linearly dependent and thus the matrix of eigenvectors could be ill-conditioned. Nonetheless, this may or may not be an issue since the Padé approximation algorithm is believed to numerically only fail to produce a relatively accurate e^{At} only when the matrix exponential condition number, $\nu(A,t)$ for $t=1$ is correspondingly large [11, page 576]. Precisely what restrictions this places on the compartmental system is unclear. We thus propose a method of estimation involving the Laplace transform, which does not directly depend on the numerical approximation of the matrix exponential, e^{At}, and has the interesting property that it uses the structure of A, namely the diagonal dominance. We first introduce the general principle behind transform estimation.

2.2 Transform Estimation

We introduce transform estimation mostly for the cases when a closed form expression of a function, like $x(t)$ given by (2.2), is not known or is impractical/impossible to determine explicitly. Thus, one is forced to rely on numerical approximations to such functions. This will be the case whenever the model function involves a convolution or is the solution to a system of differential equations. In a later section, we demonstrate the practicality of transform estimation by considering the linear time invariant compartmental systems as an example, but for now we give the idea behind transform estimation.

Suppose that we are given functions $h(t,\theta)$, θ fixed but belonging to some compact space $\Theta \subset R^N$, and $g(t)$ (which may or may not depend on θ) both in some finite dimensional vector space $W(\Omega) \subset L^2(\Omega)$. If there exists a $\theta \in \Theta$ such that the

following prevails

$$g(t) = h(t, \theta), \text{ a.e. } t \in \Omega, \tag{2.5}$$

or equivalently the following prevails

$$\int_{\Omega} (g(t) - h(t, \theta)) w(t, s) dt = 0, \forall w \in W, \text{ and } s \in \mathcal{O} \subset \Im, \tag{2.6}$$

where $\Im$ denotes the complex plane, then we say that θ yields a *strong estimator* of $g(t)$ in the sense given in the following definition.

Definition 2.1 Let $\tilde{W}(\Omega \times \mathcal{O}) = \text{span}\{w_i\}_{i=1}^{\aleph(\tilde{W}_\theta)}$, where $\aleph(\tilde{W}_\theta)$ is the dimension of the space $\tilde{W} \subset L_2(\Omega \times \mathcal{O})$. If there exist a $\theta \in \Theta$, where Θ is a compact subset of R^N, such that for given functions $h(t, \theta)$ and $g(t)$ in W relationship (2.6) prevails, then we say that θ yields a *strong estimator* of $g(t)$.

Note, however, that (2.6), in light of Definition 2.1 can be relaxed to read, only for all w in the set of basis elements of W. From this, we can still conclude that (2.6) implies (2.5) (this is even more obvious if one uses an orthogonal basis).

However, a more relaxed than Definition 2.1 is necessary since (2.6) is an integral equation of the first kind and, as such, could be ill-posed. That is, there might not exist a θ such that (2.6) holds, see for instance [18]. We thus make the following definition.

Definition 2.2 The estimator $\theta \in \Theta$ is a *transformed least squares* estimator if it solves

$$\min_{\theta} \int_{\mathcal{O}} [\int_{\Omega} (g(t) - h(t, \theta)) w(t, s) dt]^2 ds, \forall w \in \tilde{W}, \tag{2.7}$$

where $\tilde{W}$, $\mathcal{O}$, $h(t, \theta)$, and $g(t)$ are as given in Definition 2.1.

Furthermore, if $||g - h||_{L^2} \leq \epsilon$, ϵ small at the computed solution of (2.7), say θ, for all elements in Θ, then it certainly follows that θ is a *least squares estimator* of $g(t)$ in the ordinary least squares sense.

In practice, one would not directly insists that $||g - h||$ be small at θ for all elements of Θ since this would be too costly and impractical to implement directly.

However, assuming the existence of a solution to (2.7), this will happen as a consequence of the Implicit function theorem, see [19], as the following claim states.

Claim 2.1 Suppose $W_\theta(\Omega)$, $g(t)$ and $h(t,\theta)$ are as given in Definition 2.1, and that there exists a $\theta \in \Theta \subset R^N$, such that (2.7) holds. Furthermore, suppose that the residual or the value of the objective function, $g-h$, at θ is small so that (2.6) holds with relatively small error, then it follows that $||g-h||_{L_2}$ must also be small at θ.

Proof: Since h and g are in $\tilde{W} = \text{span}\{w_i\}_{i=1}^{\aleph(\tilde{W})}$, where $\aleph(\tilde{W})$ is as given in Definition 2.1, then it follows that we may express both h and g as linear combinations of the basis elements w_i. So that the L_2 inner product $(g-h, w_i)_{L_2}$ being small for each i, by the Implicit function theorem, it is possible to solve for each of the coefficients in terms of the others so that one gets a polynomial involving the norm of the basis elements (this is more obvious if one uses Gram-Schmidt to orthogonalize the basis elements). It follows that $||g-h||_{L_2}$ must be small at θ. Thus our claim is established.

2.3 Generic algorithm

Motivated by Definition 2.2 and Claim 2.1, we propose that transform estimation be implemented as follows.

Suppose that we are given a discrete set of data, m_i, $i = 1, ..., n$, such that

$$m_i = h(t_i, \theta) + \epsilon_i, \tag{2.8}$$

where ϵ is the error vector which is typically assumed to have statistical expected value equaled to zero and finite variance. Then, the following are guidelines for applying transform estimation.

1. A priori fit some $g(t)$ to the data $\{m_i\}$, (the choice of fitting function is be best decided upon the given problem); nonetheless, the expectation is that in principle the agreement between $g(t)$ and $h(t,\theta)$, for θ fixed but unknown true

value, should be good in the least squares sense. That is,

$$||g(t) - h(t,\theta)|| \leq \epsilon,$$

We thus, for ease of illustration, define a function η such that

$$g(t) - h(t,\theta) = \eta(t,\theta). \tag{2.9}$$

2. We then multiply both side of (2.9) by $w(t,s)$ and integrate over Ω so that

$$\int_\Omega (g(t) - h(t,\theta))w(t,s)dt = \int_\Omega \eta(t,\theta)w(t,s)dt. \tag{2.10}$$

3. In practice, however, due to the presence of errors in the estimation of $g(t)$ or $h(t,\theta)$, we insist that (2.10) hold in the least squares sense. Moreover, since W is finite dimensional, we propose that the following modification to Definition 2.2 be employed.

   That is, suppose that there exists a sequence s_j, $j = 1, ..., \aleph(\tilde{W}(\Omega))$ such that $w(t,s_j)$ forms a basis for W, we then solve the following discretized version of (2.7)

$$\min_\theta \sum_{s_j} (\int_{\Omega_t} (g(t) - h(t,\theta))w(t,s_j)dt)^2 \Delta s_j, \tag{2.11}$$

   where Δs_j are appropriately chosen weights for the discretization and where the choice of $w(t,s)$ or the *kernel of the transformation,* is made based on the given problem whenever feasible.

The steps of the above algorithm will be illustrated with the Laplace transform applied to linear time invariant compartmental systems. Moreover, in Section 3, we will investigate relaxing the assumption that $w(t,s_j)$ be a basis set of $\tilde{W}$. For now, we present preliminaries about the Laplace transform and compartmental matrices.

2.4 The Laplace transform and compartmental models

In this section we introduce the Laplace transform of a function and briefly discuss some of its desirable features, namely its smoothing effects, its continuity, and the

fact that it turns the convolution operator, involving the matrix exponential, into an algebraic expression. For completeness, the details are provided in the Appendix. For now, however, we define the Laplace transform of a function.

We say that a function $u(t)$ is Laplace transformable if the following integral exists

$$\hat{u}(s) := \int_0^\infty e^{-st} u(t) dt, \tag{2.12}$$

and is uniformly and absolutely convergent for $Re(s) > b$, where we denote the Laplace transform of $u(t)$ by $\hat{u}(s)$. In the Appendix, we discuss the following useful facts about the Laplace transform.

Laplace transform is known to have the effect of smoothing functions with rapid and oscillatory initial growth and to be continuous from $L_2(0, \infty)$ to $L_2[\alpha, \beta]$, $\alpha \neq 0$, where $\Omega_2 = [\alpha, \beta]$ (see (5.14)). Also, its inverse L^{-1} will exist and be stable provided we restrict ourselves to smooth functions that are defined in $L_2[\alpha, \beta]$ or to smooth functions which are supported on the interval $[\alpha, \beta]$. Lastly, it is also known that the Laplace transform of a convolution is the product of the Laplace transform of each of the functions provided the Laplace transform of each of the functions in question exists. For proofs of any of the above statements, the reader is referred to [1, 20] and the Appendix.

Applying the latter statement about the Laplace transform to (2.1), which has as solution the convolution given by (2.2), we have that the Laplace transform of (2.2) is

$$\hat{y}(s) = C(sI - A)^{-1} B \hat{u}(s). \tag{2.13}$$

Note that (2.13) is well-defined provided that the forcing function $u(t)$ is Laplace transformable and provided that the determinant of $(sI - A)$ is not zero for all $Re(s) > 0$.

Claim 2.2 $(sI - A)$ is nonsingular for $Re(s) > 0$ when A is a compartmental matrix.

Proof: See Appendix.

An interesting observation, and one which we will exploit, is the fact that the matrix $(sI - A)$ remains diagonally column dominant since A is a compartmental matrix and $Re(s) > 0$. We are now ready to specify the approach to transform estimation when dealing with the compartmental system given by (2.1).

2.5 Transform estimation for compartmental models

In this section, we proceed to set up the analogue of (2.4) but in terms of the Laplace transform. To do this, we follow steps **1-3** stated in Section 2.2 and define the $p \times m$ matrix

$$r^l := y_l(t_i), \text{ for each } l = 1, ..., p \text{ and for all } i = 1, ..., m, \tag{2.14}$$

where for each fixed l, we have a vector corresponding to the m measurements made in the lth compartment at times t_i. Then for each l and all m in (2.14), we fit a curve, $r^l(t)$, to the data, which is specified by the user and that is Laplace transformable. Nonetheless, for reasons discussed in the Appendix, section 5.1, it is expected that the fit between $r^l(t)$ and the response function $y(t)$ is a good one given the true compartmental matrix, A. That is, we suppose that in principle

$$r^l(t) - y(t) \approx 0, \text{ in } L_2 - \text{norm},$$

at the unknown but true compartmental matrix A.

More generally, if we let $r^l(t) - y(t) = \eta(t)$ and proceeding along step **2** from Section 2.2 , where $w(t, s)$ is now the kernel of the Laplace transform e^{-st}, and from the definition of the Laplace transform given by (2.12) and upon rewriting (2.10) using this new notation, we seek to estimate A in the least squares sense from the model

$$\hat{g}(s) = \hat{y}(s) + \hat{\eta}(s), \tag{2.15}$$

where $\hat{\eta}(s)$ is the error function and where for notational convenience we define the $p \times 1$ vector $\hat{g}(s) := \hat{r}^l(s)$, $l = 1, ..., p$, such that $\hat{g}(s)$ is the Laplace transform of $r^l(t)$ which was obtained from fitting, without loss of generality, in the L_2–norm

sense, the vector r^l in (2.14) for each l to the data given by y^i, $i = 1, ..., m$. Likewise $\hat{y}(s) := \hat{y}_l(s)$, $l = 1, ..., p$ is the Laplace transform of $y(t)$.

Naturally, $r^l(t)$ should be chosen to reflect the true nature of problem (2.1) but subject to being Laplace transformable. Furthermore, under the assumption that the true curve is smooth enough and well approximated by polynomials, we expect that the agreement between g and y be good if the agreement between the corresponding Laplace transforms, $\hat{g}$ and $\hat{y}$ is a good one, see [1, 20] and (5.14). More precisely, we expect that when

$$||g(t) - y(t)||_{L_2(0,\infty)} \leq \epsilon, \epsilon \text{ small}, \tag{2.16}$$

whenever

$$||\hat{g}(s) - \hat{y}(s)||_{L_2(\alpha,\beta)} \leq \epsilon, \quad 0 < \alpha < \beta < \infty \text{ and } \epsilon \text{ small}. \tag{2.17}$$

Armed with the above statement and following step **3**, we solve for a given sequence $\{s_q\}_{q=1}^{\bar{q}}$ in the kernel of the Laplace transform

$$\begin{aligned} \min_{a_{kj}} \quad & \sum_{q=1}^{\bar{q}} ||\hat{g}(s_q) - \hat{y}(s_q)||^2 \\ & \text{subject to} \\ & a_{kj} \geq 0, \quad k \neq j, \text{ and} \\ & a_{kk} \leq -\sum_{j=1}^{n} a_{jk}, \quad k \neq j, \end{aligned} \tag{2.18}$$

where the constraints insure that the solution is nonnegative and stable for all time, t, see [15], since it can readily be established that if the Laplace transform $\hat{f} \geq 0$, for $s > 0$, then it must be that $f \geq 0$, for $t > 0$, and vice-versa.

The following claim, which is an immediate consequence of Claim 2.1, provides the ground basis for transform estimation when dealing with a compartmental system consisting of a forcing function which is a "bolus/one-hit" input into the system and the eigenvalues of the compartmental matrix are all real, nonzero, and distinct. However, to establish such a claim, we need for $g(t)$ to depend on the parameters we wish to estimate, namely the entries of the compartmental matrix A. Relaxing

this rather numerically undesirable requirement is the essence of Section 3, but for now we retain it.

Claim 2.3 Let the $r \times 1$ forcing function $u(t)$ consist of the Dirac delta function, $\delta(t)$, and let the compartmental matrix A be invertible with real and distinct eigenvalues. Further, suppose the sequence of $\{s_q\}_{q=1}^{\tilde{q}}$, $\tilde{q} = 1, ..., n$ is chosen such that

$$s_q = \alpha_q, \tag{2.19}$$

where $\alpha_q = |\lambda_q|$ and λ_q corresponds to the eigenvalues of A, and suppose the residual of the objective function, (2.18), at the true value of the compartmental matrix, A, is small so that (2.6) holds with small error. Then it follows that $||g - y||_2$ must also be small at A.

Proof: The proof follows immediately from Claim 2.1 once we make the observation that under the conditions of the claim it follows from (2.2) that the general i^{th} entry of the solution to the system given by (2.1) is $x(t) = e^{At}e_j$. Furthermore, to the data we are fitting the parametric function $g(t) = C\sum_{q=1}^{n} e^{-s_q t}\tilde{B}$, where $\tilde{B}$ is a $n \times 1$ vector of unknown scalars. From this we see that both $y(t)$ and $g(t)$ span the same n-dimensional subspace of L_2 whose basis elements are $e^{-s_q t}$ for $q = 1, ..., n$ which are bounded functions of t. Thus if (2.6) holds with small error, then it follows that $||g - y||_{L_2}$ must also be small at A from Claim 2.1. Thus the claim is established.

In the Appendix, section 5.4, we relax some of the assumptions of Claim 2.3 and present a numerical approach for a general compartmental matrix and forcing function; however, the theoretical results of this section need not hold.

Our next theorem states sufficient conditions under which an algorithm involving a parametric form of $g(t)$, such as (5.23), converges to the unconstrained optimizer of (2.18); naturally, provided we start within a neighborhood of the solution.

Theorem 2.1 Let $A_c := a_c$ be within a neighborhood, D, of the true compartmental matrix $A := a$, which we assume to be invertible and to have distinct eigenvalues. Moreover, let $Bu(t) = e_o\delta(t)$, where e_o is the natural basis element with a 1 in the

o^{th} position and zero else where. For an arbitrary sequence of scalars $\{s_q\}_{q=1}^{\tilde{q}} > 0$, where $\tilde{q}$ is some a priori decision about the total number of scalars, if the function $g(t)$ is a sum of exponentials corresponding to the actual eigenvalues of a_c and if there exists a $\sigma \geq 0$ such that $\sigma < \lambda$ where λ is the smallest eigenvalue of the symmetry matrix $J^T(a)J(a)$, where $J(a)$ is the Jacobian of the objective function, then the sequence of iterates in the Levenberg-Marquardt algorithm converges q-linearly to the true A in L_2-norm.

Proof: See Appendix.

We point out that fitting a parametric function, such as (5.23), per iteration is a costly and potentially sensitive numerical procedure that should cautiously be implemented. Nonetheless, it is of theoretical value as Claim 2.3 and Theorem 2.1 illustrate.

In the next section, we look at practical modifications of transform estimation. In particular, we note that if one fits a non-parametric function to the data, then not only is the Jacobian of the objective function simpler (see Appendix), but it is also possible to obtain a more general convergence result than Theorem 2.1 provided the solution to the least squares problem exists and that we meet the hypothesis of the Levenberg-Marquardt algorithm, see [21].

3 Implementation of transform estimation

The continuity of the Laplace transform and the possible numerical stability of its inverse when restricted to smooth functions, see section 5.1, suggest that a method of estimation involving the Laplace transform applied to this class of functions be considered. Thus, in this section, we advocate selecting the kernel of the Laplace operator rather arbitrarily instead of actually computing eigenvalues. Thus, the Jacobian of this particular approach will not be as computationally demanding since we will recommend a one time, non-parametric, fit to the data. The convergence results of the previous section will follow but for a larger class of functions than

those of Theorem 2.1. However, before proceeding, we need the following broader definition of an estimator.

Definition 3.1 We say that $\theta \in \Theta$ is an *weak least squares* estimator if it solves

$$\min_{\theta} \int_{\mathcal{O}} [\int_{\Omega} (g(t) - h(t,\theta))w(t,s)dt]^2 ds, \text{ for some } w \in \tilde{W}, \tag{3.1}$$

where $\tilde{W}$ is an arbitrary subset of $L_2(\Omega \times \mathcal{O})$, and we emphasize that $g(t)$ now does not depend on θ.

However, (3.1) is an infinite dimensional problem, so that a more practical definition of $\theta \in \Theta$ being a *weak least squares estimator* is that it solve the following discretized version of (3.1)

$$\min_{\theta} \sum_{s_q > 0} [\int_{\Omega} (g(t) - h(t,\theta))w(t,s_q)dt]^2 \Delta s_q, \tag{3.2}$$

for some positive sequence of scalars $\{s_q\}_{q=1}^{\tilde{q}}$ and some sequence of weight functions $w(t, s_q) \in W \subset L_2(\Omega)$. Naturally, in both (3.1) and (3.2), it is assumed that the choice of non-parametric function, $g(t)$ describes the data well in the least squares sense.

Criteria (3.1) and (3.2) can be motivated by the continuity of the Laplace transform, which was established in (5.14). Moreover, if agreement between the data and the fitted function, $g(t)$, is reasonable, that is, if

$$||y^i - g(t)|| < \epsilon, \quad \epsilon \text{ a small scalar},$$

then, it is expected that the following prevail at the true yet unknown compartmental matrix, A.

$$||g(t) - y(t)||_{L_2(0,\infty)} \leq \epsilon. \tag{3.3}$$

Hence, we have that

$$||\hat{g}(s) - \hat{y}(s)||_{L_2(\alpha,\beta)} \leq \epsilon, \quad 0 < \alpha < \beta < \infty. \tag{3.4}$$

In practice, some educated choice(s) of the sequence $\{s_q\}$ and of $w(t, s_q)$ must be made according to the problem at hand. As we will see in this section, for linear

time invariant compartmental models, the algorithmic convergence results from the previous section still hold, but the estimator is now a weak least squares estimator in the sense defined here. Nonetheless, given the lack of numerical simulations or the analysis of a real data set, we state the guidelines recommended by [1, 20] for the selection of the variable of the transformation. The details of this argument can be found in the Appendix, section 5.3.

Workers [1, 20] argue that unless there is good reason for doing otherwise one should choose $s_q \geq 1$ for each q. Thus, when no other information is available, the sequence $\{s_q\}_{q=1}^{\tilde{q}}$ should be chosen to be real and greater than or equaled to 1. In fact, [1] suggest that $\{s_q\}$ be chosen to be the sequence $1, 2, 3, ..., \tilde{q}$ when they dealt with the inverse problem described by a linear second order ordinary differential equation. At this stage, we are not prepared to recommend otherwise.

In the next section, we give a possible implementation of the algorithm for solving (2.18) regards of whether the fitted function $g(t)$ depends on the parameters or not.

3.1 Algorithm for compartmental models

The algorithm for solving (2.18) could be implemented as follows.

1. Given a current guess of the minimizing compartmental matrix a_c, which consists of the entries of the $n \times n$ compartmental matrix A_c strung along as an $n^2 \times 1$ vector, and a positive sequence $\{s_q\}$, we solve for Z the following linear system of equations.

$$(s_q I - A)Z = B\hat{u}(s_q); \tag{3.5}$$

   so that the result is stored in Z which implicitly depends on s_q. This allows us to define $\hat{y}(s_q) := CZ(s_q)$ without ever forming the inverse of $(s_q I - A)$.

2. Create $\hat{g}(s_q)$ as indicated in the arguments from (2.14) through (2.15). Note that this step and step **1** (above) involve independent operations and as such could be performed in parallel. This property is particularly attractive if one chooses to fit a parametric curve $g(t)$ to the data at every iteration.

3. Update a_c. That is, the new iterative, a_+, of the unconstrained objective function could come as the solution to the Levenberg-Marquardt update, see [21],

$$a_+ = a_c - [J(a_c)^T J(a_c) + \mu_c I]^{-1} J^T(a_c) R(a_c),$$

where $\mu_c \geq 0$ is the regularization parameter, $J(a_c)$ is the $(\tilde{q}+p) \times n^2$ Jacobian of the $(\tilde{q}+p) \times 1$ vector-valued function $R(a_c) := \hat{g}(s_q) - \hat{y}(s_q)$.

4. Return to step 1 with a_+ and possibly generate a new sequence of $\{s_q\}_{q=1}^{\tilde{q}}$, depending on whether one wishes to use eigenvalue information or some other choice.

5. The algorithm terminates when some established criterion has been met, such as when the relative gradient of the objective function is closed enough to zero or when the model functions are sufficiently closed to the data in some sense (usually in the L_2-norm sense).

For the above algorithm, we have the following convergence result, which includes a wider class of functions than those of Theorem 2.1, provided the solution exists.

Theorem 3.1 In either (2.4) or (2.18), suppose that the objective functions and their Jacobian are in C^∞ of the parameter space, that the Jacobians are in $L_2(\alpha, \beta)$, $0 < \alpha < \beta < \infty$, for each fixed invertible compartmental matrix A, and that the least squares minimizer of either (2.4) or (2.18) exists in the interior of the feasible regions. Then, provided there exists a $\sigma \geq 0$ such that $\sigma < \lambda$ (where λ is the smallest eigenvalue of the symmetry matrix $J^T(a)J(a)$ and where $J(a)$ is the Jacobian of either of the objective functions) within a neighborhood of the solution, the Levenberg-Marquardt algorithm converges q-linearly.

Proof: The conclusion readily follows from the conditions for convergence of the Levenberg-Marquardt algorithm since it can be seen that both the exponential matrix and the Laplace of the exponential matrix are in C^∞ of the parameter space for $s > 0$. Moreover, since we are dealing with compartmental matrices, both the

objective functions and their respective Jacobians are bounded in L_2 of compact sets of (α, β), $0 < \alpha < \beta < \infty$. So that the Jacobian of the objective functions are Lipschitz continuous. Thus our theorem is established.

4 Discussion

In this paper, we introduced transform estimation as an alternative approach to standard estimation methodology to situations when a closed form expression of a model function, such as $x(t)$ in (2.2), is not known or is impractical to determine explicitly. Thus, one is forced to rely on the accuracy of numerical approximations. We illustrated the numerical and analytical appeal of transform estimation with the Laplace transform applied to the nontrivial inverse problem of estimating the flow rates in linear time invariant compartmental models from a given set of data. To this, we proposed convergent algorithms for the implementation of transform estimation that also exploited the structure of compartmental matrices such as their diagonal dominance. We conclude with the following comments pertaining to future research avenues including other applications of transform estimation.

Other interesting applications of transform estimation, particularly those involving the Laplace transform or the Z-transform, are in differential-difference equations or other branches of mathematics which necessitate the numerical approximation to e^{At} or A^k, k a positive integer. Thus, transform estimation in conjunction to time-continuous or discrete-time Markov processes should be investigated.

Our last remark pertains to the interesting observation the when $s_k \neq s_{k'} \neq 0$ for $k \neq k'$ and $w(t, s_q) = e^{-s_q t}$, $q = 1, 2, ..., \tilde{q}$, that W, defined in problem 3.2, is readily seen to be of dimension $\tilde{q}$, since the $w(t, s_q)$ are independent and thus span a $\tilde{q}$-dimensional space. This in conjunction to weak least squares estimation and Claim 2.3, should be further considered but from a theoretical and computational view point. Nonetheless, it would be very interesting to see how these methods perform in practice through simulations and the analysis of a real data set.

The first author thanks Andrew E. Smith, Gilbert G. Walter, and Charles McCulloch for valuable conversations.

5 Appendix

For completeness, we start this section by providing the Jacobians of (2.4) and (2.18). Proofs of claims and theorems and the material on the Laplace transform is appended.

Note that a typical entry of the Jacobian of the unconstrained versions of (2.4) or (2.18) involves, respectively,

$$C\int_0^{t_i} e^{A(t_i-\tau)}A_{a_{kj}}(t_i-\tau)Bu(\tau)d\tau,$$

and

$$C(s_qI-A)^{-2}A_{a_{kj}}B\hat{u}(s_q),$$

where $A_{a_{kj}}$ is as defined in the proof of Theorem 2.1 (see (5.6)). Note that through Laplace transform estimation, we are avoiding the computation of the matrix exponential; moreover, in a similar fashion to step **1** of the algorithm in section 3.1, there is no need to invert (s_qI-A), but rather we could solve for ξ the following linear system of equations,

$$(s_qI-A)^2\xi = A_{akj}B\hat{u}(s_q).$$

Note that if $(s_qI-A)^{-1}$ and A_{akj} commute then we can just use the value of Z computed in step **1**, see (3.5).

Proof of Claim 2.2:

Since A is a compartmental matrix, then its diagonal entries are non-positive, its off-diagonal entries are nonnegative, and it is diagonally column dominant. Thus

$$|a_{kk}| \geq \sum_{j=1}^{n}|a_{jk}|, \quad j\neq k, \text{ and } \quad k=1,2,...,n.$$

These facts together with Gerschgorin's circle theorem, see [11], imply that for each k, the disk $D_k := \{z : |z - a_{kk}| \leq |a_{kk}|\}$ contain at least one eigenvalue of A and that such eigenvalues have a non positive real part and none are purely imaginary. Hence, if we let $Q^H AQ = T$ be the Schur decomposition of A, where T is upper triangular, so that A is similar to T and where Q is a unitary matrix, it follows that

$$(sI - A) = (QsIQ^H - QTQ^H) = Q(sI - T)Q^H.$$

Thus, by the assumptions made on the eigenvalues of A and since Q is unitary, it gives that $(sI - T)^{-1}$ exists, consequently $(sI - A)^{-1}$ exists for all $Re(s) > 0$. Thus Claim 2.2 is established.

Proof of Theorem 2.1:

For the convergence of the sequence of iterates in the Levenberg-Marquardt algorithm, see [21], we only need to prove that the objective function is twice continuously differentiable in D and the Jacobian, $J(a)$, of the objective function is Lipschitz continuous for all $a \in D$, where D is an open convex subset of R^N and $||J(a)||_{L_2} \leq \alpha$, $\alpha > 0$, $\forall a \in D$.

We start by establishing the Lipschitz continuity of the Jacobian. Under the conditions of the theorem, we have that $g(t)$ is a parametric approximation to

$$y(t) = Ce^{A_c t}e_o,$$

where, for clarity, we let

$$g(t) = C^a e^{A_c t}e_o,$$

where C^a is the parametric matrix obtained from fitting the data to a sum of exponentials per iteration. Moreover, the Laplace transform of $y(t)$ is

$$\hat{y}(s) = C(sI - A)^{-1}e_o.$$

So that the iterative sequence for problem (2.18) is

$$a_+ = a_c - [J(a_c)^T J(a_c) + \mu_c I]^{-1} J^T(a_c)R(a_c),$$

where $\mu_c \geq 0$ is the regularization parameter $J(a_c)$ is the $(\tilde{q}+p) \times n^2$ Jacobian of the $(\tilde{q}+p) \times 1$ vector-valued function $R(a_c) := \hat{g}(s_q) - \hat{y}(s_q)$.

To show that the Jacobian is Lipschitz continuous, let us analyze what a typical entry of the Jacobian of the above iterative sequence entails. Without loss of generality, we analyze the (k, j) entry of $J(a)$ which is the derivative with respect to a of the first component of $R(a)$. This is given by

$$\frac{\partial C^a}{\partial a_{kj}}(sI - A)^{-1}e_o + C^a \frac{\partial (sI - A)^{-1}}{\partial a_{kj}} e_o - C(s_q I - A)^{-2} A_{a_{kj}} e_o, \tag{5.6}$$

where

$$\begin{aligned} A_{a_{kj}} &= 1 \text{ in } kj^{th} \text{ position} \\ &= 0 \text{ else where.} \end{aligned}$$

In (5.6), we see that if C^a is twice continuously differentiable and bounded in L_2, that every entry of the Jacobian will be likewise Lipschitz continuous and bounded . This follows from the fact that $(sI - A)^{-1}$ is well-defined when A is a compartmental matrix and $s > 0$, (so that its derivative with respect to the entries of A exist and are bounded) and from the fact that the eigenvalues of a matrix, depend continuously (in theory) on the entries of their matrix, see for instance [11].

To show that C^a is twice continuously differentiable and bounded in L_2, we recall that C^a is obtained from solving uniquely the linear least squares problem given by (5.22), where C^a plays the role of $\hat{D}_{\hat{i}}$ and $y(t_i)$ is as given in the hypothesis of the theorem.

Hence we are solving for C^a in the least squares sense from the model

$$\mathbf{X}(C^a)^T = \mathbf{Y} + \epsilon, \tag{5.7}$$

where $\mathbf{X} := x(t_1)...x(t_m)$ is the $n \times m$ matrix where each column consists of the solution of the compartmental system evaluated at the indicated times, $\mathbf{Y} := y^1...y^m$ is the $p \times m$ matrix whose columns consists of the data gathered in the p compartments at the indicated times, and ϵ is just the $m \times p$ error vector. From (5.7), we

see that $(C^a)^T$ will be uniquely determined provided the columns of $\mathbf{X}$ are linearly independent.

Since the projection of $\mathbf{Y}$ onto the column space of $\mathbf{X}$ yields the normal equations

$$\mathbf{X}^T\mathbf{X}(C^a)^T = \mathbf{X}^T\mathbf{Y},$$

which, if the matrix $\mathbf{X}$ is of full rank, the unique solution is

$$(C^a)^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}, \tag{5.8}$$

which continuously depends on a. That is, it is a well known fact that the "forward" problem depends continuously on a, so that slight perturbations on a, cause slight perturbations of the solution to the compartmental system, $x(t)$, which in turn cause slight perturbations on $\mathbf{X}^T\mathbf{X}$ (in principle, since numerically, the condition number of $\mathbf{X}^T\mathbf{X}$ must be reasonable for slight perturbations on a to yield slight perturbations on the computed solution of C^a, but we do not approach this issue in this paper).

Thus, assuming that (5.8) is well-defined, we see that C^a is indeed at least twice continuously differentiable and bounded in $L_2(0,\infty)$, since it is a product of C^∞ and $L_2(0,\infty)$ bounded functions of a. That is, $\mathbf{X}$ and $\mathbf{Y}$ involve the C^∞ term, e^{At}, which is also in $L_2(0,\infty)$ since A is an invertible compartmental matrix, so that it is a stable system for all time t, see [15].

Prior to showing that (5.8) is well-defined, we point out that by the above statement, by (5.6), and by (5.14) that the Jacobian of the objective function is in $L_2(\eta,\beta)$, $0 < \eta < \beta < \infty$, for each $a \in D$ since each component of the Jacobian is in L_2, so that it is certainly true for sums of the elements of the Jacobian. Hence, the theorem will follow once we show that (5.8) is well-defined.

To show that (5.8) is well-defined, we show that the solution to the system of differential equations span an n dimensional space, so that $\mathbf{X}$ will be of full rank.

Let $A = U\Lambda U^{-1}$ where $\Lambda = \mathrm{diag}(\lambda_i)_{i=1}^n$ such that $\lambda_i \neq \lambda_j$ for $i \neq j$ then the general $(i)^{th}$ entry of the solution to the compartmental system, $x(t) = Ue^{\Lambda t}U^{-1}e_o$

is

$$\sum_{q=1}^{n} \alpha_{iq} e^{\lambda_q t}, \tag{5.9}$$

for some scalars α_{ik}. For clarity of presentation, we let

$$x(t) := \{f_i\}_{i=1}^{n}. \tag{5.10}$$

Then, suppose there exists a k such that $f_k = \alpha_1 f_{k+1}$, for some scalar $\alpha_1 \neq 0$, and note that $U^{-1}e_o$ just picks up the oth column of U^{-1}, so that in (5.10), if $f_k \neq 0$ for all t and arbitrary k, then we see that the kth and $k+1^{th}$ equations give

$$\begin{aligned} \frac{1}{f_k} b_k e^{\lambda_k t} &= \beta, \\ \frac{1}{f_{k+1}} b_{k+1} e^{\lambda_{k+1} t} &= \tilde{\beta}, \end{aligned} \tag{5.11}$$

for some nonzero scalars b_k, b_{k+1}, β, and $\tilde{\beta}$. However, $f_k = \alpha_1 f_{k+1}$ by hypothesis, but this cannot happen if $\lambda_k \neq \lambda_{k+1}$. In a similar argument, we can show that for any i, the f_i's cannot be equaled. Thus, we have that $\{f_i\}_{i=1}^{n}$ span an n-dimensional manifold, as desired. Hence, (5.8) is well-defined and the theorem is established.

5.1 The Laplace transform

In this section, for easy of reading we present material on the Laplace transform and some of its properties that we invoked throughout this paper. For further details, the reader is referred to [1, 20].

The Laplace transform of a continuous function, $u(t)$, exists if

$$|u(t)| \leq a e^{bt},$$

for some constants a and b as $t \to \infty$ and if

$$\int_0^T |u(t)| dt < \infty$$

for every finite T. These two assumptions imply that the integral,

$$\hat{u}(s) := \int_0^\infty e^{-st} u(t) dt, \tag{5.12}$$

exists and is uniformly and absolutely convergent for $Re(s) > b$.

Next, we have that the inverse of the Laplace transform, L^{-1}, exists provided $\hat{u}(s)$ is analytic for $Re(s) > b$ and behaves like $\frac{1}{s}$ for $|s|$ large; that is,

$$\hat{u}(s) = \frac{c_i}{s} + O(\frac{1}{|s|^2}) \text{ as } |s| \to \infty \text{ along } s = b + id, \quad b > a.$$

Then

$$u(t) = \frac{1}{2\pi i} \int_{(C)} \hat{u}(s) e^{st} ds, \quad t > 0,$$

where (C) is the contour in the region of analyticity.

Furthermore, it can be easily seen that the Laplace transform is a contractive map from $L_1(0, \infty)$ to $L_\infty(0, \infty)$ since $|e^{-st}|$ for $Re(s) \geq 0$ and $t > 0$, that is,

$$||\hat{u}(s)||_{L_\infty(0,\infty)} \leq ||u||_{L_1(0,\infty)}. \tag{5.13}$$

However, L^{-1}, while it exists, it is not a continuous operator since it is not stable under reasonable perturbations. That is, since the Laplace transform has the effect of smoothing functions with rapid and oscillatory initial growth these cannot be well-approximated from numerical values of $\hat{u}(s)$ alone. However, if we restrict ourselves to functions, $u(-\log r)$, which can be well-approximated in any sense by a polynomial in r, $0 < r \leq 1$, then L^{-1} is well-behaved for these quite smooth functions, see [20]. Nonetheless, in principle, by the Weierstrass theorem [19], we know that it is always possible to approximate any continuous function by polynomials in the L_2-norm sense; moreover, so that we may pass the limit under the integral sign, we will restrict ourselves to integrable functions, see [19] .

In the following section we continue to analyze the Laplace transform but now in conjunction with compartmental matrices.

5.2 Continuity of the Laplace transform and compartmental matrices

Unfortunately, it is not true that the Laplace transform is continuous from L_1 to L_1 on all of the interval $(0, \infty)$; nor is it continuous from L_2 to L_2 on all of $(0, \infty)$

either. Nonetheless, it is an immediate application of Cauchy-Schwarz inequality and Fubini's theorem [22] to show, that when s is restricted to real numbers, the Laplace transform is continuous from $L_2(0,\infty)$ to $L_2(\Omega_1)$, where Ω_1 is a compact subset of $(0,\infty)$ which is bounded away from zero. A similar statement can be made about the L_1-norm. We briefly outline the proof for the former statement since in this paper, we will deal exclusively with the L_2-norm. That is, by Cauchy-Schwarz

$$\begin{aligned}(\int_0^\infty e^{-st/2}e^{-st/2}h(t)dt)^2 &\leq (\int_0^\infty e^{-st}dt)(\int_0^\infty e^{-st}h^2(t)dt)\\ &\leq \frac{1}{s}\int_0^\infty h^2(t)dt,\end{aligned}$$

and by Fubini's theorem we have

$$\begin{aligned}||\hat{h}(s)||^2_{L_2(\Omega_2)} &\leq \int_0^\infty (\int_{\Omega_2}\frac{1}{s}ds)h^2(t)dt \\ &\leq (ln(\beta)-ln(\alpha))||h(t)||^2_{L_2(\Omega_1)},\end{aligned} \tag{5.14}$$

where $ln(x)$ is the natural log of x. Hence, L, or the Laplace transform is continuous from $L_2(0,\infty)$ to $L_2[\alpha,\beta]$, $\alpha \neq 0$, where $\Omega_2 = [\alpha,\beta]$.

Similarly, L^{-1} will exist and be stable provided we restrict ourselves to smooth functions that are defined in $L_2[\alpha,\beta]$ or to smooth functions which are supported on the interval $[\alpha,\beta]$, see [20, 1].

5.3 A choice of the variable of the transformation

Workers [1, 20] argue that unless there is good reason for doing otherwise one should choose $s_q \geq 1$ for each q. To see this, suppose $\hat{f}(s)$ is the Laplace transform of some function $f(t)$, which we re-state here for convenience,

$$\hat{f}(s) = \int_0^\infty e^{-st}f(t)dt, \quad s \in \Im. \tag{5.15}$$

Then naturally, if s is complex-valued than $\hat{f}(s)$ will also be complex-valued. Furthermore, through the change of variable $r = e^{-t}$, we see that

$$\hat{f}(s) = \int_0^1 r^{s-1}g(r)dr,$$

where $g(r) = f(-\log(r))$ and $0 < r \leq 1$.

Next, if we let $s = \sigma + i\tau$ and using the fact that, through its Taylor series expansion, $e^{ix} = \cos x + i \sin x$, we can write $\hat{f}(s)$ in terms of its real and imaginary part, then

$$\begin{aligned} \hat{f}(s) &= a(\sigma,\tau) + ib(\sigma,\tau) \quad \text{where} \\ a(\sigma,\tau) &= \int_0^1 r^{\sigma-1}\cos(\tau \log r) g(r) dt, \text{ and} \\ b(\sigma,\tau) &= \int_0^1 r^{\sigma-1}\sin(\tau \log r) g(r) dt. \end{aligned} \tag{5.16}$$

From (5.16), we see that unless $\tau = 0$, we are faced with solving two integral equations. Moreover, if $\tau \neq 0$ then, when $\sigma < 1$, the integrand could have rapid oscillations near the origin and in fact it has a singularity at $r = 0$. Given this information, we will recommend that, when no other information is available, that the sequence $\{s_q\}_{q=1}^{\tilde{q}}$ be chosen to be real and greater than or equaled to 1. In fact, [1] suggest that $\{s_q\}$ be chosen to be the sequence $1, 2, 3, ..., \tilde{q}$ when they dealt with the inverse problem described by a linear second order ordinary differential equation.

5.4 General forcing function

Although Claim 2.3 holds under rather stringent requirements, nonetheless, for a general forcing function $u(t)$, sampling matrix, B, and diagonalizable A, we illustrate how one would fit a parametric function $g(t)$, to the data, even though the theoretical results presented in the paper need not hold.

Suppose we have the spectral decomposition of $A := U\Lambda U^{-1}$, where $\Lambda = diag\{\lambda_q\}_{q=1}^{n}$ and U has as its columns the eigenvectors corresponding to each λ_q, respectively. From this, together with the series expansion of e^{At}, it follows that

$$e^{A(t-\tau)} = Ue^{\Lambda(t-\tau)}U^{-1}; \tag{5.17}$$

Applying (5.17) to (2.2) at each time t_i, $i = 1, ..., m$ and pulling C inside the integral, we have

$$y(t_i) = \int_0^{t_i} CUe^{\Lambda(t_i-\tau)}U^{-1}Bu(\tau)d\tau. \tag{5.18}$$

Discretizing (5.18) and letting $\tilde{C} := CU$, $\tilde{B} := U^{-1}B$ and $\tilde{u}^{\tau_k} := u(\tau)\Delta_k$, where $k = 1, ..., m_i$, and where we have chosen the weights Δ_k appropriately so that we can say that the following discretization is a good approximation to (5.18), and where m_i is the total number of partition points within the interval $(0, t_i)$, then

$$y(t_i) \approx \sum_{k}^{m_i} \tilde{C} e^{(t_i - \tau_k)\Lambda} \tilde{B} \tilde{u}^{\tau_k}, \tag{5.19}$$

where $\tilde{C}$ is a $p \times n$ matrix and $\tilde{B}$ is an $n \times r$ matrix and $\tilde{u}^{\tau_k}$ is an $r \times 1$ vector. It is a straight forward exercise in matrix multiplication to conclude that (5.19) yields

$$y(t_i) \approx \sum_{k=1}^{m_i} [\sum_{\hat{i}}^{n} D_{\hat{i}} V^{\hat{i}}], \tag{5.20}$$

where $D_{\hat{i}}$ is a $p \times r$ matrix for each $\hat{i}$ and $V^{\hat{i}}$ is an $r \times 1$ vector for each $\hat{i}$ such that

$$V^{\hat{i}} = e^{(t_i - \tau_k)\lambda_{\hat{i}}} \tilde{u}_{\hat{i}}^{\tau_k}, \quad \hat{i} = 1, ..., r. \tag{5.21}$$

So an attempt to fit the data to (2.2) necessitates that we estimate prn entries once the eigenvalues are provided; that is, we solve

$$\min_{D_{\hat{i}}} \sum_{i=1}^{m} ||y(t_i) - y^i|| := \hat{D}_{\hat{i}}, \quad \hat{i} = 1, ..., n. \tag{5.22}$$

Now we can say that we have fitted the data at a discrete set of times and at each iteration, or for any given guess of the minimizer of (2.18), to a function of the form

$$g(t) = \sum_{k=1}^{m_i} (\sum_{\hat{i}}^{n} \hat{D}_{\hat{i}} e^{(t - \tau_k)\lambda_{\hat{i}}} \tilde{u}_{\hat{k}}^{\tau_k}), \quad \hat{k} = 1, ..., r; \tag{5.23}$$

Now taking the Laplace transform of (5.23), we have

$$\hat{g}(s) = \sum_{k=1}^{m_i} (\sum_{\hat{i}}^{n} \hat{D}_{\hat{i}} \frac{e^{-\tau_k \lambda_{\hat{i}}} \tilde{u}_{\hat{k}}^{\tau_k}}{s - \lambda_{\hat{i}}}), \quad \hat{k} = 1, ..., r; \tag{5.24}$$

Note that for (5.22) to have a unique solution, we need prn to exceed the number of unknowns.

References

[1] R.E. Bellman, R.E. Kalaba and J.A. Lockett, *Numerical inversion of the Laplace transform*, Elsevier Publishing Company, New York, NY, 1966, pp. 87-89.

[2] H.J. Schuh and R.L. Tweedie, Parameter estimation using transform estimation in time-evolving models, *Math. Biosci.* 45:37-67 (1979).

[3] P.D. Feigin, R.L. Tweedie and C. Belyea, Weighted area techniques for explicit parameter estimation in multi-stage models, *Austral. J. Statist.* 25:1:1-16 (1983).

[4] C.E. Wells and J.L. Bryant, An adaptive estimation procedure using the Laplace transformation, *IIE Transactions,* 17:3:242-251 (1984).

[5] A. Feuerverger and P. McDunnough, On statistical transform methods and their efficiency, *The Canadian Journal of Statistics,* 12:4:303-317 (1984).

[6] U. Bagchi, J.K. Ord and R.S. Sullivan, Parameter estimation for convolutions and compound distributions, *Operations Research Letters,* 4:6:301-308 (1986).

[7] A.F. Laurence and B.J.T. Morgan, Selection of the transformation variable in the Laplace transform method of estimation, *Austral. J. Statist.* 29:2:113-127 (1987).

[8] M. Contreras and G. Casella, Integral transform parameter estimation, Tech. Report BU-1363-M, Biometrics Unit, Cornell University, Ithaca, New York 14853 (1996).

[9] J. Jacques, *Compartmental analysis in biology and medicine, 2nd Edition*, The University of Michigan Press, Ann Arbor, Michigan, 1985, pp. 3-8.

[10] D.H. Anderson, *Compartmental modeling and tracer kinetics*, Lecture Notes in Biomathematics, Springer-Verlag, New York, NY, 1983, pp. 36-38.

[11] G.H. Golub and C. Van Loan, *Matrix Computations, 3rd. Edition*, Johns Hopkins, Baltimore, Maryland, 1996, pp. 320, 572-576.

[12] A. Berman, R.J. Plemmons, *Nonnegative matrices in the mathematical sciences*, SIAM, Philadelphia, PA, 1994, pp. 132-133.

[13] C. Cobelli and G. Romanin-Jacur, Structural identifiability of strongly connected biological compartmental systems, *Medical and Biological Engineering,* pp. 831-838 (1975).

[14] D.M. Bates, D.A. Wolf, and D.G. Watts, Nonlinear least squares and first-order kinetics, *Computer Science and Statistics: The Interface*, D.M. Allen, Ed., Elsevier Science Publishers B.V., North-Holland, 1986, pp. 71-81.

[15] J.Z. Hearon, Theorems of linear systems, *Ann. N.Y. Acad. Sci.* 108:36-68 (1963).

[16] M.S. Grewal and K. Glover, Identifiability of linear and nonlinear systems, *IEEE Trans. Auto. Contr.* 21:833-837 (1976).

[17] S. Kesavan, *Topics in functional analysis and applications*, John Wiley & Sons, New Delhi, India, 1989, pp. 131-133.

[18] C.W. Groetsch, *Inverse problems in the mathematical sciences*, Vieweg, Braunschweig/Wiesbaden, Germany, 1993, pp. 35-39, 84-90, 119.

[19] W. Rudin, *Principles of mathematical analysis*, McGraw Hill, New York, NY, 1976, pp. 159-160, 224-227.

[20] R.E. Bellman and R.S. Roth, *The Laplace transform*, World Scientific, Farrer Road, Singapore, 1984, pp.1-27, 135-154.

[21] J.E. Jr. Dennis and R. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall Series in computational Mathematics, Englewood Cliffs, NJ, 1983, pp. 218-228.

[22] H.L. Royden, *Real Analysis, 2nd ed.*, Macmillan Publishing Co., Inc., New York, NY, 1968, pp. 307-309.