Characterization of Matrix-exponential Distributions

Mark William Fackrell

Thesis submitted for the degree of

Doctor of Philosophy

in

Applied Mathematics

at

The University of Adelaide

(Faculty of Engineering, Computer and Mathematical Sciences)

School of Applied Mathematics

November 18, 2003

Contents

List of Figures

Signed Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED: ..... DATE: .....

Acknowledgements

I would like to extend my sincere thanks to my two supervisors Prof Peter Taylor and Dr Nigel Bean for their tireless support and encouragement over the past four and a half years.

Thanks also go to Dr David Green, Dr Andre Costa, and Kate Kennedy for their patient assistance in many matters throughout the course of this PhD.

The staff and students of the Teletraffic Research Centre have provided a supportive and friendly environment in which to study and I wish to express my gratitude to them.

Thanks go to Associate Prof Andrew Eberhard of the Department of Mathematics and Statistics, RMIT University, Melbourne, for suggesting the semi-infinite programming approach that led to Chapters 8 and 9. I would also like to express my gratitude to Prof Lang White of the Department of Electrical and Electronic Engineering, University of Adelaide, for his advice and encouragement, particularly in the vital, early stages of candidature.

The funding for this PhD research was provided by a Federal Government Australian Postgraduate Award scholarship and a Teletraffic Research Centre top-up scholarship. I am grateful to both funding bodies for their financial assistance without which this research would not have been possible.

I would also like to thank the two examiners of this thesis who provided prompt feedback and valuable reports.

And last, but certainly not least, a big thankyou to my wife Jenny and son Matthew for enduring much throughout the course of this PhD.

Dedication

This thesis is dedicated to Associate Professor William (Bill) Henderson (1943–2001) who was a truly inspirational applied probabilist.

Abstract

A random variable that is defined as the absorption time of an evanescent finite-state continuous-time Markov chain is said to have a *phase-type* distribution. A phase-type distribution is said to have a *representation* $(\boldsymbol{\alpha}, \boldsymbol{T})$ where $\boldsymbol{\alpha}$ is the initial state probability distribution and $\boldsymbol{T}$ is the infinitesimal generator of the Markov chain. The distribution function of a phase-type distribution can be expressed in terms of this representation. The wider class of *matrix-exponential* distributions have distribution functions of the same form as phase-type distributions, but their representations do not need to have a simple probabilistic interpretation. This class can be equivalently defined as the class of all distributions that have rational Laplace-Stieltjes transform. There exists a one-to-one correspondence between the Laplace-Stieltjes transform of a matrix-exponential distribution and a representation $(\boldsymbol{\beta}, \boldsymbol{S})$ for it where $\boldsymbol{S}$ is a companion matrix.

In order to use matrix-exponential distributions to fit data or approximate probability distributions the following question needs to be answered:

> *"Given a rational Laplace-Stieltjes transform, or a pair* $(\boldsymbol{\beta}, \boldsymbol{S})$ *where* $\boldsymbol{S}$ *is a companion matrix, when do they correspond to a matrix-exponential distribution?"*

In this thesis we address this problem and demonstrate how its solution can be applied to the abovementioned fitting or approximation problem.

Chapter 1

Introduction

This thesis is concerned with the problem of fitting data and approximating probability distributions with phase-type and matrix-exponential distributions. A random variable that is defined as the absorption time of an evanescent finite-state continuous-time Markov chain is said to have a *phase-type* (PH) distribution. The distribution and density functions of a PH distribution can be expressed in terms of the $1 \times p$ initial state distribution vector $\boldsymbol{\alpha}$ and the $p \times p$ infinitesimal generator matrix $\boldsymbol{T}$ of the underlying Markov chain. The pair $(\boldsymbol{\alpha}, \boldsymbol{T})$ is known as a *representation* of *order* p of the PH distribution. The wider class of *matrix-exponential* (ME) distributions have distribution functions of the same form as PH distributions but their representations do not need to have a simple probabilistic interpretation.

PH distributions and their point process counterparts, *Markovian arrival processes* (MAPs), are integral to the branch of computational probability known as *matrix-analytic methods*. *Computational probability* was described by Neuts [101] as

> *" . . . the study of stochastic models with a genuine added concern for algorithmic feasibility over a wide, realistic range of parameter values."*

Matrix-analytic methods deals with the analysis of stochastic models, particularly queueing systems, using a matrix formalism to develop algorithmically tractable solutions. The ever-increasing ability of computers to perform numerical calculations has supported the growing interest in this area.

Although *ME* distributions do not strictly belong to the realm of matrix-analytic methods some of what has been achieved with *PH* distributions carries over to *ME* distributions, see Asmussen and Bladt [10], and Bean and Nielson [19]. Stochastic models that use *ME* distributions in place of *PH* distributions have greater flexibility and generality but at the expense of simple probabilistic interpretations.

Before the advent of fast computers, problems in stochastic modelling, particularly queueing theory, relied on the Laplace-Stieltjes transform and the methods of complex analysis for their solution, see, for example, Cohen [37]. Often, analytical expressions for the performance measures of stochastic models were given in closed form and could not readily be implemented in algorithms. Not only this, but frequently such expressions gave little qualitative or probabilistic insight into the systems being analysed.

Since the building blocks of matrix-analytic methods, *PH* distributions and *MAP*s, are defined in terms of Markov chains, highly versatile stochastic models that exhibit an underlying Markov structure can be analysed. Quantities of interest can very often be given a meaningful probabilistic interpretation. In addition, since the matrices that represent *PH* distributions and *MAP*s consist entirely of real entries, performance measures, which are expressed in terms of these matrices and their exponentials, can be implemented in algorithms relatively easily. The field of computational probability and its progeny matrix-analytic methods have redefined the meaning of a solution to a problem in stochastic modelling: an implementable algorithm that adds insight into the system being analysed. The number of such systems that can now be modelled stochastically has increased significantly.

Over the last two decades there has been a phenomenal increase in the theory and application of matrix-analytic methods. The complexity of the stochastic models that can be analysed has grown alongside the improvement in computing power. Areas of application have included scheduling (Squillante [134], and Sethuraman and Squillante [128]), insurance risk (Asmussen and Rolski [14], Møller [98], and Asmussen [9]), machine maintenance (Green, Metcalfe, and Swailes [64]), survival

analysis (Aalen [1]), reliability theory (Bobbio, Cumani, Premoli, and Saracco [26], and Chakravarthy [33]), and drug kinetics (Faddy [49] and [50]). The greatest research activity, however, given the explosion in data traffic that we have witnessed over the last few years, has undoubtedly been in the performance analysis of telecommunications systems. The telecommunications and electronic engineering literature is awash with applications of matrix-analytic methods. For recent advances in the theory and application of matrix-analytic methods we refer the reader to the proceedings of the discipline's four conferences Chakravarthy and Alfa [34], Alfa and Chakravarthy [4], and Latouche and Taylor [83] and [84], and the references therein, and to Neuts [102] which contains an extensive bibliography on the subject.

Despite the remarkable growth in the theory and application of matrix-analytic methods, one area that has been considerably under-explored is that of statistical fitting and approximation. In order to use *PH* distributions and *MAP*s in stochastic modelling their parameters need to be selected so that they best describe, in some sense, the processes they are modelling.

Moment matching algorithms for fitting mixtures of Erlang distributions (which are particular *PH* distributions) to independent and identically distributed data have been developed by Johnson [73] and Schmickler [124]. Bobbio and Cumani [24], and Horváth and Telek [72] used maximum likelihood methods to fit data, and approximate probability distributions, respectively, with Coxian distributions (*PH* distributions whose generator matrix $\boldsymbol{T}$ has only real eigenvalues). Asmussen, Nerman, and Olsson [15] developed an expectation-maximization (*EM*) algorithm to fit general *PH* distributions to data.

Fitting with *MAP*s is more difficult because the data from an arrival stream are not necessarily independent and identically distributed. A number of moment matching methods for fitting *Markov-modulated Poisson processes* (*MMPP*s - a subclass of *MAP*s) have been developed and were briefly discussed in Rydén [123]. These methods, however, were restricted to *MMPP*s of order two or a specific structure. Meier-Hellstern [96] gave a method based on maximum likelihood for *MMPP*s of

order two but the parameter estimators were asymptotically biased. Rydén [118] proved the consistency of the maximum likelihood estimators for *MMPP*s of arbitrary order. He also compared the performance of three algorithms used to find the maximum likelihood estimates when an order two *MMPP* was fitted to some simulated data. The consistency and asymptotic normality of an estimator closely related to the maximum likelihood estimator for *MMPP*s was shown in Rydén [119]. In Rydén [121] an *EM* algorithm for *MMPP*s was developed and compared with a number of other algorithms. Diamond and Alfa [47] gave a method for approximating a *MAP* of arbitrary order with an order two *MAP* by matching the autocorrelation decay parameter and the first two or three moments. Breuer [29] developed a maximum likelihood-based method for estimating the parameters of a particular class of *batch Markovian arrival processes* (*BMAP*s - *MAP*s which allow batch arrivals), and the ideas were extended to general *BMAP*s in Breuer and Gilbert [30].

In Chapter 2 *PH* distributions are formally defined and their properties, representation, and characterization are discussed.

Chapter 3 contains a more detailed discussion of some of the existing methods developed for fitting data and approximating distributions with *PH* distributions and the problems associated with them. The main difficulties, according to Lang and Arthur [82], are that

1 the fitting or approximation problem is highly nonlinear,

2 the number of parameters to be estimated or selected is often large,

3 *PH* representations are typically not unique, and

4 the relationship between the parameters and the shape of a *PH* distribution is generally nontrivial.

Most algorithms developed used Coxian distributions (or particular subclasses of them) to circumvent the second and third difficulties. A Coxian representation of

order p is parameterized by only $2p$ parameters instead of the general PH representation's $p^2 + p$ parameters. Also, a unique canonical representation can be given for Coxian distributions. It is not clear, however, whether this restricted class is adequate, in general, for statistical fitting and approximation although some authors, for example Horváth and Telek [72], believe that it is.

In order to avoid the second difficulty, and possibly the first and third ones, we propose in Chapter 4 that the fitting or approximation with general PH distributions be carried out in the Laplace-Stieltjes transform (LST) domain. The LST of a PH distribution with a representation of order p (which is a rational function) has $2p$ parameters. A number of authors have used the idea of transform fitting or approximation but we discuss in detail two related methods given in Harris and Marchal [66] because they specifically use rational LSTs. Their methods are very simple to implement because they only require the solution of a system of linear equations. The procedure, however, has two major drawbacks. First, there is no guarantee that the final LST corresponds to a probability distribution, PH or otherwise. Harris and Marchal [66] gave no means for determining whether or not a given rational LST corresponds to a PH distribution. Second, if the LST does happen to correspond to a PH distribution it is not clear how to find a PH representation for it.

In Chapter 5, in order to tackle the two problems posed at the end of Chapter 4, the class of ME distributions is introduced. The second problem, with respect to ME distributions, is solved by using a ME representation theorem from Asmussen and Bladt [10]. The representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ they gave is such that $\boldsymbol{\alpha}$ is the vector of coefficients of the rational LST's numerator polynomial, and $\boldsymbol{T}$ is the companion matrix of the denominator polynomial. This (one-to-one) correspondence between the LST of a ME distribution and a representation of this form means that any statement about one will also be true for the other. If we define the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ to be the coefficients of the numerator and denominator polynomials, respectively, then the first problem can be stated as follows:

"When do a pair of vectors $\boldsymbol{a}$*,* $\boldsymbol{b} \in \mathbb{R}^p$ *correspond to a ME distribution?"*

This problem, although easy to state, is very difficult to solve. A necessary condition is that the polynomial defined by $\boldsymbol{b}$ must have a zero of maximal real part that is real and negative. Given a suitable vector $\boldsymbol{b}$ we define a set (or region) in terms of an uncountably infinite number of linear constraints that contains all vectors (thought of as points) $\boldsymbol{a}$ that correspond to *ME* distributions.

In Chapter 6 we derive a complete analytical description of the region when the order of the *ME* distribution is three. Some discussion is devoted to the case when the order is greater than three but a complete description has not yet been found.

We present in Chapter 7 an algorithm, based on an approach due to Dehon and Latouche [45], that determines whether or not a given vector $\boldsymbol{a}$ is contained in the region determined by a suitable vector $\boldsymbol{b}$. Since the algorithm, however, requires the global minimization of a single variable function over the nonnegative real line, it is potentially computer intensive especially when the *ME* distribution has high order. In addition, because of the relative simplicity of the order three case, we give an alternative analytical description of the region in that case.

In Chapter 8 we present a *semi-infinite programming* algorithm to determine if a given vector $\boldsymbol{a}$ is contained in the region defined by a suitable vector $\boldsymbol{b}$. The problem becomes one of minimizing a convex objective function over a (convex) feasible region which is defined by an infinite number of constraints.

The real merit in the semi-infinite programming approach, however, is not in the *ME* identification problem, but in using *ME* distributions to fit data or approximating probability distributions. This is discussed in Chapter 9. Given a suitable vector $\boldsymbol{b}$, a unique vector $\boldsymbol{a}$ can be found that maximizes the (convex) loglikelihood function over the feasible region. Combining this algorithm with the Nelder-Mead flexible polyhedron search (which updates the vector $\boldsymbol{b}$) we have a method for finding maximum likelihood parameter estimates when fitting *ME* distributions to data. The algorithm can be used to approximate distributions by choosing appropriate sample points. The chapter concludes with two examples that illustrate the algorithm.

Chapter 10 concludes the thesis and proposes some directions for future research.

Chapter 2

Phase-type Distributions

2.1 Introduction

Since their introduction by Neuts [100] in 1975, phase-type (PH) distributions have been used in a wide range of stochastic modelling applications in areas as diverse as telecommunications, teletraffic modelling, biostatistics, queueing theory, drug kinetics, reliability theory, and survival analysis. Asmussen and Olsson [13] stated that

> *"... there has been a rapidly growing realization of PH (phase-type) distributions as a main computational vehicle of applied probability."*

PH distributions have enjoyed such popularity because they constitute a very versatile class of distributions defined on the nonnegative real numbers that lead to models which are algorithmically tractable. Their formulation also allows the Markov structure of stochastic models to be retained when they replace the familiar exponential distribution.

Erlang [48], in 1917, was the first person to extend the familiar exponential distribution with his "method of stages". He defined a nonnegative random variable as the time taken to move through a fixed number of stages (or states), spending an exponential amount of time with a fixed positive rate in each one. Nowadays we refer to distributions defined in this manner as *Erlang* distributions. In 1955 Cox [41]

(see also Cox [40]) generalized Erlang's notion by allowing complex "rates". This construction, despite often having no simple probabilistic interpretation, defines the class of *distributions with rational Laplace-Stieltjes transform*, of which the class of *PH* distributions is a proper subset. These distributions are nowadays also known as *matrix-exponential* distributions which shall be discussed in detail in Chapter 5. Neuts [100] generalized Erlang's method of stages in a different direction. He defined a *phase-type* random variable as the time taken to progress through the states of a finite-state evanescent continuous-time Markov chain, spending an exponential amount of time with a positive rate in each one, until absorption. The class of *PH* distributions is hence a very flexible class of distributions that have a simple probabilistic interpretation.

PH distributions are indeed a versatile class of distributions. First, they are dense in the class of all distributions defined on the nonnegative real numbers. However, as remarked by Neuts [101, page 79], there are a number of simple distributions (for example the delayed exponential distribution) where a reasonable approximation by a *PH* distribution would require a prohibitive number of states. On the other hand, because of the flexibility of the parameters of the continuous-time Markov chain that define the *PH* distribution, they can potentially exhibit quite versatile behaviour. For example, as mentioned in O'Cinneide [108], it is known that tri-modal *PH* distributions of order five exist.

Second, the use of *PH* distributions in stochastic models often enables algorithmically tractable solutions to be found. Quantities of interest, such as the distribution and density functions, the Laplace-Stieltjes transform, and the moments of *PH* distributions are expressed simply in terms of the initial phase distribution $\boldsymbol{\alpha}$ and the exponential or powers of the infinitesimal generator $\boldsymbol{T}$ of the defining Markov chain. Since $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ consist of only real entries many of the quantitative performance measures required when using *PH* distributions in stochastic modelling (for example the waiting time distributions and mean queue lengths in queues) can be computed relatively easily given a suitable software package (for example MATLAB®). Also,

qualitative performance measures can be established in stochastic models where PH distributions are used. For example, Takahashi [138] showed that the tail of the waiting time distribution for the $PH/PH/c$ queue is exponential. See Shaked and Shanthikumar [129, pages 713–714] for a list of further examples.

Third, stochastic models, particularly where the exponential distribution is used to model quantities (for example interarrival times, service times, or lifetimes) because of its simplicity, can now be extended by using PH distributions with little extra complication. Often the exponential distribution can simply be replaced with a PH distribution while preserving the underlying Markov structure of the model. For example, the $M/M/1$ queue can be generalized to the $PH/PH/1$ queue which can be analyzed in an analogous manner.

Finally, since the class of PH distributions is closed under a variety of operations (for example finite mixture and convolution, see Section 2.5) systems with PH inputs often have PH outputs. For example, the stationary waiting time distribution in a $M/PH/1$ queue is PH, see Neuts [101, page 21]. Also, Asmussen [7] showed that the waiting time distribution in a $GI/PH/1$ queue is PH. Refer to Shaked and Shanthikumar [129, pages 713–714] for more examples. It seems, however, that it is not always the case that PH inputs produce PH outputs. For example, Olivier and Walrand [109] *conjectured* that the departure process of $MAP/PH/1$ queue is not a MAP unless the queue is a stationary $M/M/1$ queue. Therefore, it is possible that the departure process of a $PH/PH/1$ queue is not a PH renewal process (which is a particular type of MAP). Bean, Green, and Taylor [20] gave an example of a $PH/M/1$ queue where it could not be established that the departure process is a MAP.

In Section 2.2 we define PH distributions, their representation, and order, list some of their important properties, and give some examples. Section 2.3 is an analogous section on discrete PH distributions. In Section 2.4 we address the problem of characterizing (continuous) PH distributions by asking the two questions: when

does a function of the form

$$f(u) \;=\; \sum_{i=1}^{n} q_i(u) e^{-\lambda_i u},$$

where the q_i's are polynomials, correspond to the density function of a *PH* distribution; and if it does, what is a minimal representation for it? Section 2.5 contains a discussion on the closure properties of the class of *PH* distributions. Some concluding remarks are made in Section 2.6.

For a comprehensive treatment of *PH* distributions see Neuts [101, Chapter 2]. Latouche and Ramaswami [85, Chapter 2] is a very readable introduction to the topic. The literature on the theory and applications of *PH* distributions is vast and both of the abovementioned books provide extensive bibliographies. The two entries in the Encyclopedia of Statistical Science on *PH* distributions, Shaked and Shanthikumar [129], and Asmussen and Olsson [13], also provide excellent introductions to the subject.

2.2 Continuous Phase-type Distributions

Consider an evanescent continuous-time Markov chain $\{Y_u\}$, with $u \geq 0$, on a finite phase (state) space $S = \{0, 1, 2, \ldots, p\}$ where phase 0 is absorbing. Let the initial phase probability distribution be $(\alpha_0, \boldsymbol{\alpha}) = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ (with $\sum_{i=0}^{p} \alpha_i = 1$) and the infinitesimal generator be $\boldsymbol{Q}$. The random variable X, defined as the time to absorption, is said to have a *continuous phase-type* (*PH*) distribution.

The infinitesimal generator for the Markov chain can be written in block-matrix form as

$$\boldsymbol{Q} \;=\; \begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{t} & \boldsymbol{T} \end{pmatrix}.$$

Here, $\mathbf{0}$ is a $1 \times p$ vector of zeros, $\boldsymbol{t} = (t_1, t_2, \ldots, t_p)'$ where, for $i = 1, 2 \ldots p, t_i = Q_{i0} \geq 0$ is the absorption rate from phase i, and $\boldsymbol{T} = [T_{ij}]$ is a $p \times p$ matrix where,

for $i, j = 1, 2, \ldots, p$, with $i \neq j$,

$$T_{ij} \;\geq\; 0,$$

and, for $i = 1, 2, \ldots, p$,

$$T_{ii} \;<\; 0 \qquad \text{with} \qquad T_{ii} \;\leq\; -\sum_{\substack{j=1 \\ j\neq i}}^{p} T_{ij}.$$

Note that $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$ where $\boldsymbol{e}$ is a $p \times 1$ vector of ones. The *PH* distribution is said to have a *representation* $(\boldsymbol{\alpha}, \boldsymbol{T})$ of *order* p. The matrix $\boldsymbol{T}$ is referred to as a *PH-generator*. The component α_0, which is completely determined by $\boldsymbol{\alpha}$ and therefore does not need to appear in the expression for the representation, is known as the *point mass at zero*.

To ensure absorption in a finite time with probability one, we assume that every nonabsorbing state is transient. This statement is equivalent to $\boldsymbol{T}$ being nonsingular, see Neuts [101, Lemma 2.2.1, page 45], or Latouche and Ramaswami [85, Theorem 2.4.3, page 43]. An additional requirement on the *PH* representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is that there are no superfluous phases. A condition for there to exist no such phases can be derived as follows. Assume that as soon as absorption takes place in the Markov chain with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$, the process is started anew with the same parameters. The resulting point process is called a *PH-renewal process*. The distribution of interevent times of this process is a *PH* distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$. There will be no superfluous phases in the process if every nonabsorbing phase can be reached from every other phase with probability one. This occurs if the matrix

$$\boldsymbol{Q}^* \;=\; \boldsymbol{T} - (1 - \alpha_0)^{-1}\boldsymbol{T}\boldsymbol{e}\boldsymbol{\alpha},$$

which is the infinitesimal generator of the *PH*-renewal process, is *irreducible*. For the definition of an *irreducible matrix* see Seneta [127, Section 1.3 and page 46]. We then say that the representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is *irreducible*, see Neuts [101, page 48]. If a representation includes some superfluous phases they can be deleted. The resulting *PH*-renewal process and its corresponding representation will then both be irreducible in their respective senses.

A *PH* distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ has distribution function, defined for $u \geq 0$, given by

$$F(u) \;=\; \begin{cases} \alpha_0, & u = 0 \\ 1 - \boldsymbol{\alpha} \exp(\boldsymbol{T}u)\boldsymbol{e}, & u > 0. \end{cases} \tag{2.2.1}$$

For a proof see Neuts [101, Lemma 2.2.2, page 45], or Latouche and Ramaswami [85, Theorem 2.4.1, page 41]. Differentiating (2.2.1) with respect to u gives the corresponding density function, defined for $u > 0$,

$$f(u) \;=\; -\boldsymbol{\alpha} \exp(\boldsymbol{T}u)\boldsymbol{T}\boldsymbol{e}.$$

The Laplace-Stieltjes transform (*LST*) of (2.2.1), which is defined for $\lambda \in \mathbb{C}$ such that $\Re(\lambda) > -\delta$ where δ is a positive number, is given by

$$\begin{aligned} \phi(\lambda) \;&=\; \int_0^\infty e^{-\lambda u} dF(u) \\ &=\; -\boldsymbol{\alpha}(\lambda \boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{T}\boldsymbol{e} + \alpha_0. \end{aligned} \tag{2.2.2}$$

The *LST* $\phi(\lambda)$ can be expressed as the ratio of two irreducible polynomials where the degree of the numerator is less than or equal to the degree of the denominator. Following O'Cinneide [104], the *algebraic degree* of the *PH* distribution is defined to be the degree of the denominator. For $k = 1, 2, \ldots$, differentiating (2.2.2) k times with respect to λ and letting $\lambda = 0$ gives the kth noncentral moment

$$m_k \;=\; (-1)^k k! \boldsymbol{\alpha}\boldsymbol{T}^{-k}\boldsymbol{e}.$$

We now give some examples of *PH* distributions.

1. The exponential distribution with density function $f(u) = \lambda e^{-\lambda u}$ has a representation

$$\begin{aligned} \boldsymbol{\alpha} \;&=\; \begin{pmatrix} 1 \end{pmatrix} \\ \boldsymbol{T} \;&=\; \begin{pmatrix} -\lambda \end{pmatrix}. \end{aligned}$$

2. The hyperexponential distribution with density function

$$f(u) = \sum_{i=1}^{p} \alpha_i \lambda_i e^{-\lambda_i u}$$

where, for $i = 1, 2, \ldots, p$, $\alpha_i > 0$ and $\sum_{i=1}^{p} \alpha_i = 1$, has a representation

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & 0 & \ldots & 0 \\ 0 & -\lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -\lambda_p \end{pmatrix}.$$

3. The p-phase Erlang distribution with density function

$$f(u) = \frac{\lambda^p u^{p-1} e^{-\lambda u}}{p!}$$

has a representation

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda & \lambda & 0 & \ldots & 0 \\ 0 & -\lambda & \lambda & \ldots & 0 \\ 0 & 0 & -\lambda & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & -\lambda \end{pmatrix}.$$

4. The p-phase *Coxian* distributions have representations of the form

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \ldots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \ldots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & -\lambda_p \end{pmatrix},$$

where $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$.

5. The *acyclic*, or *triangular PH* (*TPH*), distributions have *PH*-generators that are upper triangular matrices.

6. The p-phase *unicyclic PH* distributions have representations of the form

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \ldots & 0 & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\lambda_{p-1} & \lambda_{p-1} \\ \mu_1 & \mu_2 & \mu_3 & \ldots & \mu_{p-1} & -\lambda_p \end{pmatrix},$$

where for $i = 1, 2, \ldots, p-1$, $\mu_i \geq 0$, $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$, and $\lambda_p > \sum_{i=1}^{p-1} \mu_i$, see O'Cinneide [108, Section 7].

In general, representations for *PH* distributions are not unique. Consider the following which is derived from an example in Botta, Harris, and Marchal [28]. The *PH* distribution with density

$$f(u) = \frac{2}{3}e^{-2t} + \frac{1}{3}e^{-5t}$$

has representations $(\boldsymbol{\alpha}, \boldsymbol{T})$, $(\boldsymbol{\beta}, \boldsymbol{S})$, and $(\boldsymbol{\gamma}, \boldsymbol{R})$ given by

$$\boldsymbol{\alpha} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \qquad \boldsymbol{T} = \begin{pmatrix} -5 & 0 \\ 0 & -2 \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \frac{1}{5} & \frac{4}{5} \end{pmatrix} \qquad \boldsymbol{S} = \begin{pmatrix} -2 & 2 \\ 0 & -5 \end{pmatrix},$$

and

$$\boldsymbol{\gamma} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad \boldsymbol{R} = \begin{pmatrix} -3 & 1 & 1 \\ 1 & -4 & 2 \\ 1 & 0 & -6 \end{pmatrix}.$$

It is apparent from this example that representations for *PH* distributions do not necessarily have the same order. In fact, there must be a representation that has a smallest or *minimal order*. A representation that has minimal order is called a *minimal representation*. The representations $(\boldsymbol{\alpha}, \boldsymbol{T})$ and $(\boldsymbol{\beta}, \boldsymbol{S})$ above are minimal representations for the given *PH* distribution. Our example also shows that minimal representations are not necessarily unique. The *order of a PH distribution* is defined to be the order of any minimal representation.

2.3 Discrete Phase-type Distributions

Even though our discussion almost entirely concerns continuous *PH* distributions we present in this section an introduction to their discrete-time counterparts for completeness. For a more thorough treatment see Neuts [101, Chapter 2], or Latouche and Ramaswami [85, Section 2.5].

A *discrete phase-type* (PH_d) random variable is defined as the absorption time of an evanescent discrete-time Markov chain $\{Y_n\}$, with $n = 0, 1, 2, \ldots$, on a finite phase space $S = \{0, 1, 2, \ldots, p\}$ where phase 0 is absorbing. As for the continuous-time case we let the initial phase probability distribution be $(\alpha_0, \boldsymbol{\alpha}) = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ (with $\sum_{i=0}^{p} \alpha_i = 1$) and the phase transition probability matrix be $\boldsymbol{Q}$. In block matrix form the phase transition probability matrix for the Markov chain can be written as

$$\boldsymbol{Q} = \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{t} & \boldsymbol{T} \end{pmatrix}.$$

Here, $\mathbf{0}$ is a $1 \times p$ vector of zeros, $\boldsymbol{t} = (t_1, t_2, \ldots, t_p)'$ where, for $i = 1, 2 \ldots p, t_i = Q_{i0}$ is the absorption probability from phase i, and $\boldsymbol{T} = [T_{ij}]$ is a $p \times p$ matrix consisting of the transition probabilities, for $i, j = 1, 2, \ldots, p$, from phase i to j. Note that $\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}$. The PH_d distribution is said to have a *representation* $(\boldsymbol{\alpha}, \boldsymbol{T})$ of *order* p. As with continuous *PH* distributions, to ensure absorption with probability one, it is assumed that $\boldsymbol{I} - \boldsymbol{T}$ is nonsingular. Also, to ensure that there are no superfluous

phases, we assume that the matrix

$$\boldsymbol{Q}^* \;=\; \boldsymbol{T} + (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}\boldsymbol{\alpha}$$

is irreducible.

A PH_d distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ has probability mass function $\{p_k\}$ given by

$$\begin{aligned} p_0 &= \alpha_0 \\ p_k &= \boldsymbol{\alpha}\boldsymbol{T}^{k-1}(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}, \qquad k \geq 1. \end{aligned}$$

The distribution function, defined for $k = 0, 1, 2, \ldots$, is given by

$$F_k \;=\; 1 - \boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{e}.$$

The probability generating function, defined for $|z| \leq 1$, is given by

$$\begin{aligned} G(z) &= \sum_{k=0}^{\infty} p_k z^k \\ &= z\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{T})^{-1}(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e} + \alpha_0, \end{aligned} \tag{2.3.1}$$

which is a rational function. For $k = 1, 2, \ldots$, differentiating (2.3.1) k times with respect to z and letting $z = 1$ gives the kth factorial moment

$$m_k^* \;=\; k!\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{T})^{-k}\boldsymbol{T}^{k-1}\boldsymbol{e}.$$

Some examples of PH_d distributions are the geometric, mixture of geometric, and negative binomial distributions. Also, any distribution with finite support $\{p_0, p_1, \ldots, p_m\}$ is a PH_d distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ of order m with

$$\begin{aligned} \boldsymbol{\alpha} &= \begin{pmatrix} p_1 & p_2 & \ldots & p_m \end{pmatrix} \\ \boldsymbol{T} &= \boldsymbol{O}, \end{aligned}$$

where $\boldsymbol{O}$ is a $m \times m$ matrix of zeros. Thus, the binomial and hypergeometric distributions are PH_d distributions. The Poisson distribution, however, is not a PH_d distribution since it does not have a rational generating function.

2.4 Characterization of Phase-type Distributions

In this section we motivate a discussion of the characterization of *PH* distributions by addressing the following two problems:

P1. Given a function, defined for $u > 0$, of the form

$$f(u) \;=\; \sum_{i=1}^{n} q_i(u) e^{-\lambda_i u} \tag{2.4.1}$$

where, for $i = 1, 2, \ldots, n$, $q_i(u)$ is a real polynomial of degree n_i, and $\Re(\lambda_i) > 0$, when does it correspond to the density function of a *PH* distribution?

P2. If the function defined by (2.4.1) does correspond to the density function of a *PH* distribution, what is a minimal representation for it?

Alternatively, the two problems can be stated in terms of *LST*s:

P1′. Given a function, defined for $\lambda \in \mathbb{C}$ such that $\Re(\lambda) > -\delta$ where δ is a positive number, of the form

$$\phi(\lambda) \;=\; \frac{a_p\lambda^{p-1} + a_{p-1}\lambda^{p-2} + \ldots + a_1}{\lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} + \ldots + b_1} + \alpha_0, \tag{2.4.2}$$

where $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_p$ are all real and $0 \leq \alpha_0 < 1$, when does it correspond to the *LST* of a *PH* distribution?

P2′. If the function defined by (2.4.2) does correspond to the *LST* of a *PH* distribution, what is a minimal representation for it?

Neither of these two problems have been solved in complete generality in the literature. Generally, progress has only been made for particular classes of *PH* distributions such as the Coxian distributions, and then, usually only for small order. For example, O'Cinneide [107] answered P1 for a particular class of order three Coxian distributions. Dehon and Latouche [45] answered P1 for the class of all generalized hyperexponential distributions of algebraic degree three. In Chapter 7 we present an

algorithm that solves the first problem. The second problem, first posed by Neuts [101], has proven to be more difficult to solve.

Arguably, the most far-reaching *PH* characterization result is due to O'Cinneide [104].

Theorem 2.1 *A distribution defined on* $[0, \infty)$ *is a PH distribution if and only if*

1 it is the point mass at zero, or

2 it has

(a) a strictly positive density on $(0, \infty)$*, and*

(b) has a rational LST such that there exists a pole of maximal real part $-\gamma$ *that is real, negative, and such that* $-\gamma > \Re(-\xi)$ *where* $-\xi$ *is any other pole.*

The proof of the necessity of Theorem 2.1 was relatively elementary compared to the proof of its sufficiency. The proof of the theorem's sufficiency relied on a geometric construction of a *PH* representation given the conditions (a) and (b). O'Cinneide [104, Section 4] gave an example of such a construction which also served to demonstrate that the order of a *PH* distribution is greater than or equal to its algebraic degree. The same paper also contained an analogous result to Theorem 2.1 for PH_d distributions.

Maier [93] gave an alternative, algebraic proof of Theorem 2.1. He first proved the discrete version of the theorem and then, by utilizing a change of variable that transformed generating functions into *LST*s, showed the result to be true for *PH* distributions. Again, the proof was constructive and an example was given in the appendix.

Thus, given a *PH* distribution defined by either its density function or *LST*, it is possible, via an involved process, to construct a representation for it. This representation will not necessarily be minimal but would give an upper bound for the order of the *PH* distribution.

Aldous and Shepp [3] showed that the *PH* distribution of order p that has the smallest *coefficient of variation*, or ratio of variance to the square of the mean

$$c = \frac{m_2 - m_1^2}{m_1^2}, \tag{2.4.3}$$

is the Erlang distribution of order p and rate $\lambda > 0$. In this case $c = p^{-1}$. Consequently, a lower bound for the order of any *PH* distribution is c^{-1}.

O'Cinneide [105] showed that if the *LST* of a *PH* distribution has a pole of maximal real part $-\lambda_1$ and complex conjugate poles $-\lambda_2 \pm i\theta$ with $\theta > 0$, then the order of the *PH* distribution p satisfies

$$p \geq \frac{\pi\theta}{\lambda_2 - \lambda_1}. \tag{2.4.4}$$

As a result, the order of a *PH* distribution increases without bound as the real part of a pair of complex conjugate poles approaches the pole of maximal real part from below. In addition, O'Cinneide [105] conjectured that as the parameters of a *PH* distribution are altered so that its density function approaches the horizontal axis its order increases without bound.

Commault and Chemla [38] completely characterized all *PH* distributions that have *LST*s of the form

$$\phi(\lambda) = \frac{\lambda_1(\lambda_2^2 + \theta^2)}{(\lambda + \lambda_1)(\lambda + \lambda_2 + i\theta)(\lambda + \lambda_2 - i\theta)}. \tag{2.4.5}$$

They proved that (2.4.5) is the *LST* of a *PH* distribution if and only if $\lambda_2 > \lambda_1$. Furthermore, they showed that (2.4.5) is the *LST* of an order three *PH* distribution if and only if

$$\theta \leq \frac{\lambda_2 - \lambda_1}{\sqrt{3}}.$$

Commault and Chemla [38] proved a number of other results which stated, or placed lower bounds on, the order of a *PH* distribution given its *LST*. The results, however, were restricted to specific cases. In particular, they showed that the difference in degrees between the denominator and the numerator of the *LST* of a *PH*

distribution equals the minimum number of transient states visited before absorption in the Markov chain governed by $\boldsymbol{\alpha}$ and $\boldsymbol{T}$. This places a lower bound on the order of any *PH* distribution but if the difference is small little can be said about it.

More recently, Commault and Mocanu [39] showed that any order p *PH* representation of some prespecified structure is a minimal representation for a *PH* distribution of algebraic degree p for almost all admissible nonzero parameter values of the representation. The set of all parameter values giving rise to *PH* distributions of algebraic degree less than p therefore has measure zero. Consequently, any *PH* distribution that has order greater than its algebraic degree would have arisen not from a particular structure of higher order representation, but rather from particular parameter values. To illustrate this, Commault and Mocanu [39] considered the *PH* distribution with *LST*

$$\phi(\lambda) \;=\; \frac{5}{(\lambda+1)(\lambda^2+4\lambda+5)},$$

which has poles $\lambda_1 = -2+i$, $\lambda_2 = -2-i$, and $\lambda_3 = -1$. The algebraic degree of the *PH* distribution is three, but (2.4.4) implies that its order must be greater than three. In fact, an order-four representation is

$$\boldsymbol{\alpha} \;=\; \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{T} \;=\; \begin{pmatrix} -2 & 2 & 0 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & -2 & 2 \\ \frac{1}{8} & 0 & 0 & -2 \end{pmatrix},$$

which has a unicyclic structure. It is these particular parameter values of the representation that give an algebraic degree of three for the *PH* distribution. If the nonzero parameters are perturbed slightly (keeping the same unicyclic structure) by letting, for example, for all admissible $\epsilon > 0$,

$$\boldsymbol{\alpha} \;=\; \begin{pmatrix} \frac{1}{3}-\epsilon & \frac{2}{3}+\epsilon & 0 & 0 \end{pmatrix},$$

then the *PH* distribution with such a representation has an algebraic degree of four.

Before stating the characterization theorem equivalent to Theorem 2.1 for Coxian distributions we state the following rather remarkable result due to Cumani [42], and Dehon and Latouche [45].

Theorem 2.2 *The classes of TPH distributions, Coxian distributions, and mixtures of convolutions of exponential distributions are identical.*

Later, O'Cinneide [103] proved the same result using the concepts of *PH*-simplicity and *PH*-majorization. A *PH*-generator $\boldsymbol{T}$ is said to be *PH-simple* if every *PH* distribution that has $\boldsymbol{T}$ as its generator has a unique representation of the form $(\boldsymbol{\alpha}, \boldsymbol{T})$. A *PH*-generator $\boldsymbol{T}$ is said to *majorize* another *PH*-generator $\boldsymbol{S}$ if any *PH* distribution with generator $\boldsymbol{S}$ has a representation of the form $(\boldsymbol{\alpha}, \boldsymbol{T})$. Both Cumani [42] and O'Cinneide [103] gave an algorithm for finding, from a *TPH* representation, a Coxian representation of the same order. Coxian representations are very useful because they can be defined with only $2p$ parameters, their generators are *PH*-simple, and they are dense in the class of all distributions defined on the nonnegative real numbers.

The following theorem is due to O'Cinneide [106].

Theorem 2.3 *A distribution defined on* $[0, \infty)$ *is a Coxian distribution if and only if*

1 it is the point mass at zero, or

2 it has

(a) a strictly positive density on $(0, \infty)$*, and*

(b) has a rational LST with only real, negative poles.

O'Cinneide [107] defined the *triangular order* of a Coxian distribution to be the order of its minimal Coxian representation. The minimal Coxian representation is

unique because, as remarked above, Coxian generators are *PH*-simple. The triangular order of a Coxian distribution does not, however, necessarily equal its order as the following example demonstrates. Botta, Harris, and Marchal [28] showed that the *PH* distribution with representation

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -5 & 0 & \frac{1}{8} \\ 4 & -4 & 0 \\ 0 & 1 & -1 \end{pmatrix},$$

whose *LST* has only real poles, can only have a Coxian representation of order greater than three. Thus, in general, all that can be said about a *PH* distribution whose *LST* has only real poles is that it is a Coxian distribution of *some* order. We therefore have for Coxian distributions

$$\text{algebraic degree} \quad \leq \quad \text{order} \quad \leq \quad \text{triangular order.}$$

O'Cinneide [107] completely characterized the class of all Coxian distributions with density function, defined for $u > 0$, of the form

$$f(u) = (c_1 \frac{u^2}{2} + c_2 u + c_3) e^{-\mu u}. \tag{2.4.6}$$

where $\mu > 0$.

Theorem 2.4 *A Coxian distribution with density function of the form (2.4.6) is a PH distribution if and only if*

1 $c_1 + \mu c_2 + \mu^2 c_3 = \mu^2(1 - \alpha_0)$,

2 $c_1, c_3 \geq 0$, *and*

3 $c_2 > -\sqrt{2c_1 c_3}$.

Furthermore, if $c_2 \geq 0$ *then the triangular order* p *of the distribution is three, otherwise it is given by*

$$p = 3 + \left\lceil \frac{c_2^2}{2c_1 c_3 - c_2^2} \right\rceil,$$

where $\lceil x \rceil$ *denotes the least integer greater than or equal to* x*.*

As a corollary to Theorem 2.4, O'Cinneide [107] showed that the Coxian distribution with density function given by

$$f(u) \;=\; \frac{((u-a)^2+\epsilon)e^{-u}}{a^2-2a+2+\epsilon},$$

where $a, \epsilon > 0$, has triangular order

$$p \;=\; 3+\left\lceil \frac{a^2}{\epsilon} \right\rceil,$$

which increases without bound as $\epsilon \to 0$. In this example we have, as the parameter ϵ approaches zero, the density function approaching the horizontal axis and the triangular order of the *PH* distribution becoming arbitrarily large.

2.5 Closure Properties of Phase-type Distributions

To complete our introduction to *PH* distributions in this section we discuss the closure properties of the class of *PH* distributions.

Theorem 2.5 *Suppose that* F *and* G *are PH distributions with representations* $(\boldsymbol{\alpha}, \boldsymbol{T})$ *of order* p*, and* $(\boldsymbol{\beta}, \boldsymbol{S})$ *of order* q*, respectively. Then we have the following.*

1. *The convolution* $F * G$ *is a PH distribution with a representation* $(\boldsymbol{\gamma}, \boldsymbol{R})$ *of order* $p+q$ *where*

$$\boldsymbol{\gamma} \;=\; \begin{pmatrix} \boldsymbol{\alpha} & \alpha_0\boldsymbol{\beta} \end{pmatrix}$$

$$\boldsymbol{R} \;=\; \begin{pmatrix} \boldsymbol{T} & -\boldsymbol{T}\boldsymbol{e}\boldsymbol{\beta} \\ \mathbf{0} & \boldsymbol{S} \end{pmatrix},$$

and $\mathbf{0}$ *is a* $p \times q$ *matrix of zeros.*

2. *The mixture* $\theta F + (1-\theta)G$, *where* $0 \leq \theta \leq 1$, *is a PH distribution with a representation* $(\boldsymbol{\gamma}, \boldsymbol{R})$ *of order* $p+q$ *where*

$$\boldsymbol{\gamma} = \begin{pmatrix} \theta\boldsymbol{\alpha} & (1-\theta)\boldsymbol{\beta} \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{pmatrix},$$

and $\boldsymbol{0}$ *is the matrix of zeros of appropriate dimension.*

3. *If* F^{*k} *denotes the k-fold convolution of* F *and* $\{p_k\}$ *is a* PH_d *distribution with a representation* $(\boldsymbol{\delta}, \boldsymbol{N})$ *of order* n, *the infinite mixture of convolutions*

$$H \equiv \sum_{k=0}^{\infty} p_k F^{*k}$$

is a PH distribution with a representation $(\boldsymbol{\gamma}, \boldsymbol{R})$ *of order* pn *where*

$$\boldsymbol{\gamma} = \boldsymbol{\alpha} \otimes \boldsymbol{\delta}(\boldsymbol{I} - \alpha_0 \boldsymbol{N})^{-1} \tag{2.5.1}$$

$$\boldsymbol{R} = \boldsymbol{T} \otimes \boldsymbol{I} - \boldsymbol{T}\boldsymbol{e}\boldsymbol{\alpha} \otimes (\boldsymbol{I} - \alpha_0 \boldsymbol{N})^{-1}\boldsymbol{N}. \tag{2.5.2}$$

Here, $\boldsymbol{I}$ *is the* $n \times n$ *identity and* $\otimes$ *denotes the Kronecker product which is defined in Steeb [135, page 55].*

Proof. See Neuts [101].

The proof in Neuts [101] is a formal one. Latouche and Ramaswami [85, Section 2.6] gave a more intuitive proof for the discrete case by considering the distribution of the absorption time of the underlying Markov chain associated with each of the three operations defined in Theorem 2.5. The proof of the continuous case was not given but is similar. Statement 3 in Theorem 2.5 is not necessarily true if the discrete distribution is not PH_d. Latouche and Ramaswami [85, page 56] provided an example where F is the exponential distribution and the discrete distribution is defined, for $k = 1, 2, \ldots$, by

$$p_k = \frac{1}{k} - \frac{1}{k+1}.$$

The resultant distribution is not *PH* and does not even have a rational *LST*.

Assaf and Langberg [16] showed that any *PH* (Coxian) distribution is a proper mixture (that is, $0 < \theta < 1$ in Statement 2 of Theorem 2.5) of two distinct *PH* (respectively, Coxian) distributions. Thus, the class of all *PH* (Coxian) distributions contains no extreme distributions.

Maier and O'Cinneide [94] proved the following *PH* characterization result:

Theorem 2.6 *The class of all PH distributions is the smallest class of distributions defined on* $[0, \infty)$ *that*

1 contains the point mass at zero and all exponential distributions,

2 is closed under the operations of finite convolution and mixture, and

3 is closed under the operation

$$H \equiv \sum_{k=0}^{\infty} (1-\xi)^k \xi F^{*(k+1)}, \tag{2.5.3}$$

where F^{*l} *denotes the l-fold convolution of the PH distribution* F *and* $0 < \xi \leq 1$.

Maier and O'Cinneide [94] also proved an analogous result for PH_d distributions.

Assaf and Levikson [17] proved the corresponding result to Theorem 2.6 for Coxian distributions:

Theorem 2.7 *The class of all Coxian distributions is the smallest class of distributions defined on* $[0, \infty)$ *that*

1 contains the point mass at zero and all exponential distributions, and

2 is closed under the operations of finite convolution and mixture.

Starting with the point mass at zero and the set of all exponential distributions any Coxian distribution can be constructed from a finite sequence of convolution and mixture operations. In order to construct a *PH* distribution that is not Coxian

we must also include operations of the type (2.5.3) in the sequence. Consider the following. Let $(\boldsymbol{\alpha}, \boldsymbol{T})$ be a Coxian representation of order p. That is,

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_p \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_p \end{pmatrix}$$

where $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$. Let $(\boldsymbol{\delta}, \boldsymbol{N})$ be the minimal PH_d representation for the geometric distribution, that is, $\boldsymbol{\delta} = (1-\xi)$ and $\boldsymbol{N} = (1-\xi)$ where $0 < \xi \leq 1$. Applying the operation defined by (2.5.3) with $(\boldsymbol{\alpha}, \boldsymbol{T})$ and $(\boldsymbol{\delta}, \boldsymbol{N})$ gives, using (2.5.1) and (2.5.2), a unicyclic PH representation $(\boldsymbol{\gamma}, \boldsymbol{R})$ with

$$\boldsymbol{\gamma} = (1-\xi)(1-\alpha_0(1-\xi))^{-1}\boldsymbol{\alpha}$$

$$\boldsymbol{R} = \boldsymbol{T} - (1-\xi)(1-\alpha_0(1-\xi))^{-1}\boldsymbol{T}\boldsymbol{e}\boldsymbol{\alpha}$$

$$= \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_{p-1} & \lambda_{p-1} \\ \zeta\lambda_p\alpha_1 & \zeta\lambda_p\alpha_2 & \zeta\lambda_p\alpha_3 & \dots & \zeta\lambda_p\alpha_{p-1} & -\lambda_p(1-\zeta\alpha_p) \end{pmatrix},$$

where $\zeta = (1-\xi)(1-\alpha_0(1-\xi))^{-1}$. The representation $(\boldsymbol{\gamma}, \boldsymbol{R})$ requires only $2p+1$ parameters. It is also a minimal representation since every phase in the underlying Markov chain is used in contributing to the total absorption time.

O'Cinneide [108, Conjecture 4] conjectured that every PH distribution of order p has a unicyclic representation of the same order. So far this conjecture has been established only for PH distributions of order three.

A final result in this line was proved by Mocanu and Commault [97]. They showed that every *PH* distribution is a mixture of *monocyclic generalized Erlang* distributions. Monocyclic generalized Erlang distributions are constructed from convolutions of Erlang and *feedback Erlang* distributions. A feedback Erlang distribution has a representation $(\boldsymbol{\gamma}, \boldsymbol{R})$, where for $\lambda > 0$ and $0 < z < 1$,

$$\boldsymbol{\gamma} = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_p \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} -\lambda & \lambda & 0 & \dots & 0 & 0 \\ 0 & -\lambda & \lambda & \dots & 0 & 0 \\ 0 & 0 & -\lambda & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda & \lambda \\ z\lambda & 0 & 0 & \dots & 0 & -\lambda \end{pmatrix}.$$

2.6 Concluding Remarks

In this chapter we have introduced and discussed *PH* distributions, a versatile class of distributions defined on the nonnegative real numbers that add flexibility to stochastic modelling in many different areas. We have also seen that even though much has already been achieved in characterizing *PH* distributions there is still a lot more to be done. O'Cinneide [108] gave a survey of *PH* distributions and presented some open *PH* characterization problems. In fact, one of the problems, Conjecture 3, the *"steepest increase conjecture"* has already been proved by Yao [149]. The conjecture, now a theorem, is stated as follows:

> *"For any PH distribution of order p, with density function $f(u)$, $\frac{f(u)}{u^{p-1}}$ is nonincreasing for $u > 0$."*

In the next chapter we look at the problem of selecting the parameters of *PH* distributions when they are used to fit data or approximate probability distributions. As we shall see this important area is also under-explored and there are still many avenues to be investigated.

Chapter 3

Parameter Estimation and Distribution Approximation with Phase-type Distributions

3.1 Introduction

In this chapter we present a review of the literature concerned with the problem of using *PH* distributions to either fit empirical data or approximate probability distributions. In the first case it is assumed that the empirical data set, say, $\{z_1, z_2, \ldots, z_n\}$, is a collection of n independent realizations from a *PH* distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$. The aim of the *fitting* procedure is to estimate the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ so that they best fit the data in some sense. In *approximating* a probability distribution with a *PH* distribution, the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ need to be selected so that a predetermined function of the *approximated* distribution and the *approximating PH* distribution is minimized. Such a function measures the "distance" between the two distributions in some sense.

To date, the most common techniques used in estimating or selecting the parameters of *PH* distributions have been the methods of *maximum likelihood*, *moment*

matching, and *least squares*. For a description of these methods see Rice [117], Wackerly, Mendenhall, and Scheaffer [145], or any other elementary text on mathematical statistics. Two particularly good references on the method of least squares are Spiegel [132] and the Open University study guide on Least-Squares Approximation [141].

When using *PH* distributions for modelling, the phases can be thought of in two different ways. First, they can be viewed as purely fictitious, in which case the class of *PH* distributions provide a versatile, dense, and algorithmically tractable class of distributions defined on the nonnegative real numbers. Second, the phases, or blocks of phases, can represent something physical. In this case the model often determines the structure of the *PH* representation to be used. For example, Faddy [49] represented the time spent in a compartmental model, where a "particle" or "token" moves through a system of compartments, with a Coxian distribution. Compartmental models are used in drug kinetics where each compartment represents a body organ or system. The model used in Faddy [49] allowed for Erlang residency times in each compartment which could represent the amount of time it took a drug to clear the organ or system. An example was given where a two-compartment system was used to model the outflow of labelled red blood cells injected into a rat liver. The flexibility of *PH* distributions, however, allows for more complex models. In Faddy [51] a slightly more complex compartmental arrangement which allowed for some cycling was used to model diffusion and clearance of a drug in body organs. Faddy [52] also used a compartmental model to describe the failure and repair times of a power station's coal pulveriser. Each phase in the fitted Coxian distribution could be interpreted as a stage in the life of the machine or its repair process. Here we have an example where the phases are really fictitious but can be given a physical interpretation, see also Faddy and McClean [55]. Aalen [1] also presented a number of compartmental models used in survival analysis.

In order to standardize the performance evaluation of *PH* parameter estimation and distribution approximation algorithms the *Aalborg benchmark* was developed.

This benchmark originated at an international workshop on fitting *PH* distributions, held in Aalborg, Denmark, in February 1991, and was extended in Bobbio and Telek [25]. The extended benchmark consisted of nine distributions: two Weibull, three lognormal, and two uniform distributions, as well as a shifted exponential, and a matrix-exponential distribution. Five goodness of fit measures were also included: the area distance between the densities, the negative of the cross entropy, and the relative errors in the mean, standard deviation, and coefficient of skewness. For a description of the extended benchmark see Bobbio and Telek [25], or Horvath and Telek [72].

In Section 3.2 we describe some of the methods for *PH* parameter estimation and distribution approximation found in the literature. Section 3.3 contains a discussion on the problems encountered when using the current algorithms. We also discuss the work of Lang and Arthur [82] where two moment matching and two maximum likelihood algorithms were compared. We conclude the chapter in Section 3.4 and propose that some of the problems with *PH* fitting and approximation methods can be overcome by performing the estimation or approximation in the Laplace-Stieltjes transform domain.

3.2 Parameter Estimation and Distribution Approximation Methods for Phase-type Distributions

This section contains a brief description of some *PH* parameter estimation and distribution approximation methods. The survey is by no means complete and we refer the reader to the comprehensive reference lists given in Bobbio and Cumani [24], Johnson [73], and Asmussen, Nerman, and Olsson [15].

Asmussen, Nerman, and Olsson [15] (see also Asmussen [8]) developed an expectation-maximization (*EM*) algorithm (named EMPHT) to calculate maximum

likelihood parameter estimates for general *PH* distributions when fitted to empirical data. They adapted the algorithm so that it could also be used for distribution approximation with *PH* distributions. In a companion paper Olsson [110] extended the algorithm so that it could be used with right-censored and interval-censored data. The original and extended algorithms are available as the downloadable package EMpht[1], which is written in C.

The *EM* algorithm, explained in full generality in the seminal paper by Dempster, Laird, and Rubin [46], is an iterative scheme that finds maximum likelihood parameter estimates when there are incomplete data. The maximum likelihood estimation problem is formulated in such a way, that if the data were complete, then the calculation of the parameter estimates that maximize the loglikelihood (M-step) would be possible. But since the data are incomplete the sufficient statistics for the parameter estimates are replaced with their expected values (E-step). Starting with some initial values for the sufficient statistics the iterations alternate between the two steps until convergence, defined through some stopping criterion, is reached. For a comprehensive treatment of the *EM* algorithm and its applications see McLachlan and Krishnan [95].

Asmussen, Nerman, and Olsson [15] considered the whole sample path in an evanescent continuous-time Markov chain as a *complete* realization or observation of the process. Such an observation keeps a record of each state visited, in order, and the sojourn times in each one, until absorption. Each element of the empirical data set, however, is only the time to absorption of the process and is hence an *incomplete* observation. Given a set of complete observations it is relatively simple to derive the sufficient statistics needed to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{T}$. These are

1 the total number of observations starting in each phase,

2 the total time spent in each phase, and

3 the total number of jumps from one phase to another.

[1] http://www.maths.lth.se/matstat/staff/asmus/pspapers.html

From these sufficient statistics the maximum likelihood estimates for the *PH* parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ (M-step) can be calculated relatively easily. Calculating the expected values of the sufficient statistics (E-step) in order to perform the M-step proved to be much more involved and required the solution of a complicated set of differential equations. Their numerical solution needed the implementation of a Runge-Kutta method of fourth order, see Kreyszig [81, pages 947–949], or Tenembaum and Pollard [139, pages 653–658]. The related distribution approximation algorithm minimized the relative entropy between the approximated density and the approximating *PH* density. The implementation was similar to that of the data fitting algorithm. A number of examples where densities from the Aalborg benchmark were approximated with *PH* distributions of varying orders was given, as well as a number of examples of fits to empirical data. Plots of the approximating (or fitted) densities against the approximated density (respectively, histogram) were given for each example but no performance evaluation using the benchmark's goodness of fit measures was done.

Bobbio and Cumani [24] developed an algorithm to calculate maximum likelihood parameter estimates. They chose to restrict themselves to the class of Coxian distributions because

1 their representations are unique,

2 the number of parameters that need to be estimated is only $2p-1$ where p is the order of the representation (they assumed that there was no point mass at zero), and

3 the partial derivatives of the loglikelihood function, with respect to the distribution's parameters, are able to be calculated easily.

In order to choose the parameters that maximized the loglikelihood function the resulting nonlinear program was solved by combining a linear program with a line search at each iteration. The algorithm was developed to fit Coxian distributions

to empirical data with the option of including right-censored data. Continuous distribution functions could also be approximated by choosing suitable sample points. The package, written in FORTRAN, was named MLAPH. Bobbio and Telek [25] evaluated MLAPH against the extended Aalborg benchmark. They gave plots of each approximated density with accompanying approximating *PH* densities of orders 2, 4, and 8. The five performance measures mentioned in Section 3.1 were tabulated for each case and the results discussed.

Horvath and Telek [72] developed a method which separately approximated the main part and the tail of an arbitrary distribution defined on the nonnegative real numbers with a *PH* distribution. The main part of the distribution was approximated with a Coxian distribution by minimizing any distance (goal) function of the approximated and approximating densities. A nonlinear programming procedure similar to that of Bobbio and Cumani [24] was used to perform the minimization. The authors also stated that their method could be used with general *PH* distributions but they believed that Coxian distributions were just as flexible in practice and much easier to compute with (refer to points 1–3 in the previous paragraph). The tail was approximated with a hyperexponential distribution using a method proposed by Feldman and Whitt [58]. The algorithm was tested by using three separate distance functions against the extended Aalborg benchmark and two Pareto density functions. The three distance functions chosen were

1 the relative entropy,

2 the L_1 distance, and

3 the relative area distance

between the main part of the approximated density and the approximating Coxian density. Both Pareto distributions, and a uniform and a Weibull distribution from the Aalborg benchmark, were evaluated graphically. The performance measures for all of the distribution approximations were tabulated in the appendix and discussed.

They also gave two examples that compared the queue length distribution for the $M/G/1$ queue with that of the approximating $M/PH/1$ queue. The service time distributions used were the two abovementioned Pareto distributions.

Faddy [51], [52], and [53], Faddy and McClean [55], and Hampel [65] used maximum likelihood estimation to fit Coxian distributions to real data. They used existing MATLAB® or S-PLUS® routines (for example the Nelder-Mead algorithm in MATLAB®) to perform the required parameter estimation. Harris and Sykes [67] developed an algorithm to fit empirical data with generalized hyperexponential distributions using maximum likelihood estimation.

Johnson [73] (see also Johnson and Taaffe [74], [75], and [76] for the underlying theory) developed an algorithm MEFIT, written in FORTRAN, that matched the first three moments of a mixture of Erlang distributions to the respective moments of empirical data or a distribution. The fit or approximation could be improved by also matching up to six moments, up to 10 values of either the distribution or density functions, or up to 10 values of the Laplace transform. The nonlinear optimization program, which resulted from the parameter estimation or distribution approximation technique, was solved using the sequential quadratic programming package NPSOL, see Gill, Murray, Saunders, and Wright [60]. To illustrate the algorithm several examples where distributions were approximated with mixtures of Erlang distributions were given. The selection of examples were not from the Aalborg benchmark (probably due to the fact that most of the work was done prior to 1991) but included a lognormal and a uniform distribution, two Weibull distributions, and a mixture of two lognormal distributions. Each example was assessed with a plot of the approximated and approximating density functions (and corresponding distribution functions), and a quantile-quantile plot. Three performance measures, the area between the density functions, the area between the distribution functions, and the maximum deviation between the distribution functions, were also used in the evaluation. In addition, the $GI/M/1$ queue, with each of the abovementioned approximated distributions used as the interarrival-time distribution, was compared

with the respective approximating $PH/M/1$ queue. The performance measure used in the comparison was the steady-state mean queue length. Results for traffic intensities of 0.5 and 0.7 were given.

Schmickler [124] also developed a moment matching algorithm where the first three moments of a mixture of two or more Erlang distributions were matched exactly to the respective moments of an empirical distribution function. Higher order moments were matched approximately by minimizing the difference in area between the empirical and fitting distributions. This algorithm, unlike those discussed so far where the user needed to preselect the order of the fitting or approximating *PH* distribution, had the added feature of being able to determine the order of the fitting *PH* distribution. The Flexible Polyhedron Search method (that is, the Nelder-Mead algorithm) was used to solve the resulting nonlinear program. The fitting package, written in PASCAL, was named MEDA. Some examples of fits to empirical distributions were given.

Bux and Herzog [32] developed an algorithm that fitted Coxian distributions with a uniform rate to empirical data. They matched the first two moments and minimized the deviation between the fitting Coxian distribution function and the empirical cumulative distribution function at the data points. The authors noted that while their algorithm was efficient, the number of phases required for a close fit could be very large.

Faddy [49] and [50] used least squares to fit Coxian distributions to real sample data in order to estimate the parameters for compartmental models used in drug kinetics.

3.3 Problems with Phase-type Parameter Estimation and Distribution Approximation Methods

In this section we discuss some of the problems encountered when estimating or selecting the parameters of *PH* distributions using the various methods described in the previous section.

The literature concerned with comparing the performance evaluation of *PH* parameter estimation and distribution approximation algorithms is scant. Khoshgoftaar and Perros [78] compared three methods (maximum likelihood, moment matching, and minimizing a distance measure) to find the parameters of an order two Coxian distribution when approximating a distribution with coefficient of variation greater than one. They found that the moment matching method worked best for this particular problem, but when the technique was used to fit empirical data the other two methods performed better. Madsen and Nielsen [92] fitted *PH* distributions to two empirical data sets of holding times for traffic streams from the Danish packet-switched network PAXNET. They fitted mixtures of Erlang distributions using MEDA, Coxian distributions using a method due to Bobbio, Cumani, Premoli, and Saracco [26] (the precursor to MLAPH), and mixtures of Erlang distributions with identical rates by minimizing the sum of the deviations between the empirical and fitting distributions. They evaluated the distribution function fits graphically and with five performance measures: the sum of the deviations, the sum of the deviations squared, the maximum deviation, the area between the empirical and fitting distributions, and the first two moments. Another notable advance in the area of evaluating the performance of *PH* parameter estimation and distribution approximation methods is the work of Lang and Arthur [82].

Lang and Arthur [82] conducted a comprehensive evaluation of the programs EMPHT, MLAPH, MEFIT, and MEDA by comparing their performance when used

to approximate the distributions in the extended Aalborg benchmark. For each package they plotted the approximated densities of the Aalborg benchmark with approximating *PH* densities of varying orders. They evaluated each algorithm using the benchmark's five performance measures and gave detailed tables of results. In addition, the algorithms were assessed by using some qualitative measures. These were:

1. *Generality* - How well the algorithm coped with a variety of distribution approximation problems.

2. *Reliability* - Whether the algorithm worked properly or not.

3. *Stability* - Whether slightly altered starting values adversely affected the parameter estimates.

4. *Accuracy* - Whether errors were introduced due to rounding and/or iterations terminating.

5. *Efficiency* - How long the algorithm took to run.

They found that no particular *PH* parameter estimation or distribution approximation algorithm performed better than any other in all tested cases except that EMPHT took a lot longer to converge than any of the other algorithms. All of the methods approximated distributions that exhibited *PH* behaviour relatively well with *PH* distributions of low order. However, no method fitted non-*PH* distributions well even using *PH* distributions of high order.

Lang and Arthur [82] stated four main problems with using *PH* distributions to fit data or approximate distributions. These were:

1. The fitting or approximation problem is highly nonlinear.

2. The number of parameters that need to be estimated or selected is often large.

3. Representations of *PH* distributions are typically not unique.

4. The relationship between the parameters and the shape of a *PH* distribution is generally nontrivial.

The first problem is evident because the algorithms MLAPH, MEFIT, and MEDA all required complicated nonlinear programming routines to solve the resulting likelihood or moment equations. Also, EMPHT required a computer intensive E-step which used a Runge-Kutta method of fourth order.

The second problem is well known in the literature. Not only is the number of parameters to be estimated large for *PH* distributions even of modest order, their representations are generally overparameterized. The *LST* of a general *PH* distribution of order p has, in general, $2p$ parameters. Since every *PH* distribution has a unique *LST* (see Feller [59, page 430]) a general *PH* distribution of order p can be parameterized with $2p$ parameters. Asmussen [8] also demonstrated this fact with an argument using moments. Since the general *PH* representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ of order p has $p^2 + p$ parameters, general *PH* distributions are considerably overparameterized. This problem has implications for general *PH* fitting methods, such as EMPHT, which need to fit a higher number of parameters than is necessary. All of the other authors mentioned in Section 3.2 bypassed the problem of overparameterization by restricting themselves to Coxian distributions, or in the case of the tail approximation in Horvath and Telek [72], to hyper-exponential distributions whose representations also require only $2p$ parameters.

To complicate matters, given the *LST* of a *PH* distribution that has algebraic degree p, it is unknown, in all except the simplest cases, how to determine a *PH* representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ of minimal order for it. In fact, the *PH* distribution's order may be greater than p but still depend on only $2p$ parameters. In Section 2.4 we saw for Coxian distributions that

$$\text{algebraic degree} \;\leq\; \text{order} \;\leq\; \text{triangular order},$$

and that the example immediately following Theorem 2.4 gave a family of Coxian distributions that have algebraic degree three but arbitrary triangular order. It is

not known what happens to the order of such a family of Coxian distributions as the triangular order increases, except that it cannot exceed the triangular order. These facts suggest, albeit rather weakly, that a fitted general *PH* distribution may do just as well as, if not better than, a Coxian distribution of higher order. This ties in with the third problem, the nonuniqueness of *PH* representations, which is not well understood. Two distinct *PH* representations can be identified by simply comparing their Laplace-Stieltjes transforms. However, given a *PH* distribution in terms of its density function, Laplace-Stieltjes transform, or representation, it is not possible, in general, to determine a minimal representation for the distribution. A method that could fit general *PH* distributions of algebraic degree p (by estimating only $2p$ parameters) would be desirable, especially if in addition the *PH* representation of minimal order (with order greater than or equal to p) could be constructed from the $2p$ estimated parameters.

Faddy [51] and [53], and Hampel [65] found that there is even overparametrization when fitting Coxian distributions to data using maximum likelihood estimation, but in a different, practical sense. This overparameterization occurred when Coxian distributions with a number of free parameters were fitted to data using maximum likelihood estimation and then compared with Coxian fits that had fewer free parameters (but defined on the same parameter space).

Consider the following. Suppose a distribution, defined on the m-dimensional parameter space Θ, is fitted to a data set $\{z_1, z_2, \ldots, z_n\}$ which consists of n realizations of the independent and identically distributed random variables $Z_1, Z_2, \ldots, Z_n$. Write $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)$. Let $\theta \in \Theta$ and $\mathcal{L}(\theta, \boldsymbol{Z})$ be the loglikelihood function. Suppose that $\Theta_0 \subset \Theta_1$ are subsets of Θ with respective dimensions m_0 and m_1 with $m_0 < m_1 \leq m$. We say that Θ_0 is a *submodel* of Θ_1. The *likelihood ratio statistic*, which tests the null hypothesis H_0: $\theta \in \Theta_0$ versus the alternative hypothesis H_1: $\theta \in \Theta_1 \backslash \Theta_0$, is defined as

$$\lambda(\boldsymbol{Z}) \;=\; \frac{\max\limits_{\theta \in \Theta_0} \mathcal{L}(\theta, \boldsymbol{Z})}{\max\limits_{\theta \in \Theta_1} \mathcal{L}(\theta, \boldsymbol{Z})}.$$

Wilks [147] showed that under H_0, $-2\log\lambda(\boldsymbol{Z})$ has a $\chi^2_{m_1-m_0}$ distribution, see also Strawderman [136].

In Faddy [51], when Coxian distributions were fitted to data using maximum likelihood estimation, it was found that some of the estimated parameters were nearly identical and others nearly equal to zero. Upon fitting a Coxian distribution with a structure that constrained these parameter values accordingly (the submodel), the loglikelihood did not decrease appreciably. For example, when an order three Coxian distribution with five free parameters was fitted to a particular data set the loglikelihood was -496.96. The Coxian fit where two of the parameters were constrained to be equal (a 4-parameter model) gave a loglikelihood of -497.15. Hampel [65] fitted the same data set with an order three 5-parameter Coxian distribution and then proceeded to look for parameter redundancies. He then fitted a number of 4-parameter submodels, and after performing an hypothesis test for each one, selected the model with the largest p-value (from the appropriate χ^2 distribution). After repeating the process another two times an order three 2-parameter fit with a loglikelihood of -497.36 was achieved. This compared with an order two 3-parameter fit with a loglikelihood of -497.52. Although this difference may not be significant, it suggests that more flexibility in fitting Coxian and *PH* distributions may be achieved by increasing the order of the representation rather than its number of free parameters.

Faddy [53] further illustrated this last point by fitting a Coxian distribution to a data set that contained the inter-eruption times of the Old Faithful geyser in Yellowstone National Park (see Silverman [130] for the data), which according to Asmussen, Nerman, and Olsson [15] "... is a notoriously difficult example in density estimation ...". The fitted Coxian distribution had an order of 397 but with only four free parameters! The loglikelihood was -100.1937. The fit was visually very good. More recently Faddy [56] achieved a fit to the same data set with a mixture of three Erlang distributions of total order 482 and five free parameters. The loglikelihood was slightly less than -100.1937. In fact, in this case a mixture

of three gamma distributions were fitted to the data (loglikelihood -100.0717) and the "a" parameters were rounded to give the number of phases in each Erlang distribution.

This particular type of overparameterization, so far, has only been investigated experimentally in the literature, and then in a limited way. Realizing this inherent parameter redundancy Faddy [54] developed a penalized maximum likelihood fitting method which penalized Coxian distributions that had disparate eigenvalues. Some examples using the technique with varying penalty parameters were used to fit real data with Coxian distributions.

The only *PH* parameter estimation or density approximation algorithm described in Section 3.2 that estimated the order of the *PH* distribution was MEDA. The other algorithms relied on the user to preselect the *PH* distribution's order. The literature addressing this problem is meagre. Rydén [122] developed a method to estimate the order of a *PH* distribution when fitted to data. His approach used a penalized likelihood method such as the Akaike information criterion (see Akaike [2]) or the Bayesian information criterion (see Schwarz [126]). He found that asymptotically the procedure never underestimated the order. A similar technique was also developed for Markov-modulated Poisson processes in the same paper. The work relied on an earlier paper by Rydén [120] where a similar penalized likelihood method was developed to estimate the order of hidden Markov models.

There is scant literature addressing Problem 4, the relationship between the parameters and the shape of the *PH* distribution. One notable exception, however, is Johnson and Taaffe [75]. They considered the four classes of *PH* distributions:

1. Mixtures of Erlangs of common order.

2. Mixtures of Erlangs.

3. Coxian.

4. General *PH*.

They applied the moment matching techniques developed in Johnson and Taaffe [74] and [76] to investigate how the order and class affected the shape of the *PH* density function. The aim of their investigation was to explore the flexibility of each class of *PH* distribution. They found that the first two classes exhibited sufficient flexibility to be on a par with the other two classes, but with fewer parameters.

3.4 Concluding Remarks

In this chapter we have given a survey of the methods used to estimate or select the parameters for *PH* distributions and discussed some of the shortcomings associated with them, in particular the problem of overparameterization.

In order to solve some of the problems when using *PH* distributions to fit empirical data or approximate probability distributions we propose that the estimation or approximation be carried out in the Laplace-Stieltjes transform (LST) domain. As mentioned in Section 3.3 *PH* distributions of algebraic degree p are parameterized by only $2p$ parameters. In order to overcome the overparameterization problem, at least in the representation sense, parameter estimation or distribution approximation in the LST domain may be a viable option. Also, since the LST of a *PH* distribution is a rational function, a technique that is linear may be able to be developed. The problem of determining a *PH* representation from the estimated or approximated LST still remains. However, the fitting or approximation problem, and the representation problem have been separated into two stages and can therefore be expected to be simpler. In the next chapter we shall explore the possibility of *PH* parameter estimation and distribution approximation in the LST domain.

Chapter 4

Parameter Estimation and Distribution Approximation in the Laplace-Stieltjes Transform Domain

4.1 Introduction

When using *PH* distributions to fit data or approximate distributions we proposed at the end of Chapter 3 that the estimation or approximation be carried out in the Laplace-Stieltjes transform (LST) domain. The motivation for this proposal is twofold. First, the number of parameters required to completely define the LST of a *PH* distribution (and hence any representation for the distribution) is $2p$ where p is its order. Thus, the problem of the first type of overparameterization discussed in Section 3.3 is avoided. Second, since the LST of a *PH* distribution is a rational function, a method for finding the parameters is likely to require only the solution of linear equations and hence be relatively easy to implement. In this chapter, after a brief literature review, we explain in detail two LST parameter estimation, or

distribution approximation, methods found in Harris and Marchal [66] and discuss some of the problems associated with them.

The idea of parameter estimation and distribution approximation in the transform (including the *LST*) domain is not new. Often such techniques have been employed when standard methods are intractable, or at least difficult to implement, in some manner. Bar-Lev, Barkan, and Langberg [18] stated that there are essentially two approaches to take when finding transform parameters:

> *"In one approach, the estimator for the unknown parameter is chosen to minimize a certain distance measure between the theoretical and empirical transforms. In the second approach, the estimator is taken to be the solution of an equation obtained by equating the theoretical transform with its empirical counterpart."*

The first approach is akin to the methods of maximum likelihood and least squares, whereas the second approach is similar to the method of moment matching. In fact, moment matching is essentially a transform estimation or approximation technique because the *LST* or characteristic function of a probability distribution is closely related to its moment generating function. Bar-Lev, Barkan, and Langberg [18] also realized that the success of transform estimation methods based on the second approach rely on the sample size and the transform variables at which the empirical and theoretical transforms are equated.

The literature on *LST* and general transform estimation and approximation is sizable, and the brief selection presented here is by no means complete. The majority of the papers cited contain their own literature reviews and should be consulted for a thorough treatment of the topic.

Paulson, Holcomb, and Leitch [112] estimated the parameters for the stable laws, given a randomly generated sample, by minimizing a weighted L_2 distance between the empirical and theoretical characteristic functions. The approach was extended in Heathcote [69] where some properties of the parameter estimators were established. Bryant and Paulson [31] also used a weighted L_2 distance between the characteristic

functions to estimate mixing proportions. Titterington, Smith, and Makov [142, Section 4.6] gave a detailed discussion on fitting various transforms to empirical transforms using a weighted least squares distance function. Their particular area of interest was in fitting mixtures of distributions to data. They also gave some asymptotic results for the transform estimators. There are a number of other references that address parameter estimation using the abovementioned "first" approach with transforms, particularly with characteristic functions, see Titterington, Smith, and Makov [142], Bar-Lev, Barkan, and Langberg [18], or Yao and Morgan [148] for comprehensive reference lists.

It appears that in the literature the second approach is used mostly with parameter estimation procedures that use the Laplace transform rather than the characteristic function. Schuh and Tweedie [125] developed three methods where the Laplace transform of a time-evolving model was used to estimate the model's parameters. They gave a number of examples to illustrate their approach including a numerical example where a time-evolving model was used to describe the lifetime of a sheep parasite. Feigin, Tweedie, and Belyea [57] further developed the Laplace transform estimation procedure to incorporate weighted Laplace transforms. The procedure was named *weighted area estimation*. The two particular weight functions they chose resulted in the Laplace transform estimators of Schuh and Tweedie [125], and moment estimators. Examples of parameter estimation for multi-stage time-evolving models were given including the previously mentioned sheep parasite case. Also, the asymptotic efficiency of weighted area estimation was shown to compare favourably with maximum likelihood estimation. In Tweedie, Zhu, and Choy [143] the weighted area technique was used to estimate the interarrival and service rates in a transient $M/M/1$ queue by only observing the queue length at fixed times. Hoeting and Tweedie [71] developed estimators for the parameter variances when weighted area estimation was used in multi-stage developmental models. The problem of selecting the transform variables at which to perform the weighted area estimation was addressed in Laurence and Morgan [86]. Yao and Morgan [148] applied the tech-

nique to estimate the parameters for indexed stochastic models. Bar-Lev, Barkan, and Langberg [18] used the second approach to develop a moment estimator for an exponential family defined on the real numbers.

In this chapter we will focus on two Laplace transform estimation or approximation procedures due to Harris and Marchal [66] because they specifically use rational LSTs. It is interesting to note, however, that Harris and Marchal [66] did not cite any of the abovementioned references on transform estimation in their bibliography. They preferred to draw upon the literature concerned with PH parameter estimation and distribution approximation (see Chapter 3), and fitting sums of exponential functions to data, see, for example, Parsons [111], Yeramian and Claverie [150], and Kammler [77]. In Section 4.2 we define the terminology needed for the rest of the chapter. Sections 4.3 and 4.4 detail Harris and Marchal's two methods, and a discussion of some of the associated problems appears in Section 4.5.

4.2 Preliminaries

In this section we define the various LSTs required for the subsequent sections of this chapter and make some observations.

The LST, defined for $\lambda \in \mathbb{C}$ with $\Re(\lambda) > -\delta$ where δ is a positive number, of a nonnegative random variable X that has distribution function $F(u)$, is given by

$$\begin{aligned} \phi(\lambda) &= \mathbb{E}[e^{-\lambda X}] \\ &= \int_0^\infty e^{-\lambda u} dF(u) \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} m_k \lambda^k, \end{aligned} \tag{4.2.1}$$

where $\mathbb{E}$ denotes the expectation operator, m_0 is defined to be one, and for $k = 1, 2, \ldots$, m_k is the kth noncentral moment of X. For $k = 1, 2, \ldots$, the kth noncentral moment can be calculated from the kth derivative of $\phi(\lambda)$ evaluated at $\lambda = 0$. That

is,

$$m_k \;=\; (-1)^k \phi^{(k)}(0).$$

If the random variable X has a *PH* distribution then the *LST* can be expressed as the ratio of two irreducible polynomials. That is,

$$\phi(\lambda) \;=\; \frac{1 + c_1\lambda + c_2\lambda^2 + \ldots + c_N\lambda^{N-1}}{1 + d_1\lambda + d_2\lambda^2 + \ldots + d_N\lambda^N}, \tag{4.2.2}$$

where the numerator and denominator have no factors in common, N is a positive integer, and $d_N \neq 0$. The constant terms in the numerator and denominator are both one to ensure that the corresponding distribution is not defective. Equation (4.2.2) is equivalent to (2.4.2) with the point mass at zero α_0 equal to 0, which we assume to be the case throughout this chapter. Consequently, we will refer to (4.2.2) as a rational *Laplace* transform (*RLT*). If a positive point mass at zero is required the two parameter estimation or distribution approximation methods described in this chapter can be easily adapted. This form for the *RLT* of a *PH* distribution has been chosen because it is the one used in Harris and Marchal [66]. As in Chapter 2 we refer to N as the *algebraic degree* of the *RLT*.

Not every *RLT* of the form (4.2.2) corresponds to a distribution, *PH* or otherwise. In addition to the conditions stated immediately after (4.2.2) we also require that

1 there exists a pole of maximal real part that is real and negative, and

2 the corresponding density function $f(u)$ is nonnegative for $u > 0$,

see Zemanian [151]. Also, Condition 2 implies, for $\lambda \geq 0$, that

$$\phi(\lambda) \;=\; \int_0^\infty e^{-\lambda u} f(u)du \;>\; 0.$$

The inequality is strict because $f(u) > 0$ for some $u > 0$ as $\int_0^\infty f(u)du = 1$. Also, every *RLT* of the form (4.2.2) that corresponds to a distribution does not necessarily correspond to a *PH* distribution. For *PH* distributions we also require that

1 there exists a pole of maximal real part $-\gamma$ that is real, negative, and such that $-\gamma > \Re(-\xi)$ where $-\xi$ is any other pole, and

2 the corresponding density function $f(u)$ is positive for $u > 0$,

see Theorem 2.1. Consequently, we will be performing the parameter estimation or distribution approximation within the class of distributions that have rational Laplace transform (see Cox [41]) rather than within the class of *PH* distributions. Of course the fitted or approximated density may be *PH* but this will not necessarily be the case.

The *empirical cumulative distribution function* of a data set $\{z_1, z_2, \ldots, z_n\}$, defined for $u \geq 0$, is given by

$$G(u) \;=\; \frac{1}{n}\#\{z_i | z_i \leq u\},$$

where # denotes the cardinality of the given set. The empirical cumulative distribution function is a right continuous step function that gives the proportion of data that are less than or equal to $u \geq 0$. The *empirical Laplace-Stieltjes transform* (*ELST*) of $G(u)$, defined for $\lambda \in \mathbb{C}$ such that $\lambda > -\delta$ where δ is a positive number, is given by

$$\begin{aligned}
\psi(\lambda) \;&=\; \int_0^\infty e^{-\lambda u} dG(u) \\
&=\; \sum_{i=1}^{n} e^{-\lambda z_i}(G(z_i) - G(\lim_{u \to z_i^-} u)) \\
&=\; \frac{1}{n}\sum_{i=1}^{n} e^{-\lambda z_i}. \qquad (4.2.3)
\end{aligned}$$

4.3 Harris and Marchal's Method 1

In this section we describe Method 1 of Harris and Marchal [66], discuss an example, and highlight some problems with the method.

In order to find the parameters of a *RLT* of the form (4.2.2) when fitting the *ELST* of a data set or approximating the *LST* of a probability distribution, we first calculate, for $k = 1, 2, \ldots, 2N-1$, the kth noncentral moment m_k. Next, we truncate (4.2.1) after $2N$ terms and equate it to (4.2.2). That is,

$$\sum_{k=0}^{2N-1} \frac{(-1)^k}{k!} m_k \lambda^k = \frac{\sum_{k=0}^{N-1} c_k \lambda^k}{\sum_{k=0}^{N} d_k \lambda^k}. \tag{4.3.1}$$

Note that $c_0 = d_0 = 1$. Upon multiplying (4.3.1) through by the denominator of the right hand side, expanding, and equating the coefficients of $\lambda, \lambda^2, \ldots, \lambda^{2N-1}$ (equating the constant terms just gives $1 = 1$), we get a system of $2N-1$ linear equations in the $2N-1$ unknowns $c_1, c_2, \ldots, c_{N-1}, d_1, d_2, \ldots, d_N$. The terms with $\lambda^{2N}, \lambda^{2N+1}, \ldots, \lambda^{3N-1}$ are ignored since only $2N-1$ equations are required to find the $2N-1$ parameters. Now, for $i = 0, 1, \ldots, N-1$,

$$\sum_{j=0}^{N-1} \frac{(-1)^{j+i}}{(j+i)!} m_{j+i} d_{N-j} = -\frac{(-1)^{N+i}}{(N+i)!} m_{N+i}, \tag{4.3.2}$$

and for $i = 1, 2, \ldots, N-1$,

$$c_i = \sum_{j=0}^{i} \frac{(-1)^j}{j!} m_j d_{i-j}. \tag{4.3.3}$$

For $k = 1, 2, \ldots, 2N-1$, differentiating the left hand side of (4.3.1) k times and letting $\lambda = 0$ gives m_k. Thus, Method 1 matches the first $2N-1$ moments of the empirical data or distribution with the first $2N-1$ derivatives of the fitting or approximating *RLT* evaluated at zero. Harris and Marchal [66] did not address the possibility of (4.3.2) and (4.3.3) not having a unique solution but altering N ought to remedy the problem in most cases.

In order to determine the minimum algebraic degree N of the *RLT* used to achieve a reasonable fit or approximation to the target *LST*, Harris and Marchal [66] proposed that N be increased until there is no appreciable gain in the algebraic

degree of the fitting or approximating *RLT*. This would be determined by either the cancellation of factors in the numerator and denominator or by the coefficient d_N becoming close to zero.

To illustrate Method 1, Harris and Marchal [66] approximated the truncated *LST*s of two gamma distributions with *RLT*s. The two approximations were assessed graphically and both looked quite good. Despite this, one of the approximating *RLT*s did not correspond to a distribution - the pole of maximal real part was not real. This problem will be further highlighted in the next example.

In order to demonstrate Method 1 when used for fitting data we fitted a sample consisting of 100 randomly generated realizations from the *PH* distribution with representation

$$\begin{aligned} \boldsymbol{\alpha} &= \begin{pmatrix} 0.6 & 0.2 & 0.2 \end{pmatrix} \\ \boldsymbol{T} &= \begin{pmatrix} -5 & 5 & 0 \\ 0 & -3 & 2 \\ 1 & 0 & -2 \end{pmatrix}, \end{aligned}$$

with a *RLT* of algebraic degree five. The *LST* for the *PH* distribution is

$$\phi(\lambda) = \frac{1 + \frac{11}{25}\lambda + \frac{1}{50}\lambda^2}{1 + \frac{31}{20}\lambda + \frac{1}{2}\lambda^2 + \frac{1}{20}\lambda^3},$$

which has zeros at -19.426 and -2.5739, and poles at -0.8663 and $-4.5669 \pm 1.4938i$. A histogram of the data is shown in Figure 4.3.1 and the empirical cumulative distribution is shown in Figure 4.3.2.

The *RLT* fitted by Method 1 was

$$\phi(\lambda) = \frac{1 + 0.1455\lambda + 0.0166\lambda^2 - 0.1335\lambda^3 - 0.0027\lambda^4}{1 + 1.2585\lambda + 0.4273\lambda^2 - 0.1386\lambda^3 - 0.1366\lambda^4 - 0.0294\lambda^5}$$

The *RLT* has zeros at $2.1532, -49.9833$, and $-1.0240 \pm 1.5566i$, and poles at $2.1543, -1.5060 \pm 1.3777i$, and $-1.8924 \pm 0.4523i$. Positive poles are disallowed but in this case the nearly identical factors, $(\lambda - 2.1532)$ and $(\lambda - 2.1543)$, in the numerator and denominator, respectively, can be cancelled. The occurrence of nearly

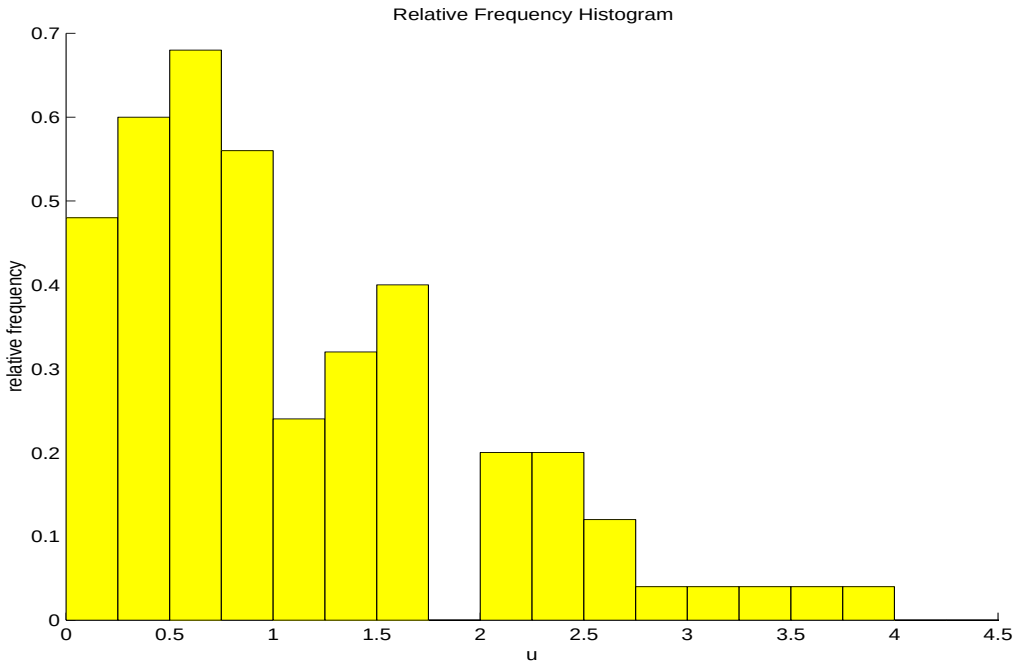

Figure 4.3.1: Histogram of the *PH* data

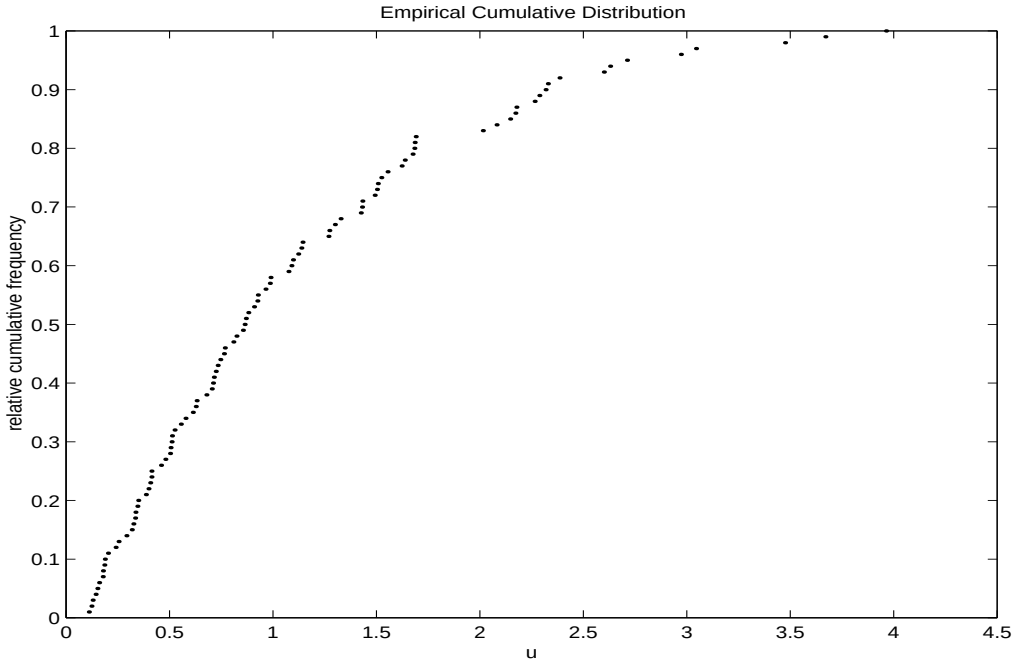

Figure 4.3.2: Empirical cumulative distribution of the *PH* data

identical factors seems to occur frequently in the examples given in Harris and Marchal [66] but the authors only state, without justification, that "Should the poles be found in the positive half-plane, because of analyticity they will be matched by roots in the positive half-plane, and each factor can be divided out of the numerator and denominator ...". After cancelling the factors and dividing the numerator and denominator through by their respective constant terms (to ensure that $m_0 = 1$) the resultant *RLT* of algebraic degree four is

$$\phi(\lambda) = \frac{1 + 0.6099\lambda + 0.2999\lambda^2 + 0.0058\lambda^3}{1 + 1.7227\lambda + 1.2270\lambda^2 + 0.4309\lambda^3 + 0.0634\lambda^4}. \tag{4.3.4}$$

Figure 4.3.3 shows the fitted *RLT* plotted with the *ELST* of the *PH* data. The fit looks very good. However, since the pole of maximal real part of the *RLT* is not real it *cannot* correspond to a distribution, *PH* or otherwise.

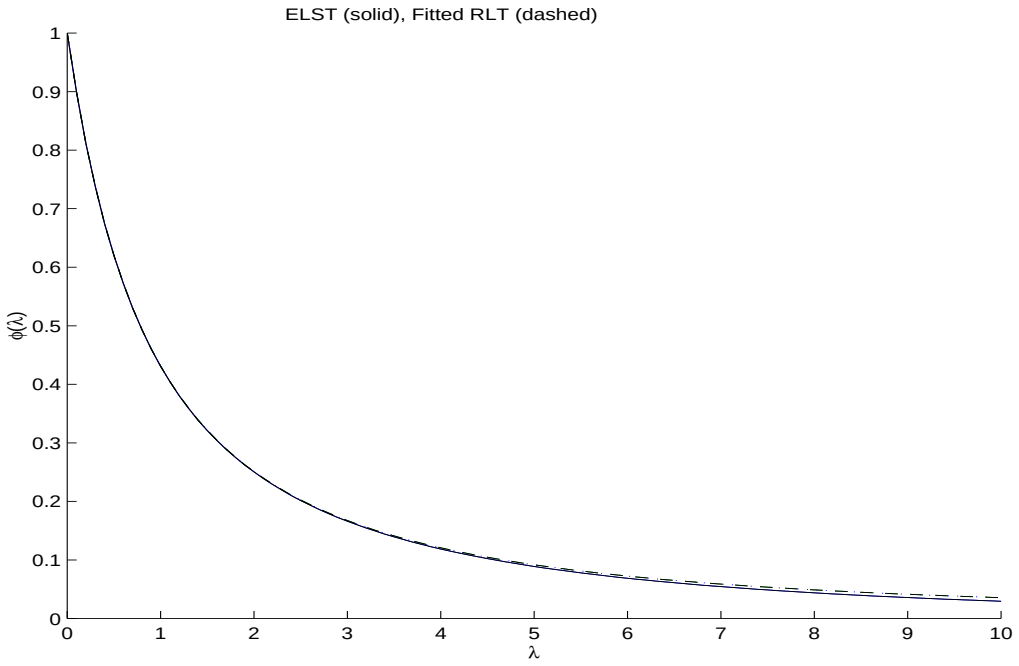

Figure 4.3.3: *ELST* of the *PH* data and fitted *RLT*

Harris and Marchal [66] recognized the problem of having fitting or approximating *RLT*s that do not correspond to distributions and detailed a method to resolve the difficulty. We will explain the method with the same example that they used.

Consider the *RLT*

$$\phi(\lambda) \;=\; \frac{1+\frac{7}{6}\lambda+\frac{5}{6}\lambda^2}{1+\frac{11}{6}\lambda+\lambda^2+\frac{1}{6}\lambda^3} \tag{4.3.5}$$

which has poles $-1, -2$, and -3. Since the pole of maximal real part is real it is possible that the *RLT* corresponds to a distribution. However, upon inverting (4.3.5) we get the function, defined for $u \geq 0$,

$$g(u) \;=\; 2e^{-u} - 6(2e^{-2u}) + 5(3e^{-3u})$$

which is negative when $0.5739 \leq u \leq 1.4410$. In order to adjust $g(u)$ so that it is nonnegative Harris and Marchal [66] stated " ... first find the value u^*, which minimizes ... $g(u)$, and then change the value of the largest damping factor so that $g(u^*) = 0$." (We have replaced their t^* with u^* to be consistent with our notation). This approach, as stated, is unlikely to work unless $g'(u^*) = 0$, also. This is unlikely to occur in practice. If, however, the procedure is repeated iteratively until both $g(u^*) = 0$ and $g'(u^*) = 0$ (u^* will be updated at each iteration) the function can be adjusted so that it is nonnegative, at least around the new value of u^*. After applying the procedure four times the resolved density is

$$g(u) \;=\; 2e^{-u} - 6(2e^{-2u}) + 5(2.7279e^{-2.7279u})$$

which has a minimum at $u^* = 0.9278$ where $g(u^*) = 2.22\times10^{-16}$. Harris and Marchal [66] gave a value of 2.7283 for the adjusted damping factor but this inaccuracy was probably due to rounding errors. Given a *RLT* it is not clear, though, how it could be determined whether or not its inverse is negative for some $u > 0$ unless a graphical or numerical procedure is used. Harris and Marchal [66] did not address this problem. In the above example if the new, resolved $g(u)$ is still negative for some other values of $u > 0$, Harris and Marchal [66] suggested that the procedure be repeated by adjusting the next highest damping factor. No justification as to whether the adjusted density is still a good fit or approximation was given. The whole procedure is rather ad hoc!

If complex conjugate poles result from the *RLT* fitting or approximation process, as is the case with our example, Harris and Marchal [66] suggested, again without justification, that either the imaginary parts of all complex zeros and poles are dropped, or they are replaced with a real number of the same modulus. The resultant *RLT* can then be made positive, if necessary, using the abovementioned procedure. In our case, upon inverting the fitted *RLT* (4.3.4), we get

$$\begin{aligned} f(u) &= (0.5326 + 1.1904i)e^{-(1.5060-1.3777i)u} + (0.5326 - 1.1904i)e^{-(1.5060+1.3777i)u} \\ &- (0.4871 + 7.9074i)e^{-(1.8924+-0.4523i)u} - (0.4871 + 7.9074i)e^{-(1.8924+0.4523i)u}. \end{aligned}$$

Dropping the imaginary parts of all complex numbers gives

$$g(u) = 1.0652e^{-1.5060u} - 0.9742e^{-1.8924u}.$$

Dividing through by

$$\int_0^\infty g(u)du = 0.1925$$

gives the function

$$h(u) = 3.6742(1.5060)e^{-1.5060u} - 2.6742(1.8924e^{-1.8924u}),$$

which is nonnegative for $u \geq 0$ and is hence a density function.

The *LST* fit is illustrated in Figure 4.3.4, and the corresponding density and distribution fits are shown in Figures 4.3.5 and 4.3.6, respectively. This fit does not look very good.

4.4 Harris and Marchal's Method 2

Harris and Marchal described a second *LST* fitting or approximation method which allowed for more flexibility. Let $\psi(\lambda)$ be either

1 the *LST* of a distribution to be approximated, or

2 the *ELST* (4.2.3) of some empirical data.

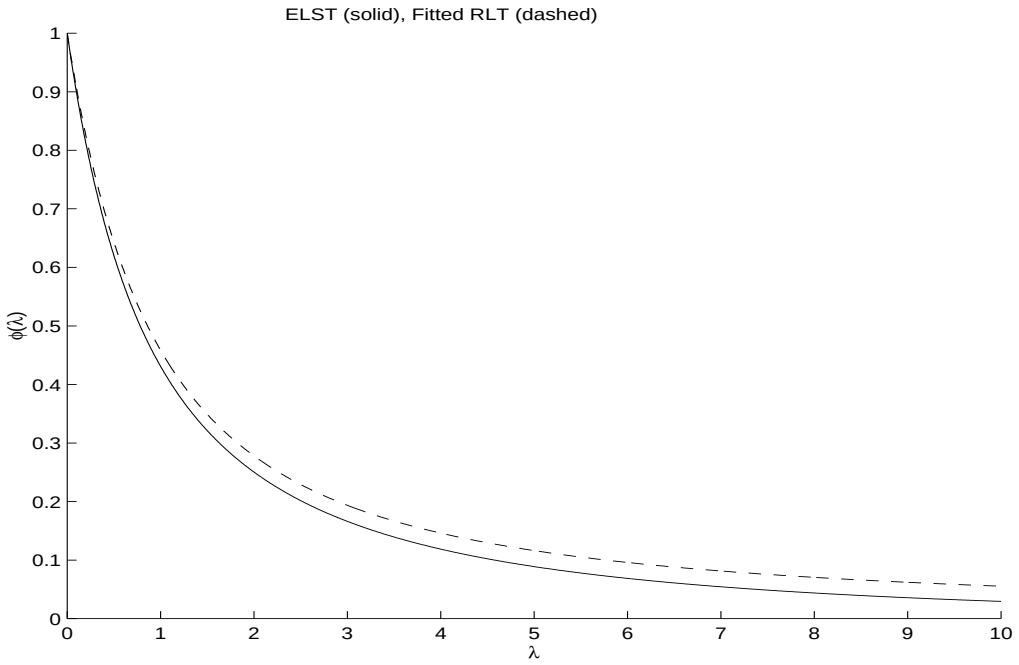

Figure 4.3.4: Adjusted transform fit

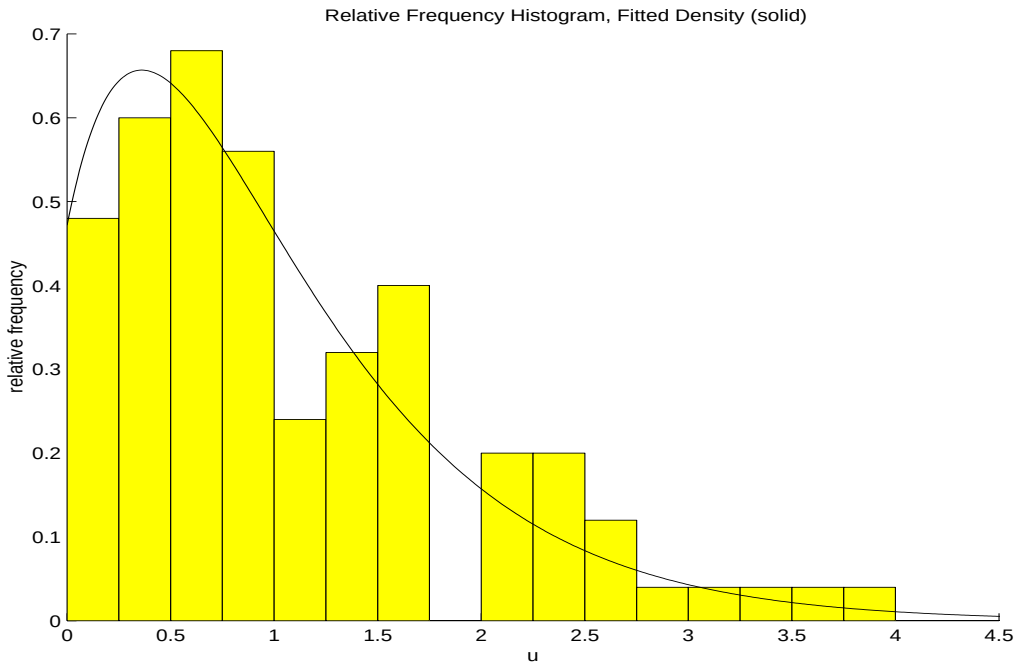

Figure 4.3.5: Adjusted density fit

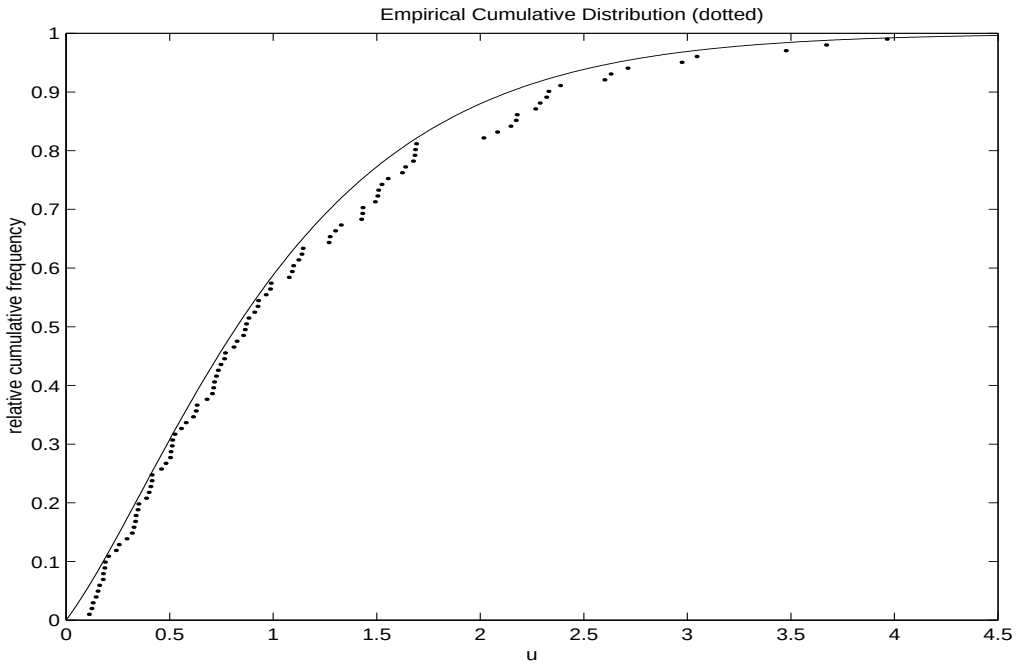

Figure 4.3.6: Adjusted distribution fit

Next, select $2N-1$ *Chebyshev-spaced interpolation points* $\lambda_1, \lambda_2, \ldots, \lambda_{2N-1}$ and substitute them into (4.3.1). After some rearrangement we get the system of $2N-1$ linear equations, for $j = 1, 2, \ldots, 2N-1$,

$$\sum_{k=0}^{N-1} c_k \lambda_j^k - \psi(\lambda_j) \sum_{k=0}^{N} d_k \lambda_j^k \;=\; 0, \tag{4.4.1}$$

in the $2N-1$ unknowns $c_1, c_2, \ldots, c_N, d_1, d_2, \ldots, d_N$, see the Open University's study guide on Chebyshev Approximation [140]. Harris and Marchal [66] did not say how the Chebyshev-spaced interpolation points could be chosen and selection of the $2N-1$ points seemed to be arbitrary. With this method if more than $2N-1$ points are chosen an overdetermined (more equations than unknowns) system of equations is likely to result. Even though, in general, no solution exists for such a system of equations the least squares solution can be found using, for example, MATLAB®.

In order to illustrate Method 2 we fitted the *ELST* of the same *PH* data set as before with a *RLT* of algebraic degree five using the nine points $\{1, 2, \ldots, 9\}$. The

fitted *RLT* was

$$\phi(\lambda) \;=\; \frac{1 + 1.9884\lambda - 0.1871\lambda^2 + 0.0217\lambda^3 - 0.0006\lambda^4}{1 + 3.0711\lambda + 2.3797\lambda^2 + 0.0931\lambda^3 + 0.0048\lambda^4 + 0.0021\lambda^5}.$$

A plot of the fitted *RLT* and *ELST* is shown in Figure 4.4.1. The fit looks very good. However, since $\phi(\lambda) < 0$ when $\lambda \gtrsim 33$, it does not correspond to a density. Choosing the set of matching points to be $\{1, 2, \ldots, 100\}$ gave a similar result. Again, Harris and Marchal [66] suggested using the same procedure explained in Section 4.3 to overcome the problem of having the inverted Laplace transform negative for some values of u.

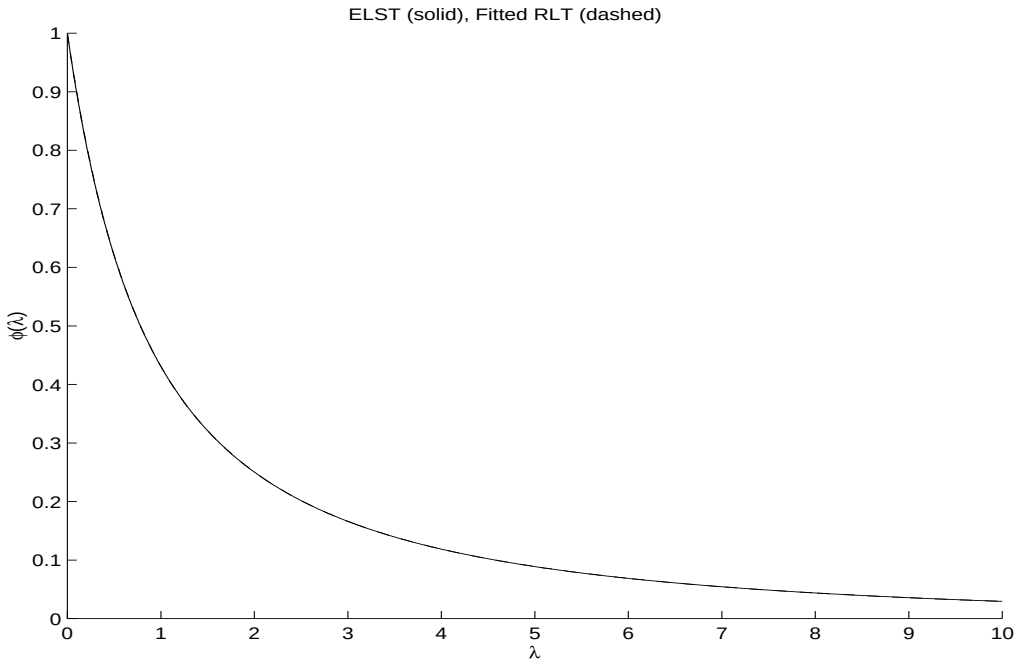

Figure 4.4.1: *ELST* of the *PH* data and fitted *RLT*

Harris and Marchal [66] used the first of their two methods to approximate the densities in the extended Aalborg benchmark. Most of the approximations were graphically quite good. However, in only two out of four cases where the approximating density function was plotted with the approximated density could it be established that it was actually a probability density. In the next section we

discuss in more detail the problems associated with finding the parameters of *RLT*s when used to fit or approximate the *LST*s of empirical or probability distributions.

4.5 Problems With Parameter Estimation and Distribution Approximation in the Laplace-Stieltjes Transform Domain

The main problem with the two parameter estimation or density approximation methods described in Sections 4.3 and 4.4 was that the fitted or approximating *RLT* did not necessarily correspond to a distribution. The reason for this is that there is nothing inherent in the *RLT* (4.2.2) which ensures that it corresponds to a distribution apart from the constant terms in the numerator and denominator being equal to one. Both methods fitted or approximated the *LST* of a distribution with a *RLT* that did *not* necessarily correspond to a distribution. It ought then to come as no surprise that the fits to the two *LST*s we got (see Figures 4.3.3 and 4.4.1) were good, but that their corresponding *RLT*s did not correspond to distributions.

It was precisely this problem that necessitated the ad hoc procedure, described in Section 4.3, that was used to adjust the "density" so that it was nonnegative. It was also unclear how to determine whether or not an inverted *RLT* corresponded to a distribution. If conditions on the parameters of the fitted or approximating *RLT* could be determined so that it does correspond to a distribution then these difficulties can be overcome. It is this problem that we address in the next four chapters.

Even if we were able to characterize the parameters of a *RLT* so that it did correspond to a distribution there still remains the problem of how good a fit or approximation we can achieve. Two *LST*s may be "close" but their inverted counterparts may not be. To clarify this point suppose that $\phi = \mathcal{L}(f)$ and $\psi = \mathcal{L}(g)$ are

two LSTs such that for a large, positive value of λ,

$$\phi(\lambda) - \psi(\lambda) \;=\; \int_0^\infty e^{-\lambda u}(f(u) - g(u))du$$

is small. It may be, however, that for particular values of u, $|f(u) - g(u)|$ is large. Consequently, a good fit or approximation in the LST domain does not necessarily guarantee that the fit or approximation is a good one in the distribution domain. Paulson, Holcomb, and Leitch [112] recognised this problem when estimating the parameters for the stable laws using characteristic functions and proposed a method which standardized the empirical data to overcome it. Longman [91] addressed this problem for Laplace transforms. He developed a method for approximating an arbitrary square integrable function with a sum of exponential functions by considering the approximation in the transform domain.

In the next chapter we ask the question: Given the rational LST of the form (4.2.2) when does it correspond to a distribution, PH or otherwise? As we shall see, the answer is not that simple!

Chapter 5

Matrix-exponential Distributions

5.1 Introduction

The Laplace transform estimation or approximation methods described in Chapter 4 had two major drawbacks. First, the estimated or approximated *LST* did not necessarily correspond to a probability distribution. Second, if the *LST* did happen to correspond to a *PH* distribution, it was not clear how to obtain a *PH* representation, whether minimal or not, for it.

In order to tackle these two problems we consider the wider class of *matrix-exponential* (*ME*) distributions which have already been mentioned, as the class of all distributions with rational *LST*, in Section 4.2. *ME* distributions have distribution functions of the same form as *PH* distributions but their representations do not need to have a simple probabilistic interpretation. Given a rational *LST* that corresponds to a distribution a *ME* representation of minimal order can be easily found. Thus, the second problem mentioned above, with respect to *ME* distributions, is not present. The first problem, that of recognizing whether or not a particular rational *LST* corresponds to a *ME* distribution, is addressed in this and the next three chapters.

The problem of characterizing *ME* distributions has been addressed in two dif-

ferent ways in the literature. In the first approach it was presupposed that the *ME* distribution was already defined by a distribution or density function, *LST*, or moments, and a minimal representation (see Section 5.3 for the definition) was found. Lipsky and Ramaswami [89] characterized *ME* distributions by finding a minimal representation given the distribution function or the *LST*. Asmussen and Bladt [10] determined a representation with only real parameters given the *LST*. Van de Liefvoort [144] developed an algorithm for finding a minimal *ME* representation given the distribution's moments. These works, particularly that of Asmussen and Bladt [10], will help us to solve the second problem stated in the opening paragraph of this section.

In the second approach particular classes of *ME* distributions were characterized by determining necessary and/or sufficient conditions on the zeros and poles of their *LST*s. Zemanian [151] and [152] developed some sufficient conditions for the class of all bounded, nondecreasing functions (which can be normalized to obtain *ME* distribution functions) whose *LST* has the number of real poles greater than or equal to the total number of complex poles and zeros. Sumita and Masuda [137] gave some necessary and sufficient conditions for the class of all *ME* distributions whose *LST* has real zeros and poles. Harris, Marchal and Botta [68] restricted themselves to the class of generalized hyperexponential distributions and gave some sufficient conditions.

Dehon and Latouche [45] approached the problem of characterizing the class of generalized hyperexponential distributions geometrically. They considered each hyperexponential distribution as a mixture (not necessarily convex) of exponential distributions, and determined the set of all permissible mixing coefficients for hyperexponential distributions of order three. It is an approach similar to this that we will use to characterize *ME* distributions in Chapter 7.

In Section 5.2 we define *ME* distributions and their representations, state some analytical characteristics, and give some examples. Section 5.3 contains the descriptions of two physical interpretations for *ME* distributions: one due to Bladt and

Neuts [23], the other due to Asmussen and Bladt [11]. A discussion about the non-uniqueness of *ME* representations, including a useful representation theorem proved in Asmussen and Bladt [10], is given in Section 5.4. In Section 5.5, by considering the definition of a probability distribution function, we determine some necessary and sufficient conditions on rational *LST*s so that they correspond to *ME* distributions. In Section 5.6, by adopting a geometric point of view, we discuss the *ME* characterization problem in more detail.

5.2 Matrix-exponential Distributions

In this section we introduce the class of *matrix-exponential* (*ME*) distributions. For a excellent discussion of the topic see Lipsky [90, Chapter 3], and Asmussen and O'Cinneide [12].

A nonnegative random variable X is distributed according to a *ME* distribution if its distribution function, defined for $u \geq 0$, has the form

$$F(u) \;=\; \begin{cases} \alpha_0, & u = 0 \\ 1 + \boldsymbol{\alpha} \exp(\boldsymbol{T}u)\boldsymbol{T}^{-1}\boldsymbol{t}, & u > 0 \end{cases} \tag{5.2.1}$$

where, for $p \geq 1$, $\boldsymbol{\alpha}$ is a $1 \times p$ row vector, $\boldsymbol{T}$ is a $p \times p$ matrix, and $\boldsymbol{t}$ is a $p \times 1$ column vector, all possibly with complex entries. It is immediately clear that $0 \leq \alpha_0 \leq 1$. A further stipulation on the parameters $\boldsymbol{\alpha}$, $\boldsymbol{T}$, and α_0 is that the distribution function (5.2.1) is right-continuous for $u = 0$. That is,

$$\lim_{u \to 0^+} (1 + \boldsymbol{\alpha} \exp(\boldsymbol{T}u)\boldsymbol{T}^{-1}\boldsymbol{t}) \;=\; \alpha_0. \tag{5.2.2}$$

The parameter α_0 is known as the *point mass at zero*. We will not consider the case when $\alpha_0 = 1$ as this gives the trivial distribution function. The distribution is said to have a *representation* $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{t})$ of *order* p. The corresponding density function, defined for $u > 0$, is given by

$$f(u) \;=\; \boldsymbol{\alpha} \exp(\boldsymbol{T}u)\boldsymbol{t}.$$

The Laplace-Stieltjes transform (LST) of (5.2.1), defined for $\lambda \in \mathbb{C}$ such that $\Re(\lambda) > -\delta$ where δ is a positive number, is given by

$$\begin{aligned} \phi(\lambda) &= \int_0^\infty e^{-\lambda u} dF(u) \\ &= \boldsymbol{\alpha}(\lambda \boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{t} + \alpha_0. \end{aligned} \tag{5.2.3}$$

Differentiating (5.2.3) k times with respect to λ and letting $\lambda = 0$ gives, for $k = 1, 2, \ldots$, the kth noncentral moment

$$m_k = (-1)^{k+1} k! \boldsymbol{\alpha} \boldsymbol{T}^{-(k+1)} \boldsymbol{t}.$$

Lipsky and Ramaswami [89], and Asmussen and Bladt [10] (see also Lipsky and Fang [88]), showed that the class of all distributions with rational LST is the same as the class of all ME distributions. Thus, as for PH distributions, the LST of a ME distribution can be expressed as a rational function of the form

$$\phi(\lambda) = \frac{a_p \lambda^{p-1} + a_{p-1}\lambda^{p-2} + \ldots + a_1}{\lambda^p + b_p \lambda^{p-1} + b_{p-1}\lambda^{p-2} \ldots + b_1} + \alpha_0$$

where $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_p$ are all real. We refer to the polynomials $a(\lambda) = a_p\lambda^{p-1} + a_{p-1}\lambda^{p-2} + \ldots + a_1$ and $b(\lambda) = \lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} \ldots + b_1$ as the *numerator* and the *denominator of the LST*, respectively. We define the *zeros of the LST* to be the zeros of $a(\lambda)$ and the *poles of the LST* to be the zeros of $b(\lambda)$. We remark here that for the remainder of the thesis we adopt the ordering $\Re(\lambda_1) \geq \Re(\lambda_2) \geq \ldots \geq \Re(\lambda_p) > 0$ for the poles of $\phi(\lambda)$ (equivalently the eigenvalues of $\boldsymbol{T}$, see the second remark after Theorem 5.1), which is the same as that found in Dehon and Latouche [45].

Below are some examples of ME distributions.

1. Phase-type (PH) distributions. Every PH distribution has a ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, -\boldsymbol{T}\boldsymbol{e})$ where $\boldsymbol{\alpha}$ is the initial state probability vector and $\boldsymbol{T}$ is the infinitesimal generator of an evanescent finite-state continuous-time Markov chain.

2. Generalized hyperexponential (GH) distributions, see Botta and Harris [27], and Botta, Harris, and Marchal [28]. They have distribution functions, defined for $u \geq 0$, of the form

$$F(u) \;=\; \sum_{i=1}^{p} a_i(1 - e^{-\lambda_i u})$$

where $a_1, a_2, \ldots, a_p$, are all real with $\sum_{i=1}^{p} a_i = 1$, and $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$. Every GH distribution has a ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{t})$, where

$$\boldsymbol{\alpha} \;=\; \begin{pmatrix} a_1 & a_2 & \ldots & a_p \end{pmatrix}$$

$$\boldsymbol{T} \;=\; \begin{pmatrix} -\lambda_1 & 0 & \ldots & 0 \\ 0 & -\lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -\lambda_p \end{pmatrix}$$

$$\boldsymbol{t} \;=\; \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{pmatrix}.$$

3. Distributions whose LST is the reciprocal of a polynomial. These distributions, introduced by Smith [131], have rational LST by definition and are therefore ME distributions. Since they are defined in terms of their LST it is not immediately clear how to determine a ME representation for them.

5.3 The Physical Interpretation of Matrix-exponential Distributions

PH distributions have a simple probabilistic interpretation in that they are defined as the absorption time of an evanescent finite-state continuous-time Markov chain.

A physical interpretation is not so straightforward, in general, for *ME* distributions. However, physical interpretations can be given to them. In this section we discuss two such interpretations for *ME* distributions. The first, due to Bladt and Neuts [23], uses random stopping times of deterministic flows. The second, due to Asmussen and Bladt [11], considers piecewise deterministic Markov processes.

Asmussen and Bladt [10] showed that any *ME* distribution has a representation of the form $(\boldsymbol{\beta}, \boldsymbol{S}, -\boldsymbol{S}\boldsymbol{e})$. It is also true, although not explicitly stated in their paper, that such a representation has $\boldsymbol{\beta}\boldsymbol{e} = 1$. It is this representation that Bladt and Neuts [23] used in their physical interpretation of *ME* distributions.

Consider p containers that may hold any positive, negative, or zero amount of fluid. The fluid is allowed to flow deterministically from container to container. For $i = 1, 2, \ldots, p$, let the initial amount of fluid in container i be $\beta_i \in \mathbb{R}$. We also have a zeroth container which initially contains an amount of fluid β_0 where $0 \leq \beta_0 < 1$. Assume that $\beta_0 + \beta_1 + \ldots + \beta_p = 1$. For $i, j = 1, 2, \ldots, p$, with $i \neq j$ fluid flows from container i to container j with constant rate $S_{ij} \in \mathbb{R}$. For $i = 1, 2, \ldots, p$, fluid flows into container 0 with constant rate $s_i \in \mathbb{R}$. Define, for $i = 1, 2, \ldots, p$,

$$S_{ii} \;=\; -\sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij} - s_i.$$

We let $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ and $\boldsymbol{S} = [S_{ij}]$. Observe that s_i is the ith component of $-\boldsymbol{S}\boldsymbol{e}$.

Bladt and Neuts [23] defined a *valid flow* with parameters $(\boldsymbol{\beta}, \boldsymbol{S})$ as one where the amount of fluid in Container 0 does not decrease over time and where eventually all of the fluid flows into this container. More precisely, if, for $t \geq 0$, $v_0(t)$ is the amount of fluid in Container 0 at time t, $v_0(t)$ is a nondecreasing function of t and $\lim_{t \to \infty} v_0(t) = 1$. Bladt and Neuts [23] remarked that there exist parameters $(\boldsymbol{\beta}, \boldsymbol{S})$ which generate valid flows. Moreover, the parameters of any valid flow correspond to a *ME* distribution. The most important result they proved was that any *ME* random variable, with a representation $(\boldsymbol{\beta}, \boldsymbol{S})$ with $\boldsymbol{\beta}\boldsymbol{e} = 1$, is distributed as the random stopping time of a valid flow with parameters $(\boldsymbol{\beta}, \boldsymbol{S})$. That is, the time taken

for $v_0(t)$ to reach a level U where U is a uniformly distributed random variable on $[0, 1)$, is distributed as a *ME* distribution with representation $(\boldsymbol{\beta}, \boldsymbol{S})$ with $\boldsymbol{\beta e} = 1$.

Bladt and Neuts [23] re-proved a number of established results from *ME* renewal theory and a new one in risk theory, by considering *ME* distributions as the random stopping times of valid deterministic flows. Their physical interpretation of *ME* distributions, like that of *PH* distributions, is appealing and ought to lead to a deeper understanding of systems that are modelled with *ME* distributions. However, unlike that for *PH* distributions, it is not easily possible to determine whether a given flow with parameters $(\boldsymbol{\beta}, \boldsymbol{S})$ is valid or not. This problem is the main focus of the present and next three chapters.

Asmussen and Bladt [11] developed a physical interpretation for *ME* distributions in the wider context of *rational arrival processes* (*RAP*s). In the same way that *ME* distributions are an extension of *PH* distributions by relaxing the probabilistic constraints on $\boldsymbol{\alpha}$ and $\boldsymbol{T}$, *RAP*s are an extension of *Markovian arrival processes* (*MAP*s). A *MAP* is a Markovian point process defined on a finite phase space where transitions between phases *without* an arrival or event occurring are governed by a generator matrix $\boldsymbol{C}$, and transitions *with* an arrival or event occurring are governed by another matrix $\boldsymbol{D}$. If $\boldsymbol{\alpha}$ is the initial phase probability distribution then the triple $(\boldsymbol{\alpha}, \boldsymbol{C}, \boldsymbol{D})$ is a *representation* of the *MAP*. For a comprehensive treatment of *MAP*s refer to Neuts [101], or Latouche and Ramaswami [85], and the references therein. A *RAP* is a point processes that has a joint density function of the same form as that of a *MAP* but without the probabilistic constraints on the representation $(\boldsymbol{\alpha}, \boldsymbol{C}, \boldsymbol{D})$. Asmussen and Bladt [11] gave some necessary conditions on $(\boldsymbol{\alpha}, \boldsymbol{C}, \boldsymbol{D})$ so that it corresponds to a *RAP*.

The motivation for the physical interpretation of *RAP*s came from the characterization of *ME* distributions due to O'Cinneide [104]. Consider the following. Let a random variable X defined on the nonnegative real numbers have distribution function F. For $t \geq 0$, let $F_t(x) = F(x + t) - F(x)$ be the (defective) distribution of the *residual life* $X - t$. If span(F) denotes the vector space of signed measures

consisting of all linear combinations of F_t, where $t \geq 0$, then we have (Proposition 1.1 in Asmussen and Bladt [11]) that

> *"a distribution F is a ME distribution if and only if* $\text{span}(F)$ *is finite-dimensional."*

It was this finite-dimensionality property that led Asmussen and Bladt [11] to represent a *RAP* as a p-dimensional *piecewise deterministic Markov process* (*PDMP*) $\boldsymbol{A}(t) = (A_1(t), A_2(t), \ldots, A_p(t))$, where $t \geq 0$, see Davis [43]. A *PDMP* evolves deterministically according to a multidimensional differential equation and at random time points it changes state. The state space is the set of all possible vectors $\boldsymbol{A}(t)$ where $t \geq 0$. In between jumps the *PDMP* that represents the *RAP* evolves according to the differential equation

$$\frac{d}{dt}(\boldsymbol{a}(t)) \;=\; \boldsymbol{a}(t)C - (\boldsymbol{a}(t)C\boldsymbol{e})\boldsymbol{a}(t), \tag{5.3.1}$$

where $\boldsymbol{a}(t) = (a_1(t), a_2(t), \ldots, a_p(t))$ is the state of the process at time $t \geq 0$ with $\boldsymbol{a}(t)\boldsymbol{e} = 1$. The solution to the differential equation (5.3.1), which is defined for $t \geq 0$, is given by

$$\boldsymbol{a}(t) \;=\; \frac{\boldsymbol{a}(0)\exp(\boldsymbol{C}t)}{\boldsymbol{a}(0)\exp(\boldsymbol{C}t)\boldsymbol{e}}.$$

When the piecewise deterministic Markov process is in state $\boldsymbol{a}(t)$, a random jump occurs with intensity $\boldsymbol{a}(t)\boldsymbol{D}$. If this jump occurs when $t = t^*$ the process that evolves deterministically according to (5.3.1) then starts anew with $\dfrac{\boldsymbol{a}(t^*)\boldsymbol{D}}{\boldsymbol{a}(t^*)\boldsymbol{D}\boldsymbol{e}}$ as its initial state. Asmussen and Bladt [11] represented the evolution of a *RAP* with an *orbit representation*, that is, as a trajectory on the $(p-1)$-dimensional subspace $x_1 + x_2 + \ldots + x_p = 1$ of $\mathbb{R}^p$ governed by the parametric equation $\boldsymbol{x}(t) = \boldsymbol{A}(t)$. The entire trajectory $\boldsymbol{x}(t)$, for $t \geq 0$, is a succession of orbits on which the vector $\boldsymbol{x}(t)$ satisfies (5.3.1). When a jump occurs the trajectory changes to another orbit. If the *RAP* is a renewal process the trajectory will restart on the same orbit at the same starting point immediately after each jump.

As with *MAP*s the state of a *RAP* at a particular time gives knowledge about its future evolution. In a *MAP* if the phase of the process is known at a particular time

t, its future evolution is independent of its past history. The same is true for *RAP*s. If the state of the process at time t, that is, $\boldsymbol{a}(t)$ is known, its future evolution is independent of its past.

When the *RAP* is a *MAP* $\boldsymbol{a}(t) = (a_1(t), a_2(t), \ldots, a_p(t))$ can be thought of as giving the probabilities of being in each phase of the underlying Markov chain at time t given that no jump has occurred. If the *RAP* is a *PH*-renewal process with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$, we would have jumps occurring at intensity $f(t) = -\boldsymbol{\alpha}\exp(\boldsymbol{T}t)\boldsymbol{T}\boldsymbol{e}$ which is the density function for the *PH* interarrival times. Also, the piecewise deterministic Markov process would restart in state $\boldsymbol{\alpha}$ immediately after each jump. If the *RAP* is a *ME*-renewal process with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ where $\boldsymbol{\alpha}\boldsymbol{e} = 1$ the jump intensity and restart state would be given by the same expressions as for the *PH* renewal process. In this case $\boldsymbol{a}(t)$ can be thought of as being the amount of fluid in each container as discussed above.

Bean and Nielsen [19] used this physical interpretation to develop and analyse *quasi-birth-and-death processes with RAP components*, the *RAP* equivalent to the traditional quasi-birth-and-death processes. As an example, they compared the "traditional" Markov chain physical interpretation of the $\boldsymbol{G}$ matrix for the $M/PH/1$ queue with the *PDMP* physical interpretation of the $\boldsymbol{G}$ matrix for the $M/ME/1$ queue.

5.4 Matrix-exponential Representations

In general, representations for *ME* distributions are not unique. For example, consider the *ME* distribution with density

$$f(u) \;=\; 2e^{-u} - 6e^{-2u} + 6e^{-3u},$$

and *LST*

$$\phi(\lambda) \;=\; \frac{2\lambda^2 + 4\lambda + 6}{\lambda^3 + 6\lambda^2 + 11\lambda + 6}.$$

This distribution has the following three distinct representations $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{t})$, $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s})$ and $(\boldsymbol{\gamma}, \boldsymbol{R}, \boldsymbol{r})$ given by

$$\boldsymbol{\alpha} = \begin{pmatrix} -1 & 1 & -1 \end{pmatrix} \qquad \boldsymbol{T} = \begin{pmatrix} -4 & 0 & -1 \\ 2 & -1 & 0 \\ 2 & 0 & -1 \end{pmatrix} \qquad \boldsymbol{t} = \begin{pmatrix} 0 \\ -4 \\ -6 \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} 6 & 4 & 2 \end{pmatrix} \qquad \boldsymbol{S} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{pmatrix} \qquad \boldsymbol{s} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and

$$\boldsymbol{\gamma} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix} \qquad \boldsymbol{R} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & -3 & 3 \\ 0 & 0 & 0 & -4 \end{pmatrix} \qquad \boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

This example, as with *PH* distributions, illustrates the fact that representations for *ME* distributions do not necessarily have the same order. A representation that has *minimal order* will be called a *minimal representation*. As we shall see, Theorem 5.3 asserts that the representations $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{t})$ and $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s})$ are both minimal. Therefore, minimal representations are not necessarily unique. The *order of the ME distribution* is defined to be the order of any minimal representation. If the *LST* of the *ME* distribution is expressed as the ratio of two irreducible polynomials then the degree of the denominator is called the *algebraic degree* of the *ME* distribution. The terminology used here is due to O'Cinneide [104].

In the introduction to this chapter we mentioned that given a rational *LST* that corresponds to a *PH* distribution, it is difficult to find a minimal *PH* representation, or even *any PH* representation, for it. This problem, first posed by Neuts [101], is still unsolved. A discussion of the problem was given in Section 2.4 of this thesis. The situation, however, is somewhat simpler for *ME* distributions. Given a rational *LST* that corresponds to a *ME* distribution we can easily find a *ME* representation

for it. The following theorem, which is a paraphrased version of Proposition 2.3 in Asmussen and Bladt [10], gives such a *ME* representation.

Theorem 5.1 *If the LST of a ME distribution is expressed as*

$$\phi(\lambda) \;=\; \frac{a_p\lambda^{p-1} + a_{p-1}\lambda^{p-2} + \ldots + a_1}{\lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} \ldots + b_1} + \alpha_0 \tag{5.4.1}$$

where $p \geq 1$, $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_p$ *are all real, and* $0 \leq \alpha_0 < 1$, *then the ME distribution has a representation* $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{t})$ *where*

$$\boldsymbol{\alpha} \;=\; \begin{pmatrix} a_1 & a_2 & \ldots & a_p \end{pmatrix} \tag{5.4.2}$$

$$\boldsymbol{T} \;=\; \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -b_1 & -b_2 & -b_3 & \ldots & -b_{p-1} & -b_p \end{pmatrix} \tag{5.4.3}$$

$$\boldsymbol{t} \;=\; \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \;=\; \boldsymbol{e}_p. \tag{5.4.4}$$

Note that this representation is not necessarily minimal.

Since $\boldsymbol{T}$ is known as a *companion matrix* we shall refer to this form of *ME* representation as a *companion form* representation.

Theorem 5.1 is important for a number of reasons. First, it shows that a representation with only real parameters can be found for any *ME* distribution. Second, because the characteristic equation of a matrix of the form (5.4.3) is

$$\lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} \ldots + b_1 \;=\; 0,$$

the eigenvalues of $\boldsymbol{T}$ are identical to the poles of the *LST*. Third, and most importantly, there is a direct one-to-one correspondence between the *LST* of a *ME* distribution and a *minimal* representation of the form (5.4.2)–(5.4.4) for it. The following lemma is required before we prove this fact.

Lemma 5.2 *If the LST of the form (5.4.1) corresponds to a ME distribution then*

$$\alpha_0 \;=\; 1 - \frac{a_1}{b_1}. \tag{5.4.5}$$

Proof.

It can be shown that

$$\boldsymbol{T}^{-1} \;=\; \begin{pmatrix} -\frac{b_2}{b_1} & -\frac{b_3}{b_1} & -\frac{b_4}{b_1} & \dots & -\frac{b_p}{b_1} & -\frac{1}{b_1} \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

By (5.2.2), (5.4.2), and (5.4.4) we have that

$$\begin{aligned} \alpha_0 \;&=\; \lim_{u \to 0^+} F(u) \\ &=\; 1 + \boldsymbol{\alpha} \boldsymbol{T}^{-1} \boldsymbol{e}_p \\ &=\; 1 - \frac{a_1}{b_1}. \end{aligned} \tag{5.4.6}$$

■

Theorem 5.3 *Every ME distribution has a unique minimal representation of the form (5.4.2) – (5.4.4).*

Proof.

Consider a *ME* distribution. Suppose that its *LST* is expressed in the form (5.4.1) where $a(\lambda)$ and $b(\lambda)$ have no factors in common. The *LST* of this form is

unique since every (*ME*) distribution has a unique *LST*, see Feller [59, page 430]. The companion form representation for the given *ME* distribution (5.4.2) – (5.4.4) is minimal because the algebraic degree of the *LST*, which cannot be any smaller, is equal to the dimension of $\boldsymbol{T}$, that is, the order of the representation. Using (5.4.5) it can be seen that every parameter set that defines the *LST* (5.4.1) also defines the *ME* representation (5.4.2) – (5.4.4). Therefore, there is a one-to-one correspondence between the *LST*s of *ME* distributions and their companion form representations. ■

The representation $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s})$ for the *ME* distribution given at the beginning of this section is an example of a companion form representation.

Theorem 5.3 asserts that any conditions on the poles of the *LST* (5.4.1) are equivalent to conditions on the eigenvalues of $\boldsymbol{T}$ given by (5.4.3), and that any conditions on the numerator of (5.4.1) are equivalent to conditions on $\boldsymbol{\alpha}$ given by (5.4.2). For the remainder of the thesis we state our results in terms of conditions on the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ where

$$\boldsymbol{a} = \begin{pmatrix} a_1 & a_2 & \dots & a_p \end{pmatrix} \tag{5.4.7}$$

and

$$\boldsymbol{b} = \begin{pmatrix} b_1 & b_2 & \dots & b_p \end{pmatrix}. \tag{5.4.8}$$

Due to the above correspondence, $\boldsymbol{a}$ and $\boldsymbol{b}$ apply equally well to *LST*s and companion form representations.

We now have a solution to the second problem posed in the opening paragraph of this chapter, at least if we are prepared to accept minimal *ME* representations in preference to minimal *PH* representations. In the next section we will begin to address the first problem. That is, given a pair of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ when do they correspond to a *ME* (or *PH*) distribution?

5.5 Distribution Functions

In this section, by considering the criteria for a function to be the distribution function of a nonnegative random variable, we determine some necessary and sufficient conditions on the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ so that they correspond to a *ME* distribution. For the following definition refer to Moran [99, Section 5.1].

Definition 5.4 *A function F is the* distribution function *of a nonnegative random variable X if*

1. *$F(u)$ is nonnegative and nondecreasing for $u \geq 0$,*
2. *$F(u)$ is right-continuous. That is, for $u \geq 0$, $\lim_{h \to 0^+} F(u+h) = F(u)$, and*
3. *$\lim_{u \to \infty} F(u) = 1$.*

The following lemma will simplify the proof of the characterization theorem for *ME* distributions, Theorem 5.6.

Lemma 5.5 *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be defined by (5.4.7) and (5.4.8), respectively. Let*

$$b(\lambda) \;=\; \lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} + \ldots + b_1$$

and

$$\boldsymbol{B} \;=\; \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -b_1 & -b_2 & -b_3 & \ldots & -b_{p-1} & -b_p \end{pmatrix}. \tag{5.5.1}$$

If, for $u > 0$,

$$f(u) \;=\; \boldsymbol{a}\exp(\boldsymbol{B}u)\boldsymbol{e}_p \;\geq\; 0 \tag{5.5.2}$$

and

$$\int_0^\infty f(u)du \;<\; \infty, \tag{5.5.3}$$

then there exists a zero of $b(\lambda)$ of maximal real part that is real and negative.

Proof.

Let the distinct zeros of $b(\lambda)$ be $\xi_1, \xi_2, \ldots, \xi_q$, with respective multiplicities $n_1, n_2, \ldots, n_q$. We assume, without loss of generality, that $\Re(\xi_1) \leq \Re(\xi_2) \leq \ldots \leq \Re(\xi_q)$.

Suppose that there exists no zero of $b(\lambda)$ of maximal real part that is real. Then there will be r, with $1 \leq r \leq \lfloor \frac{q}{2} \rfloor$ ($\lfloor x \rfloor$ denotes the greatest integer less than or equal to x), complex conjugate pairs that are the zeros of maximal real part. They are, for $j = q-2r+1, q-2r+2, \ldots, q$, given by

$$\xi_j \;=\; \gamma + i\beta_j,$$

where γ and $\beta_j \neq 0$ are real. Note that, for $j = q-2r+1, q-2r+3, \ldots, q-1$, $\beta_j = -\beta_{j+1} > 0$ without loss of generality, and that all of the β_j's are distinct. The remaining zeros, $\xi_1, \xi_2, \ldots, \xi_{q-2r}$, all have real part less than γ.

Since the zeros of $b(\lambda)$ are identically the eigenvalues of $\boldsymbol{B}$ we have that

$$f(u) \;=\; \sum_{j=q-2r+1}^{q} P_j(u) e^{(\gamma+i\beta_j)u} + \sum_{j=1}^{q-2r} P_j(u) e^{\xi_j u},$$

where for $j = 1, 2, \ldots, q$, $P_j(u)$ is a possibly complex polynomial of degree $n_j - 1$. Let $n = \max(n_{q-2r+1}, n_{q-2r+2}, \ldots, n_q) - 1$. Consider

$$u^{-n} e^{-\gamma u} f(u) \;=\; \sum_{j=q-2r+1}^{q} u^{-n} P_j(u) e^{i\beta_j u} + \sum_{j=1}^{q-2r} u^{-n} P_j(u) e^{-(\gamma-\xi_j)u}.$$

Now, for any $\epsilon_1 > 0$, there exists a $K_1 > 0$ such that, for $u > K_1$,

$$\Big| \sum_{j=1}^{q-2r} u^{-n} P_j(u) e^{-(\gamma-\xi_j)u} \Big| \;<\; \epsilon_1, \tag{5.5.4}$$

since, for $j = 1, 2, \ldots, q-2r$, $\gamma > \Re(\xi_j)$. Also, for any $\epsilon_2 > 0$, there exists a $K_2 > 0$ such that, for $u > K_2$,

$$\Big| \sum_{j=q-2r+1}^{q} u^{-n} P_j(u) e^{i\beta_j u} \;-\; \sum_{j=q-2r+1}^{q} A_j e^{i\beta_j u} \Big| \;<\; \epsilon_2, \tag{5.5.5}$$

where, for $j = q-2r+1, q-2r+2, \ldots, q$, A_j is the (possibly zero) coefficient of u^n in $P_j(u)$. Now, consider

$$\sum_{j=q-2r+1}^{q} A_j e^{i\beta_j u} \;=\; \sum_{\substack{j=q-2r+1 \\ \text{steps of 2}}}^{q-1} B_j \cos\beta_j u + C_j \sin\beta_j u, \tag{5.5.6}$$

where, for $j = q-2r+1, q-2r+3, \ldots, q-1$, B_j and C_j are real. Consider the situation when for $j = q-2r+1, q-2r+3, \ldots, q-1$, β_j is rational, that is,

$$\beta_j \;=\; \frac{r_j}{s_j}$$

where r_j and s_j are integers. Now, for any $u_1 > \max(K_1, K_2)$, there exists a $u_2 > u_1$ such that

$$\int_{u_1}^{u_2} \Big(\sum_{\substack{j=q-2r+1 \\ \text{steps of 2}}}^{q} B_j \cos\beta_j u + C_j \sin\beta_j u \Big) du \;=\; 0 \tag{5.5.7}$$

as the integrand is periodic. Choosing

$$u_2 \;=\; u_1 + s_{q-2r+1} s_{q-2r+3} \ldots s_{q-1} \times 2\pi$$

is sufficient. Since (5.5.6) is not identically zero for $u_1 < u < u_2$ (the trigonometric functions are all linearly independent because the β_j's are all distinct and at least one of the B_j's and C_j's is nonzero) there exists some $u > \max(K_1, K_2)$ with $u_1 < u < u_2$ and $\epsilon_3 > 0$ such that

$$\sum_{\substack{j=q-2r+1 \\ \text{steps of 2}}}^{q} B_j \cos\beta_j u + C_j \sin\beta_j u \;<\; -\epsilon_3. \tag{5.5.8}$$

If any β_j is an irrational number approximating it arbitrarily closely by a rational number will still give (5.5.8).

We now choose ϵ_1 and ϵ_2 so that $\epsilon_1 + \epsilon_2 < \epsilon_3$. This can be achieved because (5.5.7) holds for *any* $u_1 > \max(K_1, K_2)$. We now have that there exists a $u > \max(K_1, K_2)$,

with $u_1 < u < u_2$, such that

$$\begin{aligned} u^{-n}e^{-\gamma u}f(u) &= \sum_{j=q-2r+1}^{q} u^{-n}P_j(u)e^{i\beta_j u} + \sum_{j=1}^{q-2r} u^{-n}P_j(u)e^{-(\gamma-\xi_j)u} \\ &< \sum_{j=q-2r+1}^{q} A_j e^{i\beta_j u} + \epsilon_2 + \epsilon_1 \\ &< -\epsilon_3 + \epsilon_2 + \epsilon_1 \\ &< 0. \end{aligned}$$

The first inequality holds because of (5.5.4) and (5.5.5), and the second because of (5.5.8). Thus, there exists a $u > 0$ such that $f(u) < 0$. This contradicts (5.5.2), therefore, there exists a zero of $b(\lambda)$ of maximal real part that is real.

Suppose that there exists a zero of $b(\lambda)$ of maximal real part that is real and nonnegative. Without loss of generality, let this zero be $\xi_q = \sigma \geq 0$. The polynomial $b(\lambda)$ may have a number of zeros that occur in complex conjugate pairs whose real part equals σ. Let the number of such conjugate pairs that are distinct be s, where $0 \leq s \leq \lfloor \frac{q-1}{2} \rfloor$. That is, for $j = q-2s, q-2s+1, \ldots, q-1$,

$$\xi_j = \sigma + i\eta_j,$$

where $\eta_j \neq 0$ is real. Note that, for $j = q-2s, q-2s+2, \ldots, q-2$, $\eta_j = -\eta_{j+1} > 0$ without loss of generality. The remaining zeros, $\xi_1, \xi_2, \ldots, \xi_{q-2s-1}$, all have real parts less than σ. We have

$$g(u) = Q_q(u)e^{\sigma u} + \sum_{j=q-2s}^{q-1} Q_j(u)e^{(\sigma+\eta_j i)u} + \sum_{j=1}^{q-2s-1} Q_j(u)e^{\xi_j u}, \tag{5.5.9}$$

where for $j = 1, 2, \ldots, q$, $Q_j(u)$ is a possibly complex polynomial of degree $n_j - 1$. If $\sigma > 0$ then (5.5.9) diverges as $u \to \infty$. If $\sigma = 0$ then (5.5.9) either diverges, oscillates, or approaches a nonzero constant as $u \to \infty$. In any case (5.5.3) is not satisfied. This is a contradiction and therefore any zero of $b(\lambda)$ of maximal real part must also be negative. ■

Theorem 5.6 *Let $f(u)$, defined by (5.5.2), be such that (5.5.3) holds. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a nontrivial ME distribution if and only if*

1 $f(u) \geq 0$ *for* $u > 0$,

2 $0 < \dfrac{a_1}{b_1} \leq 1$, *and*

3 there exists a zero of $b(\lambda)$ *of maximal real part that is both real and negative.*

Proof.

Suppose that the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a nontrivial *ME* distribution. If $\boldsymbol{B}$ is defined by (5.5.1) then the distribution has a representation $(\boldsymbol{a}, \boldsymbol{B})$. The corresponding distribution function $F(u)$, which satisfies Conditions 1–3 of Definition 5.4, is of the form (5.2.1). We also have that

$$F(u) \;=\; \int_0^u f(t)dt + \alpha_0 \tag{5.5.10}$$

where $0 \leq \alpha_0 < 1$. Now, since $F(u)$ is nondecreasing for $u \geq 0$, $f(u) \geq 0$ for $u > 0$ and Statement 1 holds. Lemmas 5.2 and 5.5 give Statements 2 and 3, respectively.

Suppose that 1 and 2 in the statement of the theorem hold. We need to show that $F(u)$, defined by (5.5.10), satisfies Conditions 1–3 in Definition 5.4.

1. Since, for $u > 0$, $f(u) \geq 0$, $F(u)$ is nondecreasing for $u \geq 0$. Also, as $F(0) = \alpha_0 = 1 - \dfrac{a_1}{b_1} \geq 0$, by Lemma 5.2 and Statement 2, $F(u)$ is also nonnegative.

2. $F(u)$ is a continuous function for $u > 0$ and is hence right-continuous for $u > 0$. Since, from (5.5.10), $\lim\limits_{u \to 0^+} F(u) = \alpha_0 = F(0)$, $F(u)$ is also right-continuous at $u = 0$.

3. Since $F(u)$ is of the form (5.2.1) its *LST* is given by

$$\phi(\lambda) \;=\; \boldsymbol{a}(\lambda \boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{e}_p + \alpha_0.$$

As (5.5.3) holds, by the *Final Value Theorem* (see Debnath [44, Section 3.8]), and (5.4.6), we have that

$$\begin{aligned}\lim_{u\to\infty} F(u) &= \lim_{\lambda\to 0^+} \phi(\lambda) \\ &= -\boldsymbol{a}\boldsymbol{B}^{-1}\boldsymbol{e}_p + \alpha_0 \\ &= 1.\end{aligned}$$

■

Lemma 5.5 asserts that Statement 3 in Theorem 5.6 is a redundant condition. However, it is included because it is a condition entirely on the vector $\boldsymbol{b}$. The condition $f(u) \geq 0$ for $u > 0$ then becomes a condition entirely on the vector $\boldsymbol{a}$. It is this two-tiered checking procedure that forms the basis for our method of determining whether or not the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution.

The corresponding theorem for *PH* distributions is due to O'Cinneide [104].

Theorem 5.7 *Let $f(u)$ be defined by (5.5.2). The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a nontrivial PH distribution if and only if*

1 $f(u) > 0$ *for* $u > 0$,

2 $0 < \frac{a_1}{b_1} \leq 1$, *and*

3 the zero of $b(\lambda)$ *of maximal real part* $-\zeta$ *is real, negative, and such that* $-\zeta > \Re(-\xi)$ *where* $-\xi$ *is any other zero.*

In contrast to Theorem 5.6 where, as we have shown in Lemma 5.5, Condition 3 is a consequence of Condition 1, for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ to correspond to a nontrivial *PH* distribution *both* Conditions 1 and 3 are required. Consider the following two examples.

1. The distribution with density function

$$f(u) = 2e^{-u} - e^{-u}\cos u - e^{-u}\sin u,$$

and *LST*

$$\phi(\lambda) \;=\; \frac{\lambda^2+\lambda+2}{\lambda^3+3\lambda^2+4\lambda+2},$$

(that is, $\boldsymbol{a}=(2,1,1)$ and $\boldsymbol{b}=(2,4,3)$) has $f(u)=e^{-u}(2-\sqrt{2}\sin(u+\frac{\pi}{4}))>0$ for $u>0$. The function $f(u)$, however, does *not* correspond to a *PH* distribution because the zeros of the polynomial $b(\lambda)$ are $-1,-1\pm i$.

2. The distribution with density function

$$g(u) \;=\; (\frac{1}{2}u^2-2u+2)e^{-u},$$

and *LST*

$$\psi(\lambda) \;=\; \frac{2\lambda^2+2\lambda+1}{(\lambda+1)^3},$$

(that is, $\boldsymbol{a}=(1,2,2)$ and $\boldsymbol{b}=(1,3,3)$) has the zero of maximal real part of $b(\lambda)$, that is $\lambda=-1$, being real, negative, and with real part greater than the real part of any other zero (there are none). The function $g(u)$, in this case, does *not* correspond to a *PH* distribution because $g(2)=0$.

Given vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ that correspond to a rational *LST* of the form (5.4.1), or a companion form representation (5.4.2)–(5.4.4), it is relatively simple to check Conditions 2 and 3 in Theorems 5.6 and 5.7. Checking the first condition is much more difficult. The remainder of this chapter and Chapters 6, 7, and 8 are devoted to the solution of this problem.

5.6 Characterization of Matrix-exponential Distributions

Suppose that the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ given by (5.4.7) and (5.4.8), respectively, satisfy Conditions 2 and 3 of Theorem 5.6. Since these conditions depend only on a_1 and $\boldsymbol{b}$, in order to satisfy Condition 1 we need to find constraints on $a_2, a_3, \ldots, a_{p-1}$.

Let $\boldsymbol{T}$ be given by (5.4.3). At this stage we define the density function $f(u)$ at $u = 0$ as

$$\begin{aligned} f(0) &= \lim_{u \to 0^+} \boldsymbol{a} \exp(\boldsymbol{T}u)\boldsymbol{e}_p \\ &= \boldsymbol{a}\boldsymbol{e}_p \\ &= a_p, \end{aligned} \tag{5.6.1}$$

because it will make some of our subsequent results slightly simpler. Now, Condition 1 of Theorem 5.6 requires that, for $u \geq 0$,

$$f(u) = \boldsymbol{a} \exp(\boldsymbol{T}u)\boldsymbol{e}_p \geq 0. \tag{5.6.2}$$

Let

$$\exp(\boldsymbol{T}u)\boldsymbol{e}_p = \begin{pmatrix} f_1(u) \\ f_2(u) \\ \vdots \\ f_p(u) \end{pmatrix} \tag{5.6.3}$$

be a $p \times 1$ column vector of functions depending on u. The inequality (5.6.2) can be written, for $u \geq 0$, as

$$f(u) = a_1 f_1(u) + a_2 f_2(u) + \ldots + a_p f_p(u) \geq 0. \tag{5.6.4}$$

Since $a_1 = b_1(1 - \alpha_0)$ by Lemma 5.2, there will be a different family of constraints for each α_0 where $0 \leq \alpha_0 < 1$. The situation can be simplified by dividing (5.6.4) through by $1 - \alpha_0$ (which is positive), and letting, for $i = 1, 2, \ldots, p-1$,

$$x_i = \frac{a_{i+1}}{1 - \alpha_0}, \tag{5.6.5}$$

obtaining, for $u \geq 0$,

$$b_1 f_1(u) + x_1 f_2(u) + \ldots + x_{p-1} f_p(u) \geq 0. \tag{5.6.6}$$

In order to satisfy Condition 1 of Theorem 5.6 we need to find constraints on $x_1, x_2, \ldots, x_{p-1}$ so that (5.6.6) holds for $u \geq 0$.

Consider the $(p-1)$-tuple $\boldsymbol{x} = (x_1, x_2, \ldots, x_{p-1})$. Each $\boldsymbol{x} \in \mathbb{R}^{p-1}$ corresponds to a function, defined for $u \geq 0$, of the form

$$g(\boldsymbol{x}, u) \;=\; b_1 f_1(u) + x_1 f_2(u) + \ldots + x_{p-1} f_p(u). \tag{5.6.7}$$

Since we require that $g(\boldsymbol{x}, u) \geq 0$ for $u \geq 0$, we need to find all $\boldsymbol{x} \in \mathbb{R}^{p-1}$ such that (5.6.6) holds for $u \geq 0$. Consequently, this defines a region in $\mathbb{R}^{p-1}$,

$$\Omega_p \;=\; \bigcap_{u \geq 0} \{\; \boldsymbol{x} \in \mathbb{R}^{p-1} \mid b_1 f_1(u) + \sum_{i=1}^{p-1} x_i f_{i+1}(u) \;\geq\; 0\}.$$

Thus, (5.6.6) implies that $f(u) = (1-\alpha_0) g(\boldsymbol{x}, u)$ is the density function of a *ME* distribution if and only if $\boldsymbol{x} \in \Omega_p$.

Concerning Ω_p we have the following.

Theorem 5.8 *Suppose a_1 and $\boldsymbol{b}$ satisfy Conditions 2 and 3 of Theorem 5.6. Then Ω_p is*

1 nonempty,

2 contained in the upper half-space $x_{p-1} \geq 0$, and

3 convex.

Proof.

1. Consider the mixture of the point mass at zero and the (defective) exponential distribution with distribution function

$$F(u) \;=\; \begin{cases} \alpha_0, & u = 0 \\ 1 - (1-\alpha_0) e^{-\lambda_p u}, & u > 0. \end{cases} \tag{5.6.8}$$

The *LST* of (5.6.8) can be expressed as

$$\begin{aligned} \phi(\lambda) \;&=\; \frac{(1-\alpha_0)\lambda_p}{\lambda + \lambda_p} + \alpha_0 && (5.6.9) \\ &=\; \frac{(1-\alpha_0)(\lambda+\lambda_1)(\lambda+\lambda_2)\ldots(\lambda+\lambda_{p-1})\lambda_p}{(\lambda+\lambda_1)(\lambda+\lambda_2)\ldots(\lambda+\lambda_{p-1})(\lambda+\lambda_p)} + \alpha_0 \\ &=\; \frac{(1-\alpha_0)[\bar{x}_{p-1}\lambda^{p-1} + \bar{x}_{p-2}\lambda^{p-2} + \ldots + \bar{x}_1\lambda + b_1]}{\lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2}\ldots + b_1} + \alpha_0, && (5.6.10) \end{aligned}$$

where for $i = 1, 2, \ldots, p-1$, $\bar{x}_i$ is the coefficient of λ^i in the expansion of $(\lambda+\lambda_1)(\lambda+\lambda_2)\ldots(\lambda+\lambda_{p-1})\lambda_p$. From (5.6.9) the distribution (5.6.8) can be represented by the vectors $\boldsymbol{a} = ((1-\alpha_0)\lambda_p)$ and $\boldsymbol{b} = (\lambda_p)$. In this case, from (5.6.3), $f_1(u) = e^{-\lambda_p u}$. From (5.6.10) another pair of vectors that represent the distribution is $\bar{\boldsymbol{a}} = (1-\alpha_0)(b_1, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{p-1})$ and $\bar{\boldsymbol{b}} = (b_1, b_2, \ldots, b_p)$. In this case, for $i = 1, 2, \ldots, p$, $f_i(u)$ is given by (5.6.3) where $\boldsymbol{T}$ is given by (5.4.3). Now, by (5.6.7), we have that, for $u \geq 0$,

$$\begin{aligned} g(\bar{\boldsymbol{x}}, u) &= b_1 f_1(u) + \bar{x}_1 f_2(u) + \bar{x}_2 f_3(u) + \ldots + \bar{x}_{p-1} f_p(u) \\ &= \lambda_p e^{-\lambda_p u} \\ &> 0. \end{aligned}$$

Thus, $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{p-1})$, which corresponds to the distribution (5.6.8), is contained in Ω_p. Therefore, Ω_p is nonempty.

2. Letting $u = 0$ in (5.6.2) gives $a_p \geq 0$ which implies that $x_{p-1} \geq 0$ by (5.6.5).

3. See Poritsky [114]. ■

A natural question to ask is whether the origin is also contained in Ω_p. This is, in fact, Conjecture 5 in O'Cinneide [108] although the conjecture is stated in terms of *LST*s. So far, this conjecture has been established in the affirmative, in general, only for $p = 1, 2, 3, 4$, and for particular cases when $p = 5, 6, \ldots$. Before proceeding we require the following lemma which appears in Commault and Chemla [38] in relation to *PH* (rather than *ME*) distributions. The proof is almost identical.

Lemma 5.9 *The LST*

$$\phi(\lambda) = \frac{\mu(\alpha^2+\beta^2)}{(\lambda+\mu)(\lambda+\alpha+i\beta)(\lambda+\alpha-i\beta)},$$

where $\mu, \alpha, \beta > 0$, corresponds to a ME distribution if and only if $\mu \leq \alpha$.

Proof.

If $\phi(\lambda)$ corresponds to a *ME* distribution then $\mu \leq \alpha$ by Theorem 5.6.

Suppose that $\mu \leq \alpha$. Let $K = \frac{\mu(\alpha^2+\beta^2)}{(\alpha-\mu)^2+\beta^2}$. The inverse *LST* of $\phi(\lambda)$ is, for $u \geq 0$,

$$\begin{aligned} f(u) &= K(e^{-\mu u} - e^{-\alpha u}\cos\beta u - \frac{\alpha-\mu}{\beta}e^{-\alpha u}\sin\beta u) \\ &= Ke^{-\mu u}(1 - e^{(\mu-\alpha)u}(\cos\beta t + \frac{\alpha-\mu}{\beta}\sin\beta u)). \end{aligned} \tag{5.6.11}$$

Now since, for $u > 0$, $\cos\beta u \leq 1$, $\sin\beta u < \beta u$, and $1 + (\alpha-\mu)u \leq e^{(\alpha-\mu)u}$ we have that

$$\begin{aligned} f(u) &\geq Ke^{-\mu u}(1 - e^{(\mu-\alpha)u}(1 + (\alpha-\mu)u)) \\ &\geq Ke^{-\mu u}(1 - e^{(\mu-\alpha)u}e^{(\alpha-\mu)u}) \\ &= 0. \end{aligned}$$

Note that the inequalities above are strict if and only if $\mu < \alpha$. ■

Theorem 5.10 *Let the zeros of*

$$b(\lambda) = \lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} + \ldots + b_1,$$

$-\lambda_1, -\lambda_2, \ldots, -\lambda_p$, *be such that* $\Re(\lambda_1) \geq \Re(\lambda_2) \geq \ldots \geq \Re(\lambda_{p-1}) \geq \lambda_p > 0$. *Suppose that there are* J *pairs of zeros that are complex conjugate pairs, where* $0 \leq J \leq \lfloor\frac{p-1}{2}\rfloor$. *If, for each* $j = 1, 2, \ldots, J$, *the pair of complex conjugate zeros* $\alpha_j \pm i\beta_j$ *can be paired with a real zero, say* μ_j, *such that* $\mu_j \leq \alpha_j$, *then the origin is contained in* Ω_p.

Proof.

The *LST* that corresponds to the origin has the form

$$\phi(\lambda) = (1-\alpha_0)\frac{\lambda_1}{\lambda+\lambda_1}\frac{\lambda_2}{\lambda+\lambda_2}\cdots\frac{\lambda_p}{\lambda+\lambda_p} + \alpha_0$$

where $0 \leq \alpha_0 < 1$. The *LST* $\phi(\lambda)$ consists of factors of the form $\frac{\mu}{\lambda+\mu}$ and $\frac{\mu(\alpha^2+\beta^2)}{(\lambda+\alpha+i\beta)(\lambda+\alpha-i\beta)}$ which both correspond to *ME* distributions. The first

corresponds to the exponential distribution and the second to a distribution of the form (5.6.11). The *LST* $\phi(\lambda)$ corresponds to the convolution of such *ME* distributions and is hence itself a *ME* distribution. Therefore, the origin is contained in Ω_p. ■

Another natural question to ask is whether Ω_p is bounded. In Section 6.1 we show that Ω_3 is bounded by deriving a complete parametric description for its boundary (Theorems 6.4 and 6.5). Although we believe that Ω_p is bounded when $p > 3$ a proof has not yet been found. A discussion on the boundedness of Ω_p will be postponed until Section 7.6.

Now, in order to determine whether or not the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a nontrivial *ME* distribution we have the following algorithm:

1. If $0 < \frac{a_1}{b_1} \leq 1$ then goto 2, else goto 6.
2. Calculate the zeros of the polynomial $b(\lambda) = \lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} + \ldots + b_1$.
3. If there exists a zero of maximal real part that is both real and negative then goto 4, else goto 6.
4. Calculate $\boldsymbol{x} = \frac{1}{1-\alpha_0}(a_2, a_3, \ldots, a_p) = \frac{b_1}{a_1}(a_2, a_3, \ldots, a_p)$.
5. If $\boldsymbol{x} \in \Omega_p$ then goto 7, else goto 6.
6. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ do not correspond to a *ME* distribution.
7. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution.

Steps 1–4 are relatively simple to carry out. In order to be able to perform the test at Step 5 we need to be able to express Ω_p in such a way that it would be computationally feasible to determine whether a $(p-1)$-tuple $\boldsymbol{x}$ belongs to it or not. We begin this task by investigating further the functions $f_1(u), f_2(u), \ldots, f_p(u)$, defined by (5.6.3).

Theorem 5.11 *Let $\boldsymbol{T}$ be defined by (5.4.3). For $i = 1, 2, \ldots, p$, and for $u \geq 0$,*

$$f_i(u) = f_1^{(i-1)}(u) \tag{5.6.12}$$

$$= \boldsymbol{e}_i' \exp(\boldsymbol{T}u)\boldsymbol{e}_p \tag{5.6.13}$$

$$= \boldsymbol{e}_1' \boldsymbol{T}^{i-1} \exp(\boldsymbol{T}u)\boldsymbol{e}_p, \tag{5.6.14}$$

where for $k = 0, 1, 2, \ldots$, $f_1^{(k)}(u)$ is the kth derivative of $f_1(u)$, and $\boldsymbol{e}_i$ is the $p \times 1$ vector with a one in the ith position and zeros elsewhere. Also, $f_1(u)$ is the unique solution to the differential equation

$$f_1^{(p)}(u) + b_p f_1^{(p-1)}(u) + \ldots + b_1 f_1(u) = 0, \tag{5.6.15}$$

with initial conditions

$$f_1(0) = 0, f_1^{(1)}(0) = 0, \ldots, f_1^{(p-2)}(0) = 0, \text{ and } f_1^{(p-1)}(0) = 1. \tag{5.6.16}$$

The derivatives of $f_1(u)$ can be defined at zero in the same way we defined $f(0)$, see (5.6.1).

Proof.

The solution to the differential equation (5.6.15) with initial conditions (5.6.16) exists and is unique, see Apostol [6, Theorem 6.3]. For an *arbitrary* vector $\boldsymbol{a} \in \mathbb{R}^p$ we have, for $u \geq 0$,

$$\boldsymbol{a} \exp(\boldsymbol{T}u)\boldsymbol{e}_p = a_1 f_1(u) + a_2 f_2(u) + \ldots + a_p f_p(u). \tag{5.6.17}$$

Differentiating (5.6.17) with respect to u gives, for $u \geq 0$,

$$\boldsymbol{a}\boldsymbol{T} \exp(\boldsymbol{T}u)\boldsymbol{e}_p = a_1 f_1^{(1)}(u) + a_2 f_2^{(1)}(u) + \ldots + a_{p-1} f_{p-1}^{(1)}(u) + a_p f_p^{(1)}(u). \tag{5.6.18}$$

Now,

$$\begin{aligned}
\boldsymbol{a}\boldsymbol{T}\exp(\boldsymbol{T}u)\boldsymbol{e}_p &= \boldsymbol{a}\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -b_1 & -b_2 & -b_3 & \dots & -b_p \end{pmatrix}\begin{pmatrix} f_1(u) \\ f_2(u) \\ \vdots \\ f_{p-1}(u) \\ f_p(u) \end{pmatrix} \\
&= \boldsymbol{a}\begin{pmatrix} f_2(u) \\ f_3(u) \\ \vdots \\ f_p(u) \\ -b_1f_1(u) - b_2f_2(u) - \ldots - b_pf_p(u) \end{pmatrix} \\
&= a_1f_2(u) + a_2f_3(u) + \ldots + a_{p-1}f_p(u) \\
&\qquad + a_p(-b_1f_1(u) - b_2f_2(u) - \ldots - b_pf_p(u)).
\end{aligned} \tag{5.6.19}$$

Since, for $i = 2, 3, \ldots, p-1$, a_i is arbitrary, equating its coefficients in (5.6.18) and (5.6.19) gives $f_i(u) = f_{i-1}^{(1)}(u)$ which leads to (5.6.12). Equating the coefficients of a_p, which likewise is arbitrary, gives

$$f_p^{(1)}(u) + b_pf_p(u) + \ldots b_2f_2(u) + b_1f_1(u) = 0,$$

and using (5.6.12) gives the differential equation (5.6.15).

Since

$$\begin{aligned}
f(0) &= \boldsymbol{a}\boldsymbol{e}_p \\
&= a_p \\
&= a_1f_1(0) + a_2f_2(0) + \ldots + a_pf_p(0)
\end{aligned}$$

for all real $a_1, a_2, \ldots, a_p$, the initial conditions are, for $i = 1, 2, \ldots p-1$,

$$f_1^{(i-1)}(0) = f_i(0) = 0,$$

and $f_1^{(p-1)}(0) = f_p(0) = 1$, giving (5.6.16).

Now, for $u \geq 0$,

$$\begin{aligned} a_1 f_1(u) + a_2 f_1^{(1)}(u) + \ldots + a_p f_1^{(p-1)}(u) &= \boldsymbol{a} \exp(\boldsymbol{T}u)\boldsymbol{e}_p \\ &= (a_1 \boldsymbol{e}_1' + a_2 \boldsymbol{e}_2' + \ldots + a_p \boldsymbol{e}_p') \exp(\boldsymbol{T}u)\boldsymbol{e}_p. \end{aligned}$$

Equating the coefficients of $a_1, a_2, \ldots, a_p$, we get for $i = 1, 2, \ldots, p$,

$$f_1^{(i-1)}(u) = \boldsymbol{e}_i' \exp(\boldsymbol{T}u)\boldsymbol{e}_p, \tag{5.6.20}$$

which gives (5.6.13). Writing

$$\boldsymbol{T} = \begin{pmatrix} & \boldsymbol{e}_2' & \\ & \boldsymbol{e}_3' & \\ & \vdots & \\ & \boldsymbol{e}_p' & \\ -b_1 & -b_2 & \ldots & -b_p \end{pmatrix}$$

it can be seen that, for $i = 1, 2, \ldots, p-1$,

$$\begin{aligned} \boldsymbol{e}_i' &= \boldsymbol{e}_{i-1}' \boldsymbol{T} \\ &= \boldsymbol{e}_{i-2}' \boldsymbol{T}^2 \\ &\vdots \\ &= \boldsymbol{e}_1' \boldsymbol{T}^{i-1} \end{aligned} \tag{5.6.21}$$

Substituting (5.6.21) into (5.6.20) gives (5.6.14). ■

In the next chapter we first look at the structure of Ω_3 in detail before turning our attention to the general case.

Chapter 6

The Region Ω_p

6.1 The Region Ω_3

In this section we explore the structure of the region Ω_3 in detail.

Consider the vector $\boldsymbol{b} = (b_1, b_2, b_3)$. Suppose that the zeros of the polynomial

$$b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1,$$

$-\lambda_1, -\lambda_2$, and $-\lambda_3$, are such that $\Re(\lambda_1) \geq \Re(\lambda_2) \geq \lambda_3 > 0$. The region in $\mathbb{R}^2$ defined by the family of linear inequalities (5.6.6) is

$$\Omega_3 \;=\; \bigcap_{u\geq 0}\{(x_1, x_2) \in \mathbb{R}^2 \mid b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) \;\geq\; 0\},$$

where $f_1(u)$ is given by Theorem 5.11. Let $\partial\Omega_3$ denote the boundary of Ω_3.

We shall first investigate the structure of Ω_3 numerically. There are six subcases which may yield different types of regions. They occur when the zeros of $b(\lambda)$ are such that

1. $\lambda_1 > \lambda_2 > \lambda_3 > 0$.

2. $\lambda_1 = \lambda_2 > \lambda_3 > 0$.

3. $\lambda_1 > \lambda_2 = \lambda_3 > 0$.

4. $\lambda_1 = \lambda_2 = \lambda_3 > 0$.

5. $\lambda_1 = \alpha + i\beta,\ \lambda_2 = \alpha - i\beta,\ 0 < \lambda_3 < \alpha,\ \beta > 0$.

6. $\lambda_1 = \alpha + i\beta,\ \lambda_2 = \alpha - i\beta,\ 0 < \lambda_3 = \alpha,\ \beta > 0$.

Figure 6.1.1 shows the lines

$$b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) \ = \ 0 \tag{6.1.1}$$

plotted for $u \in U = \{0, 0.01, 0.02, \ldots, 1, 1.1, 1.2, \ldots, 10\}$ for a particular example of each subcase.

Some of the lines near the top and right-hand side of each diagram are spaced further apart than the others because at $u = 1$ the increment in u changes from 0.01 to 0.1. This choice was made to give the clearest picture of Ω_3 possible. Since each line, for $u \geq 0$, actually represents the boundary of a half-space defined by (5.6.6), Ω_3 is the region in the centre of each diagram.

For the four cases where the zeros of $b(\lambda)$ are all real there appear to be three components to $\partial\Omega_3$: two intersecting line segments, one horizontal (which occurs when $u = 0$), and the other with positive slope (which occurs when $u = 10$), and a smooth curve joining them together. The lines occurring near $u = 10$ get closer together quite rapidly. This suggests that as $u \to \infty$ the lines approach a limiting line. Also, upon closer inspection of the curved section we see that every line, for $0 < u < 10$, is tangent to the curve.

When the zeros of $b(\lambda)$ are $-\alpha \pm i\beta$ and $-\lambda_3$ with $\lambda_3 < \alpha$, there are two components to $\partial\Omega_3$: a horizontal line segment (which occurs when $u = 0$) and a smooth curve joining the two ends of the line segment together. As with the case when all the zeros are real, some lines are tangent to Ω_3, but there are also some lines, particularly visible in the bottom right-hand corner of the diagram, that are completely outside the region.

When the zeros of $b(\lambda)$ are $-\alpha \pm i\beta$ and $-\lambda_3$ with $\lambda_3 = \alpha$ the boundary consists only of a closed curve although the horizontal line when $u = 0$ touches the curve.

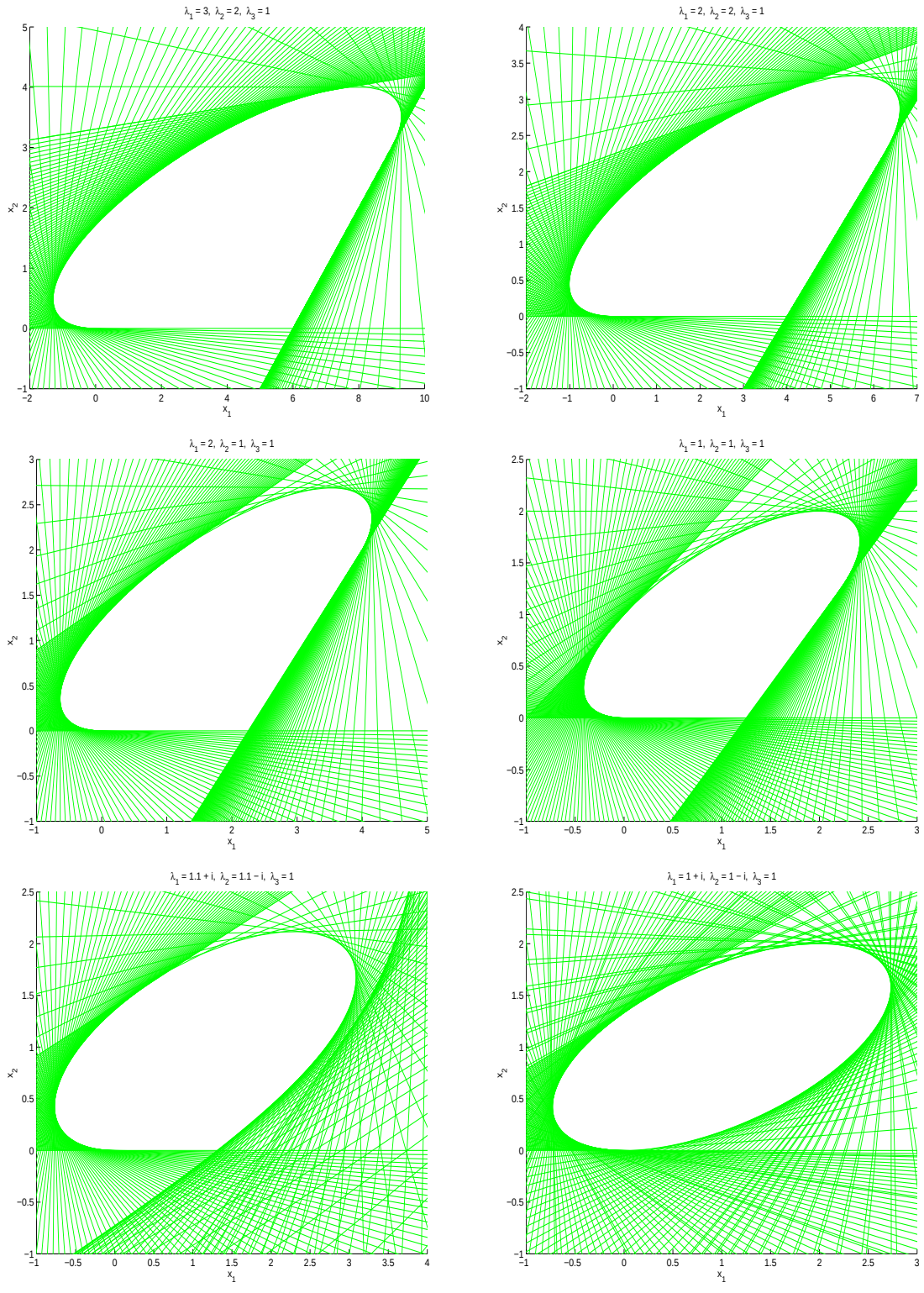

Figure 6.1.1: Plots of Ω_3 for various configurations of the zeros of $b(\lambda)$

Every line appears to be tangent to the curve. Also, some of the lines appear in pairs which would suggest some sort of periodic behaviour.

Theorems 6.4 and 6.5 give an analytical description of $\partial\Omega_3$ when the zeros of $b(\lambda)$ are all real, and two of the zeros are a complex conjugate pair, respectively. Lemmas 6.2 and 6.3 below will be required in the proof of Theorem 6.4. We first require the definition of a *Chebyshev System*, see Kreĭn and Nudel′man [80, pages 31–32 and 173–177].

Definition 6.1 *A system of continuous functions* $\{g_0(t), g_1(t), \ldots, g_n(t)\}$ *defined on* $[c, \infty)$*, with* $c \in \mathbb{R}$*, that satisfy*

$$g_n(t) \;>\; 0 \tag{6.1.2}$$

and

$$\lim_{t\to\infty} \frac{g_i(t)}{g_n(t)} \;=\; 0 \qquad (i = 0, 1, \ldots, n-1), \tag{6.1.3}$$

is called a Chebyshev System *of* order n *if the function*

$$P(t) \;=\; \sum_{i=0}^{n} \beta_i g_i(t) \qquad \textit{with} \qquad \sum_{i=0}^{n} \beta_i^2 \;>\; 0$$

has at most n *distinct zeros in* $[c, \infty)$ *and, if* $\beta_n = 0$*, at most* $n-1$ *distinct zeros.*

It can be shown that the above system of continuous functions is a Chebyshev system if and only if

$$\det \begin{pmatrix} g_0(t_0) & g_1(t_0) & \ldots & g_n(t_0) \\ g_0(t_1) & g_1(t_1) & \ldots & g_n(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ g_0(t_n) & g_1(t_n) & \ldots & g_n(t_n) \end{pmatrix} \;\neq\; 0 \tag{6.1.4}$$

for any $c \le t_0 < t_1 < \ldots < t_n < \infty$, see Kreĭn and Nudel′man [80].

Lemma 6.2 *Let*

$$h(x) \;=\; Ae^{ax} + Be^{bx} + Ce^{cx},$$

where A, B*, and* C *are real numbers, not all zero, and* $a > b > c > 0$*. If*

1 $h(0) > 0$,

2 $h'(0) = 0$,

3 $h''(0) = 0$, *and*

4 $\lim_{x\to\infty} e^{-ax}h(x) = A > 0$,

then, for $x \geq 0$, $h(x) > 0$.

Proof.

Suppose $h(x)$ is nonpositive for some $x > 0$. Then, in order to satisfy Conditions 1 and 4, there exists an $x_1 > 0$ such that $h'(x_1) = 0$. By Conditions 2 and 3 the system of linear equations,

$$\begin{array}{rcrcrcl} aA & + & bB & + & cC & = & 0 \\ a^2A & + & b^2B & + & c^2C & = & 0 \\ ae^{ax_1}A & + & be^{bx_1}B & + & ce^{cx_1}C & = & 0 \end{array} , \tag{6.1.5}$$

must be satisfied.

We shall now show that the system of functions $\{t, t^2, te^{x_1t}\}$ is a Chebyshev system of order two. First, we note that (6.1.2) and (6.1.3) are satisfied because $te^{x_1t} > 0$ when $t \geq c > 0$, and

$$\lim_{t\to\infty} \frac{t}{te^{x_1t}} = 0 \qquad \text{and} \qquad \lim_{t\to\infty} \frac{t^2}{te^{x_1t}} = 0.$$

Second, the function

$$\begin{aligned} P(t) &= \beta_0 t + \beta_1 t^2 + \beta_2 te^{x_1t} \\ &= t(\beta_0 + \beta_1 t + \beta_2 e^{x_1t}) \end{aligned}$$

has at most two distinct zeros on $[c, \infty)$ (at most one zero if $\beta_2 = 0$) since $\{1, t, e^{x_1t}\}$ is a Chebyshev system, see Kreĭn and Nudel′man [80, page 38]. Thus, $\{t, t^2, te^{x_1t}\}$ is a Chebyshev system and by (6.1.4) we have that

$$\det \begin{pmatrix} a & a^2 & ae^{ax_1} \\ b & b^2 & be^{bx_1} \\ c & c^2 & ce^{cx_1} \end{pmatrix} \neq 0$$

when $a > b > c > 0$. This means that $A = B = C = 0$ in (6.1.5) which contradicts the fact that A, B, and C are not all zero. Therefore, $h(x) > 0$ for $x \geq 0$. ■

Lemma 6.3 *Let*

$$h(x) \;=\; Ae^{-ax} + Be^{-bx} + Ce^{-cx},$$

where $A, B,$ and C are real numbers, not all zero, and $a > b > c > 0$. If

1 $h(0) = 0$,

2 $h'(0) = 0$, *and*

3 $h''(0) > 0$,

then, for $x \neq 0$, $h(x) > 0$.

Proof.

The function $h(x)$ has a local minimum at $x = 0$. Suppose $h(x)$ is nonpositive for some $x \neq 0$. Then there exists an $x_1 \neq 0$ such that $h(x_1) = 0$. Therefore, by Conditions 1 and 2 the system of linear equations,

$$\begin{array}{rcrcrcl} A & + & B & + & C & = & 0 \\ aA & + & bB & + & cC & = & 0 \\ e^{-ax_1}A & + & e^{-bx_1}B & + & e^{-cx_1}C & = & 0 \end{array}, \tag{6.1.6}$$

must be satisfied.

Now, if $x_1 > 0$ we shall show that the system of functions $\{e^{-x_1 t}, 1, t\}$ is a Chebyshev system of order two. Conditions (6.1.2) and (6.1.3) are satisfied because $t \geq c > 0$, and

$$\lim_{t\to\infty} \frac{e^{-x_1 t}}{t} = 0 \qquad \text{and} \qquad \lim_{t\to\infty} \frac{1}{t} = 0.$$

Also, the function

$$P(t) \;=\; \beta_0 e^{-x_1 t} + \beta_1 + \beta_2 t$$

has at most two distinct zeros (at most one zero if $\beta_2 = 0$) on $[c, \infty)$ by Rolle's Theorem (see Apostol [5, page 184]) and the fact that $\{e^{-x_1 t}, 1\}$ is a Chebyshev

system of order one. Similarly, we can show that if $x_1 < 0$ the system of functions $\{1, t, e^{-x_1 t}\}$ is a Chebyshev system of order two. In either case

$$\det \begin{pmatrix} 1 & a & e^{-ax_1} \\ 1 & b & e^{-bx_1} \\ 1 & c & e^{-cx_1} \end{pmatrix} \neq 0$$

when $a > b > c > 0$, and so (6.1.6) implies that $A = B = C = 0$, which contradicts the fact that A, B, and C are not all zero. Therefore $h(x) > 0$ for $x \neq 0$. ■

The following theorem gives an analytical description for Ω_3 when the three zeros of $b(\lambda)$ are real.

Theorem 6.4 *Let $\boldsymbol{b} = (b_1, b_2, b_3)$. Suppose that the zeros of $b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1$, $-\lambda_1, -\lambda_2$, and $-\lambda_3$, are all real and such that $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$. For $u \geq 0$, let $f_1(u)$ be defined by Theorem 5.11. Then $\partial\Omega_3$ consists of*

1 the line segment between $(0,0)$ and $(\lambda_1\lambda_2, 0)$,

2 the line segment between $(\lambda_1\lambda_2, 0)$ and $(\lambda_1(\lambda_2 + \lambda_3), \lambda_1)$, and

3 the parametric curve defined, for $u \geq 0$, by

$$x_1(u) = b_1 \frac{f_1^{(1)}(u) f_1^{(2)}(u) - f_1(u) f_1^{(3)}(u)}{f_1^{(1)}(u) f_1^{(3)}(u) - (f_1^{(2)}(u))^2} \tag{6.1.7}$$

$$x_2(u) = b_1 \frac{f_1(u) f_1^{(2)}(u) - (f_1^{(1)}(u))^2}{f_1^{(1)}(u) f_1^{(3)}(u) - (f_1^{(2)}(u))^2}. \tag{6.1.8}$$

Proof.

From (5.6.7) and (5.6.12) we have, for $u \geq 0$, that

$$g(x_1, x_2, u) = b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u). \tag{6.1.9}$$

The theorem will be proved in four steps.

1. First, we prove that $g(x_1, x_2, 0) = 0$ is the line $x_2 = 0$ and that as $u \to \infty$ $g(x_1, x_2, u) = 0$ approaches the line $\lambda_1\lambda_2 - x_1 + \lambda_3 x_2 = 0$. These two lines intersect at $(\lambda_1\lambda_2, 0)$.

2. Next, we determine the point $(x_1(u), x_2(u))$ which is obtained by intersecting $g(x_1, x_2, u) = 0$ with $g(x_1, x_2, u + \delta u) = 0$ and letting $\delta u \to 0$. This point is known as the *characteristic point* corresponding to u. The locus of characteristic points is called the *envelope* of the family of lines given by (6.1.1), see Bell [21, page 207] or Poritsky [114, Section 2]. Let this envelope be denoted by Σ_3.

3. Then, we check that Σ_3 meets the two line segments in the points $(0, 0)$ and $(\lambda_1(\lambda_2 + \lambda_3), \lambda_1)$.

4. Finally, to show that Σ_3 forms the *entire* curved section of $\partial\Omega_3$, we prove that, for $v \geq 0$, an arbitrary point $(x_1(v), x_2(v))$ on Σ_3 satisfies, for $u \geq 0$ with $u \neq v$, the inequality

$$g(x_1(v), x_2(v), u) \ > 0.$$

That is, any point on Σ_3 is on the "feasible" side of *all* the linear constraints $g(x_1, x_2, u) = 0$ where $u \geq 0$.

Step 1. We have from (5.6.16) that $f_1(0) = 0$, $f_1^{(1)}(0) = 0$, and $f_1^{(2)}(0) = 1$. Therefore, the equation of the line $g(x_1, x_2, 0) = 0$ is

$$x_2 \ = \ 0. \tag{6.1.10}$$

Since $f_1(u)$ is the solution to the differential equation (5.6.15) with $p = 3$, it can be written as

$$\begin{aligned}
f_1(u) &= A_{11}e^{-\lambda_1 u} + A_{12}e^{-\lambda_2 u} + A_{13}e^{-\lambda_3 u} \quad \text{when} \quad \lambda_1 < \lambda_2 < \lambda_3, \\
f_1(u) &= A_{21}e^{-\lambda_2 u} + A_{22}ue^{-\lambda_2 u} + A_{23}e^{-\lambda_3 u} \quad \text{when} \quad \lambda_1 = \lambda_2 < \lambda_3, \\
f_1(u) &= A_{31}e^{-\lambda_1 u} + A_{32}e^{-\lambda_3 u} + A_{33}ue^{-\lambda_3 u} \quad \text{when} \quad \lambda_1 < \lambda_2 = \lambda_3, \text{ or as} \\
f_1(u) &= A_{41}e^{-\lambda_3 u} + A_{42}ue^{-\lambda_3 u} + A_{43}u^2e^{-\lambda_3 u} \quad \text{when} \quad \lambda_1 = \lambda_2 = \lambda_3,
\end{aligned}$$

where for $i = 1, 2, 3, 4$, and $j = 1, 2, 3$, A_{ij} is a real constant and can be calculated using the initial conditions (5.6.16). For $i = 1$, as $u \to \infty$, the

dominant terms in $f_1(u)$, $f_1^{(1)}(u)$, and $f_1^{(2)}(u)$ are A_{13}, $-\lambda_3 A_{13}$, and $\lambda_3^2 A_{13}$, respectively. Substituting the expressions for $f_1(u)$, $f_1^{(1)}(u)$, and $f_1^{(2)}(u)$ into $g(x_1, x_2, u) = 0$, dividing through by $A_{13}e^{-\lambda_3 u}$, and letting $u \to \infty$, gives the line

$$b_1 - \lambda_3 x_1 + \lambda_3^2 x_2 \;=\; 0.$$

Since $b_1 = \lambda_1\lambda_2\lambda_3$ this is the line

$$\lambda_1\lambda_2 - x_1 + \lambda_3 x_2 \;=\; 0. \tag{6.1.11}$$

The other three cases are similar. The lines given by (6.1.10) and (6.1.11) meet at the point $(\lambda_1\lambda_2, 0)$.

Step 2. To determine the characteristic point for the family of lines given by (6.1.1) Poritsky [114] showed that this is equivalent to intersecting the two lines

$$b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) \;=\; 0 \tag{6.1.12}$$

$$b_1 f_1^{(1)}(u) + x_1 f_1^{(2)}(u) + x_2 f_1^{(3)}(u) \;=\; 0. \tag{6.1.13}$$

The solution to this system of linear equations is given by (6.1.7) and (6.1.8), which is the parametric equation for Σ_3.

Step 3. We now need to show that the endpoints of Σ_3, one at $u = 0$, the other as $u \to \infty$, meet the lines $x_2 = 0$ and $\lambda_1\lambda_2 - x_1 + \lambda_3 x_2 = 0$ in the points $(0, 0)$ and $(\lambda_1(\lambda_2 + \lambda_3), \lambda_1)$, respectively.

Letting $u = 0$ in (6.1.7) and (6.1.8) gives $x_1 = 0$ and $x_2 = 0$. The algebra for the case when $u \to \infty$ is more complicated so we shall show it only for the case when all zeros are distinct. The proofs for the other three cases are similar. When λ_1, λ_2, and λ_3 are distinct solving the differential equation (5.6.15) with initial conditions (5.6.16) gives

$$f_1(u) \;=\; \frac{e^{-\lambda_1 u}}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)} - \frac{e^{-\lambda_2 u}}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)} + \frac{e^{-\lambda_3 u}}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}. \tag{6.1.14}$$

Substituting (6.1.14) into (6.1.7) and (6.1.8) and multiplying the numerators and denominators through by $e^{(\lambda_1+\lambda_2+\lambda_3)u}$ gives

$$x_1(u) = \frac{\lambda_1\lambda_2\lambda_3((\lambda_2^2-\lambda_3^2)e^{\lambda_1 u}-(\lambda_1^2-\lambda_3^2)e^{\lambda_2 u}+(\lambda_1^2-\lambda_2^2)e^{\lambda_3 u}}{(\lambda_2^2\lambda_3-\lambda_2\lambda_3^2)e^{\lambda_1 u}-(\lambda_1^2\lambda_3-\lambda_1\lambda_3^2)e^{\lambda_2 u}+(\lambda_1^2\lambda_2-\lambda_1\lambda_2^2)e^{\lambda_3 u}} \tag{6.1.15}$$

$$x_2(u) = \frac{\lambda_1\lambda_2\lambda_3((\lambda_2-\lambda_3)e^{\lambda_1 u}-(\lambda_1-\lambda_3)e^{\lambda_2 u}+(\lambda_1-\lambda_2)e^{\lambda_3 u}}{(\lambda_2^2\lambda_3-\lambda_2\lambda_3^2)e^{\lambda_1 u}-(\lambda_1^2\lambda_3-\lambda_1\lambda_3^2)e^{\lambda_2 u}+(\lambda_1^2\lambda_2-\lambda_1\lambda_2^2)e^{\lambda_3 u}} \tag{6.1.16}$$

Since $e^{\lambda_1 u}$ is the dominant term in (6.1.15) and (6.1.16) letting $u \to \infty$ gives $x_1 = \lambda_1(\lambda_2+\lambda_3)$ and $x_2 = \lambda_1$, as required.

Step 4. Suppose that the zeros of $b(\lambda)$ are such that $\lambda_1 > \lambda_2 > \lambda_3 > 0$. The proofs for the other three cases are similar. Letting $u = v$ in (6.1.15) and (6.1.16) and substituting the point $(x_1(v), x_2(v))$ on Σ_3, where $v \geq 0$, into (6.1.9) gives, for $u \geq 0$,

$$\begin{aligned}\frac{g(x_1(v),x_2(v),u)}{b_1} &= \frac{(\lambda_2-\lambda_3)e^{-\lambda_1(u-v)}-(\lambda_1-\lambda_3)e^{-\lambda_2(u-v)}+(\lambda_1-\lambda_2)e^{-\lambda_3(u-v)}}{(\lambda_2^2\lambda_3-\lambda_2\lambda_3^2)e^{\lambda_1 v}-(\lambda_1^2\lambda_3-\lambda_1\lambda_3^2)e^{\lambda_2 v}+(\lambda_1^2\lambda_2-\lambda_1\lambda_2^2)e^{\lambda_3 v}} \qquad (6.1.17)\\ &= \frac{n(u-v)}{d(v)}\end{aligned}$$

The last expression defines $n(\cdot)$ and $d(\cdot)$ as the numerator and denominator of (6.1.17), respectively. Now, since

(a) $d(0) = (\lambda_1-\lambda_2)(\lambda_1-\lambda_3)(\lambda_2-\lambda_3) > 0$,

(b) $d'(0) = 0$,

(c) $d''(0) = 0$, and

(d) $\lim_{v\to\infty} = e^{-\lambda_1 v}d(v) = \lambda_2\lambda_3(\lambda_2-\lambda_3) > 0$,

we have that $d(v) > 0$ for $v \geq 0$, by Lemma 6.2. Also, letting $x = u - v$ we have that

(a) $n(0) = 0$,

(b) $n'(0) = 0$, and

(c) $n''(0) = (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3) > 0$,

which implies that $n(u - v) > 0$ for $u, v \geq 0$ with $u \neq v$, by Lemma 6.3. Therefore, $g(x_1(v), x_2(v), u) > 0$ for $u, v \geq 0$ with $u \neq v$. ■

The next theorem gives an analytical description for Ω_3 when two of the zeros of $b(\lambda)$ are a complex conjugate pair.

Theorem 6.5 *Let $\boldsymbol{b} = (b_1, b_2, b_3)$. Suppose that the zeros of $b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1$, $-\lambda_1, -\lambda_2$, and $-\lambda_3$, are are such that $\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$, and $\lambda_3 = \mu$ with $0 < \mu \leq \alpha$ and $\beta > 0$. For $u \geq 0$, let $f_1(u)$ be defined by Theorem 5.11. Let u^* be the minimal positive solution to*

$$f_1(u)f_1^{(2)}(u) - (f_1^{(1)}(u))^2 \;=\; 0. \tag{6.1.18}$$

Then $\partial\Omega_3$ consists of

1. *the parametric curve, defined for $0 \leq u \leq u^*$, by*

$$x_1(u) \;=\; b_1 \frac{f_1^{(1)}(u)f_1^{(2)}(u) - f_1(u)f_1^{(3)}(u)}{f_1^{(1)}(u)f_1^{(3)}(u) - (f_1^{(2)}(u))^2} \tag{6.1.19}$$

$$x_2(u) \;=\; b_1 \frac{f_1(u)f_1^{(2)}(u) - (f_1^{(1)}(u))^2}{f_1^{(1)}(u)f_1^{(3)}(u) - (f_1^{(2)}(u))^2}, \tag{6.1.20}$$

and

2. *the line segment between $(0, 0)$ and $(x_1(u^*), 0)$ where*

$$x_1(u^*) \;=\; -b_1 \frac{f_1(u^*)}{f_1^{(1)}(u^*)} \tag{6.1.21}$$

$$\;=\; -b_1 \frac{f_1^{(1)}(u^*)}{f_1^{(2)}(u^*)}. \tag{6.1.22}$$

Proof.

From (5.6.7) and (5.6.12) we have, for $u \geq 0$, that

$$g(x_1, x_2, u) \;=\; b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u). \tag{6.1.23}$$

As in Theorem 6.4 $g(x_1, x_2, 0) = 0$ is the line $x_2 = 0$. Also, the parametric equations (6.1.19) and (6.1.20) are determined in exactly the same way as in Step 2 of the proof of Theorem 6.4 and the envelope Σ_3 meets the line $x_2 = 0$ at the point $(0, 0)$. However, in this case Σ_3 intersects the line $x_2 = 0$ at least once when $u > 0$. We are interested in the point that is closest to $(0, 0)$. Let this occur when $u = u^*$. We claim that u^* is the minimal positive solution to $x_2(u) = 0$, or to (6.1.18).

Σ_3 intersects the line $x_2 = 0$ at the point $(x_1(u^*), 0)$. Substituting $x_2 = x_2(u^*) = 0$ into (6.1.12) and (6.1.13) and rearranging, gives (6.1.21) and (6.1.22), respectively. Thus, $\partial\Omega_3$ consists of Σ_3 (when $0 \leq u \leq u^*$) and the line segment between $(0, 0)$ and $(x_1(u^*), 0)$.

We prove the abovementioned claim by showing that

1 a positive solution to (6.1.18) exists,

2 if $\mu = \alpha$, for $0 \leq v \leq u^*$, the point $(x_1(v), x_2(v))$ on Σ_3 satisfies, for $u \geq 0$, $g(x_1(v), x_2(v), u) \geq 0$, and

3 if $\mu < \alpha$, for $0 \leq v \leq u^*$, the point $(x_1(v), x_2(v))$ on Σ_3 satisfies, for $u \geq 0$ with $u \neq v$, $g(x_1(v), x_2(v), u) > 0$.

Steps 2 and 3 assert that any point on Σ_3 satisfies *all* linear inequality constraints $g(x_1, x_2, u) \geq 0$. We prove steps 2 and 3 separately because if $\mu = \alpha$, $g(x_1(v), x_2(v), u) = 0$ whenever u and v differ by an integer multiple of $\frac{2\pi}{\beta}$, whereas if $\mu < \alpha$ the inequality is strict when $u \neq v$. Note that in both cases $g(x_1(u), x_2(u), u) = 0$.

Step 1. Define, for $u \geq 0$,

$$\begin{aligned} q(u) &= f_1(u)f_1^{(2)}(u) - (f_1^{(1)}(u))^2 \\ &= \frac{-\beta e^{-2\alpha u} - (\alpha-\mu)e^{-(\mu+\alpha)u}\sin\beta u + \beta e^{-(\mu+\alpha)u}\cos\beta u}{\beta((\mu-\alpha)^2+\beta^2)}. \end{aligned} \tag{6.1.24}$$

We have that

$$\begin{aligned} q\Big(\frac{\pi}{\beta}\Big) &= -\frac{e^{-\frac{(\mu+\alpha)\pi}{\beta}} + e^{-\frac{2\alpha\pi}{\beta}}}{(\mu-\alpha)^2+\beta^2} \\ &< 0, \end{aligned}$$

and, since $\mu \leq \alpha$, that

$$\begin{aligned} q\Big(\frac{2\pi}{\beta}\Big) &= \frac{e^{-\frac{2(\mu+\alpha)\pi}{\beta}} - e^{-\frac{4\alpha\pi}{\beta}}}{(\mu-\alpha)^2+\beta^2} \\ &\geq 0. \end{aligned}$$

Since $q(u)$ is continuous, by the *Intermediate Value Theorem* there exists a $w \in (\frac{\pi}{\beta}, \frac{2\pi}{\beta}]$ such that $q(w) = 0$ with $w \geq u^*$.

Step 2. Suppose $\mu = \alpha$. Replacing u with v in (6.1.19) and (6.1.20) and substituting these expressions into (6.1.23), after some rearranging we have that

$$g(x_1(v), x_2(v), u) = \frac{b_1 e^{-\alpha u}(1-\cos\beta(u-v))}{\alpha^2+\beta^2-\alpha(\beta\sin\beta v + \alpha\cos\beta v)}. \tag{6.1.25}$$

The denominator of (6.1.25) is positive for $v \geq 0$ because $\alpha(\beta\sin\beta v + \alpha\cos\beta v) \leq \alpha\sqrt{\alpha^2+\beta^2} < \alpha^2+\beta^2$. The numerator is nonnegative for $u, v \geq 0$ and zero if and only if u, v differ by an integer multiple of $\frac{2\pi}{\beta}$. Thus, for $0 \leq v \leq u^*$ and $u \geq 0$, $g(x_1(v), x_2(v), u) \geq 0$, as required.

Step 3. Suppose $\mu < \alpha$. Replacing u with v in (6.1.19) and (6.1.20) and substituting

these expressions into (6.1.23), after some rearranging we get

$$\begin{aligned}\frac{g(x_1(v),x_2(v),u)}{b_1} &= \frac{\beta e^{-\mu(u-v)} - (\alpha-\mu)e^{-\alpha(u-v)}\sin\beta(u-v) - \beta e^{-\alpha(u-v)}\cos\beta(u-v)}{\beta(\alpha^2+\beta^2)e^{\mu v} + \mu(\alpha^2-\mu\alpha-\beta^2)e^{\alpha v}\sin\beta v + \mu\beta(\mu-2\alpha)e^{\alpha v}\cos\beta v} \\ &= \frac{n(u-v)}{d(v)}.\end{aligned} \tag{6.1.26}$$

As previously, $n(\cdot)$ and $d(\cdot)$ define the numerator and denominator of (6.1.26), respectively.

From (6.1.24) we have that $q(0) = 0$, $q(u^*) = 0$ (by definition), $q'(0) = 0$, and $q''(0) = -1$. Thus, for $0 < u < u^*$, $q(u) < 0$.

Now, since $d(0) = \beta((\alpha-\mu)^2+\beta^2) > 0$ and, for $0 \le v \le u^*$,

$$\begin{aligned} d'(v) &= \mu(\alpha^2+\beta^2)[\beta e^{\mu v} + (\alpha-\mu)e^{\alpha v}\sin\beta v - \beta e^{\alpha v}\cos\beta v] \\ &= -\mu\beta(\alpha^2+\beta^2)((\alpha-\mu)^2+\beta^2)e^{(2\alpha+\mu)v}q(v) \\ &\ge 0, \end{aligned}$$

we have that $d(v) > 0$ when $0 \le v \le u^*$.

The expression for $n(u-v)$ differs from $f(u-v)$ given by (5.6.11) only by a positive multiplicative constant. Since $f(u-v) > 0$ for $u > v$ we have, for $0 \le v < u$, that $n(u-v) > 0$. Thus, for $0 \le v \le u^*$ and $u > v$, we have that $g(x_1(v), x_2(v), u) > 0$.

Replacing u with $v-u$ in (6.1.24) gives, for $0 < v-u < u^*$ (note that, $u < v$),

$$-\beta e^{-2\alpha(v-u)} - (\alpha-\mu)e^{-(\mu+\alpha)(v-u)}\sin\beta(v-u) + \beta e^{-(\mu+\alpha)(v-u)}\cos\beta(v-u) < 0. \tag{6.1.27}$$

Multiplying (6.1.27) through by $-e^{(\mu+2\alpha)(v-u)}$ gives

$$\beta e^{\mu(v-u)} + (\alpha-\mu)e^{\alpha(v-u)}\sin\beta(v-u) - \beta e^{\alpha(v-u)}\cos\beta(v-u) > 0. \tag{6.1.28}$$

Now,

$$\begin{aligned} n(u-v) &= \beta e^{-\mu(u-v)} - (\alpha-\mu)e^{-\alpha(u-v)}\sin\beta(u-v) - \beta e^{-\alpha(u-v)}\cos\beta(u-v) \\ &= \beta e^{\mu(v-u)} + (\alpha-\mu)e^{\alpha(v-u)}\sin\beta(v-u) - \beta e^{\alpha(v-u)}\cos\beta(v-u) \\ &> 0, \end{aligned}$$

by (6.1.28). Therefore, for $0 \leq v \leq u^*$ and $u \geq 0$ with $u \neq v$, $g(x_1(v), x_2(v), u) > 0$, as required. ■

Figure 6.1.2 shows $\partial\Omega_3$ for each of the examples depicted in Figure 6.1.1. All envelopes Σ_3 were plotted for $u \in U = \{0, 0.01, 0.02, \ldots, 1, 1.1, 1.2, \ldots, 10\}$ except for the case where $\lambda_1 = 1+i, \lambda_2 = 1-i, \lambda_3 = 1$, in which case the set was truncated at $u = 6.3$. The diagram for the case where $\lambda_1 = 1.1+i, \lambda_2 = 1.1-i, \lambda_3 = 1$ shows Σ_3 plotted also for values of $u \geq u^*$. This was done to illustrate the added complexity when the polynomial $b(\lambda)$ has two zeros that are a complex conjugate pair with real part less than the zero of maximal real part. We remark here that by construction Ω_3 is bounded for all possible cases.

Let $\Gamma_3 = \Sigma_3 \cap \partial\Omega_3$. We note here that when either the zeros of $b(\lambda)$ are all real, or such that the real part of the complex conjugate zeros equal the real zero, we have $\Sigma_3 \subset \partial\Omega_3$, that is $\Gamma_3 = \Sigma_3$. However, when the real part of the complex conjugate zeros is less than the real zero, we have that $\Gamma_3 \subset \Sigma_3$.

6.2 The Constraint $g(\mathbf{x}, u) = 0$ as $u \to \infty$

Before looking at the structure of Ω_p when $p > 3$, in this section we investigate the constraint $g(\boldsymbol{x}, u) = 0$ as $u \to \infty$.

In Theorem 6.4 we saw that when $p = 3$ and $b(\lambda)$ has only real zeros the constraint $g(x_1, x_2, u) = 0$ as $u \to \infty$, which simplifies to $b_1 - \lambda_3 x_1 + \lambda_3^2 = 0$, forms part of $\partial\Omega_3$. In addition, the constraint has the same form when some or all of the zeros are identical. It can be shown that when two of the zeros of $b(\lambda)$ are a complex

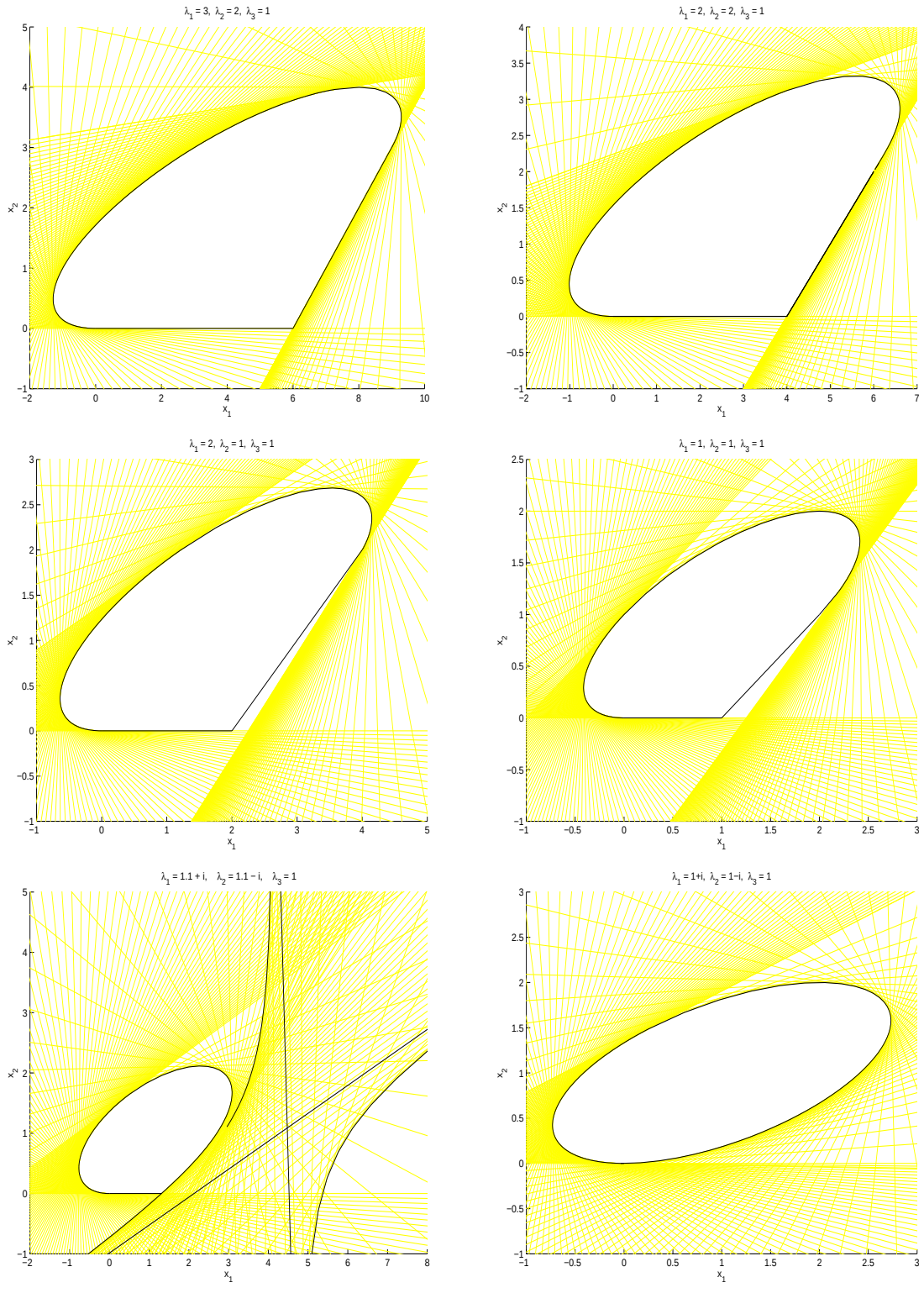

Figure 6.1.2: Plots of $\partial\Omega_3$ for various configurations of the zeros of $b(\lambda)$

conjugate pair with real part equal to λ_3, the constraint at infinity does not exist. However, when two of the zeros of $b(\lambda)$ are a complex conjugate pair with real part less than λ_3, the constraint exists but does not form part of $\partial\Omega_3$. These observations can be generalized for $p > 3$.

Theorem 6.6 *Let $\boldsymbol{b} = (b_1, b_2, \ldots, b_p)$. Suppose that the zeros of*

$$b(\lambda) \;=\; \lambda^p + b_p\lambda^{p-1} + \ldots + b_2\lambda + b_1,$$

$-\lambda_1, -\lambda_2, \ldots, -\lambda_p$, are such that $\Re(\lambda_1) \geq \Re(\lambda_2) \geq \ldots \geq \Re(\lambda_{p-1}) \geq \lambda_p > 0$. Suppose also that if any complex zero of $b(\lambda)$ has real part equal to $-\lambda_p$, it has multiplicity less than the multiplicity of $-\lambda_p$. Then, as $u \to \infty$, the hyperplane

$$g(\boldsymbol{x}, u) \;=\; b_1 f_1(u) + x_1 f_1^{(1)}(u) + \ldots + x_{p-1} f^{(p-1)}(u) \;=\; 0,$$

where $f_1(u)$ is defined by Theorem 5.11, can be expressed as

$$b_1 - \lambda_p x_1 + \lambda_p^2 x_2 + \ldots + (-1)^{p-1}\lambda_p^{p-1} x_{p-1} \;=\; 0. \tag{6.2.1}$$

Proof.

We first note that if there exists a pair of complex conjugate zeros of $b(\lambda)$ (both with multiplicity m) that have real part equal to $-\lambda_p$ (with multiplicity $n \leq m$), say $-\lambda_p \pm i\beta$, as $u \to \infty$, the dominant terms in $f_1(u)$ (and its derivatives) are $u^{m-1}e^{-\lambda_p u}\cos\beta u$ and $u^{m-1}e^{-\lambda_p u}\sin\beta u$. The term $u^{n-1}e^{-\lambda_p u}$ will also be dominant if and only if $n = m$. Under these conditions on the zeros of $b(\lambda)$, $\lim_{u\to\infty} u^{-m+1}e^{\lambda_p u} g(\boldsymbol{x}, u)$ does not exist and consequently no constraint at infinity exists.

We assume that if any complex zero of $b(\lambda)$ has real part equal to $-\lambda_p$, it has multiplicity less than n. Now, since $f_1(u), f_1^{(1)}(u), \ldots, f_1^{(p-1)}(u)$ all approach zero as $u \to \infty$, in order to obtain an expression for the constraint at infinity, we consider

$$u^{-n+1}e^{\lambda_p u} g(\boldsymbol{x}, u) \;=\; 0$$

as $u \to \infty$. We have

$$\begin{aligned}
0 &= \lim_{u\to\infty} u^{-n+1} e^{\lambda_p u} g(\boldsymbol{x}, u) \\
&= \lim_{u\to\infty} \frac{\dfrac{d^{n-1}}{du^{n-1}}(e^{\lambda_p u} g(\boldsymbol{x}, u))}{(n-1)!} \\
&= \frac{1}{(n-1)!} \lim_{\lambda\to 0^+} \lambda \Big[\lambda^{n-1} \frac{x_{p-1}(\lambda-\lambda_p)^{p-1} + x_{p-2}(\lambda-\lambda_{p-1})^{p-2} + \ldots + x_1(\lambda-\lambda_1) + b_1}{(\lambda-\lambda_p+\lambda_1)(\lambda-\lambda_p+\lambda_2)\ldots(\lambda-\lambda_p+\lambda_{p-n})(\lambda-\lambda_p+\lambda_p)^n} \\
&\qquad + \sum_{i=0}^{n-2} \lambda^{n-2-i} \frac{d^i}{du^i}(e^{\lambda_p u} g(\boldsymbol{x}, u))\Big|_{u=0} \Big] \\
&= \frac{1}{(n-1)!} \frac{x_{p-1}(-\lambda_p)^{p-1} + x_{p-2}(-\lambda_{p-1})^{p-2} + \ldots + x_1(-\lambda_1) + b_1}{(-\lambda_p+\lambda_1)(-\lambda_p+\lambda_2)\ldots(-\lambda_p+\lambda_{p-n})},
\end{aligned}$$

which gives (6.2.1). The second equality above is due to $n-1$ applications of *L'Hôpital's Rule*. The third equality is due to *Heaviside's Shifting Theorem* and the *Final Value Theorem* for Laplace transforms, and the expression for the Laplace transform of the $(n-1)$th derivative of a function, see Debnath [44, Sections 3.4 and 3.8]. ■

Corollary 6.7 *Suppose that the zeros of $b(\lambda)$ satisfy the conditions given in Theorem 6.6. The hyperplane (6.2.1) forms part of $\partial\Omega_p$ if and only if there exists an ordering of the zeros of $b(\lambda)$, $\Re(\lambda_1) \geq \Re(\lambda_2) \geq \ldots \geq \Re(\lambda_{p-1}) \geq \lambda_p > 0$, such that λ_{p-1} is real. Furthermore, any ME distribution that corresponds to a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_{p-1})$ on the hyperplane (6.2.1) has an order $p-1$ representation.*

Proof.

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_{p-1})$ satisfy (6.2.1). The *LST* corresponding to the function $g(\boldsymbol{x}, u)$ can be expressed as

$$\begin{aligned}
\phi(\lambda) &= \frac{x_{p-1}\lambda^{p-1} + x_{p-2}\lambda^{p-2} + \ldots + x_1\lambda + b_1}{(\lambda+\lambda_1)(\lambda+\lambda_2)\ldots(\lambda+\lambda_{p-1})(\lambda+\lambda_p)} \\
&= \frac{(y_{p-2}\lambda^{p-2} + y_{p-3}\lambda^{p-3} + \ldots + y_1\lambda + b_1\lambda_p^{-1})(\lambda+\lambda_p)}{(\lambda+\lambda_1)\ldots(\lambda+\lambda_{p-1})(\lambda+\lambda_p)} \\
&= \frac{y_{p-2}\lambda^{p-2} + y_{p-3}\lambda^{p-3} + \ldots + y_1\lambda + b_1\lambda_p^{-1}}{(\lambda+\lambda_1)\ldots(\lambda+\lambda_{p-1})}
\end{aligned}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_{p-2})$, since $-\lambda_p$ is a zero of the numerator of $\phi(\lambda)$. The function $g(\boldsymbol{x}, u)$ will be nonnegative (and hence correspond to the *ME* density function $f(u) = (1-\alpha_0)g(\boldsymbol{x}, u)$ where $0 \leq \alpha_0 < 1$) if and only if λ_{p-1} is real and $\boldsymbol{y}$ is contained in

$$\Omega_{p-1} \;=\; \bigcap_{u \geq 0} \{\boldsymbol{y} \in \mathbb{R}^{p-2} \mid b_1 \lambda_p^{-1} \tilde{f}_1(u) + \sum_{i=1}^{p-2} x_i \tilde{f}_1^{(i)}(u) \;\geq\; 0\},$$

with $\tilde{f}_1(u) = \boldsymbol{e}_1' \exp(\tilde{\boldsymbol{T}} u) \boldsymbol{e}_{p-1}$ where $\tilde{\boldsymbol{T}}$ is the companion matrix that corresponds to the polynomial

$$\tilde{b}(\lambda) \;=\; (\lambda + \lambda_1)(\lambda + \lambda_2) \ldots (\lambda + \lambda_{p-1}),$$

see Theorems 5.6 and 5.11. Since Ω_{p-1} is nonempty by Theorem 5.8, the hyperplane (6.2.1) forms part of $\partial\Omega_p$, and any *ME* distribution corresponding to a point on it will have an order $p-1$ representation $((1-\alpha_0)\boldsymbol{y}, \tilde{\boldsymbol{T}})$ where $0 \leq \alpha_0 < 1$. ■

6.3 The Region Ω_p

As we shall see it is a complex task to determine a complete analytical expression for $\partial\Omega_p$ when $p > 3$.

From (5.6.7) and (5.6.12) we have, for $u \geq 0$, that

$$g(\boldsymbol{x}, u) \;=\; b_1 f_1(u) + x_1 f_1^{(1)}(u) + \ldots + x_{p-1} f_1^{(p-1)}(u). \tag{6.3.1}$$

If the analysis from Step 1 in the proof of Theorem 6.4 is extended we can show that $g(\boldsymbol{x}, 0) = 0$ is the hyperplane $x_{p-1} = 0$. We established, in Theorem 5.10, that under certain conditions on the zeros of $b(\lambda)$ the origin is contained in $\partial\Omega_p$. Thus, in this case the hyperplane forms part of $\partial\Omega_p$. Also, if the zeros of $b(\lambda)$ satisfy the conditions given in Theorem 6.6 and Corollary 6.7, then as $u \to \infty$ the hyperplane $g(\boldsymbol{x}, u) = 0$, that is,

$$\lambda_1 \lambda_2 \ldots \lambda_{p-1} \;+\; \sum_{i=1}^{p-1} (-1)^i \lambda_p^{i-1} x_i \;=\; 0$$

forms part of $\partial\Omega_p$.

We now consider the case when $p = 4$. The *envelope* of the family of planes, defined for $u \geq 0$,

$$g(x_1, x_2, x_3, u) \;=\; 0 \tag{6.3.2}$$

can be defined in a manner analogous to Σ_3, see Step 2 in the proof of Theorem 6.4. The intersection of the two planes

$$g(x_1, x_2, x_3, u) \;=\; 0 \tag{6.3.3}$$

$$g^{(1)}(x_1, x_2, x_3, u) \;=\; 0 \tag{6.3.4}$$

(the differentiation being with respect to u) is called the *characteristic line* of the plane determined by u, see Spivak [133, pages 207–208, 255–263] or Wardle [146, pages 65–75]. The locus of these characteristic lines forms the envelope of the planes (6.3.2). We denote this envelope by Σ_4 and let $\Gamma_4 = \Sigma_4 \cap \partial\Omega_4$. The intersection of (6.3.3) and (6.3.4) with the plane

$$g^{(2)}(x_1, x_2, x_3, u) \;=\; 0$$

gives the *characteristic point* of the plane determined by u which lies on the corresponding characteristic line, see Spivak [133]. The locus of characteristic points, say

$$\boldsymbol{x}(u) \;=\; (x_1(u), x_2(u), x_3(u)),$$

therefore, lies on Σ_4. Under certain conditions on the functional form of the family of planes (6.3.2), Σ_4 can be expressed, for $u \geq 0$, as

$$\boldsymbol{y}(u, s) = \;\boldsymbol{x}(u) + s\boldsymbol{x}^{(1)}(u), \qquad s \in \mathbb{R}. \tag{6.3.5}$$

Equation (6.3.5) is referred to as the *tangent developable* of $\boldsymbol{x}(u)$, see Spivak [133].

In Section 6.1 we remarked that when the zeros of $b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1$ are either all real, or such that the real part of the complex conjugate zeros equals the real zero, then $\Gamma_3 = \Sigma_3$. We also observed that when the real part of the complex

conjugate zeros is less than the real zero, then $\Gamma_3 \subset \Sigma_3$. The extension to the case when $p = 4$ appears difficult. Despite having an analytic expression for Σ_4 it is unclear at this stage how to obtain an expression for Γ_4.

For the general case an analogous expression for the envelope of the family of hyperplanes (6.3.1), denoted by Σ_p, and for the locus of characteristic points can be found. However, again it is unclear how an analytic expression for $\Gamma_p = \Sigma_p \cap \partial\Omega_p$ can be derived. We refer the reader to Poritsky [114] for some work on the problem of finding the region determined by a one-parameter family of linear inequalities.

In the literature Dehon and Latouche [45], and O'Cinneide [107] also gave a geometric characterization of particular classes of *ME* distributions of arbitrary algebraic degree. The former authors considered hyperexponential distributions, the latter, Coxian distributions. We now discuss briefly the work of O'Cinneide [107] leaving a discussion of Dehon and Latouche [45] to Section 7.2.

O'Cinneide [107] used the idea of *invariant polytopes* (see O'Cinneide [104] and [105]) to characterize the class of all Coxian distributions of arbitrary algebraic degree, geometrically. In addition, a procedure to determine the triangular order of a Coxian distribution of any algebraic degree was given. As an example, the class of Coxian distributions with density function of the form, for $u \geq 0$,

$$f(u) \;=\; (x_1\frac{u^2}{2} + x_2u + x_3)e^{-u}. \tag{6.3.6}$$

was characterized. As with our characterization of *ME* distributions, O'Cinneide [107] considered distributions with a possibly nonzero point mass at zero. He found, also as in our case, that the characterization was essentially independent of the value of the point mass at zero. O'Cinneide [107] also characterized the distributions of the form (6.3.6) with triangular order k. For $k = 3, 4, \ldots$, the set

$$P_k \;=\; \{(x_1, x_2, x_3) \in \mathbb{R} \mid x_1 + x_2 + x_3 = 1, x_1, x_3 \geq 0, x_2^2 \leq \frac{2(k-3)}{k-2}x_1x_3\}$$

contains all points $(x_1, x_2, x_3) \in \mathbb{R}^3$ such that (6.3.6) corresponds to a Coxian distribution of triangular order k. Note that, for $k = 3, 4, \ldots$, P_k is a proper subset of

P_{k+1}. The set

$$P_\infty \;=\; \{(x_1, x_2, x_3) \in \mathbb{R} \mid x_1 + x_2 + x_3 = 1, x_1, x_3 \geq 0, x_2^2 \leq 2x_1x_3\}$$

consists of *all* points that correspond to distributions with density of the form (6.3.6). Figure 6.3.1 shows the sets P_3, P_4, P_5, and P_∞. Note that, P_4, for example, consists of the half-ellipse indicated by the arrow *and* the whole of P_3. The points τ_1, τ_2, and τ_3 correspond to the Erlang density functions with unit rate of orders $1, 2$, and 3, respectively. These Erlang distributions correspond to the vertices of P_3.

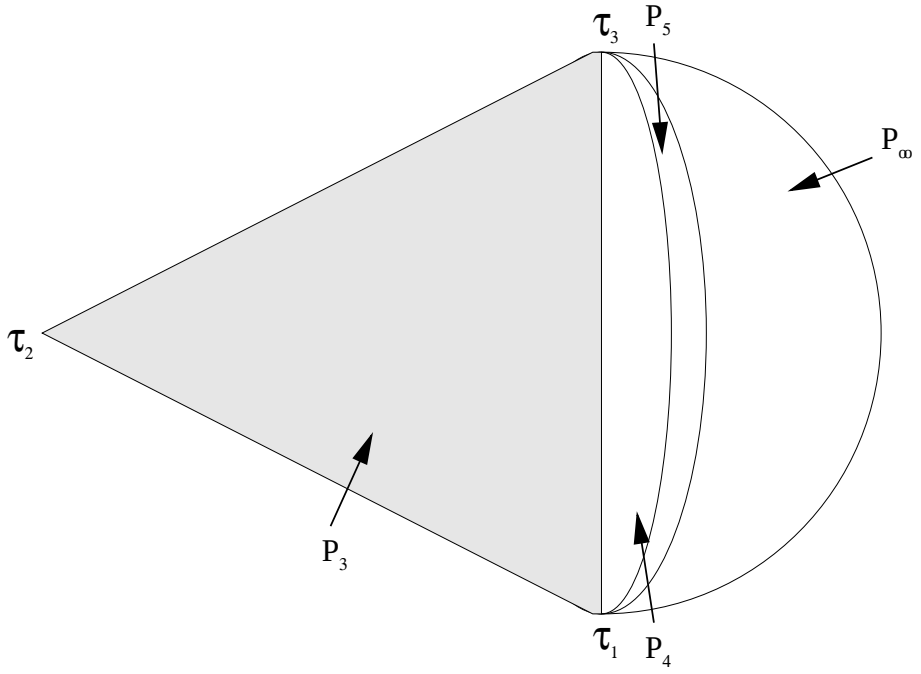

Figure 6.3.1: Diagram of the sets P_3, P_4, P_5, and P_∞

6.4 Comparing the Classes of Matrix-exponential and Phase-type Distributions

The class of *PH* distributions is a proper subset of the class of *ME* distributions, but how much larger is the latter class than the former? We conclude the chapter by answering this question.

In relation to *ME* distributions Asmussen and Bladt [10] mention "...the somewhat smaller class ... of phase-type distributions ...". Later Asmussen and O'Cinneide [12] stated that "... the family of *ME* distributions is in a sense only slightly larger than the family of *PH* distributions." We now show that the latter statement gives a better description of the situation.

Theorem 6.8 *The set of all ME distributions of algebraic degree p that are not PH distributions has measure zero in the set of all ME distributions of algebraic degree p.*

Proof.

Since any *ME* distribution of algebraic degree p can be represented by vectors $\boldsymbol{a} = (a_1, a_2, \ldots, a_p)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_p)$, the set of *all ME* distributions of algebraic degree p has dimension $2p$.

Suppose that

1 $0 < \frac{a_1}{b_1} \leq 1$, and

2 the zero of $b(\lambda)$ of maximal real part $-\zeta$ is real, negative, and such that $-\zeta > \Re(-\xi)$ where $-\xi$ is any other zero.

Then, for any fixed a_1 and $\boldsymbol{b}$ such that Conditions 1 and 2 hold, the subclass of *ME* distributions, represented by Ω_p, has dimension $p-1$. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_{p-1})$ be in the interior of Ω_p. Then, for $u \geq 0$,

$$f(u) \;=\; b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) + \ldots x_{p-1} f_1^{(p-1)}(u) \;>\; 0.$$

The vector $\boldsymbol{x}$, therefore, corresponds to a *PH* distribution by Theorem 5.7. Consequently, there are no *ME* distributions that are not *PH* distributions represented by points in the interior of Ω_p. Consider $\boldsymbol{x} \in \Gamma_p$. Then, for at least one $u > 0$,

$$f(u) \;=\; b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) + \ldots x_{p-1} f_1^{(p-1)}(u) \;=\; 0.$$

In this case, $\boldsymbol{x}$ *cannot* correspond to a *PH* distribution by Theorem 5.7. Consequently, the set of all *ME* distributions of algebraic degree p that are not *PH* distributions with a_1 and $\boldsymbol{b}$ fixed, and satisfying Conditions 1 and 2 above, is represented by Γ_p, a $(p-2)$-dimensional subset of Ω_p. It can now be seen that the set of *all ME* distributions of algebraic degree p that are not *PH* distributions has dimension $2p-1$, and hence has measure zero in the $2p$-dimensional set of all *ME* distributions.

If there exist zeros of $b(\lambda)$ of maximal real part that are not real then no $\boldsymbol{x} \in \Omega_p$ corresponds to a *PH* distribution by Theorem 5.7. Since, in this case, the zero of maximal real part that is real must equal the real part of at least one complex conjugate pair of zeros, such *ME* distributions are parameterized by $2p-1$ parameters. Thus, the set of *all* such *ME* distributions has dimension $2p-1$, which has measure zero in the $2p$-dimensional set of all *ME* distributions of algebraic degree p. ■

Chapter 7

An Algorithm for Identifying Matrix-exponential Distributions

7.1 Introduction

We saw in Chapter 6 that it is a difficult task to completely describe the region Ω_p when $p > 3$. Even for the case when $p = 3$ the analytical description of $\partial\Omega_3$ was complicated. In addition, even if $\partial\Omega_p$ could be adequately described, it is not clear how to ascertain whether a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_{p-1})$ belongs to Ω_p and hence determine if a given pair of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a ME distribution.

In this chapter we will tackle the problem of determining if a given point $\boldsymbol{x} \in \mathbb{R}^{p-1}$ belongs to Ω_p on a "case by case" basis. That is, given $\boldsymbol{x} \in \mathbb{R}^{p-1}$ we develop an algorithm to determine whether or not $\boldsymbol{x} \in \Omega_p$ without actually deriving a parametric description for the whole of $\partial\Omega_p$. In fact, for the case when $p = 3$ this approach actually does lead us to a complete description of $\partial\Omega_3$, arrived at via an approach different to that in Section 6.1.

The method we develop is related to a technique Dehon and Latouche [45] used to characterize the class of all generalized hyperexponential distributions of algebraic degree three. Section 7.2 contains a discussion of their work. In Section 7.3 we

develop the algorithm to determine if the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution. Section 7.4 contains some examples to illustrate the procedure. In Section 7.5 we develop another parameterization to completely describe the region Ω_3. Section 7.6 contains a discussion on the boundedness of Ω_p. The chapter concludes in Section 7.7 with a few remarks.

7.2 The Work of Dehon and Latouche

Dehon and Latouche [45] considered functions, defined for $u \geq 0$, of the form,

$$G(x_1, x_2, \ldots, x_p, u) \;=\; \sum_{i=1}^{p} x_i F_i(u), \tag{7.2.1}$$

where for $i = 1, 2, \ldots, p$, $F_i(u)$ is the exponential distribution function

$$F_i(u) \;=\; 1 - e^{-\lambda_i u},$$

and, without loss of generality, $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$. They investigated the problem of determining conditions on $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ such that (7.2.1) is a distribution function, approaching the characterization problem by utilizing the theory of convex sets. Their approach adds insight into our more general problem.

The region which contains all $\boldsymbol{x} \in \mathbb{R}^p$ such that (7.2.1) is a distribution function is denoted by C_p and its boundary by ∂C_p. C_p is contained in the $(p-1)$-dimensional subspace of $\mathbb{R}^p$

$$x_1 + x_2 + \ldots + x_p \;=\; 1.$$

If the distribution (7.2.1) has a nonzero point mass at zero α_0, C_p is a subset of the subspace

$$x_1 + x_2 + \ldots + x_p \;=\; 1 - \alpha_0.$$

Dehon and Latouche [45] only considered the case when $\alpha_0 = 0$.

For some integer q, where $1 \leq q \leq p$, let $I = \{i_1, i_2, \ldots, i_q\}$ be a subset of $\{1, 2, \ldots, p\}$ such that $i_1 < i_2 < \ldots < i_q$. Denote by $F_{i_1 i_2 \ldots i_q}$ (or F_I) the generalized

Erlang distribution with parameters $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_q}$ (note that $\lambda_{i_1} > \lambda_{i_2} > \ldots > \lambda_{i_q} > 0$). Let T_p denote the set of all points in C_p that correspond to distributions of the form

$$H(u) \;=\; \sum_{i=1}^{p} \zeta_i F_{12\ldots i}(u),$$

where $\zeta_i \geq 0$ with $\sum_{i=1}^{p} \zeta_i = 1$.

Let ∂T_p denote the boundary of T_p. We remark that T_p, which consists of all points that correspond to mixtures of convolutions of the exponential distributions with rates $\lambda_1, \lambda_2, \ldots, \lambda_p$, corresponds to the class of all Coxian distributions of order p whose infinitesimal generators have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$, see Cumani [42] or O'Cinneide [103].

Dehon and Latouche [45] completely characterized C_3 by developing a parametric representation for its boundary (see later). They also noted that while C_p, in general, is complicated to describe, T_p has some interesting properties:

1. T_p contains all $\boldsymbol{x} \in C_p$ that correspond to,

   (a) the exponential distribution F_i, $(i = 1, 2, \ldots, p)$,

   (b) for any $I \subset \{1, 2, \ldots, p\}$ the generalized Erlang distribution F_I, and

   (c) all hyperexponential distributions with distribution function

   $$H(u) \;=\; \sum_{i=1}^{p} x_i F_i(u),$$

   where $x_i \geq 0$.

2. T_p is a $(p-1)$-dimensional hypertetrahedron with vertices determined by the points $\boldsymbol{x}$ that correspond to the generalized Erlang distributions $F_1, F_{12}, \ldots, F_{12\ldots p}$.

3. Every $\boldsymbol{x} \in T_p$ that corresponds to a generalized Erlang distribution F_I, *except* the exponential distribution F_p, is contained in ∂T_p (and also in ∂C_p). In particular

(a) if $I \subset \{1, 2, \ldots, p-1\}$, $\boldsymbol{x} \in T_p$ that corresponds to the generalized Erlang distribution F_I lies in the $(p-2)$-dimensional hyperface of the hypertetrahedron with vertices determined by $F_1, F_{12}, \ldots, F_{12\ldots p-1}$.

(b) if $I \subset \{1, 2, \ldots, p\}$ with $\#I > 1$, $\boldsymbol{x} \in T_p$ that corresponds to the generalized Erlang distribution F_I lies in the $(p-2)$-dimensional hyperface of the hypertetrahedron with vertices determined by $F_{12}, F_{123}, \ldots, F_{12\ldots p}$.

These two hyperfaces form part of ∂C_p.

4. For $i, j \in \{1, 2, \ldots, p\}$, with $i < j$, and $u \geq 0$, we have for some ρ with $0 \leq \rho \leq 1$,

$$F_{12\ldots i-1 i+1\ldots j} \;=\; \rho F_{12\ldots j-1} + (1-\rho) F_{12\ldots j}.$$

Geometrically, the point $\boldsymbol{x} \in T_p$ representing $F_{12\ldots i-1 i+1\ldots j}$ lies on the line segment between the points in T_p representing $F_{12\ldots j-1}$ and $F_{12\ldots j}$.

Figure 7.2.1 shows C_3 and T_3. Each vertex of the triangle is labelled with the symbol denoting its corresponding distribution function. The vertices are F_1, F_{12}, and F_{123} as per note 2 above. The line segments $\overline{F_1 F_{12}}$ and $\overline{F_{12} F_{123}}$ form part of ∂C_3 in accordance with notes 3(a) and 3(b), respectively. Also, $F_2 \in \overline{F_1 F_{12}}$, and $F_{13}, F_{23} \in \overline{F_{12} F_{123}}$, as per note 4. We also have, by symmetry, that $F_3 \in \overline{F_1 F_{13}}$ and $F_3 \in \overline{F_2 F_{23}}$.

It is very likely that the characterization of T_p would be similar if some of $\lambda_1, \lambda_2, \ldots, \lambda_p$ are identical. For example, suppose that $\lambda_1 = \lambda_2 > \lambda_3 > 0$. Then T_3 would look similar to the diagram in Figure 7.2.1 but with F_1 coinciding with F_2, and F_{13} with F_{23}.

In order to describe the curved section of ∂C_3 Dehon and Latouche [45] considered, for $-\dfrac{\lambda_1(\lambda_1 - \lambda_3)}{\lambda_2(\lambda_2 - \lambda_3)} \leq \gamma \leq 0$, the family of rays emanating from F_3 with

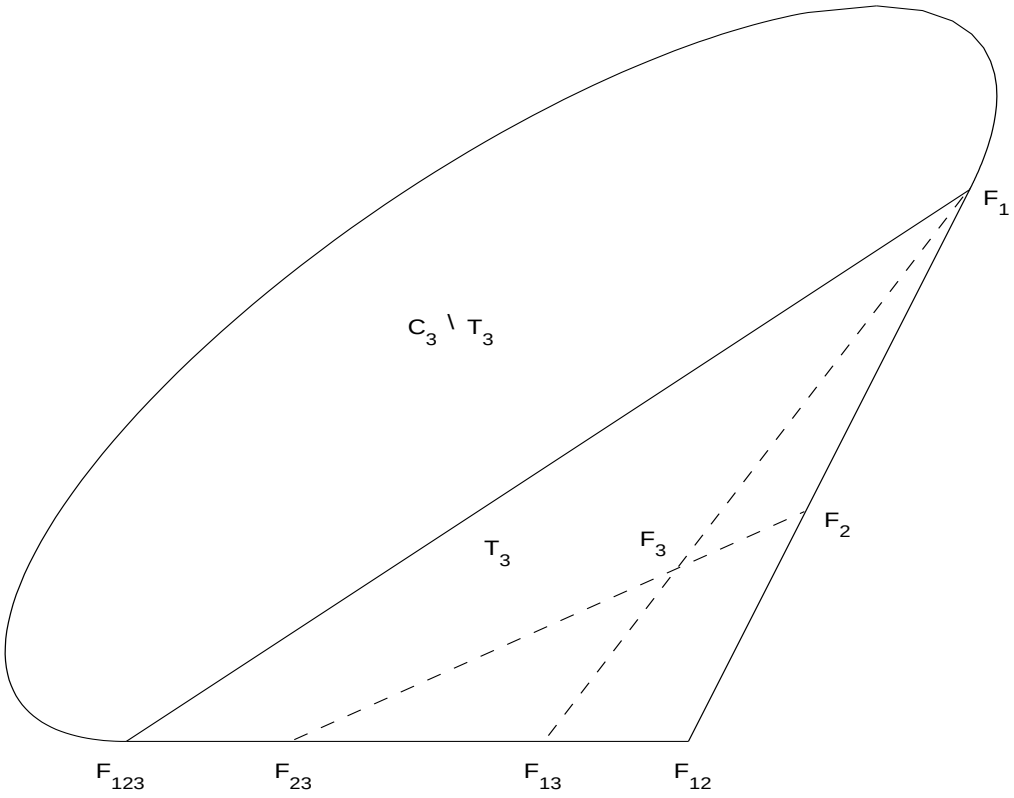

Figure 7.2.1: Diagram of C_3 showing T_3 and the arrangement of the points that represent the distributions F_1, F_2, F_3, F_{12}, F_{13}, F_{23}, and F_{123}

parametric equations

$$\begin{aligned} x_1 &= \nu \\ x_2 &= \nu\gamma \\ x_3 &= 1 - \nu - \nu\gamma, \end{aligned}$$

where $\nu \geq 0$. For each ray $\nu = 0$ corresponds to the point F_3. The ray corresponding to $\gamma = 0$ is $\overrightarrow{F_3F_1}$, and the ray corresponding to $\gamma = -\dfrac{\lambda_1(\lambda_1 - \lambda_3)}{\lambda_2(\lambda_2 - \lambda_3)}$ is $\overrightarrow{F_3F_{123}}$. For a fixed γ they determined the largest possible $\nu > 0$, say $\nu^* = \nu^*(\gamma)$, that ensures the point $(\nu^*, \nu^*\gamma, 1 - \nu^* - \nu^*\gamma)$ corresponds to a generalized hyperexponential distribution, and hence derived an expression for ∂C_3 in terms of γ.

We now state Theorem 1 of Dehon and Latouche [45] which gives the complete

description of C_3.

Theorem 7.1 *C_3 is bounded by the line segments $\overline{F_1F_{12}}$ and $\overline{F_{12}F_{123}}$, and the curve with parametric equations, for $0 \leq \beta \leq 1$,*

$$x_1(\beta) = \frac{\lambda_3}{d(\beta)} \tag{7.2.2}$$

$$x_2(\beta) = -\frac{k\lambda_3\beta}{d(\beta)} \tag{7.2.3}$$

$$x_3(\beta) = 1 - \frac{\lambda_3(1-k\beta)}{d(\beta)} \tag{7.2.4}$$

where

$$\begin{aligned} r &= \frac{\lambda_1 - \lambda_3}{\lambda_2 - \lambda_3}, \\ k &= \frac{\lambda_1(\lambda_1 - \lambda_3)}{\lambda_2(\lambda_2 - \lambda_3)}, \text{ and} \\ d(\beta) &= (-\lambda_1 + k\lambda_2)\beta^r + (1-k\beta)\lambda_3. \end{aligned}$$

Note that $\beta = -\frac{\gamma}{k}$.

We shall now show that the parameterization of ∂C_3 (7.2.2)–(7.2.4) is equivalent to the parameterization given by (6.1.15) and (6.1.16) in Theorem 6.4 when λ_1, λ_2, and λ_3 are distinct. The former description gives the set of all admissible mixing coefficients (x_1, x_2, x_3) of the exponential distributions F_1, F_2, and F_3, while the latter description gives the set of all admissible coefficients of λ and λ^2, (X_1, X_2), in the numerator polynomial of the *LST*

$$\phi(\lambda) = \frac{X_2\lambda^2 + X_1\lambda + b_1}{\lambda^3 + b_3\lambda^2 + b_2\lambda + b_1}.$$

Consider the distribution function, defined for $u \geq 0$ and $0 \leq \beta \leq 1$,

$$F_\beta(u) = x_1(\beta)(1 - e^{-\lambda_1 u}) + x_2(\beta)(1 - e^{-\lambda_2 u}) + x_3(\beta)(1 - e^{-\lambda_3 u}),$$

and its *LST*, defined for $\lambda \in \mathbb{C}$ such that $\Re(\lambda) > -\lambda_3$,

$$\begin{aligned}\phi_\beta(\lambda) &= \frac{x_1(\beta)\lambda_1}{\lambda+\lambda_1} + \frac{x_2(\beta)\lambda_2}{\lambda+\lambda_2} + \frac{x_3(\beta)\lambda_3}{\lambda+\lambda_3} \\ &= \frac{X_2(\beta)\lambda^2 + X_1(\beta)\lambda + X_0(\beta)}{(\lambda+\lambda_1)(\lambda+\lambda_2)(\lambda+\lambda_3)},\end{aligned}$$

where

$$X_0(\beta) = \lambda_1\lambda_2\lambda_3(x_1(\beta)+x_2(\beta)+x_3(\beta)) = \lambda_1\lambda_2\lambda_3, \tag{7.2.5}$$

$$X_1(\beta) = \lambda_1(\lambda_2+\lambda_3)x_1(\beta) + \lambda_2(\lambda_1+\lambda_3)x_2(\beta) + \lambda_3(\lambda_1+\lambda_2)x_3(\beta) \tag{7.2.6}$$

$$X_2(\beta) = x_1(\beta)\lambda_1 + x_2(\beta)\lambda_2 + x_3(\beta)\lambda_3. \tag{7.2.7}$$

Making the transformation $\beta = e^{-(\lambda_1-\lambda_2)u}$ in (7.2.6) and (7.2.7), after some re-arrangement, gives (6.1.15) and (6.1.16), respectively (with "X" replacing "x" throughout). Equation (7.2.5) verifies the *Final Value Theorem* for this situation since

$$\begin{aligned}\lim_{u\to\infty} F_\beta(u) &= x_1(\beta) + x_2(\beta) + x_3(\beta) \\ &= 1,\end{aligned}$$

and

$$\begin{aligned}\lim_{\lambda\to 0^+} \phi_\beta(\lambda) &= \lim_{\lambda\to 0^+} \frac{X_2(\beta)\lambda^2 + X_1(\beta)\lambda + X_0(\beta)}{(\lambda+\lambda_1)(\lambda+\lambda_2)(\lambda+\lambda_3)} \\ &= \frac{X_0(\beta)}{\lambda_1\lambda_2\lambda_3} \\ &= 1.\end{aligned}$$

It can also easily be shown that the line segments between $(0,0)$ and $(\lambda_1\lambda_2, 0)$, and $(\lambda_1\lambda_2, 0)$ and $(\lambda_1(\lambda_2+\lambda_3), \lambda_1)$ in Theorem 6.4, are equivalent to $\overline{F_{123}F_{12}}$ and $\overline{F_{12}F_1}$, respectively, in Theorem 7.1.

7.3 The Matrix-exponential Identification Algorithm

Suppose that a_1 and the vector

$$\boldsymbol{b} \;=\; \begin{pmatrix} b_1 & b_2 & \dots & b_p \end{pmatrix}$$

satisfy Conditions 2 and 3 of Theorem 5.6. Let

$$\hat{\boldsymbol{a}} \;=\; \begin{pmatrix} a_1 & \hat{a}_2 & \dots & \hat{a}_p \end{pmatrix}$$

be a fixed vector. Condition 1 of Theorem 5.6 is satisfied if and only if

$$\begin{aligned} \hat{\boldsymbol{x}} \;&=\; \begin{pmatrix} \hat{x}_1 & \hat{x}_2 & \dots & \hat{x}_{p-1} \end{pmatrix} \\ &=\; \frac{b_1}{a_1} \begin{pmatrix} \hat{a}_2 & \hat{a}_3 & \dots & \hat{a}_p \end{pmatrix} \end{aligned}$$

is contained in the set Ω_p. Let $Q \in \mathbb{R}^{p-1}$ have coordinates $\hat{\boldsymbol{x}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{p-1})$. The point Q represents the function

$$g(\hat{\boldsymbol{x}}, u) \;=\; b_1 f_1(u) + \hat{x}_1 f_1^{(1)}(u) + \hat{x}_2 f^{(2)}(u) + \ldots + \hat{x}_{p-1} f_1^{(p-1)}(u). \tag{7.3.1}$$

In order to determine if Q is in Ω_p (that is, represents a *ME* distribution) we need to ascertain whether $g(\hat{\boldsymbol{x}}, u) \geq 0$ for $u \geq 0$.

Let the point $P \in \mathbb{R}^{p-1}$, with coordinates $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{p-1})$, represent the mixture of the point mass at zero and the exponential distribution

$$F(u) \;=\; \begin{cases} \alpha_0, & u = 0 \\ 1 - (1-\alpha_0)e^{-\lambda_p u}, & u > 0, \end{cases} \tag{7.3.2}$$

where $0 \leq \alpha_0 < 1$, see Figure 7.3.1. Recall from Theorem 5.8 that $\bar{\boldsymbol{x}} \in \Omega_p$ and $\bar{x}_i$ is the coefficient of λ^i in the expansion of $(\lambda + \lambda_1)(\lambda + \lambda_2) \dots (\lambda + \lambda_{p-1})\lambda_p$. Note that, for $u \geq 0$, $g(\bar{\boldsymbol{x}}, u) = \lambda_p e^{-\lambda_p u}$. Let $\overrightarrow{PQ}$ denote the ray emanating from P and passing through Q. For $i = 1, 2, \dots, p-1$, let $\theta_i \in [0, \pi]$ be the angle $\overrightarrow{PQ}$ makes with the

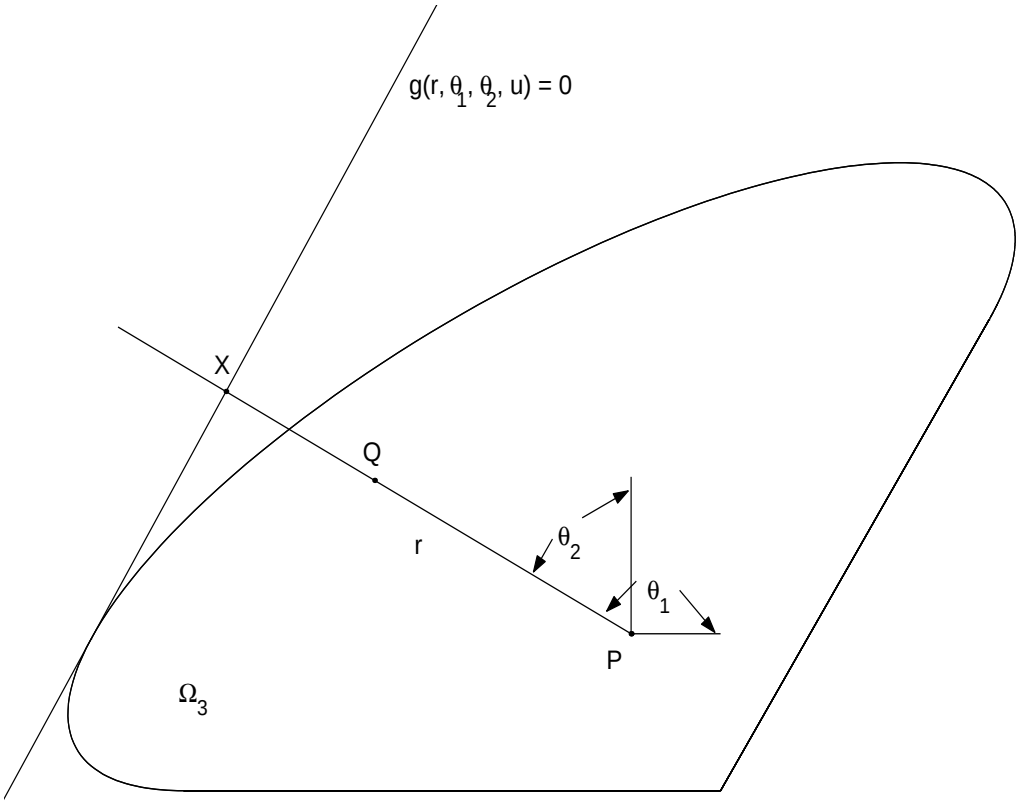

Figure 7.3.1: Diagram of Ω_3 showing the points P, Q, and X

ray emanating from P in the direction parallel to the ith positive coordinate axis. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{p-1})$. We have

$$\cos\theta_i \;=\; \frac{\hat{x}_i - \bar{x}_i}{\|\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}\|_2},$$

where $\|\cdot\|_2$ denotes the length of a vector, that is, if $\boldsymbol{y} \in \mathbb{R}^n$

$$\|\boldsymbol{y}\|_2 \;=\; \sqrt{y_1^2 + y_2^2 + \ldots + y_n^2}.$$

Any point on the line extending in both directions through P and Q has coordinates defined by, for $i = 1, 2, \ldots, p-1$, and $r \in \mathbb{R}$,

$$x_i(r, \theta_i) \;=\; \bar{x}_i + r\cos\theta_i. \tag{7.3.3}$$

For $u \geq 0$, define $r(u)$ to be the distance from the point P to the hyperplane

$$g(\boldsymbol{x}, u) \;=\; b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) + \ldots + x_{p-1} f_1^{(p-1)}(u) \;=\; 0 \tag{7.3.4}$$

measured in the direction determined by the ray $\overrightarrow{PQ}$ (or $\boldsymbol{\theta}$). Suppose that, for any given $u \geq 0$, $\overrightarrow{PQ}$ meets the hyperplane defined by (7.3.4) in the point X. Letting $r = r(u)$ and substituting (7.3.3) into (7.3.4) gives

$$b_1 f_1(u) + \sum_{i=1}^{p-1} (\bar{x}_i + r(u)\cos\theta_i) f_1^{(i)}(u) \;=\; 0. \tag{7.3.5}$$

Rearranging (7.3.5) gives

$$\begin{aligned} r(u) \;&=\; \frac{-b_1 f_1(u) - \bar{x}_1 f_1^{(1)}(u) - \bar{x}_2 f_1^{(2)}(u) - \ldots - \bar{x}_{p-1} f_1^{(p-1)}(u)}{\cos\theta_1 f_1^{(1)}(u) + \cos\theta_2 f_1^{(2)}(u) + \ldots + \cos\theta_{p-1} f_1^{(p-1)}(u)} \\ &=\; \frac{-\lambda_p e^{-\lambda_p u}}{\cos\theta_1 f_1^{(1)}(u) + \cos\theta_2 f_1^{(2)}(u) + \ldots + \cos\theta_{p-1} f_1^{(p-1)}(u)}. \end{aligned} \tag{7.3.6}$$

For $u \geq 0$, we have that $r(u) \neq 0$. If $r(u) < 0$ then $\overrightarrow{PQ}$ never meets the hyperplane $g(\boldsymbol{x}, u) = 0$. However, the ray emanating from P in the direction determined by $(\pi - \theta_1, \pi - \theta_2, \ldots, \pi - \theta_{p-1})$ does. If this ray meets $g(\boldsymbol{x}, u) = 0$ in the point Y, say, the length of the line segment $\overline{PY}$ is $-r(u)$. If $r(u)$ is infinite then $\overrightarrow{PQ}$ is parallel to $g(\boldsymbol{x}, u) = 0$. If $r(u) \leq 0$ for all $u \geq 0$ then Ω_p will be unbounded in the direction determined by $\boldsymbol{\theta}$. It is the absence of any such examples of $r(u)$ which leads us to conjecture that Ω_p is bounded, see Section 7.6.

Let $\boldsymbol{\theta}$ be fixed. Using (7.3.3), define, for $r \in \mathbb{R}$ and $u \geq 0$,

$$\begin{aligned} h(r, \boldsymbol{\theta}, u) \;&=\; g(\boldsymbol{x}(r, \boldsymbol{\theta}), u) \\ &=\; b_1 f_1(u) + \sum_{i=1}^{p-1} (\bar{x}_i + r\cos\theta_i) f_1^{(i)}(u). \end{aligned}$$

For any given $u \geq 0$, since $h(r, \boldsymbol{\theta}, u)$ is a linear function of r, we have either

$$h(r, \boldsymbol{\theta}, u) \geq 0 \quad \text{whenever } r \leq r(u), \text{ and} \tag{7.3.7}$$

$$h(r, \boldsymbol{\theta}, u) < 0 \quad \text{whenever } r > r(u), \tag{7.3.8}$$

or

$$h(r, \boldsymbol{\theta}, u) \geq 0 \quad \text{whenever } r \geq r(u), \text{ and} \tag{7.3.9}$$

$$h(r, \boldsymbol{\theta}, u) < 0 \quad \text{whenever } r < r(u). \tag{7.3.10}$$

Since, for $u \geq 0$, $h(0, \boldsymbol{\theta}, u) = \lambda_p e^{-\lambda_p u} > 0$ conditions (7.3.7) and (7.3.8) hold when $r(u) > 0$, and conditions (7.3.9) and (7.3.10) hold when $r(u) < 0$.

We require, for the calculated cosines of $\boldsymbol{\theta}$ that, for $u \geq 0$,

$$h(r, \boldsymbol{\theta}, u) \ \geq \ 0 \qquad \text{and} \qquad r(u) > 0.$$

Both inequalities are satisfied when r is less than or equal to the global positive minimum of $r(u)$, say r^*. It does not matter if this minimum is attained at more than one value of $u \geq 0$ because it is the quantity r^* that we require. If u^* is the smallest nonnegative value of u where $r(u)$ attains a global positive minimum then $r^* = r(u^*)$. Let $\hat{r} = \|\overline{PQ}\|_2 = \|\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}\|_2$. Now, if $\hat{r} \leq r^*$ then Q, or $\hat{\boldsymbol{x}}$, is in Ω_p, and the vectors $\hat{\boldsymbol{a}}$ and $\boldsymbol{b}$ correspond to a *ME* distribution, otherwise, that is, when $\hat{r} > r^*$, they do not.

We now have the following algorithm to determine whether or not the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution.

1. If $0 < \frac{a_1}{b_1} \leq 1$ then goto 2, else goto 10.
2. Calculate the zeros of the polynomial $b(\lambda) = \lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} + \ldots + b_1$.
3. If there exists a zero of maximal real part that is both real and negative then goto 4, else goto 10.
4. Calculate $\hat{\boldsymbol{x}} = \frac{b_1}{a_1}(a_2, a_3, \ldots, a_p)$.
5. Calculate the coordinates $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)$ of the point P which corresponds to the mixture of the point mass at zero and the exponential distribution (7.3.2).
6. Calculate $\hat{r} = \|\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}\|_2$.
7. For $i = 1, 2, \ldots, p-1$, calculate

$$\cos\theta_i \ = \ \frac{\hat{x}_i - \bar{x}_i}{\|\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}\|_2}.$$

8. Find the value r^* which is the global positive minimum of

$$r(u) = \frac{-\lambda_p e^{-\lambda_p u}}{\cos\theta_1 f_1^{(1)}(u) + \cos\theta_2 f_1^{(2)}(u) + \ldots + \cos\theta_{p-1} f_1^{(p-1)}(u)}.$$

where $u \geq 0$. If no such r^* exists then goto 11.

9. If $\hat{r} > r^*$ then goto 10, else goto 11.

10. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ do not correspond to a *ME* distribution.

11. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution.

At Step 8, instead of finding the global positive minimum of $r(u)$, which is discontinuous whenever the denominator is equal to zero, we can equivalently find the global minimum of the continuous function

$$\begin{aligned} d(u) &= -\frac{\lambda_p}{r(u)} \\ &= e^{\lambda_p u}(\cos\theta_1 f_1^{(1)}(u) + \cos\theta_2 f_1^{(2)}(u) + \ldots + \cos\theta_{p-1} f_1^{(p-1)}(u)). \end{aligned} \tag{7.3.11}$$

If $d^* = d(u^*)$ then $r^* = -\frac{\lambda_p}{d^*}$.

In order to minimize $d(u)$ the inbuilt routine *fminbnd* in MATLAB®, which finds the local minimum of a nonlinear convex function over a finite interval, can be applied repeatedly over intervals where the function is locally convex. In practice, the nature of $d(u)$, determined by the zeros of the polynomial $b(\lambda)$, could be used as a guide in selecting the number and size of such intervals.

According to Reemtsen and Görner [116], however, "... there does not exist an algorithm which is able to detect a global maximizer of an arbitrary continuous function with certainty." The *ME* characterization problem, in general, from an algorithmic point of view, remains a difficult problem to solve. Also, it could be argued that minimizing $d(u)$ is equivalent to minimizing $g(\boldsymbol{x}, u)$ and determining if it ever becomes negative. While this is true, the algorithm presented in this section gives a mechanism by which a function of the form (7.3.1) that does not correspond

to a *ME* distribution can be altered so that it does correspond to a *ME* distribution. That is, by altering $\boldsymbol{x}$ so that the point representing it becomes closer to the point representing the mixture of the point mass at zero and the exponential distribution (7.3.2). This method is far better than the ad hoc procedure due to Harris and Marchal [66] which was described in Section 4.3. Our method also gives us the possibility of describing Ω_p in an alternative way as we shall see for Ω_3 in Section 7.5.

7.4 Examples

We now consider the following three examples of pairs of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ and determine if they correspond to *ME* distributions.

1.

$$\boldsymbol{a} = \begin{pmatrix} 6 & 2 & 2 \end{pmatrix} \tag{7.4.1}$$

$$\boldsymbol{b} = \begin{pmatrix} 6 & 11 & 6 \end{pmatrix} \tag{7.4.2}$$

2.

$$\boldsymbol{a} = \begin{pmatrix} 2.21 & -1 & 1 \end{pmatrix} \tag{7.4.3}$$

$$\boldsymbol{b} = \begin{pmatrix} 2.21 & 4.41 & 3.2 \end{pmatrix} \tag{7.4.4}$$

3.

$$\boldsymbol{a} = \begin{pmatrix} 24 & 60 & 60 & 28 & 12 & 1 \end{pmatrix} \tag{7.4.5}$$

$$\boldsymbol{b} = \begin{pmatrix} 30 & 109 & 159 & 120 & 50 & 11 \end{pmatrix} \tag{7.4.6}$$

In the first example since $\lambda_1 = 3$, $\lambda_2 = 2$, $\lambda_3 = 1$, and $\frac{a_1}{b_1} = 1$, Conditions 2 and 3 of Theorem 5.6 are satisfied. The coordinates of P and Q are $(5, 1)$ and $(2, 2)$, respectively. We have that $\hat{r} = \sqrt{10} \approx 3.16$ and $(\cos\theta_1, \cos\theta_2) = (-\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}})$. Figure

7.4.1 shows Ω_3 with the ray $\overrightarrow{PQ}$ indicated, and Figure 7.4.2 shows the accompanying graph of $r(u)$ versus u for $0 \leq u \leq 5$.

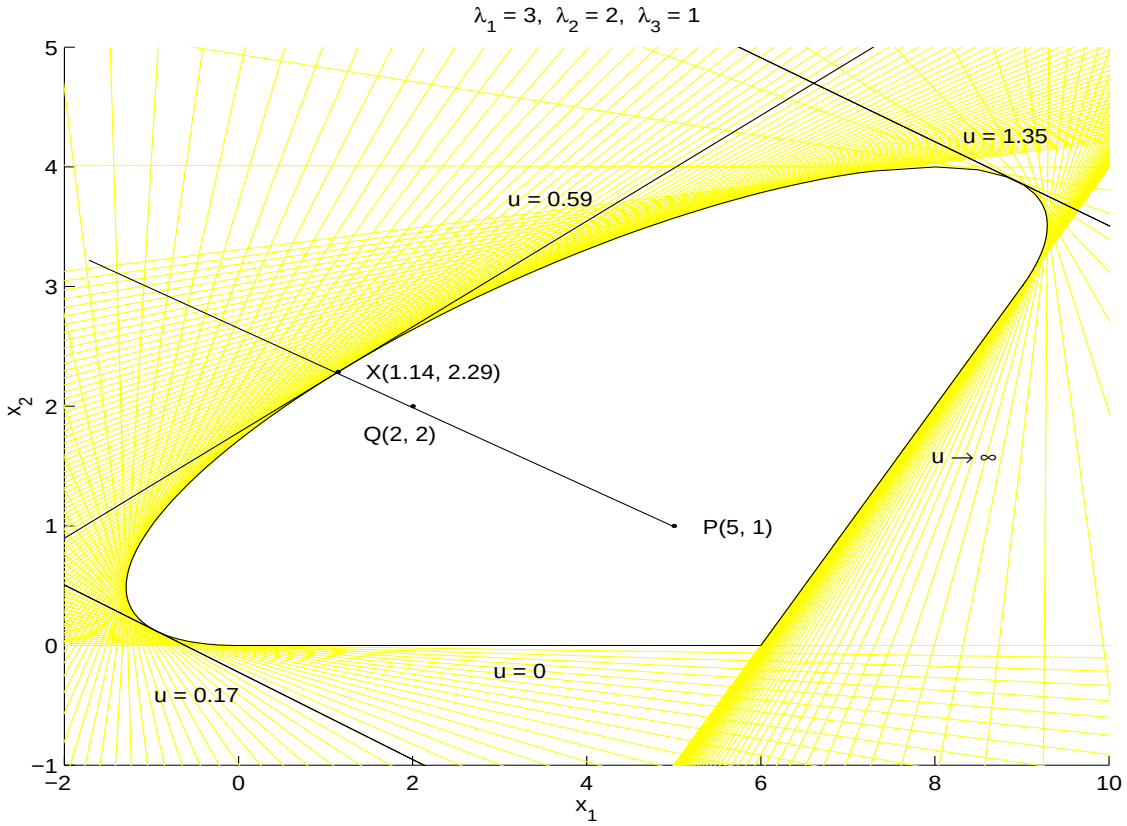

Figure 7.4.1: Diagram of Ω_3 for Example 1

We now discuss how $r(u)$ relates to Ω_3. For $u \geq 0$, $r(u)$ is the distance from P to the line $g(x_1, x_2, u) = 0$ in the direction determined by $\overrightarrow{PQ}$ or $\boldsymbol{\theta}$. When $u = 0$, $r(0) \approx -3.16$ which means that the distance P is from the line $g(x_1, x_2, 0) = x_2 = 0$ is approximately 3.16 units, but the ray emanating from P that intersects the line is in the opposite direction to $\overrightarrow{PQ}$. As u increases $r(u)$ decreases until $u \approx 0.17$ where the corresponding line (plotted on Figure 7.4.1) is parallel to $\overrightarrow{PQ}$ and $r(u)$ is infinite. This is represented by the vertical asymptote at $u \approx 0.17$ in Figure 7.4.2. As u continues to increase $r(u)$ decreases to a local minimum value of approximately 4.07 when $u \approx 0.59$. This value is the global positive minimum of $r(u)$ over $u \geq 0$. Thus, $u^* \approx 0.59$ and $r^* \approx 4.07$. The value of r^* is the maximum distance a point

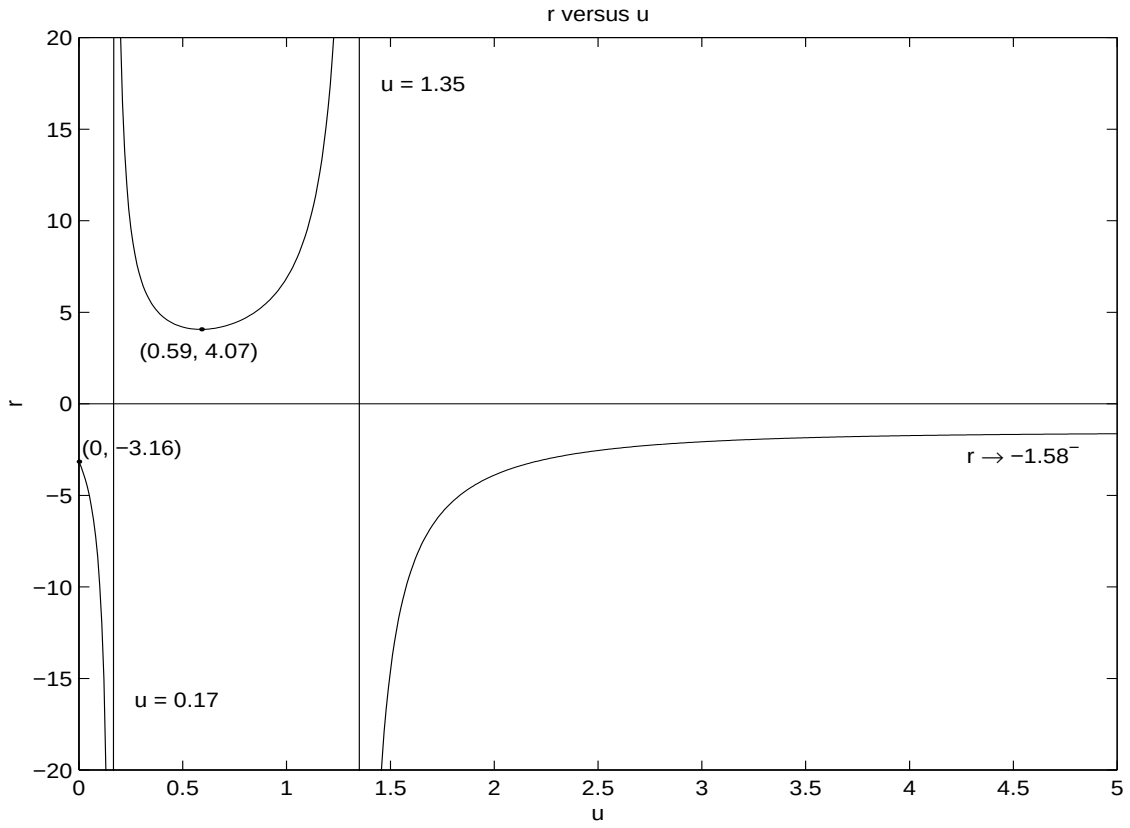

Figure 7.4.2: Graph of $r(u)$ versus u for Example 1

$X(x_1, x_2)$ can be from P in the direction of $\overrightarrow{PQ}$ such that it is in Ω_3 (or such that $g(x_1, x_2, u) \geq 0$ for $u \geq 0$). The point X that is r^* units from P occurs at the intersection of $\overrightarrow{PQ}$ and $g(x_1, x_2, u^*) = 0$ and is indicated on Figure 7.4.1. The coordinates of X are approximately $(1.14, 2.29)$. As u increases from u^*, $r(u)$ increases to infinity when $u \approx 1.35$. This is represented by the vertical asymptote at $u \approx 1.35$ in Figure 7.4.2. The corresponding line, which is parallel to $\overrightarrow{PQ}$ is plotted on Figure 7.4.1. For $u \gtrsim 1.35$, $r(u)$ is always negative and $\lim_{u\to\infty} r(u) \approx -1.58$. Thus, as $u \to \infty$, P is approximately 1.58 units from $g(x_1, x_2, u) = 6 - x_1 + x_2 = 0$ in the direction opposite to $\overrightarrow{PQ}$. Since $\hat{r} < r^*$ the point Q is contained in Ω_3 and the vectors given by (7.4.1) and (7.4.2) correspond to a *ME* distribution. In this example, to minimize $d(u)$ we used the *fminbnd* repeatedly over $[0, 2]$ in intervals of length 0.1. The run time was approximately three seconds.

In the second example $\lambda_1 = 1.1 + i$, $\lambda_2 = 1.1 - i$, $\lambda_3 = 1$, and $\frac{a_1}{b_1} = 1$. Thus, Conditions 2 and 3 of Theorem 5.6 are satisfied. The coordinates of P and Q are $(2.21, 1)$ and $(-1, 1)$, respectively. We have that $\hat{r} \approx 3.20$ and $(\cos\theta_1, \cos\theta_2) = (-1, 0)$. The global positive minimum value of $r(u)$ over $u \geq 0$ is $r^* \approx 2.65$ which occurs when $u^* \approx 0.74$. The point where this occurs is labelled X_1 on both the diagram of Ω_3, shown in Figure 7.4.3, and the graph of $r(u)$ versus u, shown in Figure 7.4.4. Since $\hat{r} > r^*$ the point Q is not contained in Ω_3 and so the vectors given by (7.4.3) and (7.4.4) do not correspond to a *ME* distribution. We also note that the graph of $r(u)$ has an asymptote at $u = 0$ since the ray $\overrightarrow{PQ}$ is parallel to the line $x_2 = 0$.

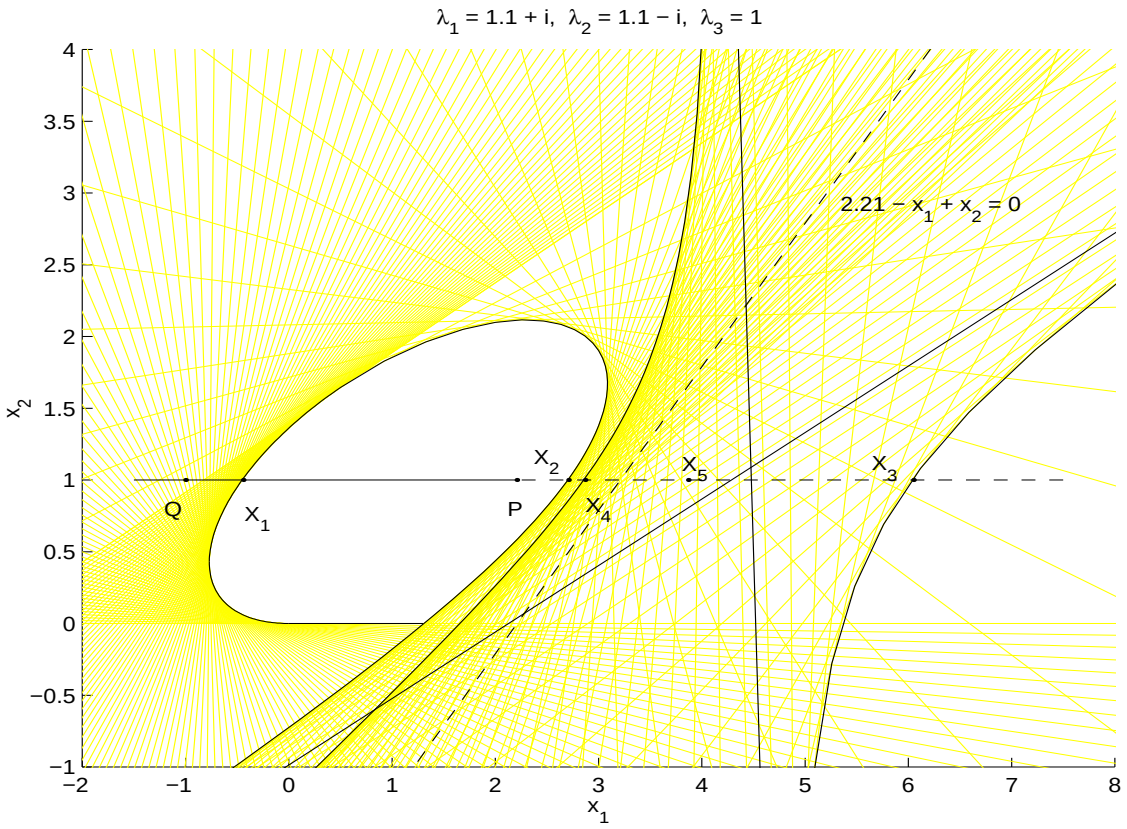

Figure 7.4.3: Diagram of Ω_3 and Σ_3 for Example 2

As u increases from u^*, $r(u)$ increases until it becomes infinite when $u \approx 1.52$. This is indicated by the asymptote in Figure 7.4.4. As u continues to increase, $r(u)$

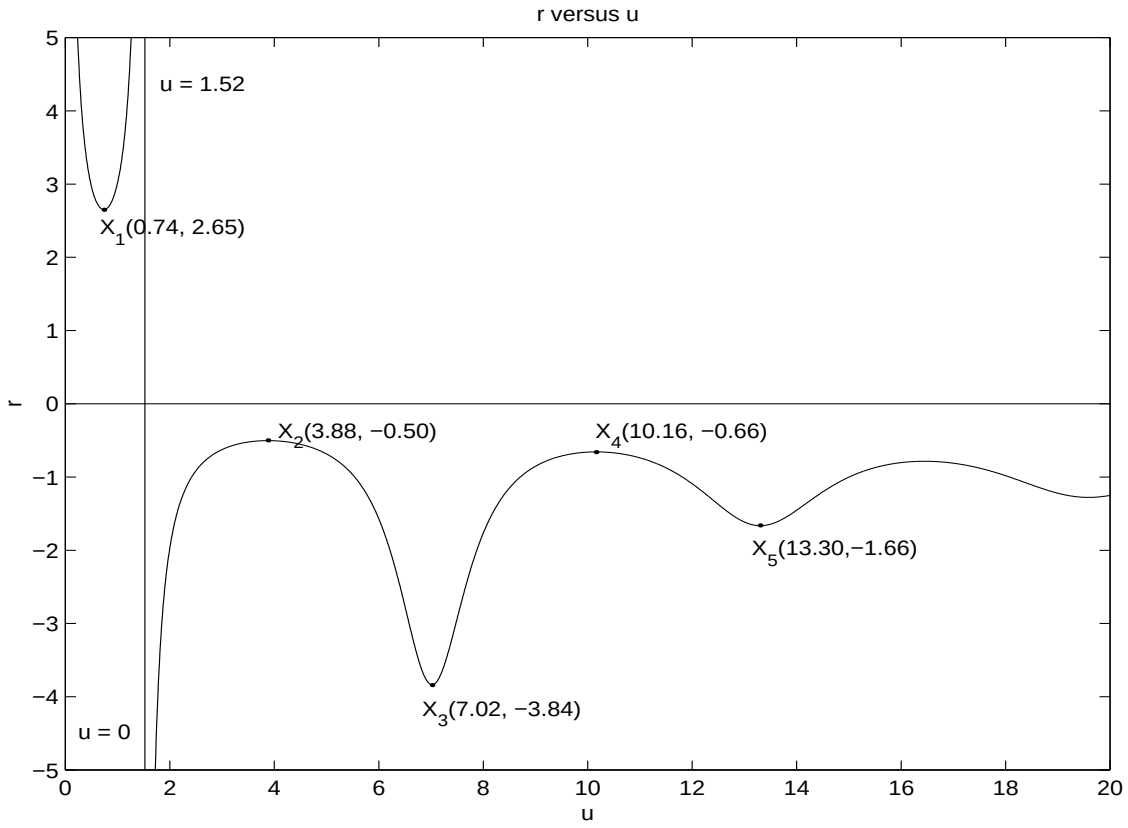

Figure 7.4.4: Graph of $r(u)$ versus u for Example 2

increases to approximately -0.50 when $u \approx 3.88$. Thus, the distance P is from the line $g(x_1, x_2, 3.88) = 0$ in the direction opposite to $\overrightarrow{PQ}$ is approximately 0.50. The point at which this occurs, in both Figures 7.4.3 and 7.4.4, is labelled X_2. As u increases further the envelope $(x_1(u), x_2(u))$ (that is, Σ_3) eventually fails to form the boundary $\partial\Omega_3$. The next three points at which the line $\overline{PX}$ intersects Σ_3 are X_3, X_4, and X_5, in order of increasing u. We can see from Figures 7.4.3 and 7.4.4, that Σ_3 exhibits complicated behaviour. As u increases the lines $g(x_1, x_2, u) = 0$ behave in such a way so that the distance from P to Σ_3 (in the opposite direction to $\overrightarrow{PQ}$) oscillates. When $u \approx 7.02$, this distance attains a local maximum of $\|\overline{PX_3}\|_2 \approx 3.84$. At $u \approx 10.16$, this distance attains a local minimum of $\|\overline{PX_4}\|_2 \approx 0.66$, and so on. The envelope Σ_3 is plotted for $u \in [0, 12]$ and since the point X_5 occurs when $u \approx 13.30$, Σ_3 is not seen to pass through it. As $u \to \infty$, $g(x_1, x_2, u) = 0$ approaches the

line $2.21 - x_1 + x_2 = 0$ (see (6.1.11) and Theorem 6.6) which is indicated by the dashed line in Figure 7.4.3. For the given zeros of $b(\lambda)$ for this example this constraint exists but lies outside of Ω_3, see Corollary 6.7. We have that $\lim_{u\to\infty} r(u) = -1.01$, which means that the point P is 1.01 units from this line in the direction opposite to $\overrightarrow{PQ}$. The solid diagonal lines in Figure 7.4.4 are asymptotes that occur when the denominator of $x_1(u)$ (or denominator of $x_2(u)$ - they are identical functions) equal zero, see (6.1.19) and (6.1.20). For this example $d(u)$ was minimized over $[0, 2]$ in intervals of length 0.1. The run time was approximately 3.5 seconds.

In the final example $\lambda_1 = 3$, $\lambda_2 = 2 + i$, $\lambda_3 = 2 - i$, $\lambda_4 = 2$, $\lambda_5 = 1$, $\lambda_6 = 1$, $\frac{a_1}{b_1} = \frac{4}{5}$. Thus, Conditions 2 and 3 of Theorem 5.6 are satisfied. The coordinates of P and Q are $(79, 80, 40, 10, 1)$ and $(75, 75, 35, 15, 1.25)$, respectively. We have that $\hat{r} \approx 9.5427$ and $(\cos\theta_1, \cos\theta_2, \cos\theta_3, \cos\theta_4, \cos\theta_5) \approx (-0.4192, -0.5240, -0.5240, 0.5240, 0.0262)$. The plot of $r(u)$ versus u is shown in Figure 7.4.5. We can see from the graph that $r^* \approx 10.05$ when $u^* \approx 4.31$. Since $\hat{r} < r^*$ the vectors given by (7.4.5) and (7.4.6) correspond to a *ME* distribution. The graph of u versus $r(u)$ gives us the distance the point P is from the hyperplane $g(\boldsymbol{x}, u) = 0$ and whether or not the intersection of the line $\overline{PQ}$ with the hyperplane is in the same direction as $\overrightarrow{PQ}$. The graph suggests that the geometry of the situation is quite complicated. For this example $d(u)$ was minimized over $[0, 5]$ in intervals of length 0.1. The run time was approximately 9.5 seconds.

7.5 Another Parameterization of Ω_3

In this section, using the method developed in Section 7.3, we give an alternative parameterization of $\partial\Omega_3$. Theorem 7.2 gives the parameterization when the zeros of the polynomial $b(\lambda)$ are all real and distinct. Similar results when some or all of the zeros are repeated are stated without proof in Theorem 7.3. Theorem 7.4 gives the parameterization when two of the zeros are a complex conjugate pair.

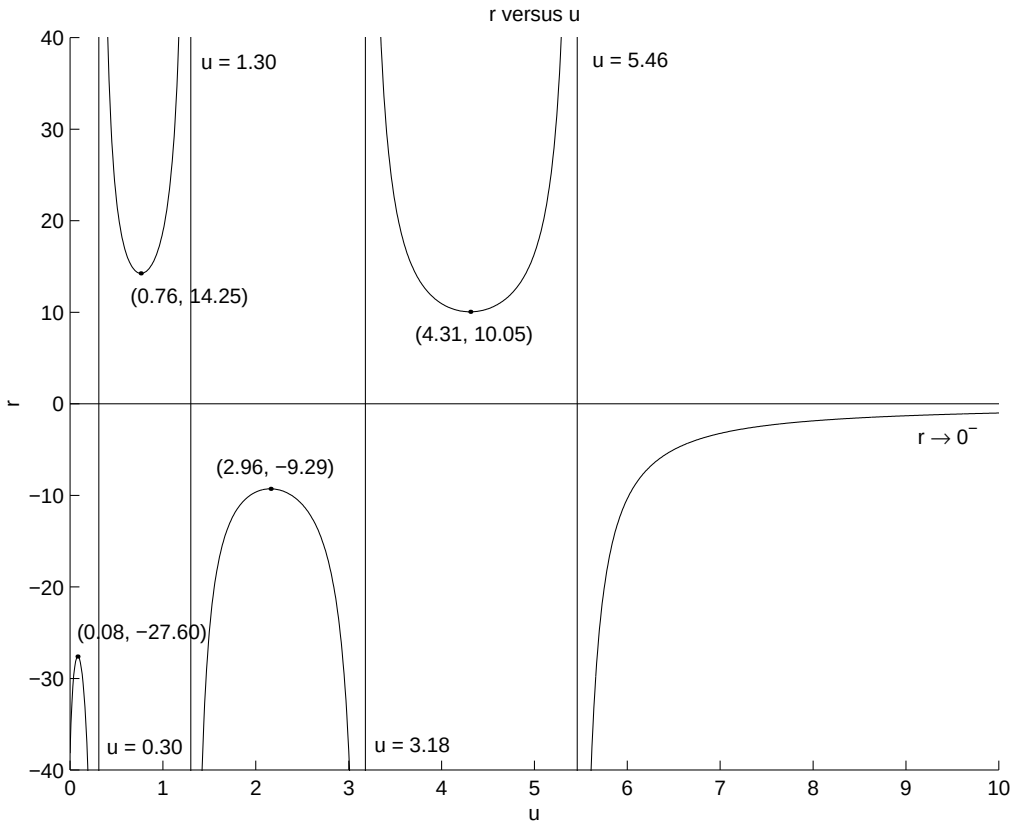

Figure 7.4.5: Graph of $r(u)$ versus u for Example 3

Theorem 7.2 *Suppose that the vectors* $\boldsymbol{a} = (a_1, a_2, a_3)$ *and* $\boldsymbol{b} = (b_1, b_2, b_3)$ *are such that*

1 $0 < \dfrac{a_1}{b_1} \leq 1$, *and*

2 *the zeros of* $b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1$, $-\lambda_1, -\lambda_2$, *and* $-\lambda_3$, *are real and such that* $\lambda_1 > \lambda_2 > \lambda_3 > 0$.

If θ_1, θ_2, *and* θ_3 *are such that*

1 $\tan\theta_1 = \dfrac{1}{\lambda_1 + \lambda_2}$ *and* $\pi < \theta_1 < \dfrac{3\pi}{2}$,

2 $\tan\theta_2 = \dfrac{\lambda_3}{\lambda_1\lambda_3 + \lambda_2\lambda_3 - \lambda_1\lambda_2}$ *and* $\pi < \theta_2 < 2\pi$, *and*

3 $\tan\theta_3 = \dfrac{1}{\lambda_2}$ *and* $0 < \theta_3 < \dfrac{\pi}{2}$,

then the parametric representation for $\partial\Omega_3$ is:

1. *For $\theta_1 < \theta \leq \theta_2$,*

$$x_1(\theta) = -\lambda_3 \cot\theta + \lambda_3(\lambda_1 + \lambda_2) \tag{7.5.1}$$

$$x_2(\theta) = 0. \tag{7.5.2}$$

2. *For $\theta_2 < \theta \leq \theta_3 + 2\pi$,*

$$x_1(\theta) = \frac{(\lambda_1\lambda_2 + \lambda_3^2)\cos\theta - \lambda_3^2(\lambda_1 + \lambda_2)\sin\theta}{\cos\theta - \lambda_3 \sin\theta} \tag{7.5.3}$$

$$x_2(\theta) = \frac{\lambda_3\cos\theta + (\lambda_1\lambda_2 - \lambda_1\lambda_3 - \lambda_2\lambda_3)\sin\theta}{\cos\theta - \lambda_3\sin\theta}. \tag{7.5.4}$$

3. *For $\theta_3 < \theta \leq \theta_1$,*

$$x_1(\theta) = \frac{\lambda_3(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)\cos\theta}{\lambda_3(\cos\theta - \lambda_3\sin\theta) - \lambda_1(\cos\theta - \lambda_1\sin\theta)\gamma(\theta)^{\frac{\lambda_3-\lambda_1}{\lambda_1-\lambda_2}}} + \lambda_3(\lambda_1 + \lambda_2) \tag{7.5.5}$$

$$x_2(\theta) = \frac{\lambda_3(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)\sin\theta}{\lambda_3(\cos\theta - \lambda_3\sin\theta) - \lambda_1(\cos\theta - \lambda_1\sin\theta)\gamma(\theta)^{\frac{\lambda_3-\lambda_1}{\lambda_1-\lambda_2}}} + \lambda_3, \tag{7.5.6}$$

where

$$\gamma(\theta) = \frac{\lambda_1(\cos\theta - \lambda_1\sin\theta)}{\lambda_2(\cos\theta - \lambda_2\sin\theta)}. \tag{7.5.7}$$

Proof.

Refer to Figure 7.5.1 below for a diagram of the situation. Recall from Theorem 6.4 that $\partial\Omega_3$ consists of

1 the line segment between $O(0,0)$ and $R(\lambda_1\lambda_2, 0)$,

2 the line segment between R and $S(\lambda_1(\lambda_2 + \lambda_3), \lambda_1)$, and

3 the curve Γ_3 which has as its endpoints O and S.

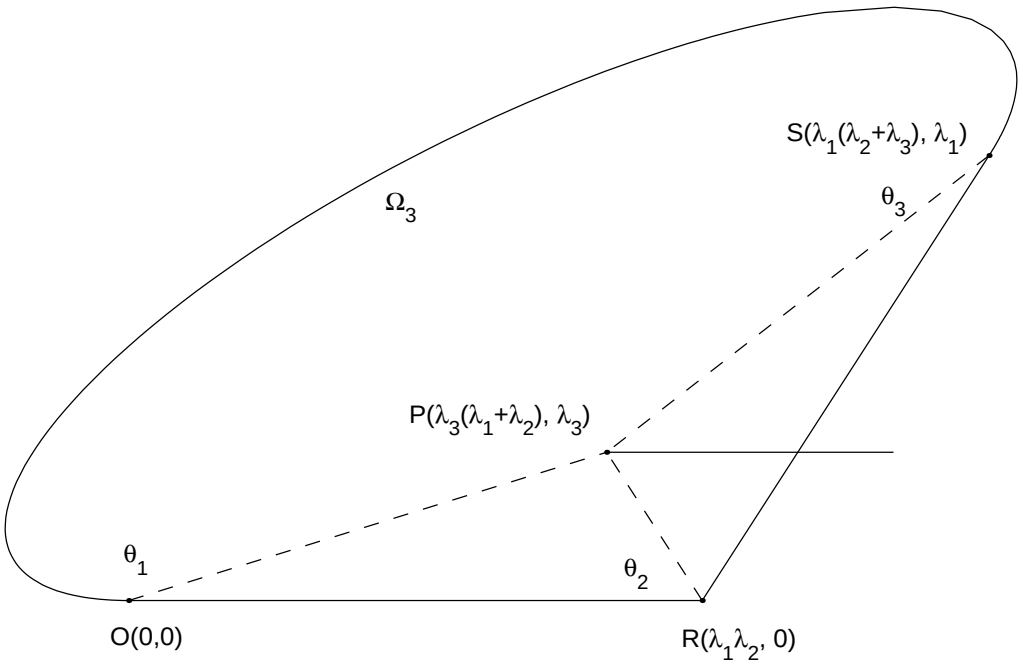

Figure 7.5.1: Diagram of Ω_3 showing the points O, P, R, and S

We show that these three sections of $\partial\Omega_3$ are parameterized by (7.5.1) and (7.5.2), (7.5.3) and (7.5.4), and (7.5.5) and (7.5.6), respectively.

Recall, also, that the mixture of the point mass at zero and the exponential distribution (7.3.2) is represented by the point $P(\lambda_3(\lambda_1 + \lambda_2), \lambda_3)$. Let θ be the angle between the ray emanating from P parallel to the positive x_1-axis and any ray emanating from P, measured in an anticlockwise direction. Let θ_1, θ_2, and θ_3 be such angles that are defined by the rays $\overrightarrow{PO}$, $\overrightarrow{PR}$, and $\overrightarrow{PS}$, respectively. We have that $\tan\theta_1 = \dfrac{1}{\lambda_1 + \lambda_2}$, and since the coordinates of P are both positive then $\pi < \theta_1 < \dfrac{3\pi}{2}$. Also, $\tan\theta_2 = \dfrac{\lambda_3}{\lambda_1\lambda_3 + \lambda_2\lambda_3 - \lambda_1\lambda_2}$, and since the x_2 coordinate of P is positive then $\pi < \theta_2 < 2\pi$. Lastly, $\tan\theta_3 = \dfrac{1}{\lambda_2}$, and since both coordinates of S are greater than their respective coordinates of P, then $0 < \theta_3 < \dfrac{\pi}{2}$.

For $p = 3$, from (7.3.6),

$$r(u) \;=\; \frac{-\lambda_3 e^{-\lambda_3 u}}{\cos\theta f_1^{(1)}(u) + \sin\theta f_1^{(2)}(u)},$$

and therefore

$$\begin{aligned} d(u) \;&=\; -\frac{\lambda_3}{r(u)} \\ &=\; e^{\lambda_3 u}(\cos\theta f_1^{(1)}(u) + \sin\theta f_1^{(2)}(u)). \end{aligned} \tag{7.5.8}$$

The function $d(u)$ attains a global minimum when one of the following three situations occur:

1 $u = 0$,

2 $u \to \infty$, or

3 $d'(u) = 0$.

1. Suppose the global minimum is achieved when $u = u^* = 0$. The minimum value of $d(u)$ is therefore

$$d^* \;=\; d(0) \;=\; \sin\theta.$$

Thus, the corresponding points on the boundary $\partial\Omega_3$, using (7.3.3), are given by

$$\begin{aligned} x_1(\theta) \;&=\; -\frac{\lambda_3}{d^*}\cos\theta + \lambda_3(\lambda_1 + \lambda_2) \\ &=\; -\lambda_3 \cot\theta + \lambda_3(\lambda_1 + \lambda_2), \end{aligned} \tag{7.5.9}$$

and

$$\begin{aligned} x_2(\theta) \;&=\; -\frac{\lambda_3}{d^*}\sin\theta + \lambda_3 \\ &=\; 0. \end{aligned}$$

This situation occurs when the ray emanating from P in the direction defined by θ intersects the line $g(x_1, x_2, 0) = 0$, that is, $\overleftrightarrow{OR}$. Thus, we have that $\theta_1 < \theta \leq \theta_2$. Note that (7.5.9) is defined for these values of θ because $\cot\theta$ is defined for $\pi < \theta < 2\pi$.

2. Suppose the global minimum is achieved as $u \to \infty$. The minimum value is therefore

$$\begin{aligned} d^* &= \lim_{u\to\infty} d(u) \\ &= \lim_{u\to\infty} e^{\lambda_3 u}(\cos\theta f_1^{(1)}(u) + \sin\theta f_1^{(2)}(u)) \\ &= \frac{-\lambda_3\cos\theta + \lambda_3^2\sin\theta}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}. \end{aligned}$$

The last equality is achieved by observing that, from (6.1.14), the dominant terms in $f_1^{(1)}(u)$ and $f_1^{(2)}(u)$, as $u \to \infty$, are $\frac{-\lambda_3 e^{-\lambda_3 u}}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}$ and $\frac{\lambda_3^2 e^{-\lambda_3 u}}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}$, respectively. Thus, the points on the boundary of $\partial\Omega_3$ are given by

$$\begin{aligned} x_1(\theta) &= -\frac{\lambda_3}{d^*}\cos\theta + \lambda_3(\lambda_1+\lambda_2) \\ &= \frac{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)\cos\theta}{\cos\theta - \lambda_3\sin\theta} + \lambda_3(\lambda_1+\lambda_2) \\ &= \frac{(\lambda_1\lambda_2+\lambda_3^2)\cos\theta - \lambda_3^2(\lambda_1+\lambda_2)\sin\theta}{\cos\theta-\lambda_3\sin\theta}, \end{aligned} \tag{7.5.10}$$

and

$$\begin{aligned} x_2(\theta) &= -\frac{\lambda_3}{d^*}\sin\theta + \lambda_3 \\ &= \frac{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)\sin\theta}{\cos\theta-\lambda_3\sin\theta} + \lambda_3 \\ &= \frac{\lambda_3\cos\theta + (\lambda_1\lambda_2 - \lambda_1\lambda_3 - \lambda_2\lambda_3)\sin\theta}{\cos\theta-\lambda_3\sin\theta}. \end{aligned} \tag{7.5.11}$$

This occurs when the ray emanating from P in the direction defined by θ intersects the line $g(x_1, x_2, u) = 0$ as $u \to \infty$, that is, $\overleftrightarrow{RS}$. Thus, $\theta_2 <$

$\theta \leq 2\pi + \theta_3$. Note that (7.5.10) and (7.5.11) are defined for such θ since $\cos\theta - \lambda_3 \sin\theta > 0$. This follows because:

(a) If $\pi < \theta_2 < \frac{3\pi}{2}$ then $\tan\theta_2 = \frac{\lambda_3}{\lambda_1\lambda_3 + \lambda_2\lambda_3 - \lambda_1\lambda_2} > 0$. Now, for $\theta_2 < \theta < \frac{3\pi}{2}$, $\tan\theta > \tan\theta_2$. Therefore, we have that

$$\begin{aligned} \cos\theta &> \frac{\lambda_1\lambda_3 + \lambda_2\lambda_3 - \lambda_1\lambda_2}{\lambda_3}\sin\theta \\ &> \lambda_3 \sin\theta, \end{aligned}$$

since $0 \leq \frac{\lambda_1\lambda_3 + \lambda_2\lambda_3 - \lambda_1\lambda_2}{\lambda_3} < \lambda_3$ and $\sin\theta < 0$.

(b) If $\frac{3\pi}{2} \leq \theta_2 < 2\pi$ then, for $\theta_2 < \theta < 2\pi$,

$$\cos\theta > \lambda_3 \sin\theta,$$

since $\cos\theta > 0$, $\sin\theta < 0$, and $\lambda_3 > 0$.

(c) If $2\pi \leq \theta \leq 2\pi + \theta_3$ then $\tan\theta \leq \tan\theta_3 = \frac{1}{\lambda_2}$. Therefore,

$$\begin{aligned} \cos\theta &\geq \lambda_2 \sin\theta \\ &> \lambda_3 \sin\theta, \end{aligned}$$

since $0 < \lambda_3 < \lambda_2$.

3. Suppose the global minimum is achieved when $d'(u) = 0$. We have, using the expression (6.1.14) for $f_1(u)$, that

$$\begin{aligned} d'(u) &= e^{\lambda_3 u}(\lambda_3 \cos\theta f_1^{(1)}(u) + (\cos\theta + \lambda_3 \sin\theta) f_1^{(2)}(u) + \sin\theta f_1^{(3)}) \\ &= e^{\lambda_3 u}\frac{(\cos\theta - \lambda_1 \sin\theta)\lambda_1 e^{-\lambda_1 u} - (\cos\theta - \lambda_2 \sin\theta)\lambda_2 e^{-\lambda_2 u}}{\lambda_1 - \lambda_2}, \end{aligned} \quad (7.5.12)$$

Solving $d'(u) = 0$ and letting the value of u when this occurs be $u^* = u^*(\theta)$ gives

$$e^{u^*(\theta)} = \Big(\frac{\lambda_1(\cos\theta - \lambda_1 \sin\theta)}{\lambda_2(\cos\theta - \lambda_2 \sin\theta)}\Big)^{\frac{1}{\lambda_1 - \lambda_2}}. \quad (7.5.13)$$

The function $d(u)$ attains a minimum when $u = u^*$. This is the case because

$$\begin{aligned} d''(u^*) &= \frac{e^{\lambda_3 u^*}}{\lambda_1 - \lambda_2}\Big((\cos\theta - \lambda_1 \sin\theta)\lambda_1\lambda_3 e^{-\lambda_1 u^*} - (\cos\theta - \lambda_2 \sin\theta)\lambda_2\lambda_3 e^{-\lambda_2 u^*} \\ &\qquad - (\cos\theta - \lambda_1 \sin\theta)\lambda_1^2 e^{-\lambda_1 u^*} + (\cos\theta - \lambda_2 \sin\theta)\lambda_2^2 e^{-\lambda_2 u^*}\Big) \\ &= \frac{e^{\lambda_3 u^*}}{\lambda_1 - \lambda_2}\Big(-(\cos\theta - \lambda_1 \sin\theta)\lambda_1^2 e^{-\lambda_1 u^*} + (\cos\theta - \lambda_2 \sin\theta)\lambda_2^2 e^{-\lambda_2 u^*}\Big) \\ &> \frac{e^{\lambda_3 u^*}}{\lambda_1 - \lambda_2}\Big(-\lambda_1(\cos\theta - \lambda_1 \sin\theta)\lambda_1 e^{-\lambda_1 u^*} + \lambda_1(\cos\theta - \lambda_2 \sin\theta)\lambda_2 e^{-\lambda_2 u^*}\Big) \\ &= 0. \end{aligned}$$

The second and third equalities are due to the fact that $d'(u^*) = 0$, and the inequality is because $0 < \lambda_2 < \lambda_1$, and for $\theta_3 < \theta \leq \theta_1$, $\cos\theta - \lambda_2 \sin\theta < 0$, see below.

Define, for $\theta_3 < \theta \leq \theta_1$,

$$\gamma(\theta) = \frac{\lambda_1(\cos\theta - \lambda_1 \sin\theta)}{\lambda_2(\cos\theta - \lambda_2 \sin\theta)}.$$

For $\theta_3 < \theta \leq \theta_1$, (7.5.13) is defined since $\gamma(\theta) > 0$. This follows because:

(a) If $\theta_3 < \theta \leq \frac{\pi}{2}$ then $\tan\theta > \tan\theta_3 = \frac{1}{\lambda_2}$. Therefore,

$$\begin{aligned} \cos\theta &< \lambda_2 \sin\theta \\ &< \lambda_1 \sin\theta, \end{aligned}$$

since $0 < \lambda_2 < \lambda_1$.

(b) If $\frac{\pi}{2} < \theta \leq \pi$, we have that

$$\begin{aligned} \cos\theta &< \lambda_2 \sin\theta \\ &< \lambda_1 \sin\theta, \end{aligned}$$

since $\cos\theta < 0$, $\sin\theta \geq 0$, and $0 < \lambda_2 < \lambda_1$.

(c) If $\pi < \theta \leq \theta_1$, we have that

$$\begin{aligned} \tan\theta &\leq \tan\theta_1 \\ &= \frac{1}{\lambda_1 + \lambda_2} \\ &< \frac{1}{\lambda_2} \\ &< \frac{1}{\lambda_1}. \end{aligned}$$

Thus, $\cos\theta < \lambda_2 \sin\theta < \lambda_1 \sin\theta$ since $0 < \lambda_2 < \lambda_1$, and $\sin\theta < 0$ and $\cos\theta < 0$.

Equating the expression for $d'(u^*)$ in (7.5.12) to zero we have that

$$\lambda_2 e^{-\lambda_2 u^*(\theta)}(\cos\theta - \lambda_2 \sin\theta) = \lambda_1 e^{-\lambda_1 u^*(\theta)}(\cos\theta - \lambda_1 \sin\theta).$$

Substituting this expression into (7.5.8), after some rearrangement, gives

$$\begin{aligned} d(u^*(\theta)) &= -\frac{\lambda_3(\cos\theta - \lambda_3 \sin\theta) - \lambda_1(\cos\theta - \lambda_1 \sin\theta)e^{(\lambda_3 - \lambda_1)u^*(\theta)}}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)} \\ &= -\frac{\lambda_3(\cos\theta - \lambda_3 \sin\theta) - \lambda_1(\cos\theta - \lambda_1 \sin\theta)\gamma(\theta)^{\frac{\lambda_3 - \lambda_1}{\lambda_1 - \lambda_2}}}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}, \end{aligned}$$

by (7.5.13). Thus, the points on $\partial\Omega_3$ are given by

$$\begin{aligned} x_1(\theta) &= -\frac{\lambda_3}{d^*}\cos\theta + \lambda_3(\lambda_1 + \lambda_2) \\ &= \frac{\lambda_3(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)\cos\theta}{\lambda_3(\cos\theta - \lambda_3 \sin\theta) - \lambda_1(\cos\theta - \lambda_1 \sin\theta)\gamma(\theta)^{\frac{\lambda_3 - \lambda_1}{\lambda_1 - \lambda_2}}} + \lambda_3(\lambda_1 + \lambda_2), \end{aligned}$$

and

$$\begin{aligned} x_2(\theta) &= -\frac{\lambda_3}{d^*}\sin\theta + \lambda_3 \\ &= \frac{\lambda_3(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)\sin\theta}{\lambda_3(\cos\theta - \lambda_3 \sin\theta) - \lambda_1(\cos\theta - \lambda_1 \sin\theta)\gamma(\theta)^{\frac{\lambda_3 - \lambda_1}{\lambda_1 - \lambda_2}}} + \lambda_3. \end{aligned}$$

This occurs when the ray emanating from P in the direction defined by θ intersects the line $g(x_1, x_2, u) = 0$ where $u > 0$. Thus, $\theta_3 < \theta \leq \theta_1$. ■

Similar parameterizations for $\partial\Omega_3$ can be can be found when the zeros of $b(\lambda)$ are real but not distinct which we state without proof.

Theorem 7.3 *Suppose that the vectors* $\boldsymbol{a} = (a_1, a_2, a_3)$ *and* $\boldsymbol{b} = (b_1, b_2, b_3)$ *are such that*

1 $0 < \frac{a_1}{b_1} \leq 1$, *and*

2 the zeros of $b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1$, $-\lambda_1, -\lambda_2$, *and* $-\lambda_3$, *are real and such that*

(a) $\lambda_1 = \lambda_2 > \lambda_3 > 0$.

If θ_1, θ_2, *and* θ_3 *are such that*

1 $\tan\theta_1 = \frac{1}{2\lambda_2}$ *and* $\pi < \theta_1 < \frac{3\pi}{2}$,
2 $\tan\theta_2 = \frac{\lambda_3}{2\lambda_2\lambda_3 - \lambda_2^2}$ *and* $\pi < \theta_2 < 2\pi$, *and*
3 $\tan\theta_3 = \frac{1}{\lambda_2}$ *and* $0 < \theta_3 < \frac{\pi}{2}$,

then the parametric representation for $\partial\Omega_3$ *is:*

1. For $\theta_1 < \theta \leq \theta_2$,

$$\begin{aligned} x_1(\theta) &= -\lambda_3 \cot\theta + 2\lambda_2\lambda_3 \\ x_2(\theta) &= 0 \end{aligned}$$

2. For $\theta_2 < \theta \leq \theta_3 + 2\pi$,

$$\begin{aligned} x_1(\theta) &= \frac{(\lambda_2^2 + \lambda_3^2)\cos\theta - 2\lambda_2\lambda_3^2 \sin\theta}{\cos\theta - \lambda_3 \sin\theta} \\ x_2(\theta) &= \frac{\lambda_3\cos\theta + (\lambda_2^2 - 2\lambda_2\lambda_3)\sin\theta}{\cos\theta - \lambda_3\sin\theta}. \end{aligned}$$

3. For $\theta_3 < \theta \leq \theta_1$,

$$\begin{aligned} x_1(\theta) &= \frac{\lambda_3(\lambda_2 - \lambda_3)^2\cos\theta}{\lambda_3(\cos\theta - \lambda_3\sin\theta) - \lambda_2(\cos\theta - \lambda_2\sin\theta)e^{(\lambda_3 - \lambda_2)\eta(\theta)}} + 2\lambda_2\lambda_3 \\ x_2(\theta) &= \frac{\lambda_3(\lambda_2 - \lambda_3)^2\sin\theta}{\lambda_3(\cos\theta - \lambda_3\sin\theta) - \lambda_2(\cos\theta - \lambda_2\sin\theta)e^{(\lambda_3 - \lambda_2)\eta(\theta)}} + \lambda_3, \end{aligned}$$

where

$$\eta(\theta) \;=\; \frac{\cos\theta - 2\lambda_2 \sin\theta}{\lambda_2(\cos\theta - \lambda_2 \sin\theta)}.$$

(b) $\lambda_1 > \lambda_2 = \lambda_3 > 0$.

If θ_1, θ_2, *and* θ_3 *are such that*

1 $\tan\theta_1 = \dfrac{1}{\lambda_1 + \lambda_3}$ *and* $\pi < \theta_1 < \dfrac{3\pi}{2}$,
2 $\tan\theta_2 = \dfrac{1}{\lambda_3}$ *and* $\pi < \theta_2 < \dfrac{3\pi}{2}$, *and*
3 $\tan\theta_3 = \dfrac{1}{\lambda_3}$ *and* $0 < \theta_3 < \dfrac{\pi}{2}$,

then the parametric representation for $\partial\Omega_3$ *is:*

1. *For* $\theta_1 < \theta \le \theta_2$,

$$x_1(\theta) \;=\; -\lambda_3 \cot\theta + \lambda_3(\lambda_1 + \lambda_3)$$

$$x_2(\theta) \;=\; 0$$

2. *For* $\theta_2 < \theta \le \theta_3 + 2\pi$,

$$x_1(\theta) \;=\; \frac{\lambda_1\lambda_3(\theta + \theta_3 - 2\theta_2 + 2\pi)}{\theta_3 - \theta_2 + 2\pi} \tag{7.5.14}$$

$$x_2(\theta) \;=\; \frac{\lambda_1(\theta - \theta_2)}{\theta_3 - \theta_2 + 2\pi}. \tag{7.5.15}$$

3. *For* $\theta_3 < \theta \le \theta_1$,

$$x_1(\theta) \;=\; \frac{\lambda_3(\lambda_1 - \lambda_3)^2 \cos\theta}{(\lambda_3 - \lambda_1)(\cos\theta - 2\lambda_3 \sin\theta) + \lambda_3(\cos\theta - \lambda_3 \sin\theta)\log\zeta(\theta)} + \lambda_3(\lambda_1 + \lambda_3)$$

$$x_2(\theta) \;=\; \frac{\lambda_3(\lambda_1 - \lambda_3)^2 \sin\theta}{(\lambda_3 - \lambda_1)(\cos\theta - 2\lambda_3 \sin\theta) + \lambda_3(\cos\theta - \lambda_3 \sin\theta)\log\zeta(\theta)} + \lambda_3,$$

where

$$\zeta(\theta) \;=\; \frac{\lambda_1(\cos\theta - \lambda_1 \sin\theta)}{\lambda_3(\cos\theta - \lambda_3 \sin\theta)}.$$

(c) $\lambda_1 = \lambda_2 = \lambda_3 > 0$.

If θ_1, θ_2, *and* θ_3 *are such that*

1 $\tan\theta_1 = \dfrac{1}{2\lambda_3}$ *and* $\pi < \theta_1 < \dfrac{3\pi}{2}$,
2 $\tan\theta_2 = \dfrac{1}{\lambda_3}$ *and* $\pi < \theta_2 < \dfrac{3\pi}{2}$, *and*
3 $\tan\theta_3 = \dfrac{1}{\lambda_3}$ *and* $0 < \theta_3 < \dfrac{\pi}{2}$,

then the parametric representation for $\partial\Omega_3$ *is:*

1. *For* $\theta_1 < \theta \leq \theta_2$,

$$x_1(\theta) = -\lambda_3 \cot\theta + 2\lambda_3^2$$

$$x_2(\theta) = 0$$

2. *For* $\theta_2 < \theta \leq \theta_3 + 2\pi$,

$$x_1(\theta) = \frac{\lambda_3^2(\theta + \theta_3 - 2\theta_2 + 2\pi)}{\theta_3 - \theta_2 + 2\pi} \tag{7.5.16}$$

$$x_2(\theta) = \frac{\lambda_3(\theta - \theta_2)}{\theta_3 - \theta_2 + 2\pi}. \tag{7.5.17}$$

3. *For* $\theta_3 < \theta \leq \theta_1$,

$$x_1(\theta) = \frac{2\lambda_3^3 \sin\theta(\lambda_3 \sin\theta - \cos\theta)}{(\cos\theta - \lambda_3 \sin\theta)^2 + \lambda_3^2 \sin^2\theta}$$

$$x_2(\theta) = \frac{\lambda_3(\cos\theta - 2\lambda_3 \sin\theta)^2}{(\cos\theta - \lambda_3 \sin\theta)^2 + \lambda_3^2 \sin^2\theta}.$$

The parametric equations (7.5.14) and (7.5.15), and (7.5.16) and (7.5.17) have a different form to (7.5.3) and (7.5.4) because the point corresponding to the mixture of the point mass at zero and the exponential distribution (7.3.2), that is, P, lies on the line segment $\overline{RS}$, see Figure 7.5.1 (in fact P and R coincide when $\lambda_1 = \lambda_2 = \lambda_3$). When $\theta_2 < \theta \leq \theta_3 + 2\pi$, $r(u)$ attains a minimum of $r^* = 0$ (d^* is infinite) when $u \to \infty$, and the parameterization given by

$$x_1(\theta) = r^* \cos\theta + \lambda_3(\lambda_1 + \lambda_3) = \lambda_3(\lambda_1 + \lambda_3)$$

$$x_2(\theta) = r^* \sin\theta + \lambda_3 = \lambda_3$$

is redundant. Consequently, $\overline{RS}$ needs to be expressed in a different way, as we have done.

We now give an alternative parameterization of $\partial\Omega_3$ when two of the zeros of $b(\lambda)$ are a complex conjugate pair.

Theorem 7.4 *Suppose that the vectors* $\boldsymbol{a} = (a_1, a_2, a_3)$ *and* $\boldsymbol{b} = (b_1, b_2, b_3)$ *are such that*

1. $0 < \frac{a_1}{b_1} \leq 1$, *and*
2. *the zeros of* $b(\lambda) = \lambda^3 + b_3\lambda^2 + b_2\lambda + b_1$, $-\lambda_1$, $-\lambda_2$, *and* $-\lambda_3$, *are such that* $\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$, *and* $\lambda_3 = \mu$ *with* $0 < \mu \leq \alpha$ *and* $\beta > 0$.

Let $f_1(u)$ *be the solution to the differential equation (5.6.15) with initial conditions (5.6.16). If* θ_1 *and* θ_2 *are such that*

1 $\tan\theta_1 = \frac{1}{2\alpha}$ *and* $\pi < \theta_1 < \frac{3\pi}{2}$, *and*

2 $\tan\theta_2 = \frac{\mu}{2\mu\alpha - x_1(u^*)}$ *where* u^* *is the minimal positive solution to*

$$f_1(u)f_1^{(2)}(u) - (f_1^{(1)}(u))^2 \;=\; 0, \tag{7.5.18}$$

and

$$x_1(u^*) \;=\; -b_1\frac{f_1(u^*)}{f_1^{(1)}(u^*)} \;=\; -b_1\frac{f_1^{(1)}(u^*)}{f_1^{(2)}(u^*)}, \tag{7.5.19}$$

and $\pi < \theta_2 < 2\pi$,

then the parametric representation for $\partial\Omega_3$ *is:*

1. *For* $\theta_1 \leq \theta < \theta_2$,

$$x_1(\theta) \;= -\mu\cot\theta + 2\mu\alpha \tag{7.5.20}$$

$$x_2(\theta) \;= 0, \tag{7.5.21}$$

2. *For* $\theta_2 \leq \theta < 2\pi + \theta_1$,

$$x_1(\theta) \;=\; \frac{\mu((\mu-\alpha)^2+\beta^2)\cos\theta}{\mu(\cos\theta - \mu\sin\theta) - w(\theta)^z\overline{w(\theta)}^{\bar{z}}} + 2\mu\alpha \tag{7.5.22}$$

$$x_2(\theta) \;=\; \frac{\mu((\mu-\alpha)^2+\beta^2)\sin\theta}{\mu(\cos\theta - \mu\sin\theta) - w(\theta)^z\overline{w(\theta)}^{\bar{z}}} + \mu, \tag{7.5.23}$$

where

$$w(\theta) \;=\; \alpha\cos\theta - (\alpha^2-\beta^2)\sin\theta + i(\beta\cos\theta - 2\alpha\beta\sin\theta), \tag{7.5.24}$$

and

$$z \;=\; \frac{1}{2} + i\frac{\alpha-\mu}{2\beta}. \tag{7.5.25}$$

The expressions $\overline{w(\theta)}$ *and* $\overline{z}$ *denote the complex conjugates of* $w(\theta)$ *and* z*, respectively.*

Proof.

Refer to Figure 7.5.2. Recall from Theorem 6.5 that $\partial\Omega_3$ consists of

1 the line segment between $O(0,0)$ and $R(x_1(u^*),0)$ where u^* is the minimal positive solution to (7.5.18), and $x_1(u^*)$ is given by (7.5.19), and

2 the curve Γ_3 which has as its endpoints O and R.

Recall also that the mixture of the point mass at zero and the exponential distribution (7.3.2) is represented by the point $P(2\mu\alpha,\mu)$. Let θ be the angle between the ray emanating from P parallel to the positive x_1-axis and any ray emanating from P, measured in an anticlockwise direction. Let θ_1 and θ_2 be such angles that are defined by the rays $\overrightarrow{PO}$ and $\overrightarrow{PR}$, respectively. We have that $\tan\theta_1 = \frac{1}{2\alpha}$, and since the coordinates of P are both positive then $\pi < \theta_1 < \frac{3\pi}{2}$. Also, $\tan\theta_2 = \frac{\mu}{2\alpha - x_1(u^*)}$, and since the x_2 coordinate of P is positive then $\pi < \theta_2 < 2\pi$.

The parametric equations for (7.5.20) and (7.5.21) are found in precisely the same way as (7.5.1) and (7.5.2), respectively, in the proof of Theorem 7.2 since λ_1, λ_2, and λ_3 are distinct. Likewise, the parametric equations for (7.5.22) and (7.5.23) are found in the same way as (7.5.5) and (7.5.6), respectively. Defining $\gamma(\theta)$ by (7.5.7) we have that

$$\begin{aligned}
\lambda_1(\cos\theta - \lambda_1\sin\theta)\gamma(\theta)^{\frac{\lambda_3-\lambda_1}{\lambda_1-\lambda_2}} &= \lambda_1(\cos\theta-\lambda_1\sin\theta)\Big(\frac{\lambda_1(\cos\theta-\lambda_1\sin\theta)}{\lambda_2(\cos\theta-\lambda_2\sin\theta)}\Big)^{\frac{\lambda_3-\lambda_1}{\lambda_1-\lambda_2}} \\
&= (\lambda_1(\cos\theta-\lambda_1\sin\theta))^{\frac{\lambda_3-\lambda_2}{\lambda_1-\lambda_2}}(\lambda_2(\cos\theta-\lambda_2\sin\theta))^{\frac{\lambda_1-\lambda_3}{\lambda_1-\lambda_2}} \\
&= w(\theta)^z\overline{w(\theta)}^{\overline{z}},
\end{aligned} \tag{7.5.26}$$

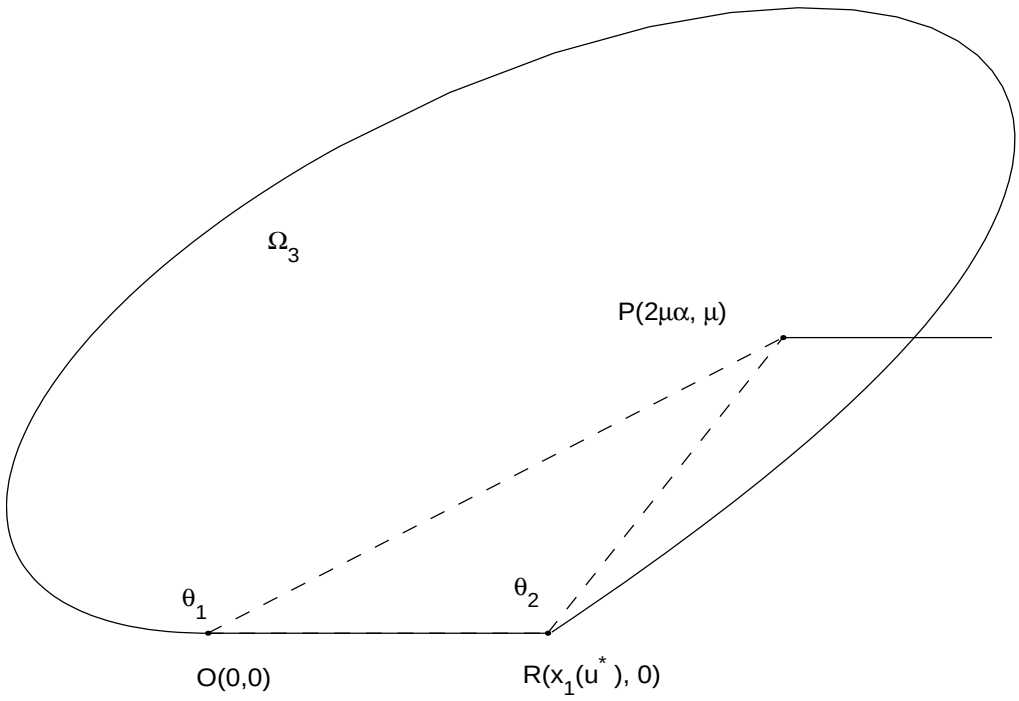

Figure 7.5.2: Diagram of Ω_3 showing the points O, P, and R

where $w(\theta)$ and z are defined by (7.5.24) and (7.5.25), respectively. Substituting the expression (7.5.26) into (7.5.5) and (7.5.6) gives (7.5.22) and (7.5.23), respectively.

■

7.6 The Boundedness of Ω_p

In this section we explore the following conjecture:

Conjecture 7.5 *Ω_p is bounded.*

In doing so we rely on the work of Kreĭn and Nudel′man [80, Chapter 1].

Recall from Section 7.3 that if Ω_p is unbounded then there exists a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{p-1})$ such that, for $u \geq 0$,

$$\begin{aligned} r(u) &= \frac{-\lambda_p e^{-\lambda_p u}}{\cos\theta_1 f_1^{(1)}(u) + \cos\theta_2 f_1^{(2)}(u) + \ldots + \cos\theta_{p-1} f_1^{(p-1)}(u)} \\ &\leq 0. \end{aligned} \tag{7.6.1}$$

Geometrically, the ray emanating from the point P whose direction is determined by $\boldsymbol{\theta}$ never intersects, for any $u \geq 0$, a hyperplane of the form

$$g(\boldsymbol{x}, u) = b_1 f_1(u) + x_1 f_1^{(1)}(u) + x_2 f_1^{(2)}(u) + \ldots + x_{p-1} f_1^{(p-1)}(u) = 0.$$

Statement (7.6.1) is equivalent to

$$\cos\theta_1 f_1^{(1)}(u) + \cos\theta_2 f_1^{(2)}(u) + \ldots + \cos\theta_{p-1} f_1^{(p-1)}(u) \geq 0.$$

Thus, if no $\boldsymbol{c} = (c_1, c_2, \ldots, c_{p-1}) \in \mathbb{R}^p$ exists such that, for all $u \geq 0$,

$$c_1 f_1^{(1)}(u) + c_2 f_1^{(2)}(u) + \ldots + c_{p-1} f_1^{(p-1)}(u) \geq 0, \tag{7.6.2}$$

then Ω_p is bounded.

Consider the vector-valued function, defined for $u \geq 0$,

$$\boldsymbol{z}(u) = \left(\begin{array}{cccc} f^{(1)}(u) & f^{(2)}(u) & \ldots & f^{(p-1)}(u)) \end{array} \right),$$

and the curve $Z = \{\boldsymbol{z}(u) \mid u \geq 0\}$. When $u = 0$, $\boldsymbol{z}(0) = (0, 0, \ldots, 1)$ by (5.6.16). Also, as $u \to \infty$, $\boldsymbol{z}(u)$ approaches the origin since the eigenvalues of $\boldsymbol{T}$ in the expression

$$f_1(u) = \boldsymbol{e}_1'(u) \exp(\boldsymbol{T}u)\boldsymbol{e}_p \tag{7.6.3}$$

all have negative real part. Let $C(Z)$ be the *convex hull* of Z, that is, the intersection of all convex sets containing Z. Since $\boldsymbol{z}(u)$ is a continuous function, if the origin is contained in the *interior* of $C(Z)$ then *any* hyperplane through the origin

$$c_1 x_1 + c_2 x_2 + \ldots + c_{p-1} x_{p-1} = 0$$

will cut the curve. If this is the case, then for any $\boldsymbol{c} = (c_1, c_2, \ldots, c_{p-1}) \in \mathbb{R}^p$, there will exist values of $u \geq 0$ such that

$$c_1 f_1^{(1)}(u) + c_2 f_1^{(2)}(u) + \ldots + c_{p-1} f_1^{(p-1)}(u) \ < \ 0$$

and (7.6.2) will not hold. Figure 7.6.1 shows Z and $C(Z)$ when the eigenvalues of $\boldsymbol{T}$ are $\lambda_1 = -3$, $\lambda_2 = -2$, and $\lambda_3 = -1$.

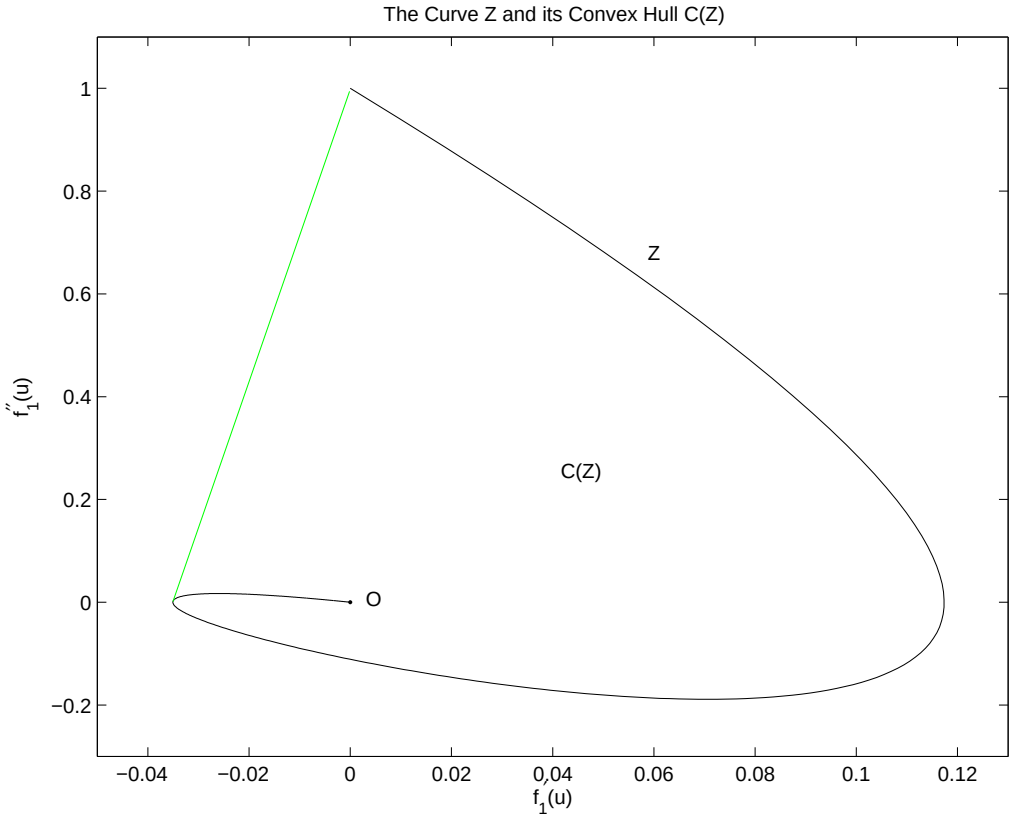

Figure 7.6.1: Diagram of the curve Z and its convex hull $C(Z)$

We now prove the following.

Theorem 7.6 Ω_3 *is bounded.*

Proof.

We need to show that a small open neighbourhood containing the origin lies entirely within $C(Z)$. Let $f_1(u)$ be given by (7.6.3) where $\boldsymbol{T}$ is a companion matrix

(5.4.3) that has an eigenvalue of maximal real part that is real and negative. Since $f_1(0) = 0$, $f_1^{(1)}(0) = 0$, and $f_1^{(2)}(0) = 1$, the function $f_1(u)$ has a local minimum of zero when $u = 0$. Also, since $\lim_{u\to\infty} f_1(u) = 0$, by the *Mean Value Theorem* there exists a $u_1 \in (0, \infty)$ such that $f_1^{(1)}(u_1) = 0$ with $f_1^{(2)}(u_1) < 0$ (a local maximum at $u = u_1$). Also by the *Mean Value Theorem*, and the fact that $f_1^{(1)}(u)$ changes from positive to negative at $u = u_1$, there exists $u_2 \in (0, u_1)$ such that $f_1^{(2)}(u_2) = 0$ with $f_1^{(1)}(u_2) > 0$, and $u_3 \in (u_1, \infty)$ such that $f_1^{(2)}(u_2) = 0$ with $f_1^{(1)}(u_3) < 0$. We have that the four points

$$(0, 1), \quad (0, f_1^{(2)}(u_1)), \quad (f_1^{(1)}(u_2), 0), \quad (f_1^{(1)}(u_3), 0) \tag{7.6.4}$$

lie on Z. If we choose $\epsilon > 0$ to be such that

$$\epsilon \ < \ \min(1, \ |f_1^{(2)}(u_1)|, \ f_1^{(1)}(u_2), \ |f_1^{(1)}(u_3)|),$$

then the open simplex defined by the four points (which are contained in $C(Z)$ since the set is convex)

$$(0, \epsilon), \quad (0, -\epsilon), \quad (\epsilon, 0), \quad (-\epsilon, 0)$$

contains the origin. Therefore, Ω_3 is bounded. ■

Proving the general case appears to be more difficult. The process of finding points on the curve Z equivalent to (7.6.4) that lead to a simplex containing the origin would become more complicated as the order of $\boldsymbol{T}$ increases. An inductive proof could be a way of showing Conjecture 7.5 as we have proven the case when $p = 3$. Further investigation is required.

7.7 Concluding Remarks

In this chapter we have reduced the problem of determining whether or not the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution to one of finding the global minimum of a continuous single-variable function $d(u)$ over the nonnegative real numbers. While, as mentioned in the last paragraph of Section 7.3, this is not necessarily a

straightforward exercise, if the size of $\boldsymbol{a}$ and $\boldsymbol{b}$ is relatively low, the optimization can be performed accurately within a reasonable time. A problem would arise, however, even for the low order case, if many applications of the algorithm are required because of the time such a task would take. This problem would arise when fitting *ME* distributions to data because the algorithm would need to be used each time the parameter values are updated. In the next chapter we develop an alternative algorithm to identify *ME* distributions and use the approach to fit *ME* distributions to data in Chapter 9.

Chapter 8

An Alternative Algorithm for Identifying Matrix-exponential Distributions

8.1 Introduction

In this chapter we make use of the fact that for a suitable vector $\boldsymbol{b}$ the vectors $\boldsymbol{a}$ that are such that $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution lie in a region that is closed and (we assume) bounded. We can then determine whether or not the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution by minimizing a distance function over Ω_p.

In Section 8.2 we set up the *ME* identification problem as a *convex semi-infinite programming* problem. More specifically, we seek to minimize a convex function over a set that is the intersection of an uncountable number of linear constraints. Section 8.3 contains a brief introduction to semi-infinite programming and a short literature review. The particular algorithm is given in Section 8.4 and we give some examples in Section 8.5.

8.2 The Matrix-exponential Identification Problem

Throughout the next two chapters we will assume that Ω_p is bounded. Ω_p is closed because it is the intersection of closed halfspaces.

Suppose the two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ satisfy conditions 2 and 3 of Theorem 5.6. Then $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution if and only if

$$\boldsymbol{c} \;=\; \frac{b_1}{a_1}\begin{pmatrix} a_2 & a_3 & \dots & a_p \end{pmatrix} \tag{8.2.1}$$

is contained in Ω_p.

Consider the function, defined for $\boldsymbol{x} = (x_1, x_2, \dots, x_{p-1}) \in \Omega_p$,

$$\begin{aligned} h(\boldsymbol{x}) \;&=\; \|\boldsymbol{x} - \boldsymbol{c}\|_2 \\ &=\; \Big(\sum_{i=1}^{p-1}(x_i - c_i)^2\Big)^{\frac{1}{2}}. \end{aligned}$$

If $\boldsymbol{c} \in \Omega_p$ then $h(\boldsymbol{x})$ attains a minimum over Ω_p when $\boldsymbol{x} = \boldsymbol{c}$. The minimum value of $h(\boldsymbol{x})$ is then $h(\boldsymbol{c}) = 0$. If $\boldsymbol{c} \notin \Omega_p$ then the minimum value of $h(\boldsymbol{x})$ occurs when $\boldsymbol{x} \neq \boldsymbol{c}$ and is greater than zero. If a suitable method of minimizing $h(\boldsymbol{x})$ over Ω_p can be developed we will be able to determine whether $\boldsymbol{c} \in \Omega_p$. We begin by investigating the function $h(\boldsymbol{x})$.

Lemma 8.1 *For $\boldsymbol{x} \in \Omega_p$ the function $h(\boldsymbol{x})$ is convex.*

Proof.

We need to show, for $\boldsymbol{x}, \boldsymbol{y} \in \Omega_p$, and $0 \leq \rho \leq 1$, that

$$h(\rho\boldsymbol{x} + (1-\rho)\boldsymbol{y}) \;\leq\; \rho h(\boldsymbol{x}) + (1-\rho)h(\boldsymbol{y}).$$

Now,

$$\begin{aligned} h(\rho\boldsymbol{x}+(1-\rho)\boldsymbol{y}) &= \|\rho\boldsymbol{x}+(1-\rho)\boldsymbol{y}-\boldsymbol{c}\|_2 \\ &= \|\rho(\boldsymbol{x}-\boldsymbol{c})+(1-\rho)(\boldsymbol{y}-\boldsymbol{c})\|_2 \\ &\leq \|\rho(\boldsymbol{x}-\boldsymbol{c})\|_2+\|(1-\rho)(\boldsymbol{y}-\boldsymbol{c})\|_2 \\ &= \rho\|(\boldsymbol{x}-\boldsymbol{c})\|_2+(1-\rho)\|(\boldsymbol{y}-\boldsymbol{c})\|_2 \\ &= \rho h(\boldsymbol{x})+(1-\rho)h(\boldsymbol{y}). \end{aligned}$$

Since $\|\cdot\|_2$ is a norm (see Apostol [6, page 48]) the inequality follows from the triangle inequality, and the second-to-last equality follows from the identity $\|\gamma(\cdot)\|_2 = |\gamma|\|\cdot\|_2$, where $\gamma \in \mathbb{R}$, and the fact that ρ and $1-\rho$ are both nonnegative. ■

Note that, $h(\boldsymbol{x})$ is not *strictly* convex because if $\boldsymbol{x}=\boldsymbol{c}$, $\boldsymbol{y}=\boldsymbol{c}$, or $\boldsymbol{x}-\boldsymbol{c}=k(\boldsymbol{y}-\boldsymbol{c})$ for some $k>0$, then, for $0<\rho<1$, $h(\rho\boldsymbol{x}+(1-\rho)\boldsymbol{y})=\rho h(\boldsymbol{x})+(1-\rho)h(\boldsymbol{y})$.

Lemma 8.2 *Under the assumption that Ω_p is bounded there exists a unique $\boldsymbol{x}^* \in \Omega_p$ that minimizes $h(\boldsymbol{x})$.*

Proof.

Since $h(\boldsymbol{x})$ is continuous and is being minimized over a closed and bounded set, $h(\boldsymbol{x})$ attains a global minimum on Ω_p, when, say, $\boldsymbol{x}=\boldsymbol{x}^*$. Let the global minimum attained be $h^*=h(\boldsymbol{x}^*)$. If $\boldsymbol{c}\in\Omega_p$ then $h(\boldsymbol{x})$ attains a global minimum when $\boldsymbol{x}^*=\boldsymbol{c}$. In this case $h^*=h(\boldsymbol{c})=0$ and $\boldsymbol{x}^*$ is unique since $\|\boldsymbol{x}^*-\boldsymbol{c}\|_2=0$ if and only if $\boldsymbol{x}^*=\boldsymbol{c}$.

Suppose that $\boldsymbol{c}\notin\Omega_p$, then $\boldsymbol{x}\in\partial\Omega_p$. This follows because if $\boldsymbol{x}$ is in the interior of Ω_p then there exists a ρ, with $0<\rho<1$, such that $(1-\rho)\boldsymbol{c}+\rho\boldsymbol{x}^*$ lies on $\partial\Omega_p$.

We now have

$$\begin{aligned} h((1-\rho)\boldsymbol{c}+\rho\boldsymbol{x}^*) &= \|(1-\rho)\boldsymbol{c}+\rho\boldsymbol{x}^*-\boldsymbol{c}\|_2 \\ &= \|\rho(\boldsymbol{x}^*-\boldsymbol{c})\|_2 \\ &= \rho\|\boldsymbol{x}^*-\boldsymbol{c}\|_2 \\ &< \|\boldsymbol{x}^*-\boldsymbol{c}\|_2 \\ &= h(\boldsymbol{x}^*), \end{aligned}$$

which contradicts the fact that the global minimum of $h(\boldsymbol{x})$ is attained when $\boldsymbol{x} = \boldsymbol{x}^* \in \Omega_p \backslash \partial\Omega_p$.

Now, suppose $h(\boldsymbol{x})$ also attains a global minimum over Ω_p when $\boldsymbol{x} = \boldsymbol{x}^{**}$ with $\boldsymbol{x}^* \neq \boldsymbol{x}^{**}$. For $0 < \rho < 1$ we have that $\rho\boldsymbol{x}^* + (1-\rho)\boldsymbol{x}^{**} \in \Omega_p$ since the set is convex. We have,

$$\begin{aligned} h(\boldsymbol{x}^*) &= \|\boldsymbol{x}^*-\boldsymbol{c}\|_2 \\ &\leq \|\rho\boldsymbol{x}^*+(1-\rho)\boldsymbol{x}^{**}-\boldsymbol{c}\|_2 \\ &= \|\rho(\boldsymbol{x}^*-\boldsymbol{c})+(1-\rho)(\boldsymbol{x}^{**}-\boldsymbol{c})\|_2 \\ &\leq \rho\|\boldsymbol{x}^*-\boldsymbol{c}\|_2+(1-\rho)\|\boldsymbol{x}^{**}-\boldsymbol{c}\|_2. \end{aligned} \tag{8.2.2}$$

The first inequality follows because $h(\boldsymbol{x})$ attains a global minimum when $\boldsymbol{x} = \boldsymbol{x}^*$. The second inequality follows from the triangle inequality. A simple argument shows that the second inequality is strict if and only if none of the following conditions hold:

1. $\boldsymbol{x}^* = \boldsymbol{c}$.
2. $\boldsymbol{x}^{**} = \boldsymbol{c}$.
3. $\boldsymbol{x}^* - \boldsymbol{c} = k(\boldsymbol{x}^{**} - \boldsymbol{c})$ for some $k > 0$.

Conditions 1 and 2 do not hold because $\boldsymbol{c} \notin \Omega_p$. If Condition 3 holds then $\boldsymbol{x}^{**} = \frac{1}{k}\boldsymbol{x}^* + (1 - \frac{1}{k})\boldsymbol{c}$. That is, $\boldsymbol{x}^{**}$ lies on the line through $\boldsymbol{x}^*$ and $\boldsymbol{c}$. This implies that one and only one of the following four situations can occur. The vector $\boldsymbol{x}^{**}$ is

1 exterior to Ω_p,

2 interior to Ω_p,

3 equal to $\boldsymbol{x}^*$, or

4 on $\partial\Omega_p$ but not equal to $\boldsymbol{x}^*$, that is, on $\partial\Omega_p$ but on the side opposite to $\boldsymbol{x}^*$.

It is immediately clear that Statements 1–3 lead to contradictions. If Statement 4 holds then $\boldsymbol{x}^* - \boldsymbol{c} = k(\boldsymbol{x}^{**} - \boldsymbol{c})$ where $0 < k < 1$. This leads to

$$\begin{aligned} \| \boldsymbol{x}^* - \boldsymbol{c} \| &= \| k(\boldsymbol{x}^{**} - \boldsymbol{c}) \| \\ &= k \| \boldsymbol{x}^{**} - \boldsymbol{c} \| \\ &< \| \boldsymbol{x}^{**} - \boldsymbol{c} \|, \end{aligned}$$

that is, $h(\boldsymbol{x}^*) < h(\boldsymbol{x}^{**})$, which contradicts the assumption that the global minimum is also obtained when $\boldsymbol{x} = \boldsymbol{x}^{**}$. Rearranging the inequality (8.2.2), that is,

$$h(\boldsymbol{x}^*) < \rho\|\boldsymbol{x}^* - \boldsymbol{c}\|_2 + (1 - \rho)\|\boldsymbol{x}^{**} - \boldsymbol{c}\|_2$$

gives $h(\boldsymbol{x}^*) < h(\boldsymbol{x}^{**})$. Thus, $\boldsymbol{x}^*$ is unique. ■

Now, given the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ we have the following algorithm to determine whether or not they correspond to a *ME* distribution.

1. If $0 < \frac{a_1}{b_1} \leq 1$ then goto 2, else goto 7.

2. Calculate the zeros of the polynomial $b(\lambda) = \lambda^p + b_p\lambda^{p-1} + b_{p-1}\lambda^{p-2} + \ldots + b_1$.

3. If there exists a zero of maximal real part that is both real and negative then goto 4, else goto 7.

4. Calculate $\boldsymbol{c} = \frac{b_1}{a_1}(a_2, a_3, \ldots, a_p)$.

5. Find h^*, the global minimum of $h(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{c}\|_2$ over Ω_p.

6. If $h^* = 0$ then goto 8, else goto 7.

7. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ do not correspond to a *ME* distribution.

8. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution.

Step 5 requires a convex semi-infinite programming problem to be solved. While this is not a trivial exercise there is a substantial literature on the subject. In the next section we give a brief discussion on the topic.

8.3 Semi-infinite Programming

In this section we present a brief introduction to *semi-infinite programming* (*SIP*) and a short literature review. This discussion draws mainly on the two references Hettich and Kortanek [70], and the more recent Reemtsen and Görner [116]. Both of these papers contain comprehensive bibliographies. We also refer the reader to the two monographs Reemtsen and Rückmann [115], and Goberna and López [61], and the references therein for a thorough treatment of the subject. In addition, the text by Polak [113] devotes a whole chapter to *SIP*.

Suppose that $\boldsymbol{x} \in \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ has infinite cardinality. The *primal SIP* problem can be stated as follows:

$$P[Y]: \quad \begin{array}{l} \text{minimize } f(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in X_P(Y), \\ \text{where } X_P(Y) \;=\; \{\boldsymbol{x} \in \mathbb{R}^n \mid g(\boldsymbol{x}, \boldsymbol{y}) \geq 0, \boldsymbol{y} \in Y\}. \end{array} \tag{8.3.1}$$

We say that $\boldsymbol{x}^*$ is a *solution* of $P[Y]$ if $f(\boldsymbol{x})$ attains either a local or global minimum at $\boldsymbol{x}^*$, given the constraints. The term "semi-infinite" refers to the fact that the objective function f depends on a finite number of variables and that it is being minimized subject to an infinite number of constraints. This is in contrast to

infinite programming where the objective function depends on an infinite number of variables. If f and $g(\cdot, \boldsymbol{y})$ are linear (convex) functions we refer to a *linear* (*convex*) *SIP* problem, otherwise we refer to a *nonlinear SIP* problem.

The theory of *SIP* was first developed by Charnes, Cooper, and Kortanek [35] in 1962 when they connected linear *SIP* problems with their dual problems. Refer to Goberna and López [62] and [63] for the formulation of the dual of a linear *SIP* problem. Kortanek [79] gave a brief history of *SIP* from 1962 to 1972. Since 1962 *SIP* has been applied to a wide variety of problems particularly those where systems are described by functions of several variables which are subject to constraints that depend continuously on time, space, or some other parameter. Areas of application have included reliability theory, control theory, digital and wavelet filter design, actuarial risk theory, air pollution control, approximation theory, game theory, and robotics. Hettich and Kortanek [70, Section 2] described a number of examples where *SIP* has been used in modelling. They also listed in their bibliography many references on the applications of *SIP*. Reemtsen and Rückmann [115, Part III], and Goberna and López [61, Part IV] have included contributions from many authors on some specific applications of *SIP*.

Given a *SIP* problem of the form (8.3.1) there are three main classes of methods used to solve them. They are

1 discretization,

2 local reduction, and

3 exchange methods.

The idea behind the method of discretization is to minimize the objective function in (8.3.1) subject to a *finite* subset of the constraints. That is, if $Y' \subset Y$ where Y' is finite, a solution of $P[Y']$ is sought. Once a solution is found, if necessary, a more accurate solution can be obtained by either refining Y' or using another method.

More precisely, suppose, for some integer $k \geq 1$, that a sequence of finite sets, $Y_0 \subset Y_1 \subset \ldots \subset Y_k \subset Y$, is used to obtain a sequence of successive approximate solutions, $\boldsymbol{x}_0^*, \boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_k^*$, to $P[Y]$. In order to do this one approach would be to use the solution of $P[Y_i]$ as the starting point for $P[Y_{i+1}]$. A problem would arise, however, if $\boldsymbol{x}_i^*$ is not feasible for $P[Y_{i+1}]$. This situation is likely to occur, especially if $\boldsymbol{x}_i^*$ lies on the boundary of $X_P(Y_i)$, because $X_P(Y_{i+1}) \subset X_P(Y_i)$. Another phase in the overall algorithm would be required to ensure that the starting point for each problem is feasible. Consequently, such a procedure to find a solution to $P[Y]$ would be computationally costly. Also, discretization methods usually produce an *outer approximation* to the solution of $P[Y]$. That is, the solution $\boldsymbol{x}_k^*$ of $P[Y_k]$ is not feasible for $P[Y]$. While the accuracy may be sufficient for most problems, if a feasible solution is required another method must be used. Hettich and Kortanek [70, Sub-sections 4.4 and 7.2], and Reemtsen and Görner [116, Sub-section 2.5] discussed conditions under which discretization algorithms converge, that is, when the successive approximate solutions converge to a solution of $P[Y]$.

The method of local reduction, like that of discretization, seeks to convert the *SIP* problem (8.3.1) into an optimization problem involving a finite number of constraints. Suppose that $Z \subset \mathbb{R}^m$ is finite. Any point $\boldsymbol{x} \in \mathbb{R}^n$ that is feasible for the finite programming problem $P[Z]$ has a neighbourhood that can be completely defined by the active constraints at $\boldsymbol{x}$. In particular, if $\boldsymbol{x}^*$ is a solution of $P[Z]$ then it is also a solution of $P[Z']$ where Z' consists of the indices for the active constraints at $\boldsymbol{x}^*$. The converse is also true, that is, if $\boldsymbol{x}^*$ is a solution of $P[Z']$ it is also a solution of $P[Z]$. The situation is not so straightforward for *SIP* problems. If Y' indexes the (finite number of) constraints that are active at the solution of $P[Y]$, it is not true, in general, that the reduced problem $P[Y']$ has the same solution as $P[Y]$. Reemtsen and Görner [116] gave a simple example to illustrate this. It is possible, however, under certain assumptions, to derive a finite number of constraints so that the reduced problem has the same solution as $P[Y]$.

For a fixed $\bar{\boldsymbol{x}} \in X_P(Y)$ consider the following *parametric optimization problem*:

$$R(\bar{\boldsymbol{x}})[Y]: \quad \text{minimize } g(\bar{\boldsymbol{x}}, \boldsymbol{y}) \text{ subject to } \boldsymbol{y} \in Y. \tag{8.3.2}$$

We assume that there are a finite number of solutions to $R(\bar{\boldsymbol{x}})[Y]$. Let the solution set be $Y(\bar{\boldsymbol{x}}) = \{\boldsymbol{y}_1(\bar{\boldsymbol{x}}), \boldsymbol{y}_2(\bar{\boldsymbol{x}}), \ldots, \boldsymbol{y}_r(\bar{\boldsymbol{x}})\}$. Consider the following *reduced optimization problem*:

$$P_{\text{red}}[Y(\bar{\boldsymbol{x}})]: \quad \begin{array}{l} \text{minimize } f(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in X_P(Y(\bar{\boldsymbol{x}})), \\ \text{where } X_P(Y(\bar{\boldsymbol{x}})) \;=\; \{\boldsymbol{x} \in \mathbb{R}^n \mid g(\boldsymbol{x}, \boldsymbol{y}) \geq 0, \boldsymbol{y} \in Y(\bar{\boldsymbol{x}})\}. \end{array} \tag{8.3.3}$$

Under certain regularity conditions the solution of $P_{\text{red}}[Y(\bar{\boldsymbol{x}})]$ is also a solution of $P[Y]$, and conversely, see Hettich and Kortanek [70, Sub-section 7.3], and Reemtsen and Görner [116, Sub-section 2.6].

The main problem with the method of reduction is that the constraints indexed by $Y(\bar{\boldsymbol{x}})$ are implicitly defined. In practice, however, the following algorithm could be used:

1. Suppose for some $i \geq 1$, $\bar{\boldsymbol{x}}_i \in \mathbb{R}^n$ (which actually does not need to be feasible) is given.

2. Find all solutions (that is, all local minima) of $R(\bar{\boldsymbol{x}}_i)[Y]$. Let $Y(\bar{\boldsymbol{x}}_i)$ be the solution set.

3. Apply a finite programming algorithm to $P_{\text{red}}[Y(\bar{\boldsymbol{x}}_i)]$ to obtain the solution $\boldsymbol{x}_i^*$.

4. Let $\bar{\boldsymbol{x}}_{i+1} = \boldsymbol{x}_i^*$ and $i = i + 1$. Return to Step 1 until the desired accuracy has been reached.

Step 2 requires the global optimization of $g(\bar{\boldsymbol{x}}_i, \boldsymbol{y})$ which may be computationally costly. Also, local reduction methods generally require a number of assumptions to ensure convergence.

The third class of methods, exchange methods, are described in Hettich and Kortanek [70, Sub-section 7.1] as any algorithm of the form:

1. Let, for some $i \geq 1$, $Y_i \subset Y$ be a finite set.

2. Find a solution $\boldsymbol{x}_i^*$ of $P[Y_i]$.

3. Find all solutions $Y(\boldsymbol{x}_i^*) = \{\boldsymbol{y}_1(\boldsymbol{x}_i^*), \boldsymbol{y}_2(\boldsymbol{x}_i^*), \ldots, \boldsymbol{y}_r(\boldsymbol{x}_i^*)\}$ of $R(\boldsymbol{x}_i^*)[Y]$.

4. If, for $j = 1, 2, \ldots, r$, $g(\boldsymbol{x}_i^*, \boldsymbol{y}_j(\boldsymbol{x}_i^*)) \geq 0$, then stop.

5. Choose $Y_{i+1} \subset Y_i \cup Y(\boldsymbol{x}_i^*)$

6. Let $i = i + 1$ and return to Step 1.

Hettich and Kortanek [70] gave conditions under which the above algorithm, using $Y_{i+1} = Y_i \cup Y(\boldsymbol{x}_i^*)$ in Step 5, converges. Reemtsen and Görner [116] preferred to use the term "semi-continuous" when referring to *SIP* algorithms that are neither based on discretization nor local reduction. The above exchange algorithm is a particular case of a semi-continuous algorithm.

Discretization and exchange methods, in general, are computationally costly and produce only outer approximations but are relatively easy to implement. Local reduction methods, on the other hand, are usually less computationally costly but require restrictive assumptions for convergence. Also, given that the constraints in the reduced optimization problem (8.3.3) are implicitly defined in terms of the solution to (8.3.2), they are more difficult to implement. In practice often a discretization or exchange method is used to produce an outer approximation with a local reduction method implemented to give a final feasible solution. Reemtsen and Görner [116, Section 3] discussed in detail how each of the three abovementioned classes of methods have been used in linear, convex, and nonlinear *SIP* problems.

8.4 The Algorithm

In this section we present a *SIP* algorithm to determine whether $\boldsymbol{c}$ given by (8.2.1) is contained in Ω_p determined by a suitable vector $\boldsymbol{b}$. A discretization method is

used because it is the simplest to implement. Also, if the parameters that index the linear constraints are chosen to be sufficiently close together, the method of Section 7.3 can be utilized to refine the optimal solution and ensure that it corresponds to a *ME* distribution. To perform the convex minimization we use a steepest projected descent method, see Chong and Żak [36, Chapter 22]. MATLAB® has an inbuilt *SIP* function *fseminf* in the optimization toolbox, but as we shall mention in Section 8.5, our algorithm takes a lot less time to run and is slightly more accurate. The likely reason for this is that *fseminf* is a multipurpose *SIP* algorithm which uses more sophisticated techniques such as quadratic and cubic interpolation, and sequential quadratic programming, see Bertsekas [22].

A detailed explanation of our algorithm is given below.

1. Choose a set of nonnegative real numbers $U = \{0 = u_1, u_2, \ldots, u_{m-1}, u_m\}$ such that the set

$$\overline{\Omega}_p \;=\; \bigcap_{j=1}^{m} \{\boldsymbol{x} \in \mathbb{R}^{p-1} \mid b_1 \boldsymbol{e}_1' \exp(\boldsymbol{T}u_j)\boldsymbol{e}_p + \sum_{i=1}^{p-1} x_i \boldsymbol{e}_{i+1}' \exp(\boldsymbol{T}u_j)\boldsymbol{e}_p \geq 0\} \quad (8.4.1)$$

approximates Ω_p "sufficiently closely". Note that $\Omega_p \subset \overline{\Omega}_p$. If the constraint at infinity needs to be included we can set $u_m = \infty$. The corresponding constraint, if it exists, is given by (6.2.1). At this stage, given the limited number of general results concerning Ω_p, choosing U is ad hoc. Selecting $u_1 = 0$ would seem to be a good choice because Conjecture 5 in O'Cinneide [108] (see also Theorem 5.10 of this thesis) suggests that the corresponding constraint is always necessary. Also, if the zeros of $b(\lambda)$ are such that any complex zero of the form $-\lambda_p + i\beta$ ($\beta \in \mathbb{R}$) has multiplicity less than the multiplicity of $-\lambda_p$, then the constraint when $u_m = \infty$ is also valid, see Theorem 6.6. In addition, Corollary 6.7 asserts that if the zeros of $b(\lambda)$ can be ordered so that $-\lambda_{p-1}$ is real, then the constraint at $u_m = \infty$ forms part of $\partial\Omega_p$ and is hence necessary. If U is too sparse then it is possible that $\overline{\Omega}_p$ is unbounded. If this is the case then the algorithm may not converge.

2. Initialize $i = 0$ and $\boldsymbol{x}^{(0)} = \boldsymbol{x}_{\text{exp}}$, the point corresponding to the mixture of the point mass at zero and the defective exponential distribution (5.6.8). This point is always contained in Ω_p by Theorem 5.8 and is hence contained in $\overline{\Omega}_p$. Let $A \subset U$ be the set of indices that correspond to the set of active constraints. Initialize $A = \emptyset$ and choose a tolerance $\epsilon > 0$.

3. Calculate the direction of steepest descent of $h(\boldsymbol{x})$ when $\boldsymbol{x} = \boldsymbol{x}^{(i)}$. That is,

$$\boldsymbol{D} \;=\; -\boldsymbol{\nabla} h(\boldsymbol{x}^{(i)}) \;=\; -\frac{\boldsymbol{x}^{(i)} - \boldsymbol{c}}{h(\boldsymbol{x}^{(i)})}.$$

4. Let q be the number of active constraints. If $A = \{u_{(1)}, u_{(2)}, \ldots, u_{(q)}\} \neq \emptyset$ determine whether the optimal point has been found at the boundary. This occurs if

$$\boldsymbol{\nabla} h(\boldsymbol{x}^{(i)}) \;=\; \sum_{j=1}^{q} \mu_{(j)} \boldsymbol{x}^{(i)} \exp(\boldsymbol{T} u_{(j)}) \boldsymbol{e}_p \tag{8.4.2}$$

has a solution where each of $\mu_{(1)}, \mu_{(2)}, \ldots, \mu_{(q)}$ are nonnegative. If this is the case then goto Step 11, otherwise proceed to Step 5.

5. Determine which constraints will become inactive if we move from $\boldsymbol{x}^{(i)}$ in the direction $\boldsymbol{D}$. These constraints correspond to those $\mu_{(j)}$'s in (8.4.2) that are negative. Remove the corresponding indices from A. Update q.

6. If $A \neq \emptyset$ then project $\boldsymbol{D}$ onto the constraints indexed by A. If the active constraints are written in matrix form as $\boldsymbol{B}\boldsymbol{x}' = \boldsymbol{d}$ where the dimensions of $\boldsymbol{B}$ and $\boldsymbol{d}$ are $q \times p$ and $q \times 1$, respectively, the projected direction of steepest descent is given by

$$\boldsymbol{D}_1 \;=\; \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B}'(\boldsymbol{B}\boldsymbol{B}')^{-1}\boldsymbol{B}),$$

where $\boldsymbol{I}$ is the $p \times p$ identity matrix, see Chong and Żak [36, Section 22.1]. Let $\boldsymbol{D} = \boldsymbol{D}_1$.

7. Determine the constraint that is first violated when moving from $\boldsymbol{x}^{(i)}$ in the direction $\boldsymbol{D}$. Let this constraint be indexed by $\tilde{u}$.

8. Perform a Golden Section line search (see Chong and Żak [36, Section 7.1]) in the direction $\boldsymbol{D}$ from $\boldsymbol{x}_i$ until the constraint indexed by $\tilde{u}$ is reached. Let $\boldsymbol{x}^{(i+1)}$ be the value that minimizes $h(\boldsymbol{x})$ in this direction.

9. If h is minimized when the constraint indexed by $\tilde{u}$ is inactive, then if $|h(\boldsymbol{x}^{(i+1)}) - h(\boldsymbol{x}^{(i)})| < \epsilon$ go to Step 11, else let $i = i + 1$ and return to Step 3.

10. If h is minimized when the constraint indexed by $\tilde{u}$ is active, then if $|h(\boldsymbol{x}^{(i+1)}) - h(\boldsymbol{x}^{(i)})| < \epsilon$ go to Step 11, else let $i = i + 1$, add $\tilde{u}$ to A, and return to Step 3.

11. Let $\boldsymbol{x}^* = \boldsymbol{x}^{(i+1)}$ and $h^* = h(\boldsymbol{x}^*)$.

12. If $h^* = 0$ then $\boldsymbol{x}^* = \boldsymbol{c}$ and $\boldsymbol{c} \in \overline{\Omega}_p$. Otherwise, $h^* > 0$ and $\boldsymbol{x}^* \neq \boldsymbol{c}$ which implies that $\boldsymbol{c} \notin \overline{\Omega}_p$.

This algorithm gives an outer approximation when minimizing $h(\boldsymbol{x})$ over Ω_p. So, if $h^* = 0$ and $\boldsymbol{x}^* \in \partial\overline{\Omega}_p$ then it is unlikely that $\boldsymbol{c} \in \Omega_p$. However, since the indices of the active constraints, $u_{(1)}, u_{(2)}, \ldots, u_{(q)}$, are known, the function $d(u)$ in the algorithm of Section 7.3 could be minimized over intervals containing these values. Ideally, the intervals would need to be small enough so that $d(u)$ is convex on each one. If $h^* = 0$ and $\boldsymbol{x}^*$ is close to $\partial\overline{\Omega}_p$, or U does not contain sufficiently many points for $\overline{\Omega}_p$ to adequately approximate Ω_p, it is also possible that $\boldsymbol{c} \notin \Omega_p$. The method of Section 7.3 would be more difficult to apply in this situation unless we have some knowledge of the indices for the constraints that are near to $\boldsymbol{x}^*$. If, however, $h^* > 0$ we can be absolutely sure that $\boldsymbol{c} \notin \Omega_p$.

8.5 Examples

To illustrate the algorithm described in Section 8.4 we determined whether or not the following four pairs of vectors correspond to *ME* distributions.

1.

$$\boldsymbol{a} = \begin{pmatrix} 6 & 6 & 2 \end{pmatrix} \tag{8.5.1}$$

$$\boldsymbol{b} = \begin{pmatrix} 6 & 11 & 6 \end{pmatrix} \tag{8.5.2}$$

2.

$$\boldsymbol{a} = \begin{pmatrix} 6 & 1 & 3 \end{pmatrix} \tag{8.5.3}$$

$$\boldsymbol{b} = \begin{pmatrix} 6 & 11 & 6 \end{pmatrix} \tag{8.5.4}$$

3.

$$\boldsymbol{a} = \begin{pmatrix} 96 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{8.5.5}$$

$$\boldsymbol{b} = \begin{pmatrix} 120 & 274 & 225 & 85 & 15 \end{pmatrix} \tag{8.5.6}$$

4.

$$\boldsymbol{a} = \begin{pmatrix} 96 & 3 & 3 & 3 & 0 \end{pmatrix} \tag{8.5.7}$$

$$\boldsymbol{b} = \begin{pmatrix} 120 & 274 & 225 & 85 & 15 \end{pmatrix} \tag{8.5.8}$$

In Examples 1 and 2, $\lambda_1 = 3$, $\lambda_2 = 2$, $\lambda_3 = 1$, and $\frac{a_1}{b_1} = 1$. Thus, Conditions 2 and 3 of Theorem 5.6 are satisfied. Applying the *SIP* algorithm with the first pair of vectors gave $h^* = 0$. In light of the comments made at the end of Section 8.4 we cannot be absolutely sure that (8.5.1) and (8.5.2) correspond to a *ME* distribution, but an application of the algorithm of Section 7.3 confirmed that they in fact do. The second pair of vectors (8.5.3) and (8.5.4), however, do not correspond to a *ME* distribution. Of this we can be sure. The algorithm gave $h^* = 0.7113$ when $\boldsymbol{x}^* = (1.2805, 2.3463)$, the active constraint being the one that corresponds to $u = 0.6$. The point $\boldsymbol{x}^*$ itself does not quite correspond to a *ME* distribution. The algorithm of Section 7.3 gave $\hat{r} = 3.9557$ and $r^* = 3.9554$ when $d(u)$ was minimized

over the intervals $[0.5, 0.6]$ and $[0.6, 0.7]$. Note that it is $\hat{r}$ that corresponds to $\boldsymbol{x}^*$ here. The point that corresponds to the exponential distribution in this case is $\bar{\boldsymbol{x}} = (5, 1)$, and $\boldsymbol{x} = (1 - \frac{r^*}{\hat{r}})\bar{\boldsymbol{x}} + \frac{r^*}{\hat{r}}\boldsymbol{x}^* = (1.2807, 2.3462)$ corresponds to a *ME* distribution. The corresponding vectors are $\boldsymbol{a} = (6, 1.2807, 2.3462)$ and $\boldsymbol{b} = (6, 11, 6)$. In Examples 1 and 2 the algorithm took less than 0.1 second to run. On the other hand, *fseminf* took over a minute in both cases. It produced the same results (to four decimal places) for Example 1, but in Example 2 it gave $h^* = 0.7115$ when $\boldsymbol{x} = (1.2832, 2.3473)$, Also, the optimal value was not achieved on a constraint.

In Examples 3 and 4, since $\lambda_1 = 5$, $\lambda_2 = 4$, $\lambda_3 = 3$, $\lambda_4 = 2$, $\lambda_5 = 1$, and $\frac{a_1}{b_1} = \frac{4}{5}$, Conditions 2 and 3 of Theorem 5.6 are satisfied. In the third example, since the point mass at zero is $\frac{1}{5}$, in order to check that (8.5.5) and (8.5.6) correspond to a *ME* distribution we need to ascertain that $\boldsymbol{c} = (\frac{5}{4}, \frac{5}{4}, \frac{5}{4}, \frac{5}{4}) \in \Omega_p$, see (8.2.1). Applying the semi-infinite programming algorithm gave $h^* = 1.0876$ when $\boldsymbol{x}^* = (1.2572, 1.2921, 1.2721, 0.1634)$ which lies on the constraint indexed by $u = 0.31$. Thus, (8.5.5) and (8.5.6) do not correspond to a *ME* distribution. As with Example 2 above, since $\boldsymbol{x}^*$ does not quite correspond to a *ME* distribution, using the method of Section 7.3 we get $\boldsymbol{x} = (1.2589, 1.2929, 1.2723, 0.1635)$ which does correspond to a *ME* distribution. This translates into the vectors $\boldsymbol{a} = (96, 1.0057, 1.0337, 1.0177, 0.1308)$ and $\boldsymbol{b} = (120, 274, 225, 85, 15)$ corresponding to a *ME* distribution. For Example 4 the algorithm gave that (8.5.7) and (8.5.8) correspond to a *ME* distribution which the algorithm of Section 7.3 also confirmed. In Examples 3 and 4 the algorithm took less than 0.2 second to run, whereas *fseminf* took approximately two minutes. In Example 3 it gave $h^* = 1.0877$ when $\boldsymbol{x} = (1.2580, 1.2911, 1.2755, 0.1634)$. The optimal value was achieved on a constraint. The same results were achieved for Example 4.

In these examples our *SIP* algorithm took less time to run than the algorithm of Section 7.3. However, given that this method was required to refine the *SIP* there is no time saved when using the *SIP* algorithm to identify *ME* distributions. The real saving will come when we apply it many times when fitting *ME* distributions to

data and the algorithm of Section 7.3 is used only once to refine the final solution. This is the focus of the next chapter.

8.6 Problems and Suggested Improvements

Despite our algorithm outperforming *fseminf* in MATLAB® a more efficient method could be sought. One method that has been suggested in the *SIP* literature is *sequential quadratic programming* (*SQP*), see Polak [113, Section 2.9] or Bertsekas [22, Section 4.3]. *SQP* is an interior point penalty function method which could be particularly useful for our problem since infeasible solutions are not suitable because they do not correspond to *ME* distributions. Lawrence and Tits [87] gave and analyzed a *SQP* algorithm for finely discretized *SIP* problems.

The overall approach of employing a discretization algorithm followed by the method of Section 7.3 (which is essentially a local reduction algorithm) is probably a good one. However, more research needs to be done into the structure of the region Ω_p which would allow a more intelligent selection of the linear constraints that define the feasible region. Once a better, more efficient algorithm is developed investigation into the properties of the *ME* estimators can be embarked upon.

Chapter 9

Fitting with Matrix-exponential Distributions

9.1 Introduction

Many algorithms that are used to fit distributions to data require the distribution's parameters to be updated at each iteration. In order to fit *ME* distributions to data we would need to check that the updated parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution at each iteration. This process would require a considerable number of applications of the method of Section 7.3 and hence become computationally infeasible. In this chapter we address this issue by making use of the fact that for a suitable vector $\boldsymbol{b}$ we have a convex, closed, and bounded region that contains all points that correspond to *ME* distributions. Given the vector $\boldsymbol{b}$, the problem of fitting a *ME* distribution to data using maximum likelihood estimation then becomes one of maximizing a convex function over a convex region.

In Section 9.2, by modifying the algorithm in Section 8.4 we develop a procedure to find maximum likelihood parameter estimates when fitting *ME* distributions to data. The algorithm can also be used to approximate probability distributions by choosing sample points to represent the distribution. To illustrate the algorithm we

present in Section 9.3 one data fitting example, and one distribution approximation example.

9.2 Fitting Matrix-exponential Distributions to Data

In this section we develop a *SIP* algorithm similar to the one in Section 8.4 to find maximum likelihood estimates when fitting *ME* distributions to data.

Let S be a sample independently and identically distributed according to a *ME* distribution with parameters $\boldsymbol{a}$ and $\boldsymbol{b}$. Suppose that $\boldsymbol{b}$ has already been selected in some manner so that it satisfies Condition 3 of Theorem 5.6. If the estimate of the point mass at zero $\hat{\alpha}_0$ is given by the proportion of zero elements in S then the estimate for a_1 is

$$\hat{a}_1 \;=\; b_1(1-\hat{\alpha}_0), \tag{9.2.1}$$

see Lemma 5.2. Denote by $\overline{S} = \{z_1, z_2, \ldots, z_n\}$ the set that consists of all nonzero elements of S. Let $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$.

Recall from (5.6.7) and Theorem 5.11 that, for $\boldsymbol{x} \in \Omega_p$ and $u \geq 0$,

$$g(\boldsymbol{x}, u) \;=\; b_1 \boldsymbol{e}_1' \exp(\boldsymbol{T}u)\boldsymbol{e}_p + \sum_{i=1}^{p-1} x_i \boldsymbol{e}_{i+1}' \exp(\boldsymbol{T}u)\boldsymbol{e}_p.$$

Define, for $\boldsymbol{x} \in \Omega_p$, the likelihood function

$$G(\boldsymbol{x}, \boldsymbol{z}) \;=\; \prod_{j=1}^{n} g(\boldsymbol{x}, z_j). \tag{9.2.2}$$

Under the assumption that Ω_p is bounded, since the set is closed and $G(\boldsymbol{x}, \boldsymbol{z})$ is a continuous function of $\boldsymbol{x}$, there exists an $\tilde{\boldsymbol{x}} \in \Omega_p$ such that $G(\boldsymbol{x}, \boldsymbol{z})$ is maximized. Define, for $u \geq 0$, the set

$$\Lambda_p(u) \;=\; \{\boldsymbol{x} \in \Omega_p \mid g(\boldsymbol{x}, u) = 0\},$$

and let $\Lambda_p(\overline{S}) \;=\; \bigcup_{j=1}^{n} \Lambda_p(z_j)$. The likelihood function (9.2.2) can never attain a maximum on $\Lambda_p(\overline{S})$ as the next lemma asserts.

Lemma 9.1 *If $\tilde{\boldsymbol{x}}$ maximizes $G(\boldsymbol{x}, \boldsymbol{z})$ then $\tilde{\boldsymbol{x}} \notin \Lambda_p(\overline{S})$.*

Proof.

If $\boldsymbol{x} \in \Lambda_p(\overline{S})$ then $G(\boldsymbol{x}, \boldsymbol{z}) = 0$. Now, choose $\boldsymbol{x}$ so that it corresponds to the exponential distribution (5.6.8). Then

$$G(\boldsymbol{x}, \boldsymbol{z}) \;=\; \prod_{j=1}^{n} \lambda_p e^{-\lambda_p z_j} \;>\; 0.$$

Therefore, the value of $\boldsymbol{x}$ that maximizes $G(\boldsymbol{x}, \boldsymbol{z})$, that is $\tilde{\boldsymbol{x}}$, cannot be in $\Lambda_p(\overline{S})$. ■

Define, for $\boldsymbol{x} \in \Omega_p$ and $u > 0$,

$$l(\boldsymbol{x}, u) \;=\; \begin{cases} \log g(\boldsymbol{x}, u), & \boldsymbol{x} \notin \Lambda_p(u) \\ -\infty, & \boldsymbol{x} \in \Lambda_p(u) \end{cases}.$$

We have the following.

Lemma 9.2 *If, for $u > 0$, $\boldsymbol{x}, \boldsymbol{y} \in \Omega_p \backslash \Lambda_p(u)$ with $\boldsymbol{x} \neq \boldsymbol{y}$, and $0 < \rho < 1$, then*

$$l(\rho\boldsymbol{x} + (1-\rho)\boldsymbol{y}, u) \;>\; \rho l(\boldsymbol{x}, u) + (1-\rho) l(\boldsymbol{y}, u).$$

Proof.

Given the conditions in the statement of the theorem we have that

$$\begin{aligned} l(\rho\boldsymbol{x} + (1-\rho)\boldsymbol{y}, u) \;&=\; \log g(\rho\boldsymbol{x} + (1-\rho)\boldsymbol{y}, u) \\ &=\; \log\left(\rho g(\boldsymbol{x}, u) + (1-\rho) g(\boldsymbol{y}, u)\right) \\ &>\; \rho \log g(\boldsymbol{x}, u) + (1-\rho) \log g(\boldsymbol{y}, u) \\ &=\; \rho l(\boldsymbol{x}) + (1-\rho) l(\boldsymbol{y}). \end{aligned}$$

The second equality holds since $g(\boldsymbol{x}, u)$ is a linear function of $\boldsymbol{x}$, and the inequality is due to the strict convexity of $-\log(\cdot)$. ■

The loglikelihood function is defined, for $\boldsymbol{x} \in \Omega_p$, as

$$L(\boldsymbol{x}, \boldsymbol{z}) \;=\; \begin{cases} \displaystyle\sum_{j=1}^{n} l(\boldsymbol{x}, z_j), & \boldsymbol{x} \notin \Lambda_p(\overline{S}) \\ -\infty, & \boldsymbol{x} \in \Lambda_p(\overline{S}) \end{cases}. \tag{9.2.3}$$

We now have

Theorem 9.3 *Under the assumption that Ω_p is bounded there exists a unique $\tilde{\boldsymbol{x}} \in \Omega_p \backslash \Lambda_p(\overline{S})$ that maximizes $L(\boldsymbol{x}, \boldsymbol{z})$.*

Proof.

The loglikelihood function cannot attain a maximum on $\Lambda_p(\overline{S})$ by Lemma 9.1. For $j = 1, 2, \ldots, n$, the function $-l(\boldsymbol{x}, z_j)$ is strictly convex on $\Omega_p \backslash \Lambda_p(z_j)$ by Lemma 9.2. Since the sum of a finite number of strictly convex functions is also strictly convex on the intersection of their domains of definition, $-L(\boldsymbol{x}, \boldsymbol{z})$ is strictly convex on $\Omega_p \backslash \Lambda_p(\overline{S})$. Consequently, as Ω_p is closed and bounded, there exists a unique $\tilde{\boldsymbol{x}} \in \Omega_p \backslash \Lambda_p(\overline{S})$ that maximizes $L(\boldsymbol{x}, \boldsymbol{z})$. ■

Given $\boldsymbol{b}$ that satisfies Condition 3 of Theorem 5.6 we can use a *SIP* algorithm very similar to that given in Section 8.4 to maximize the loglikelihood (9.2.3). The direction of steepest ascent, for $\boldsymbol{x} \in \Omega_p \backslash \Lambda_p(\overline{S})$, is given by

$$\begin{aligned} \boldsymbol{D} &= \boldsymbol{\nabla} L(\boldsymbol{x}) \\ &= \begin{pmatrix} \frac{\partial L}{\partial x_1} & \frac{\partial L}{\partial x_2} & \cdots & \frac{\partial L}{\partial x_{p-1}} \end{pmatrix} \end{aligned}$$

where, for $i = 1, 2, \ldots, p-1$,

$$\frac{\partial L}{\partial x_i} = \sum_{j=1}^{n} \frac{\boldsymbol{e}_{i+1} \exp(\boldsymbol{T} z_j) \boldsymbol{e}_p}{g(\boldsymbol{x}, z_j)}.$$

Since $\tilde{\boldsymbol{x}}$ is unique it is a function of $\boldsymbol{b}$ and we write $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}(\boldsymbol{b})$. Consequently, the loglikelihood (9.2.3) can be expressed as

$$L(\boldsymbol{b}, \boldsymbol{z}) = \sum_{j=1}^{n} \log\Big(b_1 \boldsymbol{e}_1' \exp(\boldsymbol{T} z_j) \boldsymbol{e}_p + \sum_{i=1}^{p-1} \tilde{x}_i(\boldsymbol{b}) \boldsymbol{e}_{i+1}' \exp(\boldsymbol{T} z_j) \boldsymbol{e}_p \Big). \tag{9.2.4}$$

Since $\boldsymbol{T}$ depends nonlinearly on $\boldsymbol{b}$ the function $L(\boldsymbol{b}, \boldsymbol{z})$ is *not* convex in $\boldsymbol{b}$. In order to estimate $\boldsymbol{b}$ so that (9.2.4) is maximized we require a multidimensional nonlinear optimization procedure such as simulated annealing or the Nelder-Mead flexible polyhedron search.

We need to ensure that the vector $\boldsymbol{b}$ is such that it satisfies Condition 3 of Theorem 5.6. A necessary condition for this is that its components are nonnegative.

However, this condition is *not* sufficient. For example, the polynomial x^3+x^2+x+2 has zeros that are approximately -1.3532 and $0.1766 \pm 1.2028i$. Therefore, it is insufficient to require that the components of $\boldsymbol{b}$ are positive to ensure that Condition 3 of Theorem 5.6 is satisfied when maximizing $L(\boldsymbol{b}, \boldsymbol{z})$. We can, however, ensure the condition is met by requiring that the zeros of $b(\lambda)$ are such that they all have negative real part and that there exists a zero of maximal real part that is real. Thus, we write $L(\boldsymbol{b}, \boldsymbol{z})$ as $L(\boldsymbol{\lambda}, \boldsymbol{z})$.

Since the *SIP* algorithm will produce an outer approximation (see the comments immediately after the *ME* identification algorithm in Section 8.4) the algorithm of Section 7.3 can be applied to ensure that the final parameter estimates correspond to a *ME* distribution.

In summary, in order to fit a *ME* distribution to data we have the following algorithm:

1. Calculate $\hat{\alpha}_0$ and form $\overline{S}$.

2. Calculate $\hat{a}_1$ using (9.2.1).

3. Choose an initial $\boldsymbol{\lambda}^{(0)} = (\lambda_1^{(0)}, \lambda_2^{(0)}, \ldots, \lambda_p^{(0)})$.

4. Calculate $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p)$ that maximizes $L(\boldsymbol{\lambda}, \boldsymbol{z})$.

5. Calculate $\hat{\boldsymbol{b}}$ from the coefficients of

$$b(\lambda) \;=\; (\lambda + \hat{\lambda}_1)(\lambda + \hat{\lambda}_2) \ldots (\lambda + \hat{\lambda}_p).$$

6. Calculate $\hat{\boldsymbol{x}} = \boldsymbol{x}(\hat{\boldsymbol{b}})$, the value of $\tilde{\boldsymbol{x}}$ that maximizes (9.2.3) given $\hat{\boldsymbol{b}}$.

7. Calculate $\hat{\boldsymbol{a}}$ using, for $i = 2, 3, \ldots, p$,

$$\begin{aligned} \hat{a}_i &= (1 - \alpha_0)\, \hat{x}_{i-1} \\ &= \frac{\hat{a}_1}{\hat{b}_1}\, \hat{x}_{i-1}, \end{aligned}$$

see (5.6.5) and Lemma 5.2.

8. Apply the algorithm of Section 7.3.

Since $L(\boldsymbol{\lambda}, \boldsymbol{z})$ is not convex in $\boldsymbol{\lambda}$ the algorithm will converge only to a local maximum. Therefore, there is no guarantee that the algorithm will find the maximum likelihood estimate. In practice the algorithm should be run several times with different initial parameters and the one with the greatest loglikelihood selected. It may be, however, that the best loglikelihood after Step 7 does not yield the best loglikelihood once Step 8 is applied. Given this possibility, the best loglikelihoods achieved with Steps 1–7 should be selected and then refined separately using Step 8.

9.3 Examples

In order to demonstrate the *ME* fitting algorithm described in Section 9.2 we fit three order five *ME* distributions, each one with $b(\lambda)$ having a different configuration of zeros, to a sample that was formed by shifting the Old Faithful geyser data set, mentioned in Section 3.3, 1.66 units to the left. A histogram of the data is shown in Figure 9.3.1 and its empirical cumulative distribution function is shown in Figure 9.3.2. Fitting the original data set with *PH* distributions has proven to be very difficult because there are no data between zero and 1.67, see Asmussen, Nerman, and Olsson [15]. Since *PH* density functions are positive on the positive real numbers fitting this gap well requires distributions of high order, see Faddy [53] and [56]. We anticipate a similar problem with *ME* distributions. In general, *ME* density functions may be zero for some positive real numbers, but the number of such values is usually small. For example, *ME* distributions of order three where $b(\lambda)$ either has only real zeros, or where the real zero is greater than the real part of the complex conjugate zeros, have density functions that equal zero for at most one positive value, see Theorems 6.4 and 6.5. However, as we shall see, this particular data set demonstrates the extra flexibility *ME* distributions exhibit over *PH* distributions in fitting problems.

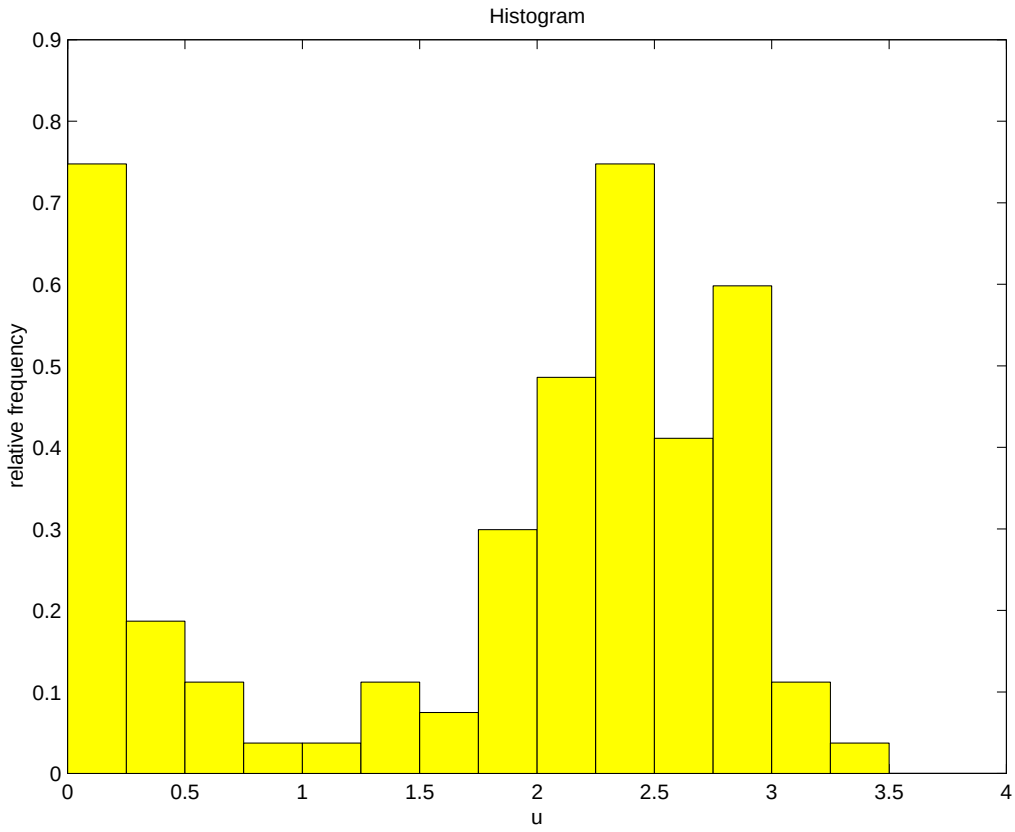

Figure 9.3.1: Histogram of the shifted inter-eruption times of the Old Faithful geyser data set

Since our implementation of the *ME* fitting algorithm is not very efficient the examples presented in this section are mainly illustrative. An in-depth study into fitting with *ME* distributions can be undertaken when a more efficient algorithm is developed. Our algorithm however, as in Section 8.5, outperformed *fseminf* in MATLAB®. An initial run using *fseminf* took considerably longer per iteration than our algorithm and eventually produced a matrix-dimensionality error which stopped the program.

In order to maximize $L(\boldsymbol{\lambda}, \boldsymbol{z})$ the Nelder-Mead algorithm was used because it is a reliable inbuilt function in MATLAB®. For each of the three fitted *ME* distributions the zeros of $b(\lambda)$ were such that

1 all five were real,

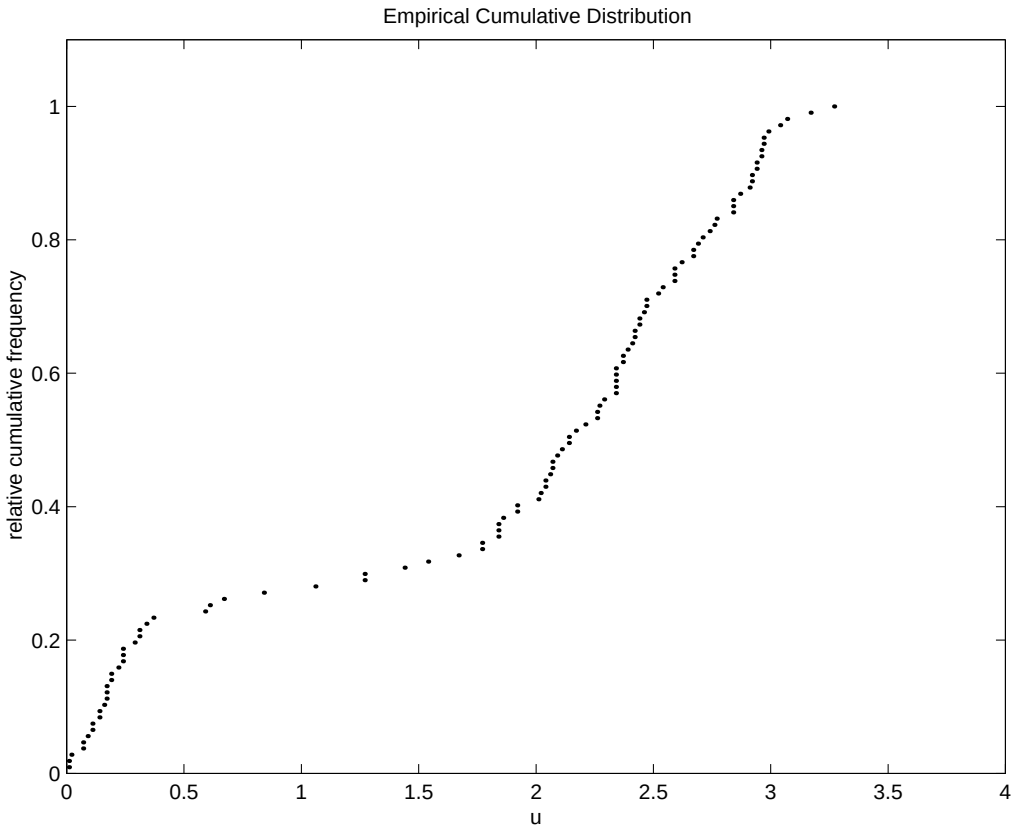

Figure 9.3.2: Empirical cumulative distribution of the shifted inter-eruption times of the Old Faithful geyser data set

2 three were real and two were a complex conjugate pair, and

3 one was real and four were complex conjugate pairs.

These three configurations of zeros were treated separately because, at each iteration in the Nelder-Mead algorithm when $\boldsymbol{\lambda}$ is updated, there is no simple way of changing a real zero into a complex one, and vice versa. It is much easier to update separately the real and imaginary (if there are any) parts of the zeros which are all real numbers.

The algorithm was run with ten different sets of starting parameters and the one with the highest loglikelihood was then selected. The final fitted parameters were refined using the method of Section 7.3. Note that the Nelder-Mead algorithm finds local minima so we are actually minimizing $-L(\boldsymbol{\lambda}, \boldsymbol{z})$ even though we speak

of maximizing $L(\boldsymbol{\lambda}, \boldsymbol{z})$. For each run $U = \{0, 0.01, \ldots, 10, \infty\}$. In each of the three cases the starting parameters were randomly chosen from a uniform distribution on $(0, 10)$. The starting parameters were the negatives of the real parts of all zeros, and either the positives or the negatives of their imaginary parts, as appropriate. Throughout it was ensured that the zero of maximal real part was real and negative.

We now discuss the results. For the first case the final loglikelihoods ranged from -266.1221 to -124.3825. Run times ranged from 21 to 43 minutes. The best loglikelihood was achieved when the zeros of $b(\lambda)$ were

$$\hat{\boldsymbol{\lambda}} \;=\; \begin{pmatrix} -2.7056 & -2.2886 & -1.9883 & -1.8489 & -0.8759 \end{pmatrix},$$

which corresponds to the estimate

$$\hat{\boldsymbol{b}} \;=\; \begin{pmatrix} 19.9375 & 59.6544 & 67.5479 & 36.7671 & 9.7073 \end{pmatrix}.$$

The estimate for $\boldsymbol{a}$ was

$$\hat{\boldsymbol{a}} \;=\; \begin{pmatrix} 19.9375 & 23.6151 & 8.6018 & 9.1515 & 0.5041 \end{pmatrix}.$$

The maximum loglikelihood occurred when the constraint indexed by $u = \infty$ was active. This solution is likely to be exact because this constraint forms part of the boundary of Ω_5, see Corollary 6.7. An application of the method of Section 7.3 confirmed that this is indeed the case. Since $\boldsymbol{\lambda}$ consists entirely of real numbers and the maximum loglikelihood was attained when $u = \infty$ the *ME* distribution represented by $\boldsymbol{a}$ and $\boldsymbol{b}$ is a *ME* distribution of order four, see Corollary 6.7. The zeros of the polynomial

$$a(\lambda) \;=\; 0.5041\lambda^4 + 9.1515\lambda^3 + 8.6018\lambda^2 + 23.6151\lambda + 19.9375$$

are -17.3183, -0.8759, and $0.0196 \pm 1.6147i$. Cancelling the identical factors in the *LST* $\phi(\lambda) = \dfrac{a(\lambda)}{b(\lambda)}$ gives the maximum likelihood parameter estimates

$$\hat{\boldsymbol{\lambda}} \;=\; \begin{pmatrix} -2.7056 & -2.2886 & -1.9883 & -1.8489 \end{pmatrix},$$

which corresponds to the estimate

$$\hat{\boldsymbol{b}} = \begin{pmatrix} 22.7629 & 42.1195 & 29.0319 & 8.8314 \end{pmatrix}.$$

The estimate for $\boldsymbol{a}$ then becomes

$$\hat{\boldsymbol{a}} = \begin{pmatrix} 22.7629 & 0.9730 & 8.7100 & 0.5041 \end{pmatrix}.$$

In the second case the final loglikelihoods ranged from -325.6831 to -114.3065. Run times ranged from 11 to 32 minutes. The best loglikelihood was achieved when the zeros of $b(\lambda)$ were

$$\hat{\boldsymbol{\lambda}} = \begin{pmatrix} -3.8163 & -1.1829 & -1.1347 \pm 1.6534i & -0.5676 \end{pmatrix}$$

which corresponds to the estimate

$$\hat{\boldsymbol{b}} = \begin{pmatrix} 10.3030 & 35.3777 & 41.6319 & 24.0063 & 7.8362 \end{pmatrix}.$$

The estimate for $\boldsymbol{a}$ was

$$\hat{\boldsymbol{a}} = \begin{pmatrix} 10.3030 & 17.3382 & 2.6518 & 6.7637 & 0.5023 \end{pmatrix}.$$

The constraint that corresponds to $u = 4.76$ was active when the maximum loglikelihood was attained. Using the method of Section 7.3 changed a_3 to 2.6519, and the loglikelihood decreased slightly to -114.3066. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution that is not a *PH* distribution because the density function is zero for $u \approx 4.76$.

For the third case the loglikelihoods ranged from -164.8511 to -110.9792. Run times ranged from 8 to 35 minutes. The best loglikelihood was achieved when the zeros of $b(\lambda)$ were

$$\hat{\boldsymbol{\lambda}} = \begin{pmatrix} -1.0079 \pm 2.0601i & -1.0079 \pm 0.7818i & -0.6512 \end{pmatrix}$$

which gives the estimate

$$\hat{\boldsymbol{b}} = \begin{pmatrix} 5.5726 & 17.5974 & 21.0122 & 13.5751 & 4.6826 \end{pmatrix}.$$

The estimate for $\boldsymbol{a}$ was

$$\hat{\boldsymbol{a}} = \begin{pmatrix} 5.5726 & 7.6548 & 2.6449 & 3.0305 & 0.7413 \end{pmatrix}.$$

The constraint that corresponds to $u = 4.54$ was active when the maximum loglikelihood was achieved. The method of Section 7.3 decreased the loglikelihood slightly to -111.0081. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ correspond to a *ME* distribution that is not a *PH* distribution because the density function is zero for $u \approx 4.54$.

In order to compare our *ME* fitting algorithm with a *PH* fitting algorithm we fitted the same data set with a general order five *PH* distribution using EMpht. The fitted *PH* distribution was a Coxian distribution of triangular order five. Its corresponding loglikelihood was -133.6592 and the parameter estimates were

$$\hat{\boldsymbol{\lambda}} = \begin{pmatrix} -4.0847 & -1.9270 & -1.9270 & -1.9270 & -1.9270 \end{pmatrix}$$

$$\hat{\boldsymbol{b}} = \begin{pmatrix} 56.3196 & 130.6962 & 119.6253 & 53.7637 & 11.7926 \end{pmatrix}$$

$$\hat{\boldsymbol{a}} = \begin{pmatrix} 56.3196 & 29.3261 & 22.8282 & 7.8978 & 1.0246 \end{pmatrix}.$$

The algorithm took approximately four seconds to perform 700 iterations, by which stage convergence had been reached. In fact, using EMpht, a Coxian distribution of order 13 needs to be fitted to the data set in order to achieve a loglikelihood greater than -111.0081. The run time for convergence for this fit was approximately one minute.

The density functions for the three *ME* fits, and the order five *PH* fit, plotted with the histogram of the data, are shown in Figure 9.3.3. The corresponding distribution functions, plotted with the empirical cumulative distribution of the data, are shown in Figure 9.3.4

It is clear from Figure 9.3.3 that the two *ME* densities whose corresponding polynomials $b(\lambda)$ have complex conjugate zeros fit the histogram the best. Their corresponding loglikelihoods are also the greatest. Although each of the four densities fit the first peak well, it is Fits 2 and 3 that fit the second, broader peak and

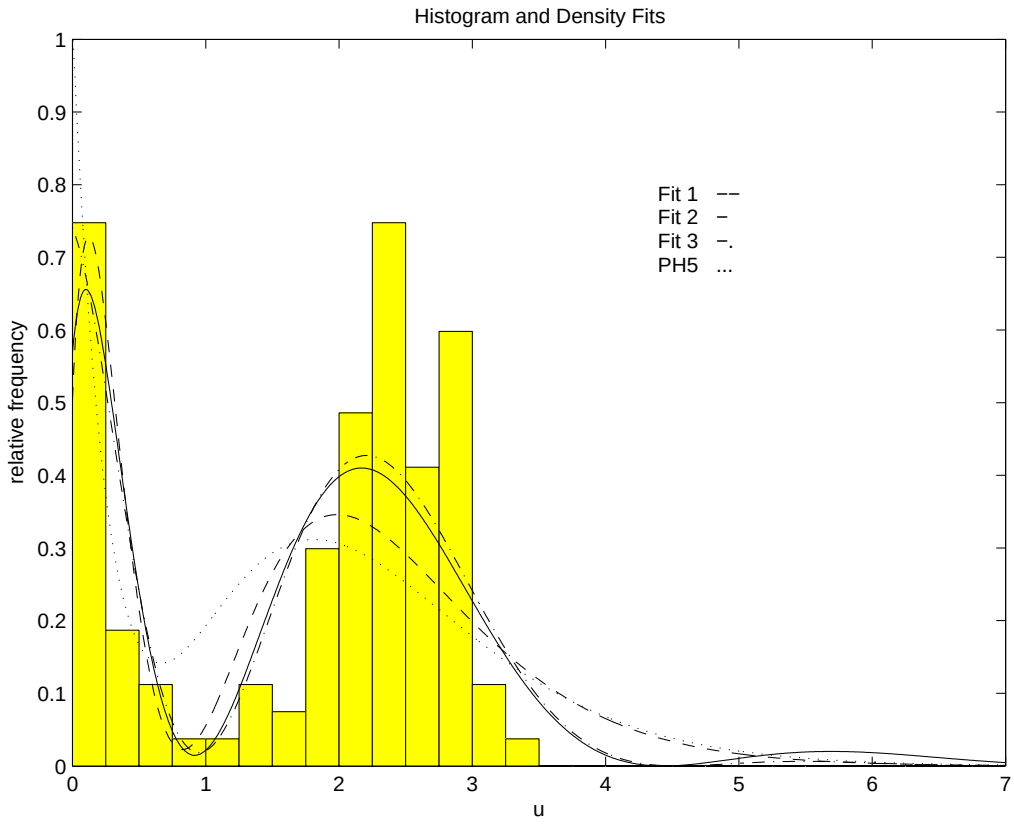

Figure 9.3.3: Density functions for the three *ME* and one *PH* fits plotted with the histogram of the data

the trough around 1 the best. Fit 3 fits the second peak better than Fit 2 and it has a shorter tail. Also, Fit 2's tail rises to a small peak around 5.5 before decaying again. Fit 1 appears to outperform the *PH* fit since it fits the trough around 1 and the second peak better.

Comparing the fitted distributions with the empirical cumulative distribution gives a better picture. Again, it is clear from Figure 9.3.3 that Fits 2 and 3 fit the empirical cumulative distribution the best. Each distribution function fits the data well up to approximately 0.3, and the *PH* fit does reasonably well up to approximately 1. Fits 2 and 3 outperform the other two fits on $[1.1, 2.3]$ but do not fit the data well on $[1.3, 2.3]$. From 2.7–4 Fits 2 and 3 do better than Fit 1 and the *PH* Fit with Fit 3 outperforming Fit 2 on this interval. On the interval $[4.3, 6.6]$ Fit 2

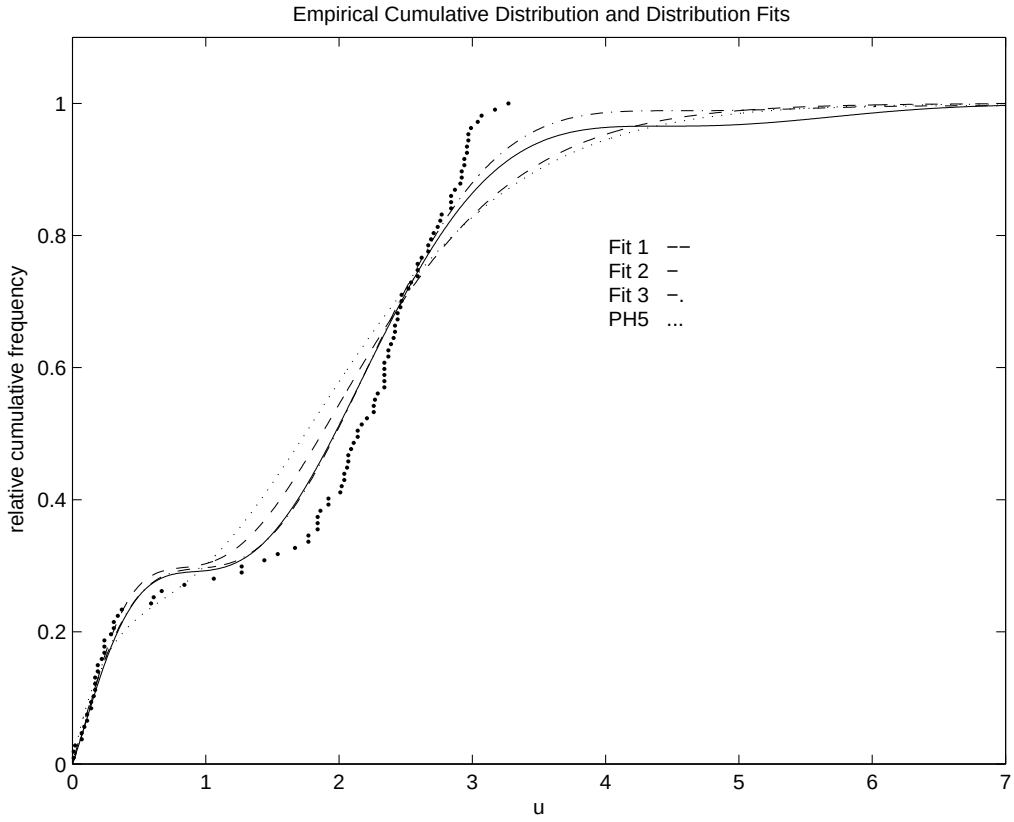

Figure 9.3.4: Distribution functions for the three *ME* and one *PH* fits with the empirical cumulative distribution of the data

performs worse than the other three fits. This corresponds with the small peak in Fit 2's density at $u \approx 5.5$ mentioned in the previous paragraph.

To further illustrate the algorithm given in Section 9.2 the uniform density on $(1, 2)$ was approximated with three order five *ME* densities. This density comes from the Aalborg benchmark, see Bobbio and Telek [25]. Each approximating *ME* density had one of the three configurations of zeros for $b(\lambda)$ used in the previous example. The sample chosen to represent the density was $S = \{1.005, 1.015, \ldots, 1.995\}$. As with the previous example we chose $U = \{0, 0.01, \ldots, 10, \infty\}$ and the starting parameters were again chosen from a uniform distribution on $(0, 10)$. In each case the algorithm was run ten times and the best loglikelihood was selected and then refined using the method of Section 7.3. However, in the case when the zeros of $b(\lambda)$ are all

real, the parameters that initially gave the best loglikelihood, when refined, actually gave a loglikelihood that was less than the second best (unrefined) loglikelihood. Refining the second best set of parameters actually produced a better loglikelihood than the first. The parameters giving rise to the best refined loglikelihood are given below. Overall, run times ranged from 2 to 37 minutes.

When $b(\lambda)$ has zeros that are all real the final parameters were

$$\begin{aligned}
\hat{\boldsymbol{\lambda}} &= \begin{pmatrix} -4.5974 & -4.2547 & -4.1718 & -4.0988 & -3.8283 \end{pmatrix}, \\
\hat{\boldsymbol{b}} &= \begin{pmatrix} 1280.5055 & 1533.3180 & 733.7806 & 175.4252 & 20.9511 \end{pmatrix}, \\
\hat{\boldsymbol{a}} &= \begin{pmatrix} 1280.5055 & -377.0691 & 83.3616 & -3.1616 & 0.4441 \end{pmatrix}.
\end{aligned}$$

The loglikelihood was -43.5593. When $b(\lambda)$ has three real zeros the final parameters were

$$\begin{aligned}
\hat{\boldsymbol{\lambda}} &= \begin{pmatrix} -6.9980 & -5.6924 & -1.6556 \pm 3.1473i & -1.6556 \end{pmatrix}, \\
\hat{\boldsymbol{b}} &= \begin{pmatrix} 834.0386 & 987.8479 & 448.8461 & 120.9943 & 17.6572 \end{pmatrix}, \\
\hat{\boldsymbol{a}} &= \begin{pmatrix} 834.0386 & -277.7983 & 60.9690 & -3.2405 & 0.3758 \end{pmatrix}.
\end{aligned}$$

The loglikelihood was -33.1678. When $b(\lambda)$ has one real zero the final parameters were

$$\begin{aligned}
\hat{\boldsymbol{\lambda}} &= \begin{pmatrix} -4.1805 \pm 1.8639i & -1.9013 \pm 3.3573i & -1.9013 \end{pmatrix}, \\
\hat{\boldsymbol{b}} &= \begin{pmatrix} 592.9764 & 699.9947 & 332.7166 & 90.7566 & 14.0648 \end{pmatrix}, \\
\hat{\boldsymbol{a}} &= \begin{pmatrix} 592.9764 & -200.2047 & 42.0915 & -2.4444 & 0.1537 \end{pmatrix}.
\end{aligned}$$

The loglikelihood was -30.9800. None of these approximating *ME* distributions is a *PH* distribution as each one has zero density for some positive values of u. Also, the second and third distributions have zeros of maximal real part that equal (at least to the above accuracy) the real part of a complex conjugate pair of zeros.

Using an order five *PH* distribution to approximate the shifted uniform density with EMpht yielded an Erlang distribution with parameters

$$\hat{\boldsymbol{\lambda}} = \begin{pmatrix} -3.3333 & -3.3333 & -3.3333 & -3.3333 & -3.3333 \end{pmatrix},$$

$$\hat{\boldsymbol{b}} = \begin{pmatrix} 411.5226 & 617.2840 & 370.3704 & 111.1111 & 16.6667 \end{pmatrix},$$

$$\hat{\boldsymbol{a}} = \begin{pmatrix} 411.523 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The loglikelihood was -61.3004. The algorithm took approximately three seconds to perform 1000 iterations, by which stage convergence had been achieved. In fact, an order 12 Erlang distribution is required to achieve an approximation that has a loglikelihood greater than -30.9800. The run time was approximately 45 seconds. Aldous and Shepp [3] showed that the *PH* distribution with the least coefficient of variation (2.4.3) is the Erlang distribution. The coefficient of variation in the order five case is 0.2. Since the shifted uniform density has a small coefficient of variation (that is, 0.0370) it would seem reasonable that the best *PH* approximation is an Erlang distribution. The coefficients of variation for the above approximating *ME* distributions are 0.1488, 0.1571, and 0.1299, respectively. This observation does not contradict the result of Aldous and Shepp [3] because they are not *PH* distributions, see also O'Cinneide [106, page 226].

The density functions for the three *ME* approximations, and the order five *PH* approximation, plotted with the shifted uniform density, are shown in Figure 9.3.5. The corresponding distribution functions, plotted with the shifted uniform distribution function, are shown in Figure 9.3.6

As with the previous example it is clear from Figure 9.3.5 that the two *ME* distributions with $b(\lambda)$ having complex conjugate zeros approximate the shifted uniform density the best. Their loglikelihoods are also the greatest. Approximations 2 and 3 approximate the peak the best but have longer tails than Approximation 1. Approximation 3 approximates the peak better than Approximation 2 and has a slightly shorter tail. All of the *ME* densities are nonzero at the origin, unlike the

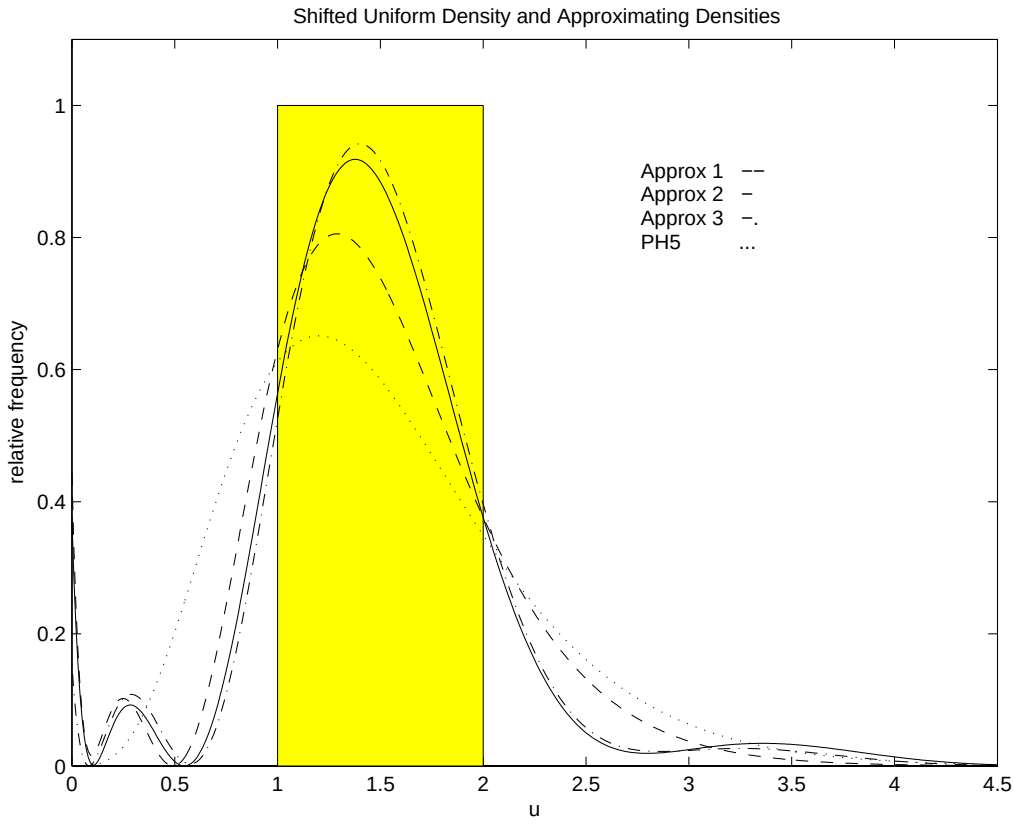

Figure 9.3.5: Density functions for the three *ME* and one *PH* approximations plotted with the density function of the uniform distribution on $(1, 2)$

Erlang density, but are able to approximate the gap $[0, 1]$ better because they are zero at two points in the interval.

It is also clear from Figure 9.3.6 that the *ME* approximations do better than the *PH* approximation, which only does better on the interval $[0, 0.5]$. Approximations 2 and 3 approximate the shifted uniform distribution function better than Approximation 1 on the intervals $[0.6, 1.4]$ and $[1.7, 2.5]$, but Approximation 1 does slightly better on $[2.8, 4]$. Approximation 3, overall, performed slightly better than Approximation 2, its logliklehood also being larger.

These examples indicate that more flexibility can be achieved when *ME* distributions are used in preference to *PH* or Coxian distributions to fit data or approximate distributions. This observation is not surprising since any *PH* distribution is a *ME*

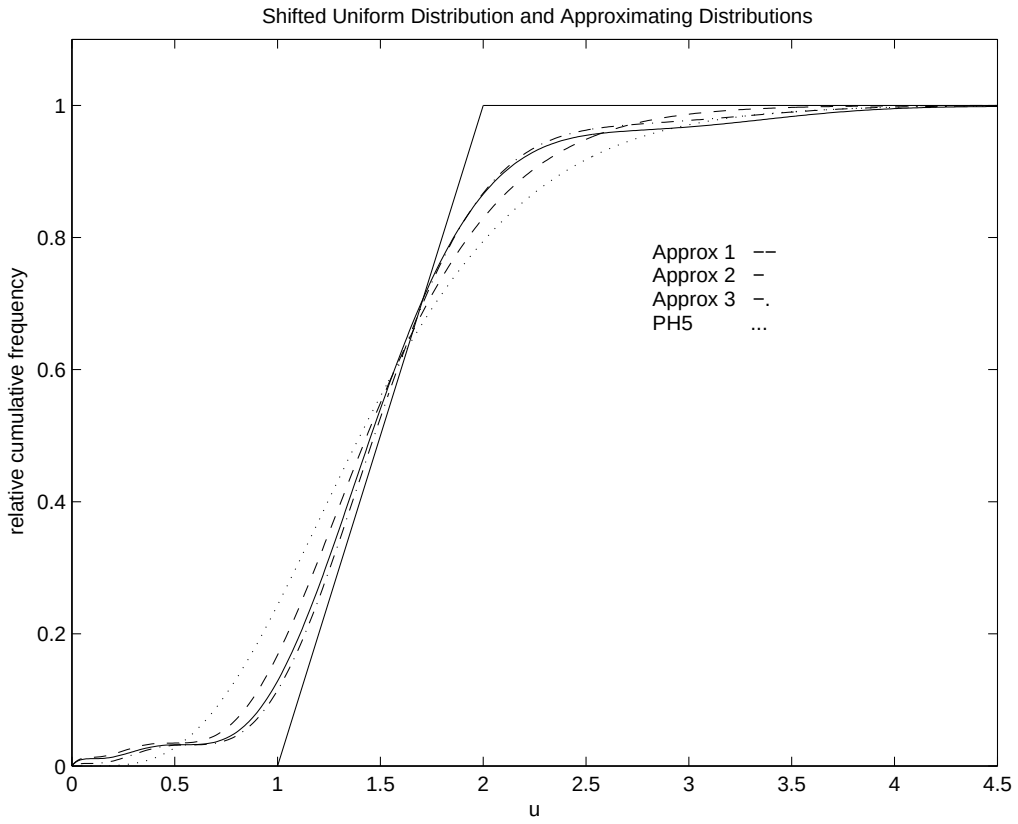

Figure 9.3.6: Distribution functions for the three *ME* and one *PH* approximations with the distribution function for the uniform distribution on $(1, 2)$

distribution and *PH* distributions of high order can be represented by *ME* distributions of low order. These examples also suggest that *ME* distributions whose corresponding polynomial $b(\lambda)$ has complex zeros may be better suited to fit data sets, or approximate distributions, that have gaps or are multimodal. This fact may be due to the sinusoidal nature of such *ME* distributions but a more in-depth study is required to establish any meaningful results.

Chapter 10

Conclusion

We began this thesis by considering *PH* distributions and how they could be used to fit data and approximate probability distributions. In order to overcome some of the problems with current methods, particularly that of overparameterization, it was proposed in Chapter 4 that the fitting or approximation be carried out in the *LST* domain. Two methods due to Harris and Marchal [66] were described in detail and some associated problems were discussed. In particular, two important questions arose:

1. When does a rational *LST* correspond to a *PH* distribution?
2. If it does correspond to a *PH* distribution, what is a *PH* representation for it?

In Chapter 5 we found that the second question could be answered easily if we considered *ME* distributions and their representations instead. The original second question, with regard to *PH* distributions, is an important one and remains unsolved. O'Cinneide's [107] characterization of the triangular order of Coxian distributions using invariant polytopes would be a natural place to start.

Given the one-to-one correspondence between a rational *LST* and the companion form representation of a *ME* distribution we stated our results in terms of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. In order to answer the first question in terms of *ME* distributions the

necessary condition on the vector $\boldsymbol{b}$ was relatively simple. For the vector $\boldsymbol{a}$ to then correspond to a *ME* distribution it needed to belong to Ω_p which was defined by an uncountably infinite set of linear constraints that depend on $\boldsymbol{b}$. In Chapter 6 we derived a complete analytical description for Ω_3, but the problem for higher orders is difficult to solve.

In Chapter 7 we developed an algorithm based on the work of Dehon and Latouche [45] which, given a suitable vector $\boldsymbol{b}$, determined whether or not the vector $\boldsymbol{a}$ belonged to Ω_p. An alternative description of Ω_3 was also given. Since the algorithm given in Section 7.3 required the global minimization of a single-variable function over the nonnegative real numbers, a potentially difficult optimization problem, in Chapter 8 we developed a *SIP* algorithm to tackle the *ME* identification problem. Even though the algorithm of Section 7.3 was required in the last stage the optimization was localized and hence easier to perform accurately. In Chapter 9 the *SIP* approach was used to develop an algorithm to fit *ME* distributions to data. The examples given towards the end of the chapter illustrated the extra flexibility *ME* distributions exhibit over *PH* distributions when used to fit data or approximate probability distributions even though our algorithm (and particularly *fseminf* in MATLAB®) took much longer to converge than EMpht. More efficient nonlinear optimization and *SIP* algorithms, to find $\boldsymbol{b}$ and $\boldsymbol{a}$, respectively, need to be developed and tested thoroughly.

Our original idea of fitting or approximating with *PH* and *ME* distributions in the *LST* domain is now possible using the *SIP* approach. An objective function which measures, in some sense, the distance between the fitting/approximating *LST* and the empirical/approximated *LST* could be minimized. It may even be, as with the loglikelihood function, that given a suitable vector $\boldsymbol{b}$ the objective function is convex in $\boldsymbol{a}$. However, given that the maximum likelihood estimation is relatively easy to perform, and that the properties of maximum likelihood estimators are well established, transform estimation is essentially unnecessary. Transform estimation has usually only been used when methods such as maximum likelihood and moment

matching have led to intractable problems.

Once a better *SIP* algorithm has been developed a thorough experimental analysis of fitting *ME* distributions to data can be carried out. Confidence intervals for the parameter estimates using, say, the bootstrap can be calculated. In the literature it is only Faddy [51] and [52], Faddy and McClean [55], and Hampel [65] that reported the asymptotic standard errors for the maximum likelihood estimates when fitting Coxian distributions to data.

On the theoretical side, properties of the maximum likelihood estimators (for example, consistency and asymptotic normality) for *PH* and *ME* distributions need to be established. Asmussen, Nerman, and Olsson [15] stated for the case with *PH* distributions

> *"... that due to the over-parameterization the situation is somewhat non-standard, although usual asymptotic distribution properties concerning estimable quantities ... should be derivable from knowledge of the existence of a sufficiently regular unique parameterization. (Candidates for such a parameterization are either the zeros and poles of the Laplace transform, or maybe a sequence of moments, ...)"*

As we have seen such a parameterization exists for *ME* distributions and therefore it seems likely that the abovementioned properties will be shown first for *ME* distributions. So far, only the consistency of the maximum likelihood estimators for *MMPP*s has been established by Rydén [118].

It may also be possible to extend the methods developed in this thesis to identify *RAP*s and then use them to develop fitting and approximation algorithms.

Concerning Ω_p we still need to establish that it is bounded and contains the origin, as well as finding a complete analytical description for the set when $p > 3$. The work of Kreĭn and Nudel′man [80] on convex sets and Chebyshev systems ought to lead to some fruitful developments in these problems.

Bibliography

[1] Aalen, O. O. Phase type distributions in survival analysis. *Scandinavian Journal of Statistics*, 22:447–463, 1995.

[2] Akaike, H. A new look at the statistical model identification. *IEEE – Transactions on Automatic Control*, 19:716–723, 1974.

[3] Aldous, D. and Shepp, L. The least variable phase type distribution is Erlang. *Communications in Statistics – Stochastic Models*, 3:467–473, 1987.

[4] Alfa, A. S. and Chakravarthy, S. R., editors. *Advances in Matrix-Analytic Methods in Stochastic Models*. Notable Publications, New Jersey, USA, 1998.

[5] Apostol, T. M. *Calculus*, volume I. John Wiley and Sons, New York, 2nd edition, 1967.

[6] Apostol, T. M. *Calculus*, volume II. John Wiley and Sons, New York, 2nd edition, 1969.

[7] Asmussen, S. Phase-type representations in random walk and queueing problems. *The Annals of Probability*, 20:772–789, 1992.

[8] Asmussen, S. Phase-type distributions and related point processes: Fitting and recent advances. In Chakravarthy, S. R. and Alfa, A. S, editors, *Matrix-analytic Methods in Stochastic Models*, volume 183 of *Lecture Notes in Pure and Applied Mathematics*, pages 137–149. Marcel Dekkar, New York, 1997.

[9] Asmussen, S. *Ruin Probabilities*, volume 2 of *Advanced Series on Statistical Science*. World Scientific Publishing Company, River Edge, New Jersey, 2000.

[10] Asmussen, S. and Bladt, M. Renewal theory and queueing algorithms for matrix-exponential distributions. In Chakravarthy, S. R. and Alfa, A. S., editors, *Matrix-analytic Methods in Stochastic Models*, volume 183 of *Lecture Notes in Pure and Applied Mathematics*, pages 313–341. Marcel Dekkar, New York, 1997.

[11] Asmussen, S. and Bladt, M. Point processes with finite-dimensional conditional probabilities. *Stochastic Processes and their Applications*, 82:127–142, 1999.

[12] Asmussen, S. and O'Cinneide, C. A. Matrix-exponential distributions. In Kotz, S., Read, C. B., and Banks, D. L., editors, *Encyclopedia of Statistical Science Update*, volume 2, pages 435–440. John Wiley and Sons, New York, 1998.

[13] Asmussen, S. and Olsson, M. Phase-type distributions (update). In Kotz, S., Read, C. B., and Banks, D. L., editors, *Encyclopedia of Statistical Science Update*, volume 2, pages 525–530. John Wiley and Sons, New York, 1998.

[14] Asmussen, S. and Rolski, T. Computational methods in risk theory: A matrix-algorithmic approach. *Insurance: Mathematics and Economics*, 10:259–274, 1992.

[15] Asmussen, S., Nerman, O., and Olsson, M. Fitting phase-type distributions via the *EM* algorithm. *Scandinavian Journal of Statistics*, 23:419–441, 1996.

[16] Assaf, D. and Langberg, N. Presentation of phase-type distributions as proper mixtures. *Journal of Applied Probability*, 22:247–250, 1985.

[17] Assaf, D. and Levikson, B. Closure of phase type distributions under operations arising in reliability theory. *The Annals of Probability*, 10:265–269, 1982.

[18] Bar-Lev, S. K., Barkan, N., and Langberg, N. A. Moment generating function based estimators with some optimal properties. *Journal of Statistical Planning and Inference*, 35:279–291, 1993.

[19] Bean, N. G. and Nielsen, B. F. Quasi-birth-and-death processes with rational arrival process components I: Definition and analysis. Submitted, 2003.

[20] Bean, N. G., Green, D. A., and Taylor, P. G. The output process of an $MMPP/M/1$ queue. *Journal of Applied Probability*, 35:998–1002, 1998.

[21] Bell, R. J. T. *An Elementary Treatise on Coordinate Geometry of Three Dimensions*. MacMillan and Company Limited, London, 3rd edition, 1956.

[22] Bertsekas, D. P. *Nonlinear Programming*. Athena Scientific, Belmont, Massachusetts, 2nd edition, 1999.

[23] Bladt, M. and Neuts, M. F. Matrix-exponential distributions: Calculus and interpretations via flows. *Communications in Statistics – Stochastic Models*, 19:115–129, 2003.

[24] Bobbio, A. and Cumani, A. *ML* estimation of the parameters of a *PH* distribution in triangular canonical form. In Balbo, G. and Serazzi, G., editors, *Computer Performance Evaluation*, pages 33–46. Elsevier Science Publishers, 1992.

[25] Bobbio, A. and Telek M. A benchmark for *PH* estimation algorithms: Results for acyclic-*PH*. *Communications in Statistics – Stochastic Models*, 10:661–677, 1994.

[26] Bobbio, A., Cumani, A., Premoli, A., and Saracco, O. Modelling and identification of non-exponential distributions by homogeneous Markov processes. In *Proceedings of the 6th Advances in Reliability Technology Symposium*, pages 373–392, Bradford, 1980.

[27] Botta, R. B and Harris, C. M. Approximation with generalized hyperexponential distributions: Weak convergence results. *Queueing Systems*, 2:169–190, 1986.

[28] Botta, R. B, Harris, C. M., and Marchal, W. G. Characterizations of generalized hyperexponential distribution functions. *Communications in Statistics – Stochastic Models*, 3:115–148, 1987.

[29] Breuer, L. Parameter estimation for a class of *BMAP*s. In Latouche, G. and Taylor, P., editors, *Advances in Algorithmic Methods for Stochastic Models*, pages 87–97. Notable Publications, New Jersey, USA, 2000.

[30] Breuer, L. and Gilbert, M. Parameter estimation for *BMAP*s. Technical report, Department IV, Computer Science, Universität Trier, D-54286 Trier, Germany, 2000.

[31] Bryant, J. L. and Paulson, A. S. Estimation of mixing proportions via distance between characteristic functions. *Communications in Statistics – Theory and Methods*, 12:1009–1029, 1983.

[32] Bux, W. and Herzog, U. The phase concept: Approximation of measured data and performance analysis. In Chandy, K. M. and Reiser, M., editors, *Computer Performance*, pages 23–38. North Holland Publishing Company, 1977.

[33] Chakravarthy, S. Reliability analysis of a parallel system with exponential life times and phase type repairs. *OR Spektrum*, 5:25–32, 1983.

[34] Chakravarthy, S. R. and Alfa, A. S., editors. *Matrix-Analytic Methods in Stochastic Models*, volume 183 of *Lecture Notes in Pure and Applied Mathematics*. Marcel Dekkar, New York, 1997.

[35] Charnes, A., Cooper, W. W., and Kortanek, K. O. Duality, Haar programs and finite sequence spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 48:783–786, 1962.

[36] Chong, E. K. P. and Żak, S. H. *An Introduction to Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley and Sons, New York, 1996.

[37] Cohen, J. W. *The Single Server Queue*, volume 8 of *North-Holland Series in Applied Mathematics and Mechanics*. North-Holland Publishing Company, Amsterdam, 1969.

[38] Commault, C. and Chemla, J. P. An invariant of representations of phase-type distributions and some applications. *Journal of Applied Probability*, 33:368–381, 1996.

[39] Commault, C. and Mocanu, S. A generic property of phase-type representations. *Journal of Applied Probability*, 39:775–785, 2002.

[40] Cox, D. R. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Proceedings of The Cambridge Philosophical Society*, 51:433–441, 1955.

[41] Cox, D. R. A use of complex probabilities in the theory of stochastic processes. *Proceedings of The Cambridge Philosophical Society*, 51:313–319, 1955.

[42] Cumani, A. On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectronics and Reliability*, 22:583–602, 1982.

[43] Davis, M. H. A. *Markov Models and Optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 1993.

[44] Debnath, L. *Integral Transforms and their Applications*. CRC Press, Boca Raton, 1995.

[45] Dehon, M. and Latouche, G. A geometric interpretation of the relations between the exponential and generalized Erlang distributions. *Advances in Applied Probability*, 14:885–897, 1982.

[46] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the *EM* algorithm. With discussion. *Journal of the Royal Statistical Society. Series B. Methodological*, 39:1–38, 1977.

[47] Diamond, J. E. and Alfa, A. S. On approximating higher order *MAP*s with *MAP*s of order two. *Queueing Systems*, 34:269–288, 2000.

[48] Erlang, A. K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *The Post Office Electrical Engineer's Journal*, 10:189–197, 1917–18.

[49] Faddy, M. J. Compartmental models with phase-type residence-time distributions. *Applied Stochastic Models and Data Analysis*, 6:121–127, 1990.

[50] Faddy, M. J. A structured compartmental model for drug kinetics. *Biometrics*, 49:243–248, 1993.

[51] Faddy, M. J. Examples of fitting structured phase-type distributions. *Applied Stochastic Models and Data Analysis*, 10:247–255, 1994.

[52] Faddy, M. J. Phase-type distributions for failure times. *Mathematical and Computer Modelling*, 22:63–70, 1995.

[53] Faddy, M. J. On inferring the number of phases in a Coxian phase-type distribution. *Communications in Statistics – Stochastic Models*, 14:407–417, 1998.

[54] Faddy, M. J. Penalized maximum likelihood estimation of the parameters in a Coxian phase-type distribution. In Latouche, G. and Taylor, P., editors, *Matrix-analytic Methods. Theory and Applications*, pages 107–114. World Scientific, New Jersey, 2002.

[55] Faddy, M. J. and McClean, S. I. Analysing data on lengths of stay of hospital patients using phase-type distributions. *Applied Stochastic Models in Business and Industry*, 15:311–317, 1999.

[56] Faddy, M. J. Private Communication, 2002.

[57] Feigin, P. D., Tweedie, R. L., and Belyea, C. Weighted area techniques for explicit parameter estimation in multi-stage models. *Australian Journal of Statistics*, 25:1–16, 1983.

[58] Feldmann A. and Whitt, W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31:245–279, 1998.

[59] Feller, W. *An Introduction to Probability Theory and its Applications*, volume II. John Wiley and Sons, New York, 2nd edition, 1971.

[60] Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H. User's guide for NPSOL (version 4.0): A FORTRAN package for nonlinear programming. Technical Report SOL 86-2, Department of Operations Research, Stanford University, 1986.

[61] Goberna, M. Á and López, M. A., editors. *Semi-Infinite Programming: Recent Advances*, volume 57 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[62] Goberna, M. Á, and López, M. A. A comprehensive survey of linear semi-infinite optimization theory. In Reemtsen, R. and Rückmann, J-J., editors, *Semi-Infinite Programming*, volume 25 of *Nonconvex Optimization and its Applications*, pages 3–27. Kluwer Academic Publishers, Boston, 1998.

[63] Goberna, M. Á, and López, M. A. *Linear Semi-Infinite Optimization*, volume 2 of *Wiley Series in Mathematical Methods in Practice*. John Wiley and Sons, Chichester, 1998.

[64] Green, D., Metcalfe, A. V., and Swailes, D. C. A matrix analytic model for machine maintenance. In Latouche, G. and Taylor, P., editors, *Matrix-analytic Methods. Theory and Applications*, pages 133–146. World Scientific, New Jersey, 2002.

[65] Hampel, K. *Modelling Phase-type Distributions*. Honours Thesis. Department of Statistics, The University of Adelaide, South Australia, 1997.

[66] Harris, C. M. and Marchal, W. G. Distribution estimation using Laplace transforms. *INFORMS Journal on Computing*, 10:448–458, 1998.

[67] Harris, C. M. and Sykes, E. A. Likelihood estimation for generalized mixed exponential distributions. *Naval Research Logistics*, 34:251–279, 1987.

[68] Harris, C. M., Marchal, W. G., and Botta, F. B. A note on generalized hyper-exponential distributions. *Communications in Statistics – Stochastic Models*, 8:179–191, 1992.

[69] Heathcote, C. R. The integrated squared error estimation of parameters. *Biometrika*, 64:255–264, 1977.

[70] Hettich, R. and Kortanek, K. O. Semi-infinite programming: Theory, methods, and applications. *SIAM Review*, 35:380–429, 1993.

[71] Hoeting, J. A. and Tweedie, R. L. Transform estimation of parameters for stage-frequency data. Unpublished manuscript, 2001.

[72] Hórvath, A. and Telek, M. Approximating heavy tailed behaviour with phase type distributions. In Latouche, G. and Taylor, P., editors, *Advances in Algorithmic Methods for Stochastic Models*, pages 191–213. Notable Publications, New Jersey, USA, 2000.

[73] Johnson, M. A. Selecting parameters of phase distributions: Combining nonlinear programming, heuristics, and Erlang distributions. *ORSA Journal on Computing*, 5:69–83, 1993.

[74] Johnson, M. A. and Taaffe, M. F. Matching moments to phase distributions: Mixtures of Erlang distributions of common order. *Communications in Statistics – Stochastic Models*, 5:711–743, 1989.

[75] Johnson, M. A. and Taaffe, M. F. Matching moments to phase distributions: Density function shapes. *Communications in Statistics – Stochastic Models*, 6:283–306, 1990.

[76] Johnson, M. A. and Taaffe, M. F. Matching moments to phase distributions: Nonlinear programming approaches. *Communications in Statistics – Stochastic Models*, 6:259–281, 1990.

[77] Kammler, D. W. Least squares approximation of completely monotonic functions by sums of exponentials. *SIAM Journal of Numerical Analysis*, 16:801–818, 1979.

[78] Khoshgoftaar, T. M. and Perros, H. G. A comparison of three methods of estimation for approximating general distributions by a Coxian distribution. In Fdida, S. and Pujolle, G., editors, *Proceedings of the 3rd International Workshop on Modelling Techniques and Performance Evaluation*, pages 81–89. Elsevier Science Publishers, 1987.

[79] Kortanek, K. O. On the 1962–1972 decade of semi-infinite programming: A subjective view. In Goberna, M. Á and López, M. A., editors, *Semi-Infinite Programming: Recent Advances*, volume 57 of *Nonconvex Optimization and its Applications*, pages 3–41. Kluwer Academic Publishers, Dordrecht, 2001.

[80] Kreĭn, M. G. and Nudel′man, A. A. *The Markov Moment Problem and Extremal Problems*, volume 50 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1977.

[81] Kreyszig, E. *Advanced Engineering Mathematics*. John Wiley and Sons, New York, 8th edition, 1997.

[82] Lang, A. and Arthur, J. L. Parameter approximation for phase-type distributions. In Chakravarthy, S. R. and Alfa, A. S., editors, *Matrix-analytic Methods in Stochastic Models*, volume 183 of *Lecture Notes in Pure and Applied Mathematics*, pages 151–206. Marcel Dekkar, New York, 1997.

[83] Latouche, G. and Taylor, P., editors. *Advances in Algorithmic Methods for Stochastic Models. Proceedings of the 3rd International Conference on Matrix Analytic Methods*, Notable Publications, New Jersey, USA, 2000.

[84] Latouche, G. and Taylor, P., editors. *Matrix-Analytic Methods – Theory and Applications. Proceedings of the Fourth International Conference on Matrix-Analytic Methods in Stochastic Models*, World Scientific, New Jersey, 2002.

[85] Latouche, G. and Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Models*. SIAM, Philadelphia, 1999.

[86] Laurence, A. F. and Morgan, B. J. T. Selection of the transformation variable in the Laplace transform method of estimation. *Australian Journal of Statistics*, 29:113–127, 1987.

[87] Lawrence, C. T. and Tits, A. L. Feasible sequential quadratic programming for finely discretized problems from *SIP*. In Reemtsen, R. and Rückmann, J-

J., editors, *Semi-Infinite Programming*, volume 25 of *Nonconvex Optimization and its Applications*, pages 159–193. Kluwer Academic Publishers, Boston, 1998.

[88] Lipsky, L. and Fang, Z. Classification of functions with rational Laplace transforms. In *Summer Computer Simulation Conference*, pages 20–25, Las Vegas, Nevada, 1986.

[89] Lipsky, L. and Ramaswami, V. A unique minimal representation of Coxian service centres. Technical report, Department of Computer Science and Engineering, University of Nebraska, Lincoln, 1985.

[90] Lipsky, L. R. *Queueing Theory: A Linear Algebraic Approach.* Macmillan Publishing Company, New York, 1992.

[91] Longman, I. M. Best rational function approximation for Laplace transform inversion. *SIAM Journal of Mathematical Analysis*, 5:574–580, 1974.

[92] Madsen, H. and Nielsen, B. F. The use of phase type distributions for modelling packet-switched traffic. In Jensen, A. and Iversen, V. B., editors, *Teletraffic and Datatraffic in a Period of Change. Proceedings of the 13th International Teletraffic Congress ITC-13*, pages 593–599. Elsevier Science Publishers, North Holland, 1992.

[93] Maier, R. S. The algebraic construction of phase-type distributions. *Communications in Statistics – Stochastic Models*, 7:573–602, 1991.

[94] Maier, R. S. and O'Cinneide, C. A. A closure characterisation of phase-type distributions. *Journal of Applied Probability*, 29:92–103, March 1992.

[95] McLachlan, G. J. and Krishnan, T. *The EM algorithm and Extensions.* Wiley Series in Probability and Statistics. John Wiley and Sons, New York, 1997.

[96] Meier-Hellstern, K. S. A fitting algorithm for Markov-modulated Poisson processes having two arrival rates. *European Journal of Operational Research*, 29:370–377, 1987.

[97] Mocanu, Ş. and Commault, C. Sparse representations of phase-type distributions. *Communications in Statistics – Stochastic Models*, 15:759–778, 1999.

[98] Møller, J. R. *On Matrix-analytic Methods and on Collective Risk in Life Insurance*. PhD thesis, Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Lund, 2000.

[99] Moran, P. A. P. *An Introduction to Probability Theory*. Oxford Science Publications. Clarendon Press, Oxford, 1984.

[100] Neuts, M. F. Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*, pages 173–206, Department of Mathematics, University of Louvain, Belgium, 1975.

[101] Neuts, M. F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The John Hopkins University Press, Baltimore, 1981.

[102] Neuts, M. F. Matrix-analytic methods in queueing theory. In Dshalalow, J. H., editor, *Advances in Queueing: Theory, Methods, and Open Problems*, Probability and Stochastics Series, pages 265–292. CRC Press, Boca Raton, Florida, 1995.

[103] O'Cinneide, C. A. On non-uniqueness of representations of phase-type distributions. *Communications in Statistics – Stochastic Models*, 5:247–259, 1989.

[104] O'Cinneide, C. A. Characterization of phase-type distributions. *Communications in Statistics – Stochastic Models*, 6:1–57, 1990.

[105] O'Cinneide, C. A. Phase-type distributions and invariant polytopes. *Advances in Applied Probability*, 23:515–535, 1991.

[106] O'Cinneide, C. A. Phase-type distributions and majorization. *The Annals of Applied Probability*, 1:219–227, 1991.

[107] O'Cinneide, C. A. Triangular order of triangular phase-type distributions. *Communications in Statistics – Stochastic Models*, 9:507–529, 1993.

[108] O'Cinneide, C. A. Phase-type distributions: Open problems and a few properties. *Communications in Statistics – Stochastic Models*, 15:731–757, 1999.

[109] Olivier, C. and Walrand, J. On the existence of finite-dimensional filters for Markov-modulated traffic. *Journal of Applied Probability*, 31:515–525, 1994.

[110] Olsson, M. Estimation of phase-type distributions from censored data. *Scandinavian Journal of Statistics*, 23:443–460, 1996.

[111] Parsons, D. H. Biological problems involving sums of exponential functions of time: An improved method of calculation. *Mathematical Biosciences*, 9:37–47, 1970.

[112] Paulson, A. S., Holcomb, E. W., and Leitch, R. A. The estimation of the parameters of the stable laws. *Biometrika*, 62:163–170, 1975.

[113] Polak, E. *Optimization. Algorithms and Consistent Approximations*, volume 124 of *Applied Mathematical Sciences*. Springer, New York, 1997.

[114] Poritsky, H. Convex spaces associated with a family of linear inequalities. In Klee, V. L., editor, *Convexity. Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society*, volume 7, pages 403–436. American Mathematical Society, Providence, Rhode Island, 1963.

[115] Reemtsen, R. and Rückmann, J-J., editors. *Semi-Infinite Programming*, volume 25 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Boston, 1998.

[116] Reemtsen, R. and Görner, S. Numerical methods for semi-infinite programming: A survey. In Reemtsen, R. and Rückmann, J-J., editors, *Semi-Infinite Programming*, volume 25 of *Nonconvex Optimization and its Applications*, pages 195–275. Kluwer Academic Publishers, Boston, 1998.

[117] Rice, J. A. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, California, 2nd edition, 1995.

[118] Rydén, T. Parameter estimation for Markov-modulated Poisson processes. *Communications in Statistics – Stochastic Models*, 10:795–829, 1994.

[119] Rydén, T. Consistent and asymptotically normal parameter estimates for Markov modulated Poisson processes. *Scandinavian Journal of Statistics*, 22:295–303, 1995.

[120] Rydén, T. Estimating the order of hidden Markov models. *Statistics*, 26:345–354, 1995.

[121] Rydén, T. An *EM* algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics and Data Analysis*, 21:431–447, 1996.

[122] Rydén, T. Estimating the order of continuous phase-type distributions and Markov-modulated Poisson processes. *Communications in Statistics – Stochastic Models*, 13:417–433, 1997.

[123] Rydén, T. Statistical estimation for Markov-modulated Poisson processes and Markovian arrival processes. In Latouche, G. and Taylor, P., editors, *Advances in Algorithmic Methods for Stochastic Models*, pages 329–350. Notable Publications, New Jersey, USA, 2000.

[124] Schmickler, L. MEDA: Mixed Erlang distributions as phase-type representations of empirical distribution functions. *Communications in Statistics – Stochastic Models*, 8:131–156, 1992.

[125] Schuh, H-J. and Tweedie, R. L. Parameter estimation using transform estimation in time-evolving models. *Mathematical Biosciences*, 45:37–67, 1979.

[126] Schwarz, G. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[127] Seneta, E. *Non-negative Matrices and Markov Chains*. Springer Series in Statistics. Springer-Verlag, New York, 2nd edition, 1981.

[128] Sethuraman, J. and Squillante, M. S. Analysis of parallel-server queues under spacesharing and timesharing disciplines. In Latouche, G. and Taylor, P., editors, *Matrix-analytic Methods. Theory and Applications*, pages 357–380. World Scientific, New Jersey, 2002.

[129] Shaked, M. and Shanthikumar, J. G. Phase-type distributions. In Kotz, S., Johnson, N. L., and Read, C. B., editors, *Encyclopedia of Statistical Science*, volume 6, pages 709–715. John Wiley and Sons, New York, 1985.

[130] Silverman, B.W. *Density Estimation for Statistics and Data Analysis*, volume 26 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, London, 1986.

[131] Smith, W. L. On the distribution of queueing times. *Proceedings of the Cambridge Philosophical Society*, 49:449–461, 1953.

[132] Spiegel, M. R. *Theory and Problems of Statistics*. Schaum's Outline Series. Schaum Publishing Company, New York, 1961.

[133] Spivak, M. *A Comprehensive Guide to Differential Geometry*, volume 3. Publish or Perish, Berkeley, 2nd edition, 1979.

[134] Squillante, M. S. Matrix-analytic methods: Applications results and software tools. In Latouche, G. and Taylor, P., editors, *Advances in Algorithmic Meth-*

ods for Stochastic Models, pages 351–374. Notable Publications, New Jersey, USA, 2000.

[135] Steeb, W-H. in collaboration with Shi, T. K. *Matrix Calculus and Kronecker Product with Applications and C++ Programs*. World Scientific, Singapore, 1997.

[136] Strawderman, W. S. Likelihood ratio tests. In Kotz, S., Johnson, N. L., and Read, C. B., editors, *Encyclopedia of Statistical Science*, volume 4, pages 647–650. John Wiley and Sons, New York, 1985.

[137] Sumita, U. and Masuda, Y. Classes of probability density functions having Laplace transforms with negative zeros and poles. *Advances in Applied Probability*, 19:632–651, 1987.

[138] Takahashi, Y. Asymptotic exponentiality of the tail of the waiting time distribution in a $PH/PH/c$ queue. *Advances in Applied Probability*, 13:619–630, 1981.

[139] Tenenbaum, M. and Pollard, H. *Ordinary Differential Equations*. Harper's Mathematics Series. Harper and Row, New York, 1963.

[140] The Open University. *Chebychev Approximation. Unit 27*. The Open University Press, Bletchley, Buckinghamshire, 1972.

[141] The Open University. *Least-squares Approximation*. The Open University Press, Bletchley, Buckinghamshire, 1972.

[142] Titterington, D. M., Smith, A. F. M., and Makov, U. E. *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Chichester, 1985.

[143] Tweedie, R. L., Zhu, Z. Y., and Choy, S. L. Parameter estimation using Laplace transforms in the $M/M/1$ queue. Unpublished manuscript, 1993.

[144] Van de Liefvoort, A. The moment problem for continuous distributions. Working paper WP-CM-1990-02. Technical report, Computer Science/Telecommunications Program, University of Missouri, Kansas City, 1990.

[145] Wackerly, D. D., Mendenhall, W., and Scheaffer, R. L. *Mathematical Statistics and Applications*. Duxbury Press, 5th edition, 1995.

[146] Wardle, K. L. *Differential Geometry*. Library of Mathematics. Routledge and Kegan Paul, London, 1965.

[147] Wilks, S. S. *Mathematical Statistics*. John Wiley and Sons, New York, 1962.

[148] Yao, Q. and Morgan, B. J. T. Empirical transform estimation for indexed stochastic models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 61:127–141, 1999.

[149] Yao, R. A proof of the steepest increase conjecture of a phase-type density. *Communications in Statistics – Stochastic Models*, 18:1–6, 2002.

[150] Yeramian, E. and Claverie, P. Analysis of multiexponential functions without a hypothesis as to the number of components. *Nature*, 326:169–174, 12 March 1987.

[151] Zemanian, A. H. On the pole and zero locations of rational Laplace transformations of non-negative functions. *Proceedings of the American Mathematical Society*, 10:868–872, 1959.

[152] Zemanian, A. H. On the pole and zero locations of rational Laplace transformations of non-negative functions. II. *Proceedings of the American Mathematical Society*, 12:870–874, 1961.