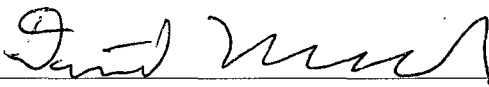



CONTROL THEORETIC APPROACH TO SAMPLING AND
APPROXIMATION PROBLEMS


By

Anna S. Bulanova

RECOMMENDED:

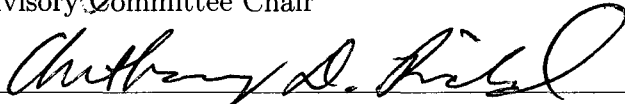








Advisory Committee Chair

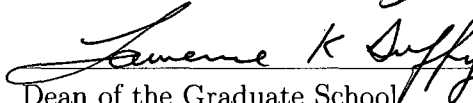


Chair, Department of Mathematics and Statistics

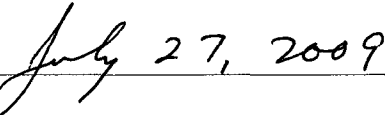
APPROVED:



Dean, College of Natural Science and Mathematics



Dean of the Graduate School



Date

**CONTROL THEORETIC APPROACH TO SAMPLING AND
APPROXIMATION PROBLEMS**

A
THESIS

Presented to the Faculty
of the University of Alaska Fairbanks
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

By
Anna S. Bulanova, B.S., M.S.

Fairbanks, Alaska

August 2009

UMI Number: 3386294

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

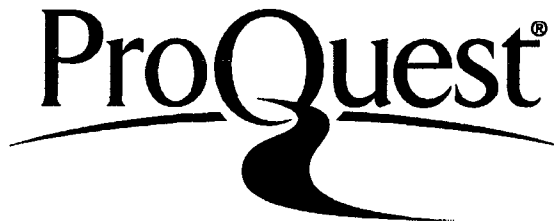
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3386294

Copyright 2010 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

We present applications of some methods of control theory to problems of signal processing and optimal quadrature problems.

The following problems are considered: construction of sampling and interpolating sequences for multi-band signals; spectral estimation of signals modeled by a finite sum of exponentials modulated by polynomials; construction of optimal quadrature formulae for integrands determined by solutions of initial boundary value problems.

A multi-band signal is a function whose Fourier transform is supported on a finite union of intervals. The approach used in Chapter I is based on connections between the sampling and interpolation problem and the problem of the controllability of a dynamical system. We prove that there exist infinitely many sampling and interpolating sequences for signals whose spectra are supported on a union of two disjoint intervals, and provide an algorithm for construction of such sequences.

There exist numerous methods for solving the spectral estimation problem. In Chapter II we introduce a new approach to this problem based on the Boundary Control method, which uses the connection between inverse problems of mathematical physics and control theory for partial differential equations. Using samples of the signal at integer moments of time we construct a convolution operator regarded as an input-output map of a linear discrete dynamical system. This system can be identified, and the exponents and amplitudes of the signal can be found from the parameters of the system. We show that the coefficients of the signal can be recovered by solving a generalized eigenvalue problem as in the Matrix Pencil method. Our method allows to consider signals with polynomial amplitudes, and we obtain an exact formula for these amplitudes.

In the third chapter we consider an optimal quadrature problem for solutions of initial boundary value problems. The problem of optimization of an error functional over the set of solutions and quadrature weights is a problem of optimal control of partial differential equations. We obtain estimates for the error in quadrature formulae and an optimality condition for quadrature weights.

Table of Contents

	Page
Signature Page	i
Title Page	ii
Abstract	iii
Table of Contents	iv
List of Tables	vii
Acknowledgements	viii
General Introduction	1
Sampling and interpolation	1
Frequency estimation	3
Approximate integration	4
Statement of contributions	5
Bibliography	7
1 Construction of sampling and interpolating sequences for multi-band signals. The two-band case	10
Abstract	10
1.1 Introduction	10
The main results	15
1.2 The Operators W , V and K	15
1.3 The invertibility of the Operator K	19
1.3.1 The case of $a - b \in \mathbb{Q}$	22
1.3.2 The case of $a - b \in \mathbb{R} \setminus \mathbb{Q}$	27
1.4 The invertibility of the Operator V	30
Appendix 1.A. The proof of Theorem 2	34
Appendix 1.B. The proof of Lemma 1	38
Bibliography	40

2	Boundary Control approach to the spectral estimation problem. The case of multiple poles	43
	Abstract	43
	2.1 Introduction	43
	2.2 Dynamical systems	46
	2.3 Controllability	48
	2.4 Operators W and R	48
	2.5 Identification	49
	2.5.1 Determining the order of the systems	51
	2.5.2 Determining eigenvalues	51
	2.5.3 Determining decompositions of vectors b and c in bases made of generalized eigenvectors of M and M^*	54
	2.6 Equivalence of dynamical systems with respect to a transformation of variable	55
	2.7 Connection with the original problem	56
	2.7.1 Matrix M	56
	2.7.2 Dynamical systems and the controllability condition	58
	2.7.3 Kernel of the response operator of system (2.32)	58
	2.7.4 Equivalence of the problem of signal decomposition for signal (2.29) to the identification problem for a dynamical system (2.32).	60
	Appendix 2.A. The proof of Lemma 1	62
	Bibliography	65
3	Optimal quadrature formulae related to solutions of initial boundary value problems	67
	Abstract	67
	3.1 Introduction	67
	3.2 A maximization problem in the case of a parabolic equation	68
	3.2.1 Control by the initial conditions	68

3.2.2	Control on the boundary	71
3.3	Minimax problem in the case of a parabolic equation	72
3.3.1	Control by the initial conditions	72
3.3.2	Control on the boundary	75
3.4	A maximization problem in the case of a hyperbolic equation	76
3.5	An example of finding coefficients for a quadrature formula	78
	Bibliography	81
	General Conclusions	83
	Sampling and interpolation	83
	Frequency estimation	84
	Approximate integration	85
	Bibliography	87

List of Tables

	Page
1.1 Summary of invertibility conditions for different combinations of coefficients	20
3.1 Numerical Example 1	79
3.2 Numerical Example 2	80

Acknowledgements

This dissertation would not have been written without the support of many people. I would like to thank my adviser Prof. Sergei Avdonin for proposed research topics and permanent attention to my work. I am grateful to my graduate committee for their work and cooperation, particularly to Dr. Maxwell for letting me sit in his Functional Analysis class and for grading my papers. My work would not have been possible without financial support from the Department of Mathematics and Statistics, Graduate School, Alaska Volcano Observatory and Seismology Laboratory. I thank Prof. Neal Carothers, Prof. John Rhodes and Mrs. Laura Bender for their useful advice. I thank all my fellow graduate students, especially Victor Mikhailov, Vasil Godabrelidze and Odile Bastille for their help and for being great officemates, and Dmitry Nicolsky for giving me a template for this thesis. I thank all my teachers who contributed to my growth. I thank my family for their help, encouragement and advice. Thanks to my boyfriend Robert for everything. While living in Fairbanks I enjoyed presence of my friends here, including all members of the Fairbanks fencing community.

General Introduction

This thesis presents a collection of papers that have been published, accepted or submitted for publishing. The overall theme of the thesis is the application of methods of Control Theory to problems in Signal Processing and Numerical Integration. The main analytical results are contained in Chapters 1 and 2. Chapter 3 completes the manuscript. We demonstrate how control theoretical ideas can be applied to

- the problem of sampling and interpolation;
- the spectral estimation problem;
- non-standard approximation problems.

Sampling and interpolation

One of the fundamental topics in Signal Processing is Sampling Theory. Sampling theory is concerned with the reconstruction of members of certain classes of functions, usually classes of band-limited functions, from sampled data.

Let $E \subset \mathbb{R}$ be a bounded set. The Paley-Wiener space L_E^2 is the space of entire functions of the form

$$s(\lambda) = \int_E e^{i\lambda t} \phi(t) dt, \quad \phi \in L^2(E),$$

with the $L^2(\mathbb{R})$ norm. If E has two or more disjoint components, a member of L_E^2 is called a multi-band function. A discrete set $\{\lambda_n\}$ is a set of stable sampling for L_E^2 if $f \in L_E^2$ implies $\|f\|_{L^2} \leq K \|f(\lambda_n)\|_{l^2}$ for a constant K independent of f . If for any sequence $\{a_n\} \in l^2$ there exists an $f \in L_E^2$ such that $f(\lambda_n) = a_n$ for all n , then $\{\lambda_n\}$ is said to be a set of interpolation. Sequences that are both sampling and interpolating are non-redundant sampling sequences: if we remove one element from $\{\lambda_n\}$, the resulting sequence is no longer sampling.

In the case of band-limited functions, or $E = [-\sigma, \sigma]$, the simplest sampling and interpolating sequence is given by the Wittaker-Shannon-Kotel'nikov sampling theorem.

Theorem 1. *Let s be a function band-limited to $[-\sigma, \sigma]$:*

$$s(\lambda) = \int_{-\sigma}^{\sigma} e^{i\lambda t} \phi(t) dt, \quad \phi \in L^2(-\sigma, \sigma).$$

Then the function s can be reconstructed from its sampled values at $\lambda_k = k\pi/\sigma$ using the formula

$$s(\lambda) = \sum_{-\infty}^{\infty} s(\lambda_k) \frac{\sin \sigma(\lambda - \lambda_k)}{\sigma(\lambda - \lambda_k)}.$$

This is an example of regular or uniform sampling.

The theory of non-uniform sampling for one interval is also well developed. Necessary and sufficient conditions for a sequence $\{\lambda_n\}$ to have a sampling and interpolating property can be stated on the basis of the results of Pavlov [23]. Book by Avdonin and Ivanov [5] and paper by Hruščev, Nikol'skii, and Pavlov [13] give a complete characterization of such sequences.

In practice spectrum of the signal may have gaps. In this case, applying results for single-band signals gives redundant sampling sequences – sequences which are sampling, but not interpolating for L_E^2 . The question of whether there exists for every finite union $E = I_1 \cup I_2 \cup \dots \cup I_n$ of finite intervals a real sampling and interpolating sequence does not have a complete answer. It is known that there exist such complex sequences lying in horizontal strips. Works on this topic include [17; 12; 8; 9; 10], which consider cases of intervals and gaps between intervals having commensurable lengths. In paper [27] Seip constructs at least one real sampling and interpolating sequence for an arbitrary union of two intervals.

There are several papers that have related construction of sampling and interpolating sequences to the invertibility of certain convolution operators. Katsnelson [16] connected a sampling and interpolating property to invertibility of a certain convolution operator, and proved its invertibility in some cases, including the case when E is a union of 2 intervals $[a_1, b_1]$, $[a_2, b_2]$ for which the gap $a_2 - b_1$ is smaller than the minimum of the lengths of two intervals. Lubaraskii and Spitkovsky [19] also construct a convolution operator and prove existence of a sampling and interpolating

sequence in a strip $\{z : |Im(z)| \leq B\}$ for any finite union of intervals. Lubarskii and Seip [18] prove that there exists a sampling and interpolating sequence of real numbers for the case of a finite union of interval of equal length; this work is based on the results of Kohlenberg [17].

In Chapter 1 we consider a problem of construction of sampling and interpolating sequences for a class of two-band signals. We construct sampling and interpolating sequences in the Paley-Wiener space using control theoretic ideas. To solve this problem we use a connection with a problem of construction of a controllable dynamical system with control supported on a union of two intervals. The original problem is reduced to invertibility of the new system's control operator. Our approach can be extended to other classes of multi-band signals: signals with spectrum supported on a union of n intervals with $n > 2$, where lengths of intervals and gaps are arbitrary. The results of this chapter are published in Avdonin, Bulanova, and Moran [2].

Frequency estimation

Another important problem in signal processing is known as frequency estimation problem. Let a signal $r(t)$ be modeled by

$$r(t) = \sum_{n=1}^K a_n(t) e^{\lambda_n t},$$

where $a_n(t)$ are polynomials and λ_n can be real or complex numbers. We need to recover the number of poles K , the polynomial amplitudes $\{a_n(t)\}$ and the exponents $\{\lambda_n\}$ knowing the observations of the signal at discrete moments of time $r(0), r(1), \dots$. The classical spectral estimation problem is to recover the coefficients a_i, λ_i of a signal $r(t) = \sum_{i=1}^N a_i e^{\lambda_i t}$ with constant amplitudes a_i , by the given observations $r(j)$, $j = 0, \dots$. This problem is very important in signal processing, there are applications in wireless communications, antenna array design, bio-medical imaging, high-speed circuit analysis and others (see [14; 25]).

There are many methods of solving spectral estimation problems. The first one developed is the method of Prony [11; 20]. This method was developed by Baron

Gaspard Riche de Prony in 1795. It reduces the frequency estimation problem to one of finding solutions of a polynomial equation. The Matrix Pencil method was developed by Hua and Sarkar in late 1980-s [15; 14; 25]. In the Matrix Pencil method the exponents λ_n are found by solving a generalized eigenvalue problem with matrices constructed from observations of the signal. There are also iterative maximum likelihood methods (see, for example, [21]); MUSIC (Multiple Signal Classification) [26], ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [24], and many others.

In Chapter 2 a new approach based on the Boundary Control method is introduced. The Boundary Control method has been developed for solving dynamical and spectral inverse problems for partial differential equations, and is based on connections between controllability and identification problems. We reduce the problem of estimating frequencies and amplitudes of the signal to an identification problem for a discrete time linear dynamical system, which can be solved using the BC method. The results of this chapter are submitted for publication in Avdonin, Bulanova, and Nicolsky [3] and Avdonin and Bulanova [1].

Approximate integration

In the last part we study an approximate integration problem for solutions of initial boundary value problems. An integral is approximated by a linear combination of the values of the integrand:

$$\int_{\Omega} y(x) dx \approx \sum_{k=1}^N c_k y(x_k), \quad x_k \in \Omega \subset \mathbb{R}^n, c_k \in \mathbb{R}.$$

Formulas of this type are usually called quadrature or cubature (when $n > 1$) formulas. Optimal quadrature formulas are quadrature formulas that are the best in some sense for a given class of functions. Usually formulas that minimize the error are considered. Let the “error” functional have the form:

$$E(y, c_k, x_k) = \left| \int_{\Omega} y(x) dx - \sum_{k=1}^N c_k y(x_k) \right|.$$

If, for a given class of functions Y , and for some $\{c_k^*\}, \{x_k^*\}$,

$$\sup_{y \in Y} E(y, c_k^*, x_k^*) = \inf_{\{c_k, x_k\}} \sup_{y \in Y} E(y, c_k, x_k),$$

then $\sum_{k=1}^N c_k^* y(x_k^*)$ is called an optimal quadrature formula for the class Y , and $\sup_{y \in Y} E(y, c_k^*, x_k^*)$ is an optimal quadrature error.

We consider parts of this problem that consist of finding $\sup_{y \in Y} E(y, c_k, x_k)$ for fixed $\{c_k\}, \{x_k\}$ and $\min_{\{c_k\}} \sup_{y \in Y} E(y, c_k, x_k)$ with fixed $\{x_k\}$, where Y is a class of solutions of a parabolic initial boundary value problem with nonzero boundary or initial condition. In this situation an optimal quadrature problem naturally becomes a problem of optimal control governed by a partial differential equation. The results of this chapter are published in Avdonin, Bulanova, and Ovsyannikov [4].

The optimal quadrature problem is a classical problem in approximate integration theory. It is covered in extensive literature and numerous papers. However, there are no results concerning the problem we are considering in this thesis.

Statement of contributions

Chapter 1 is a continuation of joint research by Avdonin and Moran (see [6]). In paper [6] Avdonin and Moran derived the convolution operator W (see formula (1.11)), invertibility of which is equivalent to sampling and interpolating property of a corresponding real sequence. My advisor Prof. S. Avdonin stated the goal of proving invertibility of W for small enough values of parameter μ , by reducing the problem to a problem of invertibility of a simpler operator. Introduction was written by S. Avdonin and W. Moran, and later edited by me and S. Avdonin. The results by Avdonin and Moran are stated in the introduction without proofs. The rest of the results and proofs in this chapter are obtained by me. Prof. S. Avdonin pointed out possible ways of proving Theorem 5 in Section 1.3.2 (invertibility conditions for operator K in irrational case).

In Chapter 2 we present a control theoretic approach to the spectral estimation problem. Prof. S. Avdonin suggested that the Boundary Control method is appli-

cable to the spectral estimation problem for signals modeled by sums of complex exponentials with polynomial coefficients, and demonstrated the scheme of such application for the case of constant coefficients. I have developed his idea by proving all the necessary facts from realization theory, and extended it to the polynomial case. I have performed all the research and writing in Chapter 2.

Chapter 3 extends joint work of S. Avdonin and D. Ovsyannikov [7; 22]. The original results by Avdonin and Ovsyannikov are presented in sections 3.2.1, 3.4, and the first part of section 3.3.1. The additional results obtained by me are in section 3.3.1 starting with the subheading “A wider class of sets U ”, sections 3.2.2, 3.3.2, 3.5; these include a more general class of initial conditions for the minimax problem in section 3.3.1, maximization and minimax problems for initial boundary value problem with nonzero boundary condition, and a numerical example. I was responsible for writing, formatting and editing of this chapter.

The main results of the thesis were presented at Joint Mathematics Meetings, Washington, DC, January 5-8, 2009; Joint Mathematics Meetings, San Diego, January 6-9, 2008; Joint Mathematics Meetings, San Antonio, January 12-15, 2006; Colloquium, Department of Mathematical Sciences, University of Alaska, Fairbanks, April 7, 2005; Colloquium, Department of Mathematical Sciences, University of Alaska, Fairbanks, Spring, 2004, and are published or submitted for publishing in [1; 2; 3; 4].

Bibliography

- [1] S. A. Avdonin and A. S. Bulanova, *Boundary control approach to the spectral estimation problem. The case of multiple poles*, submitted, 2007.
- [2] S. A. Avdonin, A. S. Bulanova, and W. Moran, *Construction of sampling and interpolating sequences for multi-band signals. The two-band case*, International Journal of Applied Mathematics and Computer Science **17** (2007), no. 2, 143–156.
- [3] S. A. Avdonin, A. S. Bulanova, and D. Nicolsky, *Boundary control approach to the spectral estimation problem. The case of simple poles*, Sampling Theory in Signal and Image Processing (2009), accepted.
- [4] S. A. Avdonin, A. S. Bulanova, and D. A. Ovsyannikov, *Optimal cubature formulae related to solutions of initial boundary value problems*, Vestnik St. Petersburg University. Series 10. Applied Mathematics, Mechanics, Control Processes (2008), no. 2.
- [5] S. A. Avdonin and S. A. Ivanov, *Families of exponentials. the method of moments in controllability problems for distributed parameter systems*, Cambridge University Press, New York, 1995.
- [6] S. A. Avdonin and W. Moran, *Sampling and interpolation of functions with multi-band spectra and controllability problems*, Optimal Control of Partial Differential Equations (K.-H. Hoffmann, G. Leugering, and Tröltzsch F., eds.), vol. 133, Birkhäuser, Basel, 1999, Internat. Ser. Numer. Math., pp. 43–51.
- [7] S. A. Avdonin and D. A. Ovsyannikov, *An approach to the construction of optimal cubature formulas*, Partial Differential Equations (1988), 153–158, (in Russian).
- [8] M. G. Beaty and M. M. Dodson, *Derivative sampling for multiband signals*, Numer. Funct. Anal. Optim. **10** (1989), 875–898.

- [9] ———, *The distribution of sampling rates for signals with equally wide, equally spaced spectral bands*, SIAM J. Appl. Math. **53** (1993), 893–906.
- [10] L. Bezuglaya and V. Katsnelson, *The sampling theorem for functions with limited multi-band spectrum, I.*, Z. Anal. Anwendungen **12** (1993), 511–534.
- [11] Baron G. R. de Prony, *Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l’alkool, à différentes températures*, Journal de l’École Polytechnique **1** (1795), no. 22, 24–76.
- [12] M. M. Dodson and A. M. Silva, *An algorithm for optimal regular sampling*, Signal Process. **17** (1989), 169–174.
- [13] S. V. Hruščev, N. K. Nikol’skii, and B. S. Pavlov, *Unconditional bases of exponentials and reproducing kernels*, Complex Analysis and Spectral Theory, Lecture Notes Math. **864** (1981), 214–335.
- [14] Y. Hua, A. B. Gershman, and Q. Cheng (eds.), *High-resolution and robust signal processing*, Marcel Dekker, New York, Basel, 2004.
- [15] Y. Hua and T. K. Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, IEEE transactions of acoustics, speech, and signal processing **38** (1990), no. 5, 814–824.
- [16] V. E. Katsnelson, *Sampling and interpolation for functions with multi-band spectrum: the mean-periodic continuation method*, Wiener-Symposium (Grossbothen, 1994) Synerg. Syntropie Nichtlineare Syst. (Leipzig), vol. 4, Verlag Wiss. Leipzig., 1996, pp. 91–132.
- [17] A. Kohlenberg, *Exact interpolation of band-limited functions*, J. Appl. Phys. **24** (1953), 1432–1436.

- [18] Yu. Lyubarskii and K. Seip, *Sampling and interpolating sequences for multiband-limited functions and exponential bases on disconnected sets*, J. Fourier Analysis Appl. **3** (1997), 597–615.
- [19] Yu. Lyubarskii and I. Spitkovsky, *Sampling and interpolation for a lacunary spectrum*, Proc. Royal. Soc. Edinburgh, vol. 126 A, 1996, pp. 77–87.
- [20] S. L. Marple, *Digital spectral analysis with applications*, Prentice-Hall, 1987.
- [21] V Nagesha and S Kay, *On frequency estimation with the IQML algorithm*, IEEE Trans. Signal Processing **42** (1994), no. 9, 2509–2513.
- [22] D. A. Ovsyannikov and S. A. Avdonin, *On construction of optimal cubature formulae*, in: Mathematical methods for the control of beams (Leningrad), Leningrad. Univ., 1980, (in Russian), pp. 281–288.
- [23] B. S. Pavlov, *Basicity of an exponential system and muckenhoupt's condition*, Soviet Math. Dokl. **20** (1979), 655–659.
- [24] R. Roy, A. Paulraj, and T. Kailath, *Multiple emitter location and signal parameter estimation*, IEEE Trans. Acoust., Speech, Signal Process. **34** (1986), no. 5, 1340–1342.
- [25] T. K. Sarkar, M. C. Wicks, M. Salazar-Palma, and R. J. Bonneau, *Smart antennas*, John Wiley & Sons, Hoboken, New Jersey, 2003.
- [26] R. O. Schmidt, *Multiple emitter location and signal parameter estimation*, IEEE Trans. Antennas Propag. **34** (1986), no. 3, 276–280.
- [27] K. Seip, *A simple construction of exponential bases in L^2 of the union of several intervals*, Proc. Edinburgh Math. Soc. **38** (1995), 171–177.

Chapter 1

Construction of sampling and interpolating sequences for multi-band signals. The two-band case¹

Abstract

Recently several papers have related the production of sampling and interpolating sequences for multi-band signals to the solution of certain kinds of Wiener-Hopf equations. Our approach is based on connections between exponential Riesz bases and the controllability of distributed parameter systems. For the case of two-band signals we derive an operator whose invertibility is equivalent to the existence of a sampling and interpolating sequence, and prove the invertibility of this operator.

Keywords: sampling and interpolation, multi-band signals, Riesz bases, families of exponentials, Wiener-Hopf equations, control, observation

1.1 Introduction

Let E be a finite union of disjoint intervals:

$$E = \bigcup_{j=1}^N I_j, \quad I_j = [a_j, b_j], \quad 0 = a_1 < b_1 < a_2 < b_2 < \dots < a_N < b_N.$$

Several papers [Avdonin and Moran, 1999; Bezuglaya and Katsnelson, 1993; Katsnelson, 1996; Lyubarskii and Seip, 1997; Lyubarskii and Spitkovsky, 1996; Moran and Avdonin, 1999; Seip, 1995] have recently appeared that discuss Riesz bases of exponentials in $L^2(E)$. All of them emphasize the importance of this problem in communication theory: if $\{e^{i\lambda_k t}\}$ forms a Riesz basis in $L^2(E)$ then $\Lambda = \{\lambda_k\}$ is a sampling and interpolating set for corresponding multi-band signals. In other words, the interpolation problem

$$s(\lambda_k) = \alpha_k, \quad \lambda_k \in \Lambda, \quad s \in L_E^2,$$

¹S.A. Avdonin, A.S. Bulanova, and W. Moran, *Construction of sampling and interpolating sequences for multi-band signals. The two-band case*, International Journal of Applied Mathematics and Computer Science, vol. 17, (2007), no. 2, 143-156.

has a unique solution for each $\{\alpha_k\} \in l^2$. Here L_E^2 is the space of entire functions of the form

$$s(\lambda) = \int_E e^{i\lambda t} \phi(t) dt, \quad \phi \in L^2(E),$$

endowed with the $L^2(\mathbb{R})$ norm. The equivalence of these two problems is well known; it follows from standard duality arguments [see, for example, Hruščev *et al.*, 1981; Lyubarskii and Seip, 1997].

It is interesting to note that papers [Avdonin and Moran, 1999; Katsnelson, 1996; Lyubarskii and Spitkovsky, 1996] have related the production of Riesz bases to the invertibility of certain convolution integral operators. The method of Katsnelson [1996] is based on the mean periodic continuation of a function with respect to a finite measure. The convolution operator in [Lyubarskii and Spitkovsky, 1996] is constructed on a union of intervals connected with the entire function generating the set Λ .

Another approach to the problem was proposed in paper [Avdonin and Moran, 1999]. It is based on connections between the controllability of a dynamical system described by a linear PDE and the Riesz basis property of a corresponding exponential family. These connections are well known and widely exploited in control theory; see, for example, an excellent survey paper [Russell, 1978] and the book [Avdonin and Ivanov, 1995]. The problem of constructing an exponential basis on several intervals gives rise to a new type of control problem with boundary control supported on these intervals of time.

More precisely, Avdonin and Moran [1999] introduced an auxiliary dynamical system described by the string equation with boundary control u :

$$\rho^2(x)y_{tt}(x, t) = y_{xx}(x, t), \quad y(0, t) = u(t), \quad y_x(l, t) = 0, \quad 0 < x < l, \quad t \in \mathbb{R}, \quad (1.1)$$

where $\rho(x)$ is a positive function on $[0, l]$ which will be determined later. Usually in control theory the function u is taken from $L^2(0, T)$ for some positive T , but for our purposes we take u from $L_{loc}^2(\mathbb{R})$ with support restricted to E and consider the initial

conditions

$$y(x, a_1) = y_0(x), \quad y_t(x, a_1) = y_1(x). \quad (1.2)$$

Eigen-frequencies $\lambda_n, n \in \mathbb{N}$, of this system can be found from the boundary value problem

$$\phi''(x) + \lambda^2 \rho^2(x) \phi(x) = 0, \quad 0 < x < l, \quad \phi(0) = \phi'(l) = 0. \quad (1.3)$$

System (1.1) is called *exactly controllable* if for any initial conditions $(y_0, y_1) \in L^2(0, l) \times H^{-1}(0, l)$ there is a unique control $u \in L^2(E)$ which brings the system to the origin at $t = b_N$:

$$y(\cdot, b_N) = y_t(\cdot, b_N) = 0.$$

Here $H^{-1}(0, l)$ is the space dual to $H_1(0, l) := \{\psi \in H^1(0, l) : \psi(0) = 0\}$.

The following statement plays a key role in this approach to construction of sampling and interpolating sequences.

Theorem 1. [Avdonin and Moran, 1999]. *System (1.1) is exactly controllable if and only if the family $\{e^{\pm i\lambda_n t}\}$ forms a Riesz basis in $L^2(E)$.*

In other words, the exact controllability of (1.1) is equivalent to the fact that $\Lambda = \{\pm\lambda_n\}$ is a sampling and interpolating sequence for L^2_E . Note that all λ_n^2 — eigenvalues of boundary value problem (1.3) — are positive and we may therefore choose λ_n to be positive.

Our problem then becomes that of constructing the function $\rho(x)$ in such a way that system (1.1) is exactly controllable. If the set E consists only of the interval $[a, b]$ and control u acts from $t = a$ to $t = b$ then, as is well known [see, e.g., Russell, 1978; Avdonin and Ivanov, 1995], system (1.1) is exactly controllable if and only if the length of the interval is equal to two optical lengths of the string:

$$b - a = 2 \int_0^l \rho(x) dx.$$

Choosing $\rho = \text{const}$ (homogeneous string), we obtain the uniform sampling and interpolating sequence for $L^2_{[a,b]}$:

$$\Lambda = \pm \frac{2\pi}{b-a} \left(n - \frac{1}{2} \right), \quad n \in \mathbb{N}. \quad (1.4)$$

Taking ρ as a smooth (from $C^2[a, b]$) non-constant function, we obtain a non-uniform sampling and interpolating sequence asymptotically close to (1.4).

In the multi-band case, we cannot (in a general situation) produce a sampling and interpolating sequence taking ρ as a constant or smooth function from $C^2[a, b]$. This fact can be understood by taking into account a necessary “geometric” condition of controllability of system (1.1): *if system (1.1) is exactly controllable, then for every $x_0 \in [0, l]$ both characteristics starting at the point $x = x_0, t = 0$ and lying in the strip $[0, l] \times \{t \geq 0\}$ of (x, t) -plane have nonempty intersection with $\{x = 0\} \times E$.* We suppose that the characteristics “reflect” from the boundaries subjecting to the geometric optics laws.

For example, if $E = [0, 1] \cup [2, 3]$, none of the smooth functions ρ satisfies the “geometric” condition. Using Theorem 1 it can also be proved that uniform sampling and interpolation of multi-band signals is possible only when very special relations exist between lengths of intervals and gaps between them. More precisely, the special case is when E is an *explosion* of an interval [Higgins, 1996, Sec. 13.1].

To satisfy the “geometric” controllability condition in the multi-band case, we should consider piecewise smooth functions ρ . More exactly, we take points $0 = x_0 < x_1 < \dots < x_N = l$ and a piecewise constant function $\rho(x)$ such that

$$\rho(x) = \rho_j, \quad \text{for } x_{j-1} < x < x_j; \quad 0 < \rho_j < \infty, \quad \rho_j \neq \rho_{j+1}, \quad (1.5)$$

$$\rho_j(x_j - x_{j-1}) = (b_j - a_j)/2, \quad j = 1, 2, \dots, N. \quad (1.6)$$

Due to the condition $\rho_j \neq \rho_{j+1}$ there are additional reflections of the waves from the boundaries $x = x_j$ of the “layers” which improve controllability of system (1.1). Notice that additional compatibility conditions are required for systems (1.1), (1.3) at points $x_i, i = 1, 2, \dots, N$ of discontinuity of function $\rho(x)$ [see Avdonin and Moran, 1999; Avdonin and Ivanov, 2008]. For system (1.1) these conditions are:

$$y(x_i - 0, t) = y(x_i + 0, t), \quad y_x(x_i - 0, t) = y_x(x_i + 0, t), \quad i = 1, 2, \dots, N; \quad (1.7)$$

for system (1.3):

$$\phi(x_i - 0) = \phi(x_i + 0), \quad \phi_x(x_i - 0) = \phi_x(x_i + 0), \quad i = 1, 2, \dots, N. \quad (1.8)$$

Analysis of the obtained control problem leads us to the following conjecture.

Conjecture 1. *Let E be a multi-band set described above. Then, for all functions $\rho(x)$ satisfying (1.5), (1.6), system (1.1) is exactly controllable.*

This conjecture was confirmed in some particular cases in [Avdonin and Moran, 1999], and we are working on its complete proof using PDE techniques.

Conjecture 1 implies that the exponential family $\{e^{\pm i\lambda_n t}\}_{n \in \mathbb{N}}$ forms a Riesz basis in $L^2(E)$ where λ_n^2 are the eigenvalues of boundary value problem (1.3) and $\rho(x)$ satisfies conditions (1.5), (1.6). It is important for applications that the sampling and interpolating set $\{\pm\lambda_n\}$ is real.

Boundary value problem (1.3), (1.5), (1.6) represents an important example of an eigenvalue problem whose spectrum generates a sampling and interpolation sequence for a multi-band signal.

In Avdonin and Moran [1999] the sampling and interpolation problem is reduced to the solution of linear functional equations, specifically, Wiener–Hopf equations of a special form. The solution of problem (1.1), (1.2) with $\rho(x)$ satisfying conditions (1.5), (1.6) can be written in an explicit although rather complicated form. Analysis of that formula leads to invertibility problems for operators connected with linear functional equations. While this method appears to extend to handle arbitrary finite unions of intervals, we illustrate it in the case of two intervals.

Only a few results concerning sampling and interpolating sequences for the case when the set E is a union of two intervals are known. Kohlenberg [1953] constructed a sampling and interpolating sequence for signals whose spectrum is restricted to the union of two intervals of the same length (*band-pass signals*). The later great impact to this field was due to Dodson and Silva [1989] and Beaty and Dodson [1989, 1993] and due to Bezuglaya and Katsnelson [1993]. In these papers the lengths of the intervals and the gaps were supposed to have special structure such as commensurability of the lengths of the intervals and the gaps. Lyubarskii and Seip [1997] remark that the method of Kohlenberg [1953] can be extended to the case when the intervals comprising E have commensurable lengths. The results of Seip [1995] are free of

arithmetic restrictions on lengths of intervals comprising the set E ; in particular, starting from the “1/4 in the mean” theorem [Avdonin, 1979] he gives a construction of at least one real sampling and interpolating sequence for an arbitrary E consisting of two intervals.

The main results

This chapter is devoted to the investigation of the convolution operator proposed in Avdonin and Moran [1999] for the cases of E being a union of two arbitrary intervals. We prove that this operator is invertible if a parameter $\mu = (\rho_2 - \rho_1)/(\rho_2 + \rho_1)$ is small enough. This is a new result in theory of linear functional equations and convolution operators. It proves existence of infinitely many real sampling and interpolating sequences for signals with the spectrum supported on two arbitrary intervals. We also give an algorithm for construction of such sequences. The former are results in sampling and interpolation theory. Also, the result on controllability of the corresponding dynamical system (1.1) follows from the invertibility of the convolution operator.

1.2 The Operators W , V and K

Let

$$E = I_1 \cup I_2, \quad I_j = [a_j, b_j], \quad |I_j| := b_j - a_j = \alpha_j, \quad j = 1, 2, \quad (1.9)$$

$$\alpha_1 + \alpha_2 = \alpha, \quad a_2 - b_1 = \alpha'. \quad (1.10)$$

Note that, without loss of generality, we can assume that α_1 is less than or equal to α_2 .

In Avdonin and Moran [1999] it was proved that the problem of construction of sampling and interpolating sequence for L^2_E can be reduced to study of the invertibility of the operator

$$W : L^2(0, \alpha_1) \mapsto L^2(\alpha', \alpha' + \alpha_1).$$

$$(Wf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) \sum_{r=0}^{\infty} \sum_{k=0}^{\infty} \sum_{q=0}^k A(r, k, q) f(t - w(r, k, q)), \quad (1.11)$$

where $w(r, k, q) = \alpha r + \alpha_1(k - q) + \alpha_2 q$,

$$A(r, k, q) = (-1)^{r+k} \mu^k \frac{(r+k)!}{r!q!(k-q)!}, \quad \mu \in (-1, 0) \cup (0, 1),$$

$$\chi_{[a,b]}(t) = \begin{cases} 1, & \text{if } t \in [a, b], \\ 0, & \text{otherwise.} \end{cases}$$

Here $\mu = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}$, where ρ_1, ρ_2 are values of a piecewise constant density function of an associated string equation (1.1) satisfying controllability conditions (1.5), (1.6).

Once we find a parameter μ for which the operator W is invertible, a sampling and interpolating sequence for signals with the spectrum supported on E can be found using the following scheme.

Algorithm 1. (a) Pick any two different values $\rho_1 > 0, \rho_2 > 0$ such that $\mu = (\rho_2 - \rho_1)/(\rho_2 + \rho_1)$.

(b) Find the numbers l and x_1 from the equations

$$\begin{aligned} \rho_1 x_1 &= \frac{\alpha_1}{2}, \\ \rho_2(l - x_1) &= \frac{\alpha_2}{2}. \end{aligned}$$

(c) Define the density function

$$\rho(x) = \begin{cases} \rho_1, & \text{when } 0 < x < x_1, \\ \rho_2, & \text{when } x_1 < x < l. \end{cases}$$

(d) Find the eigenvalues λ_n^2 of boundary value system (1.3), (1.8) with the function $\rho(x)$ and the number l found on steps (b) and (c). The sequence $\Lambda = \{\pm\lambda_n\}$ is a sampling and interpolating sequence for L_E^2 .

In formula (1.11) and in what follows it is convenient to assume that f is defined on the real axis with support in $[0, \alpha_1]$. One can see that in this case for each $t \in [\alpha', \alpha' + \alpha_1]$ the number of terms in the sum is finite, since only the terms with $t - w(r, k, q) \in [0, \alpha_1]$ are not equal to zero.

Our goal is to reduce the problem of the invertibility of the operator W to a problem of the invertibility of a simpler operator. We are going to break the sum corresponding to the operator W into two sums, $W = U + \tilde{U}$, so that the operator U is invertible and its invertibility implies the invertibility of W . We show that it is possible to make U contain no more than four terms. The invertibility of the operator U is proven in Theorems 7, 8.

Theorem 2. *For any a_j and b_j , there exists a nonnegative integer number k such that the operator W can be written in the form*

$$W = U + \tilde{U},$$

where the operator U is comprised of at most four terms whose coefficients each involve the parameter μ to a power not exceeding $k+1$, and all terms of the operator \tilde{U} contain a factor μ at a power at least equal to $k+2$. The operator U has the following form:

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [c_1 f(t - w_1) + c_2 f(t - w_1 - \alpha_1) \\ + c_3 f(t - w_2 + \alpha_1) + c_4 f(t - w_2)],$$

where one or more of the coefficients c_i may be zero, and w_1 and w_2 have the form $w(\bar{r}, \bar{k}, \bar{q})$ with \bar{r} , \bar{k} , \bar{q} depending on relative position and lengths of the intervals.

Note that the operator U may contain 2, 3 or 4 terms depending on the locations and the lengths of the intervals. There are many cases and sub-cases of the position of the intervals, so the proof is postponed to Appendix 1.A. Exact formulas for the operator U , which are important for the proof of the invertibility of U and W , are derived in the process of the proof.

We prove that for sufficiently small μ the invertibility of the operator U implies the invertibility of the operator W . This statement is based on the following lemma which is proved in Appendix 1.B.

Lemma 1. *If the operator U is invertible then $\|U^{-1}\| \leq |\mu|^{-(k+1)}C$ for small enough $|\mu|$, where $C > 0$ does not depend on μ .*

Theorem 3. *If the operator U is invertible, then for small enough μ , the operator W is also invertible.*

Proof. Theorem 2 states that the operator W can be represented as a sum of two other operators $W = U + \tilde{U}$. The operator U is made up of no more than 4 terms of W with powers smaller or equal to $k + 1$, and the operator \tilde{U} contains the rest of the terms of W .

We have noticed that the operator W has a finite number of terms. Therefore \tilde{U} also has a finite number of terms. Since \tilde{U} contains only powers of μ higher than $k + 1$, then for small enough μ ,

$$\|\tilde{U}\| \leq |\mu|^{k+2} D,$$

where D does not depend on μ .

Then from Lemma 1 it follows that for small enough μ ,

$$\|U^{-1}\| \|\tilde{U}\| < 1.$$

Note that

$$W = U + \tilde{U} = U(I + U^{-1}\tilde{U}).$$

Thus for small enough μ the operator W is invertible □

It is convenient to scale so that $\alpha_1 = 1$. After a change of variable the operator U is reduced to the operator V in $L^2(0, 1)$:

$$(Vf)(t) = \chi_{[0,1]}(t) [c_1 f(t+a) + c_2 f(t+a-1) + c_3 f(t+b) + c_4 f(t+b-1)]. \quad (1.12)$$

Here

$$0 \leq b \leq a \leq 1$$

and c_i are the corresponding coefficients $A(r, k, q)$ or 0.

To prove that the operator V is invertible, we introduce a new operator K which has the same form as the operator V , but coefficients c_i are arbitrary real numbers.

We consider two cases: the case when $a - b$ is a rational number and the case of irrational $a - b$. The case of $a - b \in \mathbb{Q}$ corresponds to the situation of $\alpha_1/\alpha_2 \in \mathbb{Q}$ and the irrational case occurs if $\alpha_1/\alpha_2 \in \mathbb{R} \setminus \mathbb{Q}$, where α_1 and α_2 are the lengths of the intervals I_1, I_2 as in (1.9). First we find the invertibility condition for the operator K , and then we show that the coefficients c_i of the operator V satisfy this condition.

1.3 The invertibility of the Operator K .

Consider the operator K in $L^2[0, 1]$:

$$(Kf)(t) = [c_1f(t+a) + c_2f(t+a-1) + c_3f(t+b) + c_4f(t+b-1)], \quad (1.13)$$

where $t \in [0, 1]$; $a, b \in [0, 1]$; $b \leq a$; $c_1 \neq 0$ or $c_4 \neq 0$. Our goal is a sufficient condition for the invertibility of K . We do not consider the case of $c_1 = c_4 = 0$: the invertibility conditions for K in this case are different from the invertibility conditions in all other cases, and we do not need the case of $c_1 = c_4 = 0$ to prove the invertibility of the operator V .

From (1.13) one can easily see that the invertibility of the operator K is equivalent to solvability for f of the following system of equations:

$$\begin{aligned} c_1f(t+a) + c_3f(t+b) &= g(t), & t \in [0, 1-a], \\ c_3f(t+b) + c_2f(t+a-1) &= g(t), & t \in (1-a, 1-b), \\ c_2f(t+a-1) + c_4f(t+b-1) &= g(t), & t \in (1-b, 1], \end{aligned} \quad (1.14)$$

where $g(t)$ is in $L^2[0, 1]$.

Let us find the conditions for the invertibility of the operator K in special cases: $c_1 = c_3 = 0$, $c_2 = c_3 = 0$, $c_2 = c_4 = 0$. The following lemma is a particular case of Theorems 4 and 5 which are proved in subsections 1.3.1, 1.3.2 respectively. We formulate it as a separate lemma because its proof is different from the proofs of the theorems.

Lemma 2. *Suppose that $c_1 = c_3 = 0$, or $c_2 = c_3 = 0$, or $c_2 = c_4 = 0$. If $a - b$ is a rational number, The operator K is invertible in $L^2[0, 1]$ if and only if*

$(-1)^n c_2^{k_1+m} c_1^{n-m-k_1} \neq c_3^{n-k_1} c_4^{k_1}$ and $(-1)^n c_2^{k_2+m} c_1^{n-m-k_2} \neq c_3^{n-k_2} c_4^{k_2}$, where $a - b = \frac{m}{n}$ ($\frac{m}{n}$ is an irreducible fraction), k_1 is the integer part of bn , and k_2 is the smallest integer such that $k_2 \geq bn$. If $a - b$ is an irrational number, the operator K is invertible if and only if $|c_3|^{1-b}|c_4|^b \neq |c_2|^a|c_1|^{1-a}$.

Proof. Assume that $c_1 = c_3 = 0$. If $a \neq 1$, the first equation of the system (1.14) becomes

$$0 = g(t), \quad t \in [0, 1 - a).$$

So, the operator K is not invertible.

If $a = 1$, then the first equation of (1.14) is defined on an interval of length zero. We get a system of two equations; this system is solvable when $c_2 \neq 0$ if $b \neq 1$, and when $c_2 + c_4 \neq 0$ if $b = 1$. We can find the invertibility conditions for the cases of $c_2 = c_3 = 0$ and $c_2 = c_4 = 0$ using the same reasoning. We summarize all the cases in Table 1.1.

Table 1.1: Summary of invertibility conditions for different combinations of coefficients

Coefficients	Invertibility condition
$c_1 = c_3 = 0 \quad a \neq 1$	not invertible
$c_1 = c_3 = 0 \quad a = 1, b \neq 1$	$c_2 \neq 0$
$c_1 = c_3 = 0 \quad a = b = 1$	$c_2 + c_4 \neq 0$
$c_2 = c_3 = 0 \quad a \neq b$	not invertible
$c_2 = c_3 = 0 \quad a = b, a \neq 1, a \neq 0$	$c_1 \neq 0$ and $c_4 \neq 0$
$c_2 = c_3 = 0 \quad a = b = 0$	$c_1 \neq 0$
$c_2 = c_3 = 0 \quad a = b = 1$	$c_4 \neq 0$
$c_2 = c_4 = 0 \quad b \neq 0$	not invertible
$c_2 = c_4 = 0 \quad b = 0, a \neq 0$	$c_3 \neq 0$
$c_2 = c_4 = 0 \quad a = b = 0$	$c_1 + c_3 \neq 0$

All the cases in the table 1.1 can be generalized by the two conditions given in the statement of this lemma. Observing this table we can see that under conditions

of this lemma operator K may be invertible only if $b = 0$, $a = 1$, or $a = b$. In the rational case these conditions correspond to $k_1 = k_2 = n - m$, $k_1 = k_2 = 0$, and $n = 1$, $m = 0$, $0 \leq k_1 \leq k_2 \leq 1$, so that zero coefficients get raised to the zero power. \square

In what follows we will need to divide the equations of system (1.14) by c_1 and c_2 or c_3 and c_4 . By proving Lemma 2, we exclude the cases when $c_1 = c_3 = 0$, $c_2 = c_4 = 0$ from consideration. Thus we can assume that ($c_1 \neq 0$ and $c_2 \neq 0$) or ($c_3 \neq 0$ and $c_4 \neq 0$).

Let $\Delta = a - b$, $\bar{\Delta} = 1 - \Delta$.

If $c_3 \neq 0$ and $c_4 \neq 0$, then system (1.14) is equivalent to

$$\begin{aligned} f(t) + \frac{c_1}{c_3} f(t + \Delta) &= \frac{1}{c_3} g(t - b), & t \in [b, \bar{\Delta}), \\ f(t) + \frac{c_2}{c_3} f(t + \Delta - 1) &= \frac{1}{c_3} g(t - b), & t \in (\bar{\Delta}, 1), \\ f(t) + \frac{c_2}{c_4} f(t + \Delta) &= \frac{1}{c_4} g(t + 1 - b), & t \in (0, b], \end{aligned}$$

which is equivalent to the equation

$$f(t) + \phi(t) f((t + \Delta) \bmod 1) = h(t), \quad t \in [0, 1], \quad (1.15)$$

where

$$\phi(t) = \begin{cases} c_2/c_4, & t \in (0, b), \\ c_1/c_3, & t \in (b, \bar{\Delta}), \\ c_2/c_3, & t \in (\bar{\Delta}, 1). \end{cases} \quad (1.16)$$

When $c_1 \neq 0$ and $c_2 \neq 0$, system (1.14) is equivalent to

$$f(t) + \psi(t) f((t + \bar{\Delta}) \bmod 1) = k(t), \quad t \in [0, 1], \quad (1.17)$$

where

$$\psi(t) = \begin{cases} c_3/c_2, & t \in (0, \Delta), \\ c_4/c_2, & t \in (\Delta, a), \\ c_3/c_1, & t \in (a, 1). \end{cases} \quad (1.18)$$

Equations of type (1.15) were investigated in [Antonevich, 1996, Th. 2.1, pp. 29–32] for the case of continuous $\phi(t)$. In the course of proof of Theorems 4, 5 we will obtain solvability conditions for equations (1.15), (1.17) for piecewise continuous $\phi(t)$ and $\psi(t)$ as defined in (1.16), (1.18).

1.3.1 The case of $a - b \in \mathbb{Q}$

Theorem 4. *Let $a - b = \frac{m}{n}$ be an irreducible fraction. The operator K is invertible in $L^2[0, 1]$ if and only if $(-1)^n c_2^{k_1+m} c_1^{n-m-k_1} \neq c_3^{n-k_1} c_4^{k_1}$ and $(-1)^n c_2^{k_2+m} c_1^{n-m-k_2} \neq c_3^{n-k_2} c_4^{k_2}$, where $a - b = \frac{m}{n}$, k_1 is the integer part of bn , and k_2 is the smallest integer such that $k_2 \geq bn$.*

The results equivalent to Theorem 4 were independently obtained by I. Spitkovsky [2006]. Theory of convolution operators in spaces of matrix valued functions can be found in the book [Böttcher *et al.*, 2002].

Proof. From Lemma 2 it follows that this theorem holds for the cases $c_1 = c_3 = 0$, $c_2 = c_3 = 0$, $c_2 = c_4 = 0$. Thus we do not need to consider these cases in the proof of Theorem 4, and we can assume that $(c_1 \neq 0$ and $c_2 \neq 0)$ or $(c_3 \neq 0$ and $c_4 \neq 0)$. We defined the operator K so that $c_1 \neq 0$ or $c_4 \neq 0$.

As we have already noted, when both c_3 and c_4 are not equal to zero, the invertibility of the operator K is equivalent to solvability of equation (1.15). If $c_3 = 0$ or $c_4 = 0$ then $c_1 \neq 0$ and $c_2 \neq 0$, and in this case the invertibility of K is equivalent to solvability of (1.17). In the first case the problem will be reduced to solvability of two algebraic systems with determinants

$$1 + (-1)^{n+1} c_2^{k_2+m} c_1^{n-m-k_2} / (c_3^{n-k_2} c_4^{k_2}) \text{ and } 1 + (-1)^{n+1} c_2^{k_1+m} c_1^{n-m-k_1} / (c_3^{n-k_1} c_4^{k_1}).$$

In the second case the problem reduces to solvability of two systems with determinants

$$1 + (-1)^{n+1} c_3^{n-k_2} c_4^{k_2} / (c_2^{k_2+m} c_1^{n-m-k_2}) \text{ and } 1 + (-1)^{n+1} c_3^{n-k_1} c_4^{k_1} / (c_2^{k_1+m} c_1^{n-m-k_1}).$$

The proofs of the last two facts are analogous, so we will only show the derivation of the first of them.

Suppose that $c_3 \neq 0$ and $c_4 \neq 0$.

Let us rewrite equation (1.15) as a family of equations defined on disjoint subintervals of the interval $[0, 1]$, choosing subintervals so that in each of those equations the function $\phi(t)$ is constant.

First we divide the interval into n pieces of the length $\frac{1}{n}$: $\{(\frac{i-1}{n}, \frac{i}{n})\}_{i=1}^n$. The number of such subintervals that are entirely inside of the interval $[0, b]$ is equal to the integer part of bn . Let us denote this number by k . Let us also introduce d – the length of the interval $[\frac{k}{n}, b]$: $d = b - k\frac{1}{n}$.

Now we divide each subinterval of the length $\frac{1}{n}$ into two smaller subintervals with the lengths of d and $\frac{1}{n} - d$ and consider two sets of subintervals:

$$J_1 = \{(\frac{i-1}{n}, \frac{i-1}{n} + d)\}_{i=1}^n \quad J_2 = \{(\frac{i-1}{n} + d, \frac{i}{n})\}_{i=1}^n.$$

The set J_1 contains all intervals of the length d , and J_2 has all intervals of the length $\frac{1}{n} - d$.

Note that $\phi(t)$ is constant on each of these subintervals ($\phi(t)$ is piecewise constant and it changes its values at points $b = \frac{k}{n} + d$ and $\bar{\Delta} = \frac{n-m}{n}$).

Now we can rewrite equation (1.15) as the following family of equations:

$$\begin{aligned} f(t) + c_2 c_4^{-1} f(t + \Delta) &= h(t) & t \in (0, d) \\ f(t) + c_2 c_4^{-1} f(t + \Delta) &= h(t) & t \in (d, \frac{1}{n}) \\ &\dots & \\ f(t) + c_2 c_4^{-1} f(t + \Delta) &= h(t) & t \in (\frac{k-1}{n} + d, \frac{k}{n}) \\ f(t) + c_2 c_4^{-1} f(t + \Delta) &= h(t) & t \in (\frac{k}{n}, \frac{k}{n} + d) = (\frac{k}{n}, b) \end{aligned} \tag{1.19}$$

$$\begin{aligned}
f(t) + c_1 c_3^{-1} f(t + \Delta) &= h(t) & t \in \left(\frac{k}{n} + d, \frac{k+1}{n}\right) &= \left(b, \frac{k+1}{n}\right) \\
f(t) + c_1 c_3^{-1} f(t + \Delta) &= h(t) & t \in \left(\frac{k+1}{n}, \frac{k+1}{n} + d\right) \\
&\dots \\
f(t) + c_1 c_3^{-1} f(t + \Delta) &= h(t) & t \in \left(\frac{n-m-1}{n}, \frac{n-m-1}{n} + d\right) \\
f(t) + c_1 c_3^{-1} f(t + \Delta) &= h(t) & t \in \left(\frac{n-m-1}{n} + d, \bar{\Delta}\right) \\
f(t) + c_2 c_3^{-1} f(t + \Delta - 1) &= h(t) & t \in \left(\bar{\Delta}, \frac{n-m}{n} + d\right) \\
f(t) + c_2 c_3^{-1} f(t + \Delta - 1) &= h(t) & t \in \left(\frac{n-m}{n} + d, \frac{n-m+1}{n}\right) \\
&\dots \\
f(t) + c_2 c_3^{-1} f(t + \Delta - 1) &= h(t) & t \in \left(\frac{n-1}{n}, \frac{n-1}{n} + d\right) \\
f(t) + c_2 c_3^{-1} f(t + \Delta - 1) &= h(t) & t \in \left(\frac{n-1}{n} + d, 1\right),
\end{aligned}$$

Note that this family has three groups of equations: the first group contains $2k + 1$ equations defined on subintervals of $(0, b)$, and the coefficient of $f(t + \Delta)$ is c_2/c_4 ; the second group contains $2n - 2m - 2k - 1$ equations with the coefficient of $f(t + \Delta)$ equal to c_1/c_3 ; in the third group there are $2m$ equations, and the coefficient is c_2/c_3 .

Let us introduce $f_i, g_i \in L_2(0, \frac{1}{n})$ for $1 \leq i \leq n$:

$$\begin{aligned}
f_i(t) &= f\left(t + \frac{i-1}{n}\right) \\
h_i(t) &= h\left(t + \frac{i-1}{n}\right).
\end{aligned}$$

Substituting $f(t)$ and $h(t)$ by $f_i(t)$ and $h_i(t)$ into each of the equations of family

(1.19), the latter can be transformed into two systems:

$$\left\{ \begin{array}{l} f_1(t) + c_2 c_4^{-1} f_{m+1}(t) = h_1(t) \\ \dots \\ f_{k+1}(t) + c_2 c_4^{-1} f_{k+m+1}(t) = h_{k+1}(t) \\ f_{k+2}(t) + c_1 c_3^{-1} f_{k+m+2}(t) = h_{k+2}(t) \\ \dots \\ f_{n-m}(t) + c_1 c_3^{-1} f_n(t) = h_{n-m}(t) \\ f_{n-m+1}(t) + c_2 c_3^{-1} f_1(t) = h_{n-m+1}(t) \\ \dots \\ f_n(t) + c_2 c_3^{-1} f_m(t) = h_n(t) \end{array} \right. \quad \text{on } t \in (0, d) \quad (1.20)$$

$$\left\{ \begin{array}{l} f_1(t) + c_2 c_4^{-1} f_{m+1}(t) = h_1(t) \\ \dots \\ f_k(t) + c_2 c_4^{-1} f_{k+m}(t) = h_k(t) \\ f_{k+1}(t) + c_1 c_3^{-1} f_{k+m+1}(t) = h_{k+1}(t) \\ \dots \\ f_{n-m}(t) + c_1 c_3^{-1} f_n(t) = h_{n-m}(t) \\ f_{n-m+1}(t) + c_2 c_3^{-1} f_1(t) = h_{n-m+1}(t) \\ \dots \\ f_n(t) + c_2 c_3^{-1} f_m(t) = h_n(t) \end{array} \right. \quad \text{on } t \in (d, \frac{1}{n}) \quad (1.21)$$

Let $x_i(t) = f_{(1+(i-1)m) \bmod n}(t)$.

Since m and n are co-prime, this substitution maps the set $\{f_i(t)\}_{i=1}^n$ into the set $\{x_i(t)\}_{i=1}^n$.

Systems (1.20) and (1.21) take the form:

$$x_i(t) + \psi_i x_{(i+1) \bmod n}(t) = h_i(t) \quad \text{on } t \in (0, d), \quad 1 \leq i \leq n, \quad (1.22)$$

$$x_i(t) + \xi_i x_{(i+1) \bmod n}(t) = h_i(t) \quad \text{on } t \in (d, \frac{1}{n}), \quad 1 \leq i \leq n, \quad (1.23)$$

where

$$\psi_i, \xi_i \in \left\{ \frac{c_2}{c_4}, \frac{c_1}{c_3}, \frac{c_2}{c_3} \right\}, \quad 1 \leq i \leq n.$$

In system (1.22), $\{\psi_i\}_{i=1}^n$ has $k+1$ occurrences of c_2/c_4 , $n-m-k-1$ occurrences of c_1/c_3 , and m occurrences of c_2/c_3 . System (1.23) has k occurrences of c_2/c_4 , $n-m-k$ occurrences of c_1/c_3 , and m occurrences of c_2/c_3 . Therefore, determinant of the first system is equal to

$$\begin{aligned} 1 + (-1)^{n+1} \prod_{i=1}^n \psi_i &= 1 + (-1)^{n+1} \left(\frac{c_2}{c_4} \right)^{k+1} \left(\frac{c_1}{c_3} \right)^{n-m-k-1} \left(\frac{c_2}{c_3} \right)^m = \\ &= 1 + (-1)^{n+1} \frac{c_2^{k+m+1} c_1^{n-m-k-1}}{c_3^{n-k-1} c_4^{k+1}}, \end{aligned} \quad (1.24)$$

determinant of the second system is:

$$1 + (-1)^{n+1} \prod_{i=1}^n \xi_i = 1 + (-1)^{n+1} \frac{c_2^{k+m} c_1^{n-m-k}}{c_3^{n-k} c_4^k}. \quad (1.25)$$

If bn is an integer, $bn = k$, then $d = 0$, and the first system lives on an empty interval. In this case the invertibility of the operator K is equivalent to solvability of system (1.23) with determinant $1 + (-1)^{n+1} c_2^{k+m} c_1^{n-m-k} / (c_3^{n-k} c_4^k)$ ($k = bn$). So K is invertible if and only if $1 + (-1)^{n+1} c_2^{k+m} c_1^{n-m-k} / (c_3^{n-k} c_4^k) \neq 0$.

Let $bn \neq k$. In this case both intervals $(0, d)$ and $(d, \frac{1}{n})$ are nonempty. Therefore the invertibility of K is equivalent to inequality to zero of determinants $1 + (-1)^{n+1} c_2^{k+m+1} c_1^{n-m-k-1} / (c_3^{n-k-1} c_4^{k+1})$ and $1 + (-1)^{n+1} c_2^{k+m} c_1^{n-m-k} / (c_3^{n-k} c_4^k)$. Here k is the integer part of bn , and $k+1$ is the least integer greater than or equal to bn .

In the formulation of Theorem 4 we defined k_1 as the integer part of bn , and k_2 as the smallest integer such that $k_2 \geq bn$. Now we can see that the determinants (1.24), (1.25) are equal to

$$1 + (-1)^{n+1} c_2^{k_1+m} c_1^{n-m-k_1} / (c_3^{n-k_1} c_4^{k_1}) \quad \text{and} \quad 1 + (-1)^{n+1} c_2^{k_2+m} c_1^{n-m-k_2} / (c_3^{n-k_2} c_4^{k_2}),$$

correspondingly, and they are not equal to zero when

$$(-1)^n c_2^{k_1+m} c_1^{n-m-k_1} \neq c_3^{n-k_1} c_4^{k_1} \quad \text{and} \quad (-1)^n c_2^{k_2+m} c_1^{n-m-k_2} \neq c_3^{n-k_2} c_4^{k_2}.$$

This proves the theorem for the case of $c_3 \neq 0$ and $c_4 \neq 0$.

We have discussed in the beginning of the proof that in the case of $c_3 = 0$ or $c_4 = 0$, the invertibility of K is equivalent to

$$1 + (-1)^{n+1} c_3^{n-k_2} c_4^{k_2} / (c_2^{k_2+m} c_1^{n-m-k_2}) \neq 0 \text{ and } 1 + (-1)^{n+1} c_3^{n-k_1} c_4^{k_1} / (c_2^{k_1+m} c_1^{n-m-k_1}) \neq 0.$$

Theorem 4 is proved. □

1.3.2 The case of $a - b \in \mathbb{R} \setminus \mathbb{Q}$

Theorem 5. *When $a - b$ is irrational, the operator K is invertible if*

$$|c_3|^{1-b} |c_4|^b \neq |c_2|^a |c_1|^{1-a}.$$

Notice that following the scheme of the proof in [Antonevich, 1996, Th. 2.1, pp. 29–32] it is possible to show that the above condition is necessary and sufficient for the invertibility of the operator K . We omit the proof of the “necessary” part since we do not use it in the application to sampling and interpolation problems.

Proof. From Lemma 2 it follows that this theorem holds for the cases $c_1 = c_3 = 0$, $c_2 = c_3 = 0$, $c_2 = c_4 = 0$. Thus we do not need to consider these cases to prove this theorem. This means that we can assume that $(c_1 \neq 0$ and $c_2 \neq 0)$ or $(c_3 \neq 0$ and $c_4 \neq 0)$.

We know that when $c_3 \neq 0$ and $c_4 \neq 0$ the operator K is invertible if equation (1.15) has a unique solution. If $c_3 = 0$ or $c_4 = 0$, we can assume that $c_1 \neq 0$ and $c_2 \neq 0$, and in this case K is invertible when (1.17) has a unique solution. In this proof we first consider the case of $c_3 \neq 0$, $c_4 \neq 0$ and $|c_2|^a |c_1|^{1-a} |c_3|^{b-1} |c_4|^{-b} < 1$. Next we turn to the proof for $c_1 \neq 0$, $c_2 \neq 0$, and $|c_2|^a |c_1|^{1-a} |c_3|^{b-1} |c_4|^{-b} > 1$ (or equivalently $|c_4|^b |c_3|^{1-b} |c_1|^{a-1} |c_2|^{-a} < 1$); it has almost no differences from the first one and leads to the same result.

Notice that the cases of $(c_3 = 0$ and $b \neq 1)$ and $(c_4 = 0$ and $b \neq 0)$ are covered by the second part of the proof, and the cases of $(c_1 = 0$ and $a \neq 1)$, $(c_2 = 0$ and $a \neq 0)$

correspond to the first part. When $c_3 = 0$, $b = 1$, $c_4 \neq 0$ (and in all other similar cases) the expression $|c_2|^a |c_1|^{1-a} |c_3|^{b-1} |c_4|^{-b}$ has no zero factors, and this case falls in one of the two categories depending on the values of c_1 , c_2 , c_4 .

Suppose that $c_3 \neq 0$, $c_4 \neq 0$ and $|c_2|^a |c_1|^{1-a} |c_3|^{b-1} |c_4|^{-b} < 1$.

In this case the invertibility of the operator K is equivalent to solvability of equation (1.15):

$$f(t) + \phi(t)f((t + \Delta) \bmod 1) = h(t), \quad t \in [0, 1],$$

where

$$\phi(t) = \begin{cases} c_2/c_4, & t \in (0, b), \\ c_1/c_3, & t \in (b, \bar{\Delta}), \\ c_2/c_3, & t \in (\bar{\Delta}, 1). \end{cases}$$

To solve equation (1.15) we can apply successive approximations

$$f_0(t) = h(t), \quad f_{n+1}(t) = -\phi(t)f_n((t + \Delta) \bmod 1) + h(t), \quad n = 0, 1, \dots$$

Then

$$\begin{aligned} f_{n+1}(t) - f_n(t) &= \\ &= \left(\prod_{j=0}^{n-1} [-\phi((t + j\Delta) \bmod 1)] \right) [f_1((t + n\Delta) \bmod 1) - f_0((t + n\Delta) \bmod 1)]. \end{aligned} \quad (1.26)$$

If $c_1 = 0$ or $c_2 = 0$, then there is l such that

$$f_{n+1}(t) - f_n(t) = 0 \quad \text{for any } n \geq l.$$

Thus, $f(l)$ is a solution of equation (1.15). Therefore, the operator K is invertible when c_1 or c_2 is equal to zero, and c_3 and c_4 both are not equal to zero.

Now, let us assume that $c_1 \neq 0$ and $c_2 \neq 0$.

Since $\ln |\phi(t)|$ is Riemann integrable, for any irrational Δ

$$\begin{aligned} \frac{1}{N} \sum_{k=0}^{N-1} \ln |\phi((t + k\Delta) \bmod 1)| &\xrightarrow{N \rightarrow \infty} \int_0^1 \ln |\phi(t)| dt = \\ &= b \ln \left| \frac{c_2}{c_4} \right| + (1 - a) \ln \left| \frac{c_1}{c_3} \right| + (a - b) \ln \left| \frac{c_2}{c_3} \right|, \end{aligned}$$

uniformly in $t \in [0, 1]$ (see, e.g., Peterson [1983], p. 156). Then,

$$\begin{aligned} \lim_{N \rightarrow \infty} \max_t \left(\prod_{j=0}^{N-1} |\phi((t + j\Delta) \bmod 1)| \right)^{1/N} &= \\ &= \exp \lim_{N \rightarrow \infty} \max_t \frac{1}{N} \sum_{k=0}^{N-1} \ln |\phi((t + k\Delta) \bmod 1)| = \\ &= \exp \left(b \ln \left| \frac{c_2}{c_4} \right| + (1 - a) \ln \left| \frac{c_1}{c_3} \right| + (a - b) \ln \left| \frac{c_2}{c_3} \right| \right) = \left| \frac{c_2}{c_4} \right|^b \left| \frac{c_1}{c_3} \right|^{1-a} \left| \frac{c_2}{c_3} \right|^{a-b}. \end{aligned}$$

Note that

$$\left| \frac{c_2}{c_4} \right|^b \left| \frac{c_1}{c_3} \right|^{1-a} \left| \frac{c_2}{c_3} \right|^{a-b} = \frac{|c_2|^a |c_1|^{1-a}}{|c_3|^{1-b} |c_4|^b} < 1.$$

Then, for any ε such that $|\frac{c_2}{c_4}|^b |\frac{c_1}{c_3}|^{1-a} |\frac{c_2}{c_3}|^{a-b} < \varepsilon < 1$, there exist such M that

$$\max_t \left(\prod_{j=0}^{N-1} |\phi((t + j\Delta) \bmod 1)| \right)^{1/N} < \varepsilon \quad (1.27)$$

for any $N \geq M$.

Then, from (1.26), (1.27) we obtain that for large enough n

$$\begin{aligned} |f_{n+p} - f_n| &\leq \sum_{k=1}^p \prod_{j=0}^{n+k-2} |\phi((t + j\Delta) \bmod 1)| \\ &\quad |f_1((t + (n + k - 1)\Delta) \bmod 1) - f_0((t + (n + k - 1)\Delta) \bmod 1)| \end{aligned}$$

and

$$\|f_{n+p} - f_n\|_{L^2}^2 \leq 2 \sum_{k=1}^{\infty} (\varepsilon^{n+k-1})^2 \|f_1 - f_0\|_{L^2}^2 = \frac{2\varepsilon^{2n}}{1 - \varepsilon^2} \|f_1 - f_0\|_{L^2}^2.$$

Thus,

$$\|f_{n+p} - f_n\|_{L^2} \leq \frac{\sqrt{2}\varepsilon^n}{\sqrt{1 - \varepsilon^2}} \|f_1 - f_0\|_{L^2}.$$

The norm $\|f_{n+p} - f_n\|_{L^2}$ can be done arbitrary small for all p taking large enough n . Therefore, the sequence $\{f_i\}_{i=1}^n$ converges to a function f , and $f(t)$ is the solution of equation (1.15).

Let now $c_1 \neq 0$, $c_2 \neq 0$ and $|c_4|^b |c_3|^{1-b} |c_1|^{a-1} |c_2|^{-a} < 1$.

When $c_1 \neq 0$ and $c_2 \neq 0$, the invertibility of the operator K is equivalent to solvability of equation (1.17):

$$f(t) + \psi(t)f((t + \bar{\Delta}) \bmod 1) = k(t), \quad t \in [0, 1],$$

where

$$\psi(t) = \begin{cases} c_3/c_2, & t \in (0, \Delta), \\ c_4/c_2, & t \in (\Delta, a), \\ c_3/c_1, & t \in (a, 1). \end{cases}$$

This time we use the same kind of successive approximations to prove solvability of second equation (1.17). We use the fact that

$$\lim_{n \rightarrow \infty} \max_t \left(\prod_{j=0}^{N-1} |\psi((t + j\bar{\Delta}) \bmod 1)| \right)^{1/N} = \left| \frac{c_4}{c_2} \right|^b \left| \frac{c_3}{c_1} \right|^{1-a} \left| \frac{c_3}{c_2} \right|^{a-b} < 1$$

to prove that the new sequence $\{f_i\}$ converges to the solution of equation (1.17).

Therefore, the operator K is invertible when $|c_3|^{1-b}|c_4|^b \neq |c_2|^a|c_1|^{1-a}$. \square

1.4 The invertibility of the Operator V .

In this section we use Theorems 4 and 5 to show that the operator V (see (1.12)) is invertible.

From formulas (1.37)–(1.46) (see Appendix 1.A), we know that there are three kinds of the operator V : 1) with $c_1 \neq 0$, $c_2 \neq 0$, and $c_3 = c_4 = 0$ (or $c_3 \neq 0$, $c_4 \neq 0$, and $c_1 = c_2 = 0$); 2) with only one of the coefficients c_i equal to zero; 3) with $c_i \neq 0$ for $1 \leq i \leq 4$.

In case 1 the conditions of Theorems 4 and 5 hold, so, the operator V is invertible.

Let us prove that if only one of the coefficients c_i is equal to zero, the operator V is invertible.

Theorem 6. *When exactly one of the coefficients c_i is equal to zero, the operator V is invertible in $L^2[0, 1]$ for small enough μ .*

Proof. The conditions of Theorems 4 and 5 hold if the zero coefficient is raised to a nonzero power. For example if $a - b$ is an irrational number, $c_3 = 0$, and $b \neq 1$, then the condition of Theorem 5 becomes $|0|^{1-b}|c_4|^b \neq |c_2|^a|c_1|^{1-a}$, or $|c_2|^a|c_1|^{1-a} \neq 0$, which is obviously true.

We will have to separately handle the cases when the zero coefficient is raised to a zero power. In the case $a - b \in \mathbb{R} \setminus \mathbb{Q}$, $c_1 = 0$, $a = 1$ the condition of Theorem 5 becomes

$$|c_3|^{1-b}|c_4|^b \neq |c_2|. \quad (1.28)$$

In the case $a - b \in \mathbb{R} \setminus \mathbb{Q}$, $c_4 = 0$, $b = 0$ the condition of Theorem 5 becomes

$$|c_3| \neq |c_2|^a|c_1|^{1-a}. \quad (1.29)$$

From the formulas for the coefficients c_i (1.38),(1.39),(1.42),(1.43),(1.45) derived in the proof of Theorem 2 in Appendix 1.A, it follows that the left-hand side and the right-hand side of the inequalities (1.28),(1.29) involve different powers of μ . Thus for small enough μ the inequalities (1.28),(1.29) hold, and the operator V is invertible.

In the case $a - b \in \mathbb{Q}$, $c_1 = 0$ the condition of Theorem 4 is true unless $n - m - k_1 = 0$ or $n - m - k_2 = 0$. When $n - m - k_i = 0$, the condition of Theorem 4 is

$$(-1)^n c_2^n \neq c_3^m c_4^{n-m}. \quad (1.30)$$

In the case $a - b \in \mathbb{Q}$, $c_2 = 0$, $k_i + m = 0$, the condition of Theorem 4 is

$$(-1)c_1 \neq c_3. \quad (1.31)$$

In the case $a - b \in \mathbb{Q}$, $c_3 = 0$, $n = k_i$, the condition of Theorem 4 is

$$(-1)c_2 \neq c_4. \quad (1.32)$$

In the case $a - b \in \mathbb{Q}$, $c_4 = 0$, $k_i = 0$, the condition of Theorem 4 is

$$(-1)^n c_2^m c_1^{n-m} \neq c_3^n. \quad (1.33)$$

One can check that if $m \neq 0$, the conditions (1.30),(1.33) hold for small enough μ , since the left-hand sides and the right-hand sides of the inequalities involve different powers of μ . Therefore the operator V is invertible in these cases.

If $m = 0$ then $n = 1$, since $\frac{m}{n}$ is an irreducible fraction. Inequalities (1.30),(1.33) take the form of inequalities (1.32),(1.31) correspondingly. To show that the inequalities (1.31),(1.32) hold, we use the fact that

$$|A(r, k, 0)| > |A(r - k + 1, k, k)| \text{ for } k > 1. \quad (1.34)$$

From the formulas for the coefficients c_i (1.38),(1.39),(1.42),(1.43),(1.45) and the relation (1.34) it follows that one of the coefficients c_1, c_3 (or c_2, c_4) is larger than the other one by absolute value, or both coefficients are positive or negative. Thus, the inequalities (1.31),(1.32) hold. \square

Let us consider the case of $c_i \neq 0$ for $1 \leq i \leq 4$. In this case the operator U is given by the following formula:

$$\begin{aligned} (Uf)(t) = & \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r_1, k_1, 0)f(t - w(r_1, k_1, 0)) \\ & + A(r_1, k_1 + 1, 0)f(t - w(r_1, k_1 + 1, 0)) \\ & + A(r_1 - k_1 + 1, k_1, k_1)f(t - w(r_1 - k_1 + 1, k_1, k_1)) \\ & + A(r_1 - k_1, k_1 + 1, k_1 + 1)f(t - w(r_1 - k_1, k_1 + 1, k_1 + 1))] \end{aligned}$$

(see the proof of Theorem 2 in Appendix 1.A).

Depending on the relations between the shifts $w(r, k, q)$ in the above formula, coefficients c_i may have two forms:

$$\begin{aligned} c_1 = A(r, k, 0) &= (-1)^{r+k} \mu^k \frac{(r+k)!}{r!k!}, \\ c_2 = A(r, k+1, 0) &= (-1)^{r+k+1} \mu^{k+1} \frac{(r+k+1)!}{r!(k+1)!}, \\ c_3 = A(r-k, k+1, k+1) &= (-1)^{r+1} \mu^{k+1} \frac{(r+1)!}{(r-k)!(k+1)!}, \\ c_4 = A(r-k+1, k, k) &= (-1)^{r+1} \mu^k \frac{(r+1)!}{(r-k+1)!k!}, \end{aligned} \quad (1.35)$$

and

$$\begin{aligned}
c_1 &= A(r-k, k+1, k+1) = (-1)^{r+1} \mu^{k+1} \frac{(r+1)!}{(r-k)!(k+1)!}, \\
c_2 &= A(r-k+1, k, k) = (-1)^{r+1} \mu^k \frac{(r+1)!}{(r-k+1)!k!}, \\
c_3 &= A(r, k, 0) = (-1)^{r+k} \mu^k \frac{(r+k)!}{r!k!}, \\
c_4 &= A(r, k+1, 0) = (-1)^{r+k+1} \mu^{k+1} \frac{(r+k+1)!}{r!(k+1)!}.
\end{aligned} \tag{1.36}$$

with some $r \geq 0$ and $k \geq 1$ (we do not have to consider cases with $k = 0$, because when $k = 0$, at least one of the coefficients c_i is zero).

Theorem 7. *When $a - b$ is rational and $c_i \neq 0$ for $1 \leq i \leq 4$, the operator V is invertible in $L^2[0, 1]$ for small enough μ .*

Proof. By Theorem 4 the operator V is invertible if and only if $(-1)^n c_2^{k_1+m} c_1^{n-m-k_1} \neq c_3^{n-k_1} c_4^{k_1}$ and $(-1)^n c_2^{k_2+m} c_1^{n-m-k_2} \neq c_3^{n-k_2} c_4^{k_2}$, where $a - b = \frac{m}{n}$, k_1 is the integer part of bn , and k_2 is the smallest integer such that $k_2 \geq bn$.

From formulas (1.35) and (1.36) we see that when $k_1 \neq \frac{n-m}{2}$, terms $c_2^{k_1+m} c_1^{n-m-k_1}$ and $c_3^{n-k_1} c_4^{k_1}$ have different powers of μ . Therefore, in this case μ can be made small enough, to make $(-1)^n c_2^{k_1+m} c_1^{n-m-k_1} \neq c_3^{n-k_1} c_4^{k_1}$. Similarly, when $k_2 \neq \frac{n-m}{2}$, for small enough μ , $(-1)^n c_2^{k_2+m} c_1^{n-m-k_2} \neq c_3^{n-k_2} c_4^{k_2}$.

If k_1 or k_2 or both are equal to $\frac{n-m}{2}$, then the corresponding inequality will have the form $(-1)^n c_2^{\frac{n+m}{2}} c_1^{\frac{n-m}{2}} \neq c_3^{\frac{n+m}{2}} c_4^{\frac{n-m}{2}}$. Now we cannot achieve the condition of Theorem 4 by making μ small, because the powers of μ are the same on both sides of inequality. So, we have to consider the specific forms of the coefficients c_i (see (1.35), (1.36)).

Since $k \geq 1$, $|A(r, k, 0)| \geq |A(r-k+1, k, k)|$ and $|A(r, k+1, 0)| > |A(r-k, k+1, k+1)|$. Then $|c_2|^{\frac{n+m}{2}} |c_1|^{\frac{n-m}{2}} \neq |c_3|^{\frac{n+m}{2}} |c_4|^{\frac{n-m}{2}}$, and therefore $(-1)^n c_2^{\frac{n+m}{2}} c_1^{\frac{n-m}{2}} \neq c_3^{\frac{n+m}{2}} c_4^{\frac{n-m}{2}}$.

□

Theorem 8. *When $a - b$ is irrational and $c_i \neq 0$ for $1 \leq i \leq 4$, the operator V is invertible in $L^2[0, 1]$ for small enough μ .*

Proof. From Theorem 5 we know that the operator V is invertible if

$$|c_3|^{1-b}|c_4|^b \neq |c_2|^a|c_1|^{1-a}.$$

Using formulas (1.35) and (1.36) we see that if $1 - b \neq a$, then terms $|c_3|^{1-b}|c_4|^b$ and $|c_2|^a|c_1|^{1-a}$ involve different powers of μ . We can choose μ such that $|c_3|^{1-b}|c_4|^b \neq |c_2|^a|c_1|^{1-a}$.

If $1 - b = a$, then the expression above will have the form: $|c_1|^{1-a}|c_2|^a \neq |c_3|^a|c_4|^{1-a}$. We will again have to look at the concrete forms of c_i . As we know, for $k \geq 1$, $|A(r - k, k + 1, k + 1)| < |A(r, k + 1, 0)|$ and $|A(r - k + 1, k, k)| \leq |A(r, k, 0)|$. Since $a - b = \Delta \in \mathbb{J}$ and $1 - b = a$, then $a \neq 0$ and $a \neq 1$. Thus, either $|c_1|^{1-a}|c_2|^a < |c_3|^a|c_4|^{1-a}$, or $|c_1|^{1-a}|c_2|^a > |c_3|^a|c_4|^{1-a}$.

Therefore, $|c_3|^{1-b}|c_4|^b \neq |c_2|^a|c_1|^{1-a}$.

This completes the proof of the invertibility of the operator V for irrational Δ . □

Appendix 1.A. The proof of Theorem 2

Now we prove Theorem 2 from Section 1.2. We single out several terms of sum (1.11) that have smallest powers of μ . Sum of those terms form the operator U . We choose the number of terms so that later it will be possible to prove the invertibility of U . In the course of this proof we show that this number of terms does not need to be greater than four.

Theorem 2. *For any a_j and b_j , there exists a nonnegative integer number k such that the operator W can be written in the form*

$$W = U + \tilde{U},$$

where the operator U is comprised of at most four terms whose coefficients each involve the parameter μ to a power not exceeding $k + 1$, and all terms of the operator \tilde{U} contain

a factor μ at a power at least equal to $k + 2$. The operator U has the following form:

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [c_1 f(t - w_1) + c_2 f(t - w_1 - \alpha_1) \\ + c_3 f(t - w_2 + \alpha_1) + c_4 f(t - w_2)],$$

where one or more of coefficients c_i may be zero, w_1 and w_2 have the form $w(\bar{r}, \bar{k}, \bar{q})$ with $\bar{r}, \bar{k}, \bar{q}$ depending on relative position and lengths of the intervals.

Proof. 1. We are looking for r, k, q with the smallest possible k such that $t - w(r, k, q) \in [0, \alpha_1]$ for some $t \in [\alpha', \alpha' + \alpha_1]$.

2. Suppose that $\alpha' - w(r, 0, 0) \in [0, \alpha_1]$ or $\alpha' + \alpha_1 - w(r, 0, 0) \in [0, \alpha_1]$ for some r .

2.1. Let $\alpha' - w(r, 0, 0) \in [0, \alpha_1]$. Then, $\alpha' + \alpha_1 - w(r, 1, 0) = \alpha' - w(r, 0, 0) \in [0, \alpha_1]$.

Also may or may not be $\alpha' + \alpha_1 - w(r, 1, 1) \in [0, \alpha_1]$.

2.1.1. If $\alpha' + \alpha_1 - w(r, 1, 1) \in [0, \alpha_1]$, then

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r, 0, 0)f(t - w(r, 0, 0)) \\ + A(r, 1, 0)f(t - w(r, 1, 0)) + A(r, 1, 1)f(t - w(r, 1, 1))]. \quad (1.37)$$

2.1.2. If $\alpha' + \alpha_1 - w(r, 1, 1) \notin [0, \alpha_1]$, then

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r, 0, 0)f(t - w(r, 0, 0)) \\ + A(r, 1, 0)f(t - w(r, 1, 0))]. \quad (1.38)$$

2.2. Let $\alpha' + \alpha_1 - w(r, 0, 0) \in [0, \alpha_1]$. Since $\alpha' > 0, r > 0$. Then $\alpha' - w(r - 1, 1, 1) = \alpha' + \alpha_1 - w(r, 0, 0) \in [0, \alpha_1]$. Also, $\alpha' - w(r - 1, 1, 0) \in [0, \alpha_1]$ may hold.

2.2.1. If $\alpha' - w(r - 1, 1, 0) \in [0, \alpha_1]$, then

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r, 0, 0)f(t - w(r, 0, 0)) \\ + A(r - 1, 1, 0)f(t - w(r - 1, 1, 0)) + A(r - 1, 1, 1)f(t - w(r - 1, 1, 1))]. \quad (1.39)$$

2.2.2. If $\alpha' - w(r - 1, 1, 0) \notin [0, \alpha_1]$, then

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r, 0, 0)f(t - w(r, 0, 0)) \\ + A(r - 1, 1, 1)f(t - w(r - 1, 1, 1))]. \quad (1.40)$$

3. Now we can assume that $t - w(r, 0, 0) \notin [0, \alpha_1]$ for any $t \in [\alpha', \alpha' + \alpha_1]$ and $r \geq 0$.

Note,

$$w(r, k, q) = w(r + q, k - 2q, 0) \quad \text{for } k \geq 2q$$

and

$$w(r, k, q) = w(r + k - q, 2q - k, 2q - k) \quad \text{for } 2q \geq k.$$

Therefore, (r, k, q) with minimal k such that $t - w(r, k, q) \in [0, \alpha_1]$ for some $t \in [\alpha', \alpha' + \alpha_1]$ will have form $(r, k, 0)$ or (r, k, k) where $k \geq 1$.

4. Let us find r_1, k_1 such that $\alpha' - w(r_1, k_1, 0) \in [0, \alpha_1]$ and k_1 is the smallest possible.

The answer is: $r_1 = \lfloor \frac{\alpha'}{\alpha} \rfloor$, $k_1 = \lfloor \frac{\alpha' - r_1 \alpha}{\alpha_1} \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x . Also, $\alpha' + \alpha_1 - w(r_1, k_1 + 1, 0) \in [0, \alpha_1]$.

Note that $k_1 + 1$ is the smallest k_2 such that $\alpha' + \alpha_1 - w(r_2, k_2, 0) \in [0, \alpha_1]$. If there is $k_2 < k_1 + 1$ with $\alpha' + \alpha_1 - w(r_2, k_2, 0) \in [0, \alpha_1]$, then $\alpha' + w(r_2, k_2 - 1, 0) \in [0, \alpha_1]$ and $k_2 - 1 < k_1$ - a contradiction.

5. Let us find r_3, k_3, r_4, k_4 such that $\alpha' + \alpha_1 - w(r_3, k_3, k_3) \in [0, \alpha_1]$ and $\alpha' - w(r_4, k_4, k_4) \in [0, \alpha_1]$ with the smallest k_3, k_4 .

It is $k_3 = \lfloor \frac{(r_1 + 1)\alpha - \alpha'}{\alpha_1} \rfloor$ and $r_3 = r_1 + 1 - k_3$;

$$k_4 = k_3 + 1, \quad r_4 = r_3 - 1.$$

6.1. Let $r_3 < 0$ or $k_3 > k_1 + 1$. Then

$$\begin{aligned} (Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [& A(r_1, k_1, 0)f(t - w(r_1, k_1, 0)) \\ & + A(r_1, k_1 + 1, 0)f(t - w(r_1, k_1 + 1, 0))] . \end{aligned} \quad (1.41)$$

6.2. Let $r_3 = 0$, $k_3 \leq k_1 + 1$.

$$\begin{aligned} (Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [& A(r_1, k_1, 0)f(t - w(r_1, k_1, 0)) \\ & + A(r_1, k_1 + 1, 0)f(t - w(r_1, k_1 + 1, 0))] + A(0, k_3, k_3)f(t - w(0, k_3, k_3)) . \end{aligned} \quad (1.42)$$

6.3. Let $r_3 > 0$ and $k_3 = k_1 + 1$. Then $r_3 = r_1 - k_1$.

$$\begin{aligned} (Uf)(t) = & \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r_1, k_1, 0)f(t - w(r_1, k_1, 0)) \\ & + A(r_1, k_1 + 1, 0)f(t - w(r_1, k_1 + 1, 0)) \\ & + A(r_1 - k_1, k_1 + 1, k_1 + 1)f(t - w(r_1 - k_1, k_1 + 1, k_1 + 1))] . \end{aligned} \quad (1.43)$$

6.4. Let $r_3 > 0$ and $k_3 = k_1$. Then

$$\begin{aligned} (Uf)(t) = & \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r_1, k_1, 0)f(t - w(r_1, k_1, 0)) \\ & + A(r_1, k_1 + 1, 0)f(t - w(r_1, k_1 + 1, 0)) \\ & + A(r_1 - k_1 + 1, k_1, k_1)f(t - w(r_1 - k_1 + 1, k_1, k_1)) \\ & + A(r_1 - k_1, k_1 + 1, k_1 + 1)f(t - w(r_1 - k_1, k_1 + 1, k_1 + 1))] . \end{aligned} \quad (1.44)$$

6.5. Let $r_3 > 0$ and $k_3 = k_1 - 1$.

$$\begin{aligned} (Uf)(t) = & \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r_1, k_1, 0)f(t - w(r_1, k_1, 0)) \\ & + A(r_1 - k_1 + 2, k_1 - 1, k_1 - 1)f(t - w(r_1 - k_1 + 2, k_1 - 1, k_1 - 1)) \\ & + A(r_1 - k_1 + 1, k_1, k_1)f(t - w(r_1 - k_1 + 1, k_1, k_1))] . \end{aligned} \quad (1.45)$$

6.6. Suppose $r_3 > 0$ and $k_3 < k_1 - 1$.

$$\begin{aligned} (Uf)(t) = & \chi_{[\alpha', \alpha' + \alpha_1]}(t) [A(r_3, k_3, k_3)f(t - w(r_3, k_3, k_3)) \\ & + A(r_3 - 1, k_3 + 1, k_3 + 1)f(t - w(r_3 - 1, k_3 + 1, k_3 + 1))] . \end{aligned} \quad (1.46)$$

We derived all possible formulas for U for various relative positions of intervals I_1 and I_2 (see (1.10)). Note that U may contains two (formulas (1.38), (1.40), (1.41), (1.46)), three (formulas (1.37), (1.39), (1.42), (1.43), (1.45)), or four (formula (1.44)) terms.

□

Appendix 1.B. The proof of Lemma 1

Lemma 1. *If the operator U is invertible, then $\|U^{-1}\| \leq |\mu|^{-(k+1)}C$ for small enough $|\mu|$, where $C > 0$ depends only on r, k, q .*

Proof. First we show that $\|Uf\| \geq |\mu|^{k+1}B\|f\|$ for any $f \in L^2[0, \alpha_1]$, where B is a positive constant. To prove this we need the formulas for the operator U from the proof of Theorem 2 given in Appendix 1.A. We consider separately cases when U consists of 2, 3 and 4 terms.

In formulas (1.38,1.40,1.41,1.46) the operator U has two terms:

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [\mu^k A_1 f(t - w) + \mu^{k+1} A_2 f(t - w \pm \alpha_1)].$$

Here A_1, A_2, w do not depend on μ , and $A_1 \neq 0, A_2 \neq 0$. Notice that since f is defined on $[0, \alpha_1]$, then the two terms are never nonzero on the same part of the interval $[0, \alpha_1]$, because distance between $t - w$ and $t - w \pm \alpha_1$ is α_1 . Then

$$\|(Uf)(t)\| \geq \min(|A_1|, |\mu A_2|) |\mu|^k \|f\| \geq |\mu|^{k+1} |A_2| \|f\|$$

for small enough μ .

Let us consider the cases when the operator U has three terms. In formulas (1.37,1.39,1.43,1.45) the operator U has the form

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [\mu^k A_1 f(t - w_1) + \mu^{k+1} A_2 f(t - w_1 \pm \alpha_1) + \mu^{k+1} A_3 f(t - w_2)].$$

For these cases

$$\|(Uf)(t)\| \geq \min(|A_1|, |\mu A_2|, |A_1 + \mu A_3|, |\mu(A_2 + A_3)|) |\mu|^k \|f\| \geq |\mu|^{k+1} \min(|A_2 + A_3|, |A_2|) \|f\|$$

for small enough μ . One can check using the exact formulas for the coefficients A_i , that $A_2 + A_3 \neq 0$.

In formula (1.42) the operator U is:

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [\mu^k A_1 f(t - w_1) + \mu^{k+1} A_2 f(t - w_1 \pm \alpha_1) + \mu^{k_3} A_3 f(t - w_2)], \quad k_3 \leq k + 1.$$

Then for small enough μ and a positive constant D we have

$$\begin{aligned} \|(Uf)(t)\| &\geq \min(|A_1|, |\mu A_2|, |A_1 + \mu^{k_3 - k} A_3|, |\mu A_2 + \mu^{k_3 - k} A_3|) |\mu|^k \|f\| \\ &\geq |\mu|^{k+1} \min(|A_2 + \mu^{k_3 - k - 1} A_3|, |A_2|) \|f\| \geq D |\mu|^{k+1} \|f\|. \end{aligned}$$

Now we consider the last case of four terms (see formula (1.44)):

$$(Uf)(t) = \chi_{[\alpha', \alpha' + \alpha_1]}(t) [\mu^k A_1 f(t - w_1) + \mu^{k+1} A_2 f(t - w_1 - \alpha_1) + \mu^k A_3 f(t - w_2) + \mu^{k+1} A_4 f(t - w_2 + \alpha_1)].$$

Then

$$\begin{aligned} \|(Uf)(t)\| &\geq \min(|A_1 + A_3|, |A_1 + \mu A_4|, |A_3 + \mu A_2|, |\mu(A_2 + A_4)|) |\mu|^k \|f\| \geq \\ &|\mu|^{k+1} |A_2 + A_4| \|f\| \end{aligned}$$

for small enough μ . Let us show that $A_2 + A_4 \neq 0$. From formula (1.44) we see that

$$A_2 = (-1)^{r+k+1} \frac{(r+k+1)!}{r!(k+1)!}, \quad A_4 = (-1)^{r+1} \frac{(r+1)!}{(r-k)!(k+1)!}$$

for some $r, k \geq 1$. Note that for $k \geq 1$

$$\frac{(r+k+1)!}{r!(k+1)!} > \frac{(r+1)!}{(r-k)!(k+1)!},$$

so $A_2 + A_4 \neq 0$.

We have shown that for any $f \in L^2([0, \alpha_1])$, $\|Uf\| \geq |\mu|^{k+1} B \|f\|$, where B is a positive nonzero constant. Since we assumed that U is invertible, then for any $g \in L^2([\alpha', \alpha' + \alpha_1])$, $\|g\| \geq |\mu|^{k+1} B \|U^{-1}g\|$. So, for every $g \in L^2([\alpha', \alpha' + \alpha_1])$: $\|U^{-1}g\| \leq |\mu|^{-(k+1)} \frac{1}{B} \|g\|$. Thus $\|U^{-1}\| \leq |\mu|^{-(k+1)} C$. \square

Bibliography

- Antonevich A. (1996): *Linear Functional Equations. Operator Approach*. Basel, Boston, Berlin: Birkhäuser.
- Avdonin S. (1979): *On Riesz bases from exponentials in L^2* . Vestnik Leningrad Univ. Math., vol. 7, pp. 203–211.
- Avdonin S. and Ivanov S. (1995): *Families of Exponentials. The Method of Moments in Controllability Problems for Distributed Parameter Systems*. New York: Cambridge University Press.
- Avdonin S. and Ivanov S. (2008): *Sampling and interpolation problems for vector valued signals in the Paley-Wiener spaces*. IEEE Transactions on Signal Processing, vol. 56, no. 11, pp. 5435–5441.
- Avdonin S. and Moran W. (1999): *Sampling and interpolation of functions with multi-band spectra and controllability problems*. In: *Optimal Control of Partial Differential Equations* (K.H. Hoffmann, G. Leugering and T. F., Eds.), vol. 133, pp. 43–51, Basel: Birkhäuser, internat. Ser. Numer. Math.
- Beatty M. and Dodson M. (1989): *Derivative sampling for multiband signals*. Numer. Funct. Anal. Optim., vol. 10, pp. 875–898.
- Beatty M. and Dodson M. (1993): *The distribution of sampling rates for signals with equally wide, equally spaced spectral bands*. SIAM J. Appl. Math., vol. 53, pp. 893–906.
- Bezuglaya L. and Katsnelson V. (1993): *The sampling theorem for functions with limited multi-band spectrum, I*. Z. Anal. Anwendungen, vol. 12, pp. 511–534.
- Böttcher A., Karlovich Y. and Spitkovsky I. (2002): *Convolution operators and factorization of almost periodic matrix functions*. Basel, Boston: Birkhäuser Verlag.

- Dodson M. and Silva A. (1989): *An algorithm for optimal regular sampling*. Signal Process., vol. 17, pp. 169–174.
- Higgins J. (1996): *Sampling theory in Fourier and signal analysis: foundations*. Oxford: Clarendon Press.
- Hruščev S., Nikol'skii N. and Pavlov B. (1981): *Unconditional bases of exponentials and reproducing kernals*. Complex Analysis and Spectral Theory, Lecture Notes Math., vol. 864, pp. 214–335.
- Katsnelson V. (1996): *Sampling and interpolation for functions with multi-band spectrum: the mean-periodic continuation method*. In: *Wiener-Symposium (Grossbothen, 1994) Synerg. Syntropie Nichtlineare Syst.*, vol. 4, pp. 91–132, Leipzig: Verlag Wiss. Leipzig.
- Kohlenberg A. (1953): *Exact interpolation of band-limited functions*. J. Appl. Phys., vol. 24, pp. 1432–1436.
- Lyubarskii Y. and Seip K. (1997): *Sampling and interpolating sequences for multiband-limited functions and exponential bases on disconnected sets*. J. Fourier Analysis Appl., vol. 3, pp. 597–615.
- Lyubarskii Y. and Spitkovsky I. (1996): *Sampling and interpolation for a lacunary spectrum*. In: *Proc. Royal Soc. Edinburgh*, vol. 126 A, pp. 77–87.
- Moran W. and Avdonin S. (1999): *Sampling of multi-band signals*. In: *Proceedings of the Fourth International Congress on Industrial and Applied Mathematics* (J. Ball and J. Hunt, Eds.), vol. 126 A, pp. 163–174.
- Peterson K. (1983): *Ergodic theory*. Cambridge: Cambridge University Press.
- Russell D. (1978): *Controllability and stabilizability theory for linear partial differential equations*. SIAM Review, vol. 20, pp. 639–739.

Seip K. (1995): *A simple construction of exponential bases in L^2 of the union of several intervals.* Proc. Edinburgh Math. Soc., vol. 38, pp. 171–177.

Spitkovsky I. (2006): Personal communication.

Chapter 2

Boundary Control approach to the spectral estimation problem. The case of multiple poles¹

Abstract

There exist many methods for solving the spectral estimation problem. This chapter proposes a new approach to this problem based on the Boundary Control method. We show that the problem of decomposition of a signal modeled by a sum of exponentials with polynomial coefficients can be reduced to an identification problem for a discrete time linear dynamical system. It follows that values of exponentials can be found solving a generalized eigenvalue problem as in the Matrix Pencil method. We also give exact formulas for the polynomial amplitudes.

Keywords: Spectral estimation, Signal Processing, Boundary Control method, Control theory, Matrix Pencil method.

2.1 Introduction

Let a signal $r(t)$ be modeled by the following expression:

$$r(t) = \sum_{n=1}^K a_n(t)e^{\lambda_n t}, \quad (2.1)$$

where $a_n(t)$ are polynomials and λ_n can be real or complex numbers. Our problem is to recover the number of poles K , the polynomial coefficients $\{a_n(t)\}$ and the exponents $\{\lambda_n\}$ knowing the observations of the signal $r(0), r(1), \dots$

Functions of the form (2.1) arise as solutions of linear homogeneous ordinary differentials equations with constant coefficients

$$x^{(d)} + A_1 x^{(d-1)} + \dots + A_d = 0 \quad (2.2)$$

¹S.A. Avdonin and A.S. Bulanova, *Boundary control approach to the spectral estimation problem. The case of multiple poles*, Mathematics of Control, Signals, and Systems, submitted, 2007.

and linear homogeneous recurrence relations with constant coefficients

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \dots + c_d a_{n-d}. \quad (2.3)$$

A general solution to equation (2.2) is

$$x(t) = \sum_{i=1}^K \sum_{j=0}^{M_i-1} a_{ij} t^j e^{z_i t}$$

where z_i is a zero of multiplicity M_i of the characteristic polynomial

$$p(z) = z^d + A_1 z^{d-1} + \dots + A_d.$$

A general solution to equation (2.3) is

$$a_n = \sum_{i=1}^K \sum_{j=0}^{M_i-1} b_{ij} n^j \lambda_i^n = \sum_{i=1}^K \sum_{j=0}^{M_i-1} b_{ij} n^j e^{(\ln \lambda_i) n}$$

where λ_i is a zero of multiplicity M_i of the characteristic polynomial

$$p(\lambda) = \lambda^d - c_1 \lambda^{d-1} - \dots - c_d.$$

Solutions of the form $\sum_{i=1}^N a_i e^{-b_i t}$ with real coefficients a_i and $b_i > 0$ appear in heat diffusion and diffusion of chemical compounds problems, time series in medicine, economics. Solutions of the form $\sum_{i=1}^N a_i \sin(b_i t + c_i)$ occur when the characteristic polynomial has complex roots, and are typical for electrical systems.

The classical spectral estimation problem is to recover the coefficients a_i , λ_i of a signal $r(t) = \sum_{i=1}^N a_i e^{\lambda_i t}$, by the given observations $r(j)$, $j = 0, \dots, 2N - 1$. This problem is very important in signal processing, the applications are in wireless communications, antenna array design, bio-medical imaging, high-speed circuit analysis and others (see [9; 16]).

There exist many methods for solving the spectral estimation problem: the method of Prony and its numerous modifications [13; 11]; the Matrix Pencil method developed by Hua and Sarkar [8; 9; 16]; iterative maximum likelihood methods (see, for

example, [12]); MUSIC (Multiple Signal Classification) [17] and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [14] algorithms, and others. Badeau et al. [4] develop a generalized ESPRIT algorithm for estimation of parameters of a signal modeled by the Polynomial Amplitude Complex Exponentials model.

We propose a new approach to this problem based on the “nonselfadjoint” version of the Boundary Control (BC) method [2]. The BC method has been recently developed for solving boundary spectral and dynamical inverse problems for partial differential equations (see, e.g., [5; 1]). The BC method reveals that the two central problems of the theory of inverse and control theory of distributed parameter systems have a direct connection with each other. The first problem is the recovery of unknown coefficients, the second problem is the controllability of the corresponding initial boundary value problem. Roughly speaking, the BC method gives the realization of R. Kalman’s idea that the controllable (or observable) part of a system can be identified. We extend this method to dynamical systems with discrete time.

In the joint paper with S. Avdonin and D. Nicol’sky [3] the BC method is applied to the problem of decomposition of a signal $r(t) = \sum_{n=1}^K a_n e^{\lambda_n t}$, where amplitudes a_n are constant. Here we consider the case of a signal with polynomial amplitudes $a_n(t)$: $r(t) = \sum_{n=1}^K a_n(t) e^{\lambda_n t}$.

Using Boundary Control method we show that the coefficients $\{\lambda_n\}$ can be obtained as in the Matrix Pencil method by solving the generalized eigenvalue problem for the matrices A and B :

$$Af = \hat{\lambda}Bf$$

$$A_{ij} = r(i + j - 1), \quad B_{ij} = r(i + j - 2), \quad i, j = 1, \dots, N$$

using this formula: $\lambda_n = \ln \hat{\lambda}_n$. Also our method gives exact formulas for computation of the amplitudes $a_n(t)$ in terms of generalized eigenvectors and eigenvalues of the

above eigenvalue problem and observations $r(t)$. Note that N may be unknown and can be found in the procedure.

This chapter is organized as follows. We introduce a pair of auxiliary dynamical systems (2.5), (2.10) (see Sec. 2.2); state a controllability condition for such systems (see Sec. 2.3), introduce the control and response operators for systems (2.5), (2.10) (Sec. 2.4). In Section 2.5, we consider the problem of identification for system (2.5): we show that parameters of system (2.5) can be recovered using values of the kernel of its response operator. Section 2.6 shows that an application of a transformation of variable to system (2.5) does not change its response operator. In Section 2.7, we show that the problem of decomposition of a signal of form (2.1) is equivalent to an identification problem for a certain system of form (2.5) and present an algorithm for signal estimation based on these ideas.

2.2 Dynamical systems

In this section we construct such a dynamical system that the function

$$r(k) = \sum_{n=1}^K a_n(k) \lambda_n^k \quad (2.4)$$

is the kernel of the input-output operator of this system. The problems of determining the coefficients λ_n and a_n for functions (2.1) and (2.4) are equivalent. In what follows it is more convenient to work with form (2.4).

Let $N = \sum_{n=1}^K M_n$, where M_n are the degrees of the polynomials $a_n(k)$.

Let us introduce an auxiliary discrete-time dynamical system:

$$x(k+1) = Mx(k) + bf(k), \quad x(k) \in \mathbb{C}^N, \quad x(0) = 0 \quad (2.5)$$

with an observation y

$$y(k) = \langle x(k), c \rangle_{\mathbb{C}^N} := c^* x(k).$$

Here M is an $N \times N$ constant matrix, f is a scalar control, $b, c \in \mathbb{C}^N$, and c^* means conjugate transpose of $N \times 1$ vector c . In general situation M is not self-adjoint.

Solving equation (2.5) for a given control $f(0), f(1), \dots, f(k), \dots$ we get

$$x^f(k) = \sum_{i=0}^{k-1} M^{k-1-i} b f(i). \quad (2.6)$$

Define the matrix $Y(k)_{N \times k}$ and the vector $F(k)_{k \times 1}$ by

$$Y(k) = (b | Mb | M^2 b | \dots | M^{k-1} b), \quad (2.7)$$

$$F(k) = (f(k-1), \dots, f(1), f(0))^T. \quad (2.8)$$

Using (2.6), (2.7) and (2.8) we obtain

$$x^f(k) = Y(k)F(k). \quad (2.9)$$

We also consider the adjoint system:

$$z(k+1) = M^* z(k) + cg(k), \quad z(k) \in \mathbb{C}^N, \quad z(0) = 0 \quad (2.10)$$

with an observation $w(k) = \langle z(k), b \rangle_{\mathbb{C}^N} = b^* z(k)$. The solution to this equation is

$$z^g(k) = \sum_{j=0}^{k-1} (M^*)^{k-1-j} cg(j). \quad (2.11)$$

Formula (2.11) can be rewritten in matrix form as

$$z^g(k) = Y^\#(k)G(k),$$

where

$$Y^\#(k) = (c | M^* c | (M^*)^2 c | \dots | (M^*)^{k-1} c),$$

$$G(k) = (g(k-1), \dots, g(1), g(0))^T.$$

We show that it is possible to solve the spectral estimation problem for signal (2.4) using systems (2.5), (2.10) and the ideas of the Boundary Control method.

2.3 Controllability

Definition 1. System (2.5) is said to be controllable, if for any given $w \in \mathbb{C}^n$ there exists a finite positive integer T and a sequence of inputs $f(0), \dots, f(T-1)$ such that $x^f(T) = w$.

Equation (2.9) connects a control $f(i)$ defined for i from 0 to $k-1$ and the state $x^f(k)$ of the system at the step k . Recall that $x^f(k)$ is an $N \times 1$ vector, $Y(k)$ is an $N \times k$ matrix, $F(k)$ is a $k \times 1$ vector. Then we can solve equation (2.9) for $F(k)$ with arbitrary $x^f(k)$ and given $Y(k)$ if $k \geq N$ and $Y(k)$ has full rank N . This constitutes the well known *Kalman's controllability condition* (see, for example, [6; 7; 10; 15]).

Proposition 1 (Kalman's controllability condition). *System (2.5) is controllable if and only if the $N \times N$ matrix $Y(N) = (b|Mb|M^2b|\dots|M^{N-1}b)$ has rank N .*

The matrix $Y(N)$ is called a *controllability matrix*. Similar condition obviously holds for adjoint system (2.10): system (2.10) is controllable if and only if

$$\text{rank}(Y^\#(N)) = \text{rank}(c|M^*c|(M^*)^2c|\dots|(M^*)^{N-1}c) = N.$$

2.4 Operators W and R

Let us introduce the *control operator* $W : \mathbb{C}^{N+1} \rightarrow \mathbb{C}^N$,

$$(Wf) := x^f(N+1)$$

and the *response operator* $R : \mathbb{C}^\infty \rightarrow \mathbb{C}^\infty$ for system (2.5),

$$(Rf)(k) := y(k) = \langle x^f(k), c \rangle, \quad k = 1, 2, \dots \quad (2.12)$$

Similarly we introduce *control* and *response* operators for adjoint system (2.10):

$$W^\# : \mathbb{C}^{N+1} \rightarrow \mathbb{C}^N, \quad (W^\#g) := z^g(N+1),$$

$$R^\# : \mathbb{C}^\infty \rightarrow \mathbb{C}^\infty, \quad (R^\#g)(k) := w(k) = \langle z^g(k), b \rangle, \quad k = 1, 2, \dots \quad (2.13)$$

Notice that

$$\langle x^f(k), c \rangle = c^* x^f(k) = c^* \sum_{i=0}^{k-1} M^{k-1-i} b f(i) = \sum_{i=0}^{k-1} [c^* M^{k-1-i} b] f(i).$$

Therefore, if we denote $c^* M^k b$ by $r(k)$ then R takes the form:

$$(Rf)(k) = \sum_{j=0}^{k-1} f(j) r(k-1-j). \quad (2.14)$$

Likewise,

$$(R^\# g)(k) = \sum_{j=0}^{k-1} g(j) \overline{r(k-1-j)}. \quad (2.15)$$

2.5 Identification

In this section we show that it is possible to obtain the eigenvalues of M and coefficients of decomposition of vectors b and c in bases of eigenvectors of matrices M and M^* respectively from values $r(k)$.

Suppose $f(0) = 0$, $f(i) = 0$ for $i \geq N+1$. Let us introduce the shift operator S

$$\tilde{f}(k) = Sf(k) = f(k+1).$$

Then $\tilde{f}(i) = 0$ for $i \geq N$.

Since M and b are both constant (do not depend on k), then $x^{\tilde{f}}(k) = x^f(k+1)$ for $1 \leq k \leq N$.

Since $\tilde{f}(N) = 0$,

$$x^{\tilde{f}}(N+1) = Mx^{\tilde{f}}(N) = Mx^f(N+1),$$

which can be rewritten as

$$WSf = MWf.$$

This is true for all controls f with support on the set $1, \dots, N$.

Consider expressions for following scalar products:

$$\begin{aligned} \langle Wf, W^\#g \rangle &= \langle x^f(N+1), z^g(N+1) \rangle = (z^g(N+1))^* x^f(N+1) \\ &= (Y^\#(N)G)^* Y(N)F = G^* (Y^\#(N))^* Y(N)F, \end{aligned}$$

and

$$\begin{aligned}\langle W\tilde{f}, W^\#g \rangle &= \langle MWf, W^\#g \rangle = \langle Mx^f(N+1), z^g(N+1) \rangle \\ &= (z^g(N+1))^* Mx^f(N+1) = (Y^\#(N)G)^* MY(N)F \\ &= G^* (Y^\#(N))^* MY(N)F.\end{aligned}$$

Here F and G are control vectors:

$$F = (f(N), \dots, f(1))^T, \quad G = (g(N), \dots, g(1))^T.$$

Let us define two $N \times N$ matrices:

$$B = (Y^\#(N))^* Y(N), \quad (2.16)$$

$$A = (Y^\#(N))^* MY(N). \quad (2.17)$$

Now we can rewrite the above scalar products as

$$\langle Wf, W^\#g \rangle = G^* BF, \quad (2.18)$$

$$\langle W\tilde{f}, W^\#g \rangle = G^* AF. \quad (2.19)$$

Since

$$B = (Y^\#(N))^* Y(N) = (c|M^*c| \dots (M^*)^{N-1}c)^*(b|Mb| \dots |M^{N-1}b),$$

then

$$B_{ij} = ((M^*)^{i-1}c)^* M^{j-1}b = c^* M^{i-1} M^{j-1}b = c^* M^{i+j-2}b = r(i+j-2). \quad (2.20)$$

Also, using (2.17), we get

$$A_{ij} = r(i+j-1). \quad (2.21)$$

Matrices A and B are nonsingular if $Y(N)$, $Y^\#(N)$, or M are nonsingular. Therefore if A or B are not of full rank, then one or both of systems (2.5), (2.10) are not controllable. If $\text{rank}(A) = \text{rank}(B) = N$, then $\text{rank}(Y(N)) = \text{rank}(Y^\#(N)) = N$, and therefore both dynamical systems are controllable. This way we can find out if both systems are controllable from values of $r(k)$, and the order of the systems N .

2.5.1 Determining the order of the systems

On the other hand we can show that it is possible to find order N of systems (2.5), (2.10), if we know that both of those systems are controllable and we are given values $r(k)$ for large enough number k .

For this we introduce a sequence of matrices $B^{1 \times 1}, B^{2 \times 2}, \dots, B^{L \times L}, \dots$ of increasing sizes. Each matrix $B^{L \times L}$ is an $L \times L$ matrix, and $B_{ij}^{L \times L} = r(i + j - 2)$. It is easy to see that $\text{rank}(B^{L \times L}) \leq \text{rank}(B^{(L+1) \times (L+1)})$. Also, notice that $B^{L \times L} = (Y^\#(L))^* Y(L)$.

Theorem 1. *Assume that both systems (2.5), (2.10) are controllable. Then the systems have order N if and only if $\text{rank}(B^{N \times N}) = N$ and $\text{rank}(B^{(N+1) \times (N+1)}) = N$.*

Proof. Suppose that systems (2.5), (2.10) have order N and are both controllable. Then

$$\text{rank}(Y^\#(N)) = \text{rank}(Y(N)) = \text{rank}(Y^\#(N+1)) = \text{rank}(Y(N+1)) = N.$$

We know that $B^{L \times L} = (Y^\#(L))^* Y(L)$, therefore $\text{rank}(B^{N \times N}) = N$, $\text{rank}(B^{(N+1) \times (N+1)}) \leq N$. Since $\text{rank}(B^{L \times L}) \leq \text{rank}(B^{(L+1) \times (L+1)})$, we have $\text{rank}(B^{(N+1) \times (N+1)}) = N$.

Let us suppose that systems (2.5), (2.10) are controllable, but their order is unknown; and there is such number N , that $\text{rank}(B^{N \times N}) = \text{rank}(B^{(N+1) \times (N+1)}) = N$. Let D be an order of systems (2.5), (2.10). Let us show that $D = N$. Suppose that $N < D$. Then $\text{rank}(B^{(N+1) \times (N+1)}) = N + 1$. This contradicts our assumption. Let $N > D$. Then $\text{rank}(B^{N \times N}) = D < N$. This contradiction completes the proof. \square

So we can find order of system (2.5) and of adjoint system (2.10) by considering square matrices $B^{L \times L}$ of increasing size until we find the first number L_0 such that the matrix $B^{L_0 \times L_0}$ is singular. Then $L_0 - 1$ is the order of both dynamical systems.

2.5.2 Determining eigenvalues

In this section we show how to find eigenvalues of the matrix M of system (2.5) using the matrices A and B as in (2.20), (2.21) (Notice that the eigenvalues of M^*

are complex conjugates of the eigenvalues of M). We assume that both systems are controllable.

Suppose that matrix M has K eigenvalues $\{\lambda_i\}_{i=1}^K$, and each eigenvalue λ_i has multiplicity M_i . The matrices M and M^* both have N generalized eigenvectors:

$$M\phi_k^{(i)} = \lambda_i\phi_k^{(i)} + \phi_{k-1}^{(i)}, \quad (2.22)$$

$$M^*\psi_k^{(i)} = \bar{\lambda}_i\psi_k^{(i)} + \psi_{k+1}^{(i)}.$$

Note, that we assume $\phi_0^{(i)} = \psi_{M_i+1}^{(i)} = 0$, so that $\phi_1^{(i)}$ and $\psi_{M_i}^{(i)}$ are eigenvectors of M and M^* in a regular sense:

$$M\phi_1^{(i)} = \lambda_i\phi_1^{(i)}, \quad M^*\psi_{M_i}^{(i)} = \bar{\lambda}_i\psi_{M_i}^{(i)}.$$

Since we assumed that both systems (2.5), (2.10) are controllable, then matrices $Y(N), Y^\#(N)$ are not singular, and we can multiply the equality (2.22) by $(Y^\#(N))^*$ from the left:

$$\begin{aligned} & (Y^\#(N))^* M (Y(N)Y^{-1}(N)) \phi_k^{(i)} \\ &= \lambda_i (Y^\#(N))^* (Y(N)Y^{-1}(N)) \phi_k^{(i)} + (Y^\#(N))^* (Y(N)Y^{-1}(N)) \phi_{k-1}^{(i)}. \end{aligned} \quad (2.23)$$

Recall that

$$A = (Y^\#(N))^* M Y(N); \quad B = (Y^\#(N))^* Y(N).$$

Then (2.23) transforms to

$$A Y^{-1}(N) \phi_k^{(i)} = \lambda_i B Y^{-1}(N) \phi_k^{(i)} + B Y^{-1}(N) \phi_{k-1}^{(i)}.$$

Therefore the generalized eigenproblem

$$(A - \lambda_i B)^k F_k^{(i)} = 0$$

has the same eigenvalues λ_i with the same multiplicities M_i as the matrix M .

The generalized eigenvectors $F_k^{(i)}$ are connected to generalized eigenvectors of M by the equality:

$$F_k^{(i)} = Y^{-1}(N) \phi_k^{(i)}.$$

Thus

$$\phi_k^{(i)} = Y(N)F_k^{(i)}.$$

Notice that for such control $f_k^{(i)}$ that $F_k^{(i)} = (f_k^{(i)}(N), \dots, f_k^{(i)}(1))^T$, $x^{f_k^{(i)}}(N+1) = Y(N)F_k^{(i)}$. Therefore the control $f_k^{(i)}$ drives our system to the state $x^{f_k^{(i)}}(N+1) = \phi_k^{(i)}$.

Using the same reasoning we can deduce that solving the generalized eigenproblem

$$(A^* - \bar{\lambda}_i B^*)^{M_i+1-k} G_k^{(i)} = 0$$

we find control vectors $G_k^{(i)}$ that take adjoint system (2.10) to $z^{g_k^{(i)}}(N+1) = \psi_k^{(i)}$.

We summarize this subsection in the following theorem:

Theorem 2. *Suppose that both systems (2.5), (2.10) are controllable and have order N . Then we can find eigenvalues of the matrix M and their multiplicities by solving the following generalized eigenproblem:*

$$(A - \lambda_i B)^k F_k^{(i)} = 0, \quad (2.24)$$

where A and B are $N \times N$ matrices,

$$A_{ij} = r(i+j-1), \quad B_{ij} = r(i+j-2), \quad 1 \leq i, j \leq N.$$

$F_k^{(i)}$ gives us a control yielding the corresponding eigenvector of matrix M : if we take a control f such that $F_k^{(i)} = (f(N), \dots, f(0))^T$, then $x^f(N+1) = \phi_k^{(i)}$. Each generalized eigenvector of our generalized eigenproblem (2.24) corresponds to an eigenvector of matrix M .

Determining generalized eigenvectors of the adjoint generalized eigenproblem

$$(A^* - \bar{\lambda}_i B^*)^{M_i+1-k} G_k^{(i)} = 0, \quad (2.25)$$

we obtain control vectors for the second system that yield $z^{g_k^{(i)}}(N+1) = \psi_k^{(i)}$.

Using Theorems 1 and 2 we can recover order of the systems and their eigenvalues with multiplicities, knowing values $r(0), \dots, r(L)$ for large enough number L .

2.5.3 Determining decompositions of vectors b and c in bases made of generalized eigenvectors of M and M^*

Here we assume that generalized eigenvectors $\psi_k^{(i)}$ of M^* are biorthogonal to generalized eigenvectors $\phi_k^{(i)}$ of M .

In this section we use controls described in the previous section. These controls $f_k^{(i)}$ and $g_k^{(i)}$ take our system to the corresponding generalized eigenvectors of M and M^* : $W f_k^{(i)} = \phi_k^{(i)}$, $W^\# g_k^{(i)} = \psi_k^{(i)}$.

Recall, that those controls can be found using Theorem 2 from generalized eigenvectors of two generalized eigenproblems (2.24), (2.25). Then, $f_k^{(i)}$ is such that $F_k^{(i)} = (f_k^{(i)}(N), f_k^{(i)}(N-1), \dots, f_k^{(i)}(1))^T$; and $g_k^{(i)}$ is such that $G_k^{(i)} = (g_k^{(i)}(N), g_k^{(i)}(N-1), \dots, g_k^{(i)}(1))^T$. To get the controls such that the eigenvectors $\phi_k^{(i)}$ and $\psi_k^{(i)}$ are biorthogonal, it is necessary to chose vectors $F_k^{(i)}$ and $G_k^{(i)}$ so that $(G_k^{(i)})^* B F_k^{(i)} = 1$, and $(G_k^{(i)})^* B F_l^{(j)} = 0$ when $j \neq i$, $k \neq l$.

Formulas (2.12), (2.14) and (2.13), (2.15) give us:

$$(R f_k^{(i)})(N+1) = \sum_{j=1}^N f_k^{(i)}(j) r(N-j) = \langle \phi_k^{(i)}, c \rangle, \quad (2.26)$$

$$(R^\# g_k^{(i)})(N+1) = \sum_{j=1}^N g_k^{(i)}(j) \overline{r(N-j)} = \langle \psi_k^{(i)}, b \rangle. \quad (2.27)$$

Notice that $\langle \psi_k^{(i)}, b \rangle$ are coefficients in a decomposition of the vector b in a basis of eigenvectors of matrix M :

$$b = \sum_{i,k} \langle \psi_k^{(i)}, b \rangle \phi_k^{(i)}.$$

Also, $\langle \phi_k^{(i)}, c \rangle$ are coefficients in a decomposition of the vector c in a basis made by eigenvectors of matrix M^* :

$$c = \sum_{i,k} \langle \phi_k^{(i)}, c \rangle \psi_k^{(i)}.$$

Therefore we can use formulas (2.26), (2.27) to find the decomposition coefficients $\langle \phi_k^{(i)}, c \rangle$, $\langle \psi_k^{(i)}, b \rangle$, if we know the kernel $r(k)$ of the response operator R (we find $f_k^{(i)}$ and $g_k^{(i)}$ using Theorem 2).

Theorem 3. *The coefficients in a decomposition of the vector $b = \sum_{ik} b_k^{(i)} \phi_k^{(i)}$ in a basis of eigenvectors of M are given by the formula:*

$$b_k^{(i)} = \sum_{j=1}^N g_k^{(i)}(j) \overline{r(N-j)}.$$

The coefficients in a decomposition of the vector $c = \sum_{ik} c_k^{(i)} \psi_k^{(i)}$ are

$$c_k^{(i)} = \sum_{j=1}^N f_k^{(i)}(j) r(N-j).$$

Theorems 1, 2, 3 allow us to extract information about dynamical systems (2.5), (2.10) from the kernel of the response operator R . We can find the order of the systems, the eigenvalues of M , and decompositions of vectors b and c in bases of eigenvectors of M and M^* respectively. However, using only values of $r(k)$ we cannot find eigenvectors of M and M^* ; instead we can find the controls $f_k^{(i)}$ for system (2.5) yielding the eigenvectors $\phi_k^{(i)}$ of M and the controls $g_k^{(i)}$ for system (2.10) yielding the eigenvectors $\psi_k^{(i)}$ of M^* .

2.6 Equivalence of dynamical systems with respect to a transformation of variable

Let us apply a transformation of variable $x(k) = Q\hat{x}(k)$ with a nonsingular matrix Q to system (2.5):

$$\begin{aligned} Q\hat{x}(k+1) &= MQ\hat{x}(k) + bf(k), \\ y(k) &= \langle x(k), c \rangle = \langle Q\hat{x}(k), c \rangle = \langle \hat{x}(k), Q^*c \rangle. \end{aligned}$$

Multiplying both sides of the first equation by Q^{-1} we get:

$$\hat{x}(k+1) = Q^{-1}MQ\hat{x}(k) + Q^{-1}bf(k). \quad (2.28)$$

The kernel of the response operator R of system (2.5) is

$$r(k) = c^* M^k b.$$

New system has the same structure, with matrix M replaced by $Q^{-1}MQ$, vector b by $Q^{-1}b$, and c by Q^*c . The kernel of the response operator of system (2.28), is

$$\hat{r}(k) = (Q^*c)^* (Q^{-1}MQ)^k Q^{-1}b = c^* M^k b = r(k).$$

Therefore the kernel $r(k)$ does not change when we apply a transformation of a variable to the system. This means that we can work with a convenient form of matrix M , for example Jordan normal form.

2.7 Connection with the original problem

In this section we show that the problem of estimation of coefficients $a_n(t)$ and λ_n of function (2.4) is equivalent to the problem of identification of parameters of system (2.5). First let us choose a form of the matrix M and vectors b and c so that the kernel $r(k) = c^* M^k b$ of response operator R (2.12) has the same form as signal

$$r(t) = \sum_{n=1}^K a_n(t) \lambda_n^t. \quad (2.29)$$

2.7.1 Matrix M

Let us assume that $N \times N$ matrix M has the following structure:

$$M = T\Lambda T^{-1}, \quad (2.30)$$

where Λ is a Jordan canonical form of M :

$$\Lambda = \text{diag}(J_1, \dots, J_K); \quad J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ \vdots & & \ddots & & \ddots \\ 0 & \dots & 0 & \lambda_i & 1 \\ 0 & & \dots & 0 & \lambda_i \end{pmatrix}.$$

By M_i denote the order of each Jordan block J_i .

$$T = \text{diag}(1, \lambda_1^{-1}, \dots, \lambda_1^{-(M_1-1)}, 1, \lambda_2^{-1}, \dots, \lambda_2^{-(M_2-1)}, \dots, 1, \lambda_K^{-1}, \dots, \lambda_K^{-(M_K-1)}).$$

From formula (2.30) it is easy to see that columns of the matrix T are generalized eigenvectors $\phi_k^{(i)}$ of M ; columns of the matrix $(T^{-1})^*$ are generalized eigenvectors $\psi_k^{(i)}$ of M^* . Notice, that eigenvectors constructed this way are biorthogonal:

$$\langle \phi_k^{(i)}, \psi_l^{(j)} \rangle = \begin{cases} 1, & \text{when } k = l \text{ and } i = j \\ 0, & \text{otherwise.} \end{cases}$$

We suppose that M has K nonzero eigenvalues λ_i , each eigenvalue has algebraic multiplicity M_i , and geometric multiplicity 1. Eigenvalue having geometric multiplicity 1 means that only one Jordan block J_j corresponds to this eigenvalue. In what follows we show that this condition is necessary for systems (2.5), (2.10) to be controllable.

It is easy to check that the matrix M has a block form:

$$M = T\Lambda T^{-1} = \text{diag}(D_1, \dots, D_k), \text{ where } D_i = \lambda_i \begin{pmatrix} 1 & 1 & 0 & \dots \\ 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 \\ 0 & \dots & 0 & 1 \end{pmatrix},$$

D_i are $M_i \times M_i$ matrices.

Then,

$$M^m = \text{diag}(D_1^m, \dots, D_K^m); \tag{2.31}$$

$$D_i^m = \lambda_i^m \begin{pmatrix} 1 & \binom{m}{1} & \binom{m}{2} & \dots & \binom{m}{M_i-1} \\ 0 & 1 & \binom{m}{1} & \binom{m}{2} & \dots \\ & & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & 1 & \binom{m}{1} \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix}.$$

Here $\binom{k}{j}$ are binomial coefficients $\binom{k}{j} = \frac{k!}{(k-j)!j!}$. We assume that $\binom{k}{j} = 0$ when $j > k$.

2.7.2 Dynamical systems and the controllability condition

We work with two linear dynamical systems of the same structure as systems (2.5), (2.10). We use matrix $M = T\Lambda T^{-1}$ as in (2.30):

$$x(k+1) = T\Lambda T^{-1}x(k) + bf(k), \quad x(0) = 0, \quad y(k) = \langle x(k), c \rangle = c^*x(k); \quad (2.32)$$

$$z(k+1) = (T^{-1})^*\Lambda^*T^*z(k) + cg(k), \quad z(0) = 0, \quad w(k) = \langle z(k), c \rangle = b^*z(k). \quad (2.33)$$

We denote entries of vectors b, c by $\beta_k^{(i)}$ and $\gamma_k^{(i)}$ in the following way:

$$b = (\beta_0^{(1)}, \dots, \beta_{M_1-1}^{(1)}, \beta_0^{(2)}, \dots, \beta_{M_2-1}^{(2)}, \dots, \beta_0^{(K)}, \dots, \beta_{M_K-1}^{(K)})^T; \quad (2.34)$$

$$c = (\gamma_0^{(1)}, \dots, \gamma_{M_1-1}^{(1)}, \gamma_0^{(2)}, \dots, \gamma_{M_2-1}^{(2)}, \dots, \gamma_0^{(K)}, \dots, \gamma_{M_K-1}^{(K)})^T. \quad (2.35)$$

Then subvectors $b^{(j)} = (\beta_0^{(j)}, \dots, \beta_{M_j-1}^{(j)})^T$ and $c^{(j)} = (\gamma_0^{(j)}, \dots, \gamma_{M_j-1}^{(j)})^T$ of b and c correspond to j -th block D_j of M .

Lemma 1. *System (2.32) is controllable if and only if all eigenvalues of M are not equal to zero ($\lambda_i \neq 0 \forall i$), have geometric multiplicity 1 ($\lambda_i \neq \lambda_j \forall i \neq j$), and $\beta_{M_i-1}^{(i)} \neq 0$ (the last entry of each subvector $b^{(i)}$ is not equal to zero).*

System (2.33) is controllable if and only if all eigenvalues of M are not equal to zero ($\lambda_i \neq 0 \forall i$), have geometric multiplicity 1 ($\lambda_i \neq \lambda_j \forall i \neq j$), and $\gamma_0^{(i)} \neq 0$ (the first entry of each subvector $c^{(i)}$ is not equal to zero).

We present a proof of this Lemma in Appendix 2.A.

2.7.3 Kernel of the response operator of system (2.32)

From section 2.4 we know that kernel $r(k)$ of response operator R for system (2.5) has the form $r(m) = c^*M^m b$. For our choice of matrix M and vectors b and c (see

(2.31), (2.34), (2.35)):

$$\begin{aligned}
r(m) &= c^* M^m b = \sum_{i=1}^K (c^{(i)})^* (D_i)^m b^{(i)} \\
&= \sum_{i=1}^K \lambda_i^m (\overline{\gamma_0^{(i)}}, \dots, \overline{\gamma_{M_i-1}^{(i)}}) \begin{pmatrix} 1 & \binom{m}{1} & \binom{m}{2} & \cdots & \binom{m}{M_i-1} \\ 0 & 1 & \binom{m}{1} & \binom{m}{2} & \cdots \\ & & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & \binom{m}{1} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} (\beta_0^{(i)}, \dots, \beta_{M_i-1}^{(i)})^T \\
&= \sum_{i=1}^K \lambda_i^m \left[\sum_{j=0}^{M_i-1} \binom{m}{j} \left(\sum_{l=0}^{M_i-1-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right) \right]. \quad (2.36)
\end{aligned}$$

Notice, that $\binom{m}{j}$ is a j -th degree polynomial in m . So, from the formula above it follows that $r(m)$ is a combination of λ_i^m with polynomial coefficients. Thus $r(m)$ has the same form as the signal $r(t)$ in (2.29). The kernel (2.36) of the response operator of system (2.32) corresponds to a signal

$$r(t) = \sum_{i=1}^K \left[\sum_{j=0}^{M_i-1} \binom{t}{j} \left(\sum_{l=0}^{M_i-1-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right) \right] \lambda_i^t.$$

Notice that the coefficient $\left[\sum_{j=0}^{M_i-1} \binom{t}{j} \left(\sum_{l=0}^{M_i-1-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right) \right]$ is a polynomial of degree $M_i - 1$ with respect to t . Therefore it is obvious that for every controllable system (2.32) there is a signal $r(t)$ of form (2.29) that is equal to the kernel $r(m) = c^* M^m b$.

We would like to show that for any signal $r(t) = \sum_{i=1}^K a_i(t) \lambda_i^t$ where $a_i(t)$ are polynomials, there exists a controllable system (2.32) such that the kernel of the response operator of system (2.32) is $r(m) = \sum_{i=1}^K a_i(m) \lambda_i^m$.

Let $a_i(t) = \sum_{j=0}^{p_i} \alpha_{ij} t^j$, for $0 \leq i \leq K$. Suppose $\alpha_{ip_i} \neq 0$, so that polynomial $a_i(t)$ has the degree p_i . Since $\binom{t}{j}$ is a j -th degree polynomial, then

$$a_i(t) = \sum_{j=0}^{p_i} \alpha_{ij} t^j = \sum_{j=0}^{p_i} w_{ij} \binom{t}{j} \quad (2.37)$$

for some set $\{w_{ij}\}$, $w_{ip_i} \neq 0$.

It is rather easy to show that for any such set $\{w_{ij}\}$ we can find two sequences $\beta_j^{(i)}, \gamma_j^{(i)}$ such that

$$w_{ij} = \sum_{l=0}^{p_i-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \quad (2.38)$$

and $\beta_{p_i}^{(i)} \neq 0, \gamma_0^{(i)} \neq 0$ (this is necessary for the systems we are constructing to be controllable). The proof of the following lemma is straightforward.

Lemma 2. *Given $p+1$ numbers $\{w_j\}_{j=0}^p$ with $w_p \neq 0$, one can find $2(p+1)$ numbers $\{\beta_j\}_{j=0}^p, \{\gamma_j\}_{j=0}^p$ such that $\beta_p \neq 0, \gamma_0 \neq 0, w_j = \sum_{l=0}^{p-j} \overline{\gamma_l} \beta_{l+j}$.*

Combining (2.29),(2.37), and (2.38) we get

$$r(t) = \sum_{i=1}^K \left[\sum_{j=0}^{p_i} \binom{t}{j} \left(\sum_{l=0}^{p_i-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right) \right] \lambda_i^t.$$

This is equivalent to (2.36) if we replace $M_i - 1$ with p_i . Therefore, if we take a matrix M as in (2.30) with λ_i the same as λ_i in our signal with multiplicities $M_i = p_i + 1$, where p_i is the degree of $a_i(t)$; and vectors b and c so that $w_{ij} = \sum_{l=0}^{p_i-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)}$, then the kernel of the response operator of system (2.32) is equal to $r(t)$.

2.7.4 Equivalence of the problem of signal decomposition for signal (2.29) to the identification problem for a dynamical system (2.32).

Returning to our original problem: given values of the signal $r(t)$ at $t = 0, 1, \dots$ we want to find number and values of the poles λ_i , and their polynomial amplitudes $a_i(t)$.

By assumption our signal satisfies

$$r(t) = \sum_{i=1}^K a_i(t) \lambda_i^t,$$

with some unknown $K, a_i(t), \lambda_i$. We need to find all these unknown values. In the previous section we showed that then $r(t)$ also satisfies

$$r(t) = \sum_{i=1}^K \left[\sum_{j=0}^{p_i} \binom{t}{j} \left(\sum_{l=0}^{p_i-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right) \right] \lambda_i^t \quad (2.39)$$

with $\gamma_0^{(i)} \neq 0$ and $\beta_{p_i}^{(i)} \neq 0$. Therefore, $r(t)$ is the kernel of the response operator of some controllable (notice that all the conditions of Lemma 1 are satisfied) system of type (2.5).

Thus, we can apply Theorems 1, 2, and 3 to find $\lambda_i, \gamma_l^{(i)}, \beta_l^{(i)}$. First we apply Theorem 1 to find the order N of the system. All that we need for that are values $r(k)$ for k from 0 to $2N$ to construct a sequence of matrices $B^{j \times j}$. After that we can apply Theorem 2; it gives us λ_i , number K , and $p_i = M_i - 1$. Using Theorem 3 we find $\langle \phi_k^{(i)}, c \rangle$ and $\langle \psi_k^{(i)}, b \rangle$. For system (2.32) $\langle \phi_k^{(i)}, c \rangle = \overline{\gamma_{k-1}^{(i)}} \lambda_i^{-(k-1)}$, $\langle \psi_k^{(i)}, b \rangle = \overline{\beta_{k-1}^{(i)}} \lambda_i^{k-1}$. Since we already found all λ_i , these equalities allow us to find $\beta_j^{(i)}, \gamma_j^{(i)}$. Now all the coefficients in (2.39) are known. Thus, we have decomposed the signal $r(t)$: we have found $\lambda_i, K, a_i(t) = \sum_{j=0}^{p_i} \binom{t}{j} \left(\sum_{l=0}^{p_i-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right)$.

We describe an algorithm that allows us to recover the parameters of the signal $r(t)$ from $2N + 1$ equispaced samples.

Algorithm 1.

step 1 Construct a sequence of matrices of increasing size $B^{p \times p}$. $B^{p \times p}$ is a $p \times p$ matrix, $B_{ij}^{p \times p} = r(i + j - 2)$, $i, j = 1, \dots, p$. Find N such that $B^{N \times N}$ is nonsingular, and $B^{(N+1) \times (N+1)}$ is singular. Then N is the order of our systems. We use this number on the next step.

step 2 Consider two $N \times N$ matrices A and B : $A_{ij} = r(i + j - 1)$, $B_{ij} = r(i + j - 2)$, $i, j = 1, \dots, N$. Find eigenvectors and eigenvalues of the generalized eigenproblem

$$(A - \lambda_i B)^k F_k^{(i)} = 0$$

and

$$(A^* - \overline{\lambda_i} B^*)^{M_i+1-k} G_k^{(i)} = 0$$

so that $(G_k^{(i)})^* B F_k^{(i)} = 1$, and $(G_k^{(i)})^* B F_l^{(j)} = 0$ when $j \neq i, k \neq l$. We obtain:

- Number and values of λ_i . Those values λ_i are the poles of our signal. Number of distinct eigenvalues is the number K in formula (2.29).

- Algebraic multiplicities M_i of λ_i correspond to degrees of the polynomials $a_i(t)$ ($p_i = M_i - 1$). They are used on the next step.
- The generalized eigenvectors $F_k^{(i)}$ and $G_k^{(i)}$ are used on the next step to find $\beta_j^{(i)}$ and $\gamma_j^{(i)}$.

step 3 Theorem 3 allows us to find $\langle \phi_k^{(i)}, c \rangle$ and $\langle \psi_k^{(i)}, b \rangle$ knowing $r(k)$ and vectors $F_k^{(i)}$, $G_k^{(i)}$.

$$\langle \psi_k^{(i)}, b \rangle = \sum_{j=1}^N g_k^{(i)}(j) \overline{r(N-j)},$$

$$\langle \phi_k^{(i)}, c \rangle = \sum_{j=1}^N f_k^{(i)}(j) r(N-j).$$

Here $F_k^{(i)} = (f_k^{(i)}(N), f_k^{(i)}(N-1), \dots, f_k^{(i)}(1))^T$,
 $G_k^{(i)} = (g_k^{(i)}(N), g_k^{(i)}(N-1), \dots, g_k^{(i)}(1))^T$.

Since for our kind of system $\langle \phi_k^{(i)}, c \rangle = \overline{\gamma_{k-1}^{(i)} \lambda_i^{-(k-1)}}$, $\langle \psi_k^{(i)}, b \rangle = \overline{\beta_{k-1}^{(i)} \lambda_i^{k-1}}$, $\beta_k^{(i)}$ and $\gamma_k^{(i)}$ are given by

$$\beta_k^{(i)} = \overline{\langle \psi_{k+1}^{(i)}, b \rangle} \lambda_i^{-k}, \quad \gamma_k^{(i)} = \overline{\langle \phi_{k+1}^{(i)}, c \rangle} \lambda_i^k.$$

Then the polynomial coefficients $a_i(t)$ can be found using the formula

$$a_i(t) = \sum_{j=0}^{M_i-1} \binom{t}{j} \left(\sum_{l=0}^{M_i-1-j} \overline{\gamma_l^{(i)}} \beta_{l+j}^{(i)} \right).$$

Appendix 2.A. The proof of Lemma 1

Lemma 1. System (2.32) is controllable if and only if all eigenvalues of M are not equal to zero ($\lambda_i \neq 0 \forall i$), have geometric multiplicity 1 ($\lambda_i \neq \lambda_j \forall i \neq j$), and $\beta_{M_i-1}^{(i)} \neq 0$ (the last entry of each subvector $b^{(i)}$ is not equal to zero).

System (2.33) is controllable if and only if all eigenvalues of M are not equal to zero ($\lambda_i \neq 0 \forall i$), have geometric multiplicity 1 ($\lambda_i \neq \lambda_j \forall i \neq j$), and $\gamma_0^{(i)} \neq 0$ (the first entry of each subvector $c^{(i)}$ is not equal to zero).

Proof. Controllability of (2.32) is equivalent to nonsingularity of matrix $Y(N) = (b|Mb| \dots |M^{N-1}b)$ (see Proposition 1 in section 2.3).

Let us find the condition for $\det(Y(N)) \neq 0$.

Let

$$\tilde{Y} = (b|(M - \lambda_1 I)b| \dots |(M - \lambda_1 I)^{N-1}b).$$

One can see that $\det(Y(N)) = \det(\tilde{Y})$. Since $M = \text{diag}(D_1, \dots, D_K)$, then

$$(M - \lambda_1 I)^j = \text{diag}((D_1 - \lambda_1 I)^j, (D_2 - \lambda_1 I)^j, \dots, (D_K - \lambda_1 I)^j).$$

We can rewrite \tilde{Y} as

$$\tilde{Y} = \begin{pmatrix} b^{(1)} & (D_1 - \lambda_1 I)b^{(1)} & \dots & (D_1 - \lambda_1 I)^{N-1}b^{(1)} \\ b^{(2)} & (D_2 - \lambda_1 I)b^{(2)} & \dots & (D_2 - \lambda_1 I)^{N-1}b^{(2)} \\ \vdots & \vdots & & \vdots \\ b^{(K)} & (D_K - \lambda_1 I)b^{(K)} & \dots & (D_K - \lambda_1 I)^{N-1}b^{(K)} \end{pmatrix}.$$

Notice, that D_1 is an $M_1 \times M_1$ upper triangular matrix with λ_1 on its main diagonal. Thus $(D_1 - \lambda_1 I)^j = 0$ for any $j \geq M_1$. The matrix \tilde{Y} can be called a "block lower triangular" matrix;

$$\begin{aligned} \det(Y(N)) &= \det(\tilde{Y}) \\ &= \det(b^{(1)}|(D_1 - \lambda_1 I)b^{(1)}| \dots |(D_1 - \lambda_1 I)^{M_1-1}b^{(1)}) \det(Y_1), \end{aligned}$$

$$Y_1 = \begin{pmatrix} (D_2 - \lambda_1 I)^{M_1}b^{(2)} & \dots & (D_2 - \lambda_1 I)^{N-1}b^{(2)} \\ \vdots & & \vdots \\ (D_K - \lambda_1 I)^{M_1}b^{(K)} & \dots & (D_K - \lambda_1 I)^{N-1}b^{(K)} \end{pmatrix}.$$

The first matrix $(b^{(1)}|(D_1 - \lambda_1 I)b^{(1)}| \dots |(D_1 - \lambda_1 I)^{M_1-1}b^{(1)})$ is a square "upper-left" triangular matrix with $\lambda_1^j \beta_{M_1-1}^{(1)}$ on its antidiagonal. Thus, the first determinant $\det(b^{(1)}|(D_1 - \lambda_1 I)b^{(1)}| \dots |(D_1 - \lambda_1 I)^{M_1-1}b^{(1)})$ is not zero if and only if $\beta_{M_1-1}^{(1)} \neq 0$ and $\lambda_1 \neq 0$.

Notice, that it is necessary that $\lambda_i \neq \lambda_1$ for $i = 2, 3, \dots, K$ in order for the second matrix Y_1 to be nonsingular, otherwise one of the rows of Y_1 contain only zeros.

Therefore matrix $Y(N)$ is nonsingular if and only if $\beta_{M_1-1}^{(1)} \neq 0$, $\lambda_1 \neq 0$, $\lambda_1 \neq \lambda_i$ for $i = 2, 3, \dots, K$, and $\det(Y_1) \neq 0$.

Let us rename columns of Y_1 in the following way:

$$\bar{b} := \begin{pmatrix} (D_2 - \lambda_1)^{M_1} b^{(2)} \\ \vdots \\ (D_K - \lambda_1)^{M_1} b^{(K)} \end{pmatrix}; \quad \bar{M} = \text{diag}((D_2 - \lambda_1 I), \dots, (D_K - \lambda_1 I)).$$

Then

$$Y_1 = (\bar{b} | \bar{M} \bar{b} | \dots | \bar{M}^{N-1-M_1} \bar{b}).$$

Notice that Y_1 is the same type of matrix as $Y(N)$, with \bar{M} being a block diagonal upper triangular matrix with diagonal entries $\lambda_i - \lambda_1$ ($i \geq 2$). If we repeat the same procedure for the new matrix Y_1 , using $\lambda_2 - \lambda_1$ instead of λ_1 , we see that $\det(Y_1) \neq 0$ if and only if

$$\beta_{M_2-1}^{(2)} \neq 0, \quad \lambda_2 \neq 0, \quad \lambda_i \neq \lambda_2 \text{ for } i \geq 3; \quad \det(Y_2) \neq 0,$$

where

$$Y_2 = (\bar{\bar{b}} | \bar{\bar{M}} \bar{\bar{b}} | \dots | \bar{\bar{M}}^{N-1-M_1-M_2} \bar{\bar{b}}),$$

$$\bar{\bar{b}} := \begin{pmatrix} (D_3 - \lambda_2)^{M_2} \bar{b}^{(3)} \\ \vdots \\ (D_K - \lambda_2)^{M_2} \bar{b}^{(K)} \end{pmatrix}; \quad \bar{\bar{M}} = \text{diag}((D_3 - \lambda_2 I), \dots, (D_K - \lambda_2 I)).$$

Using the above scheme K times we obtain the statement of the first part of the lemma. The proof of the second part is similar. \square

Bibliography

- [1] Avdonin S, Belishev M (1996) Boundary control and dynamical inverse problem for nonselfadjoint Sturm-Liouville operator. *Control and Cybernetics* 25:429–440
- [2] Avdonin S, Lenhart S, Protopopescu V (2002) Schrödinger equation by the Boundary Control method. *Inverse Problems* 18:41–57
- [3] Avdonin S, Bulanova A, Nicolsky D (2009) Boundary control approach to the spectral estimation problem. the case of simple poles. *Sampling Theory in Signal and Image Processing* Accepted
- [4] Badeau R, David B, Richard G (2006) High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials. *IEEE transactions on signal processing* 54(4):1341–1350
- [5] Belishev M (1997) Boundary Control method in reconstruction of manifolds and metrics (the BC method). *Inverse Problems* 13:R1–R45
- [6] Costa O, Fragoso M, Marques R (2005) *Discrete-Time Markov Jump Linear Systems*. Springer-Verlag, London
- [7] Elaydi SN (1999) *An Introduction to Difference Equations*, 2nd edn. Springer-Verlag, New York
- [8] Hua Y, Sarkar TK (1990) Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE transactions of acoustics, speech, and signal processing* 38(5):814–824
- [9] Hua Y, Gershman AB, Cheng Q (eds) (2004) *High-Resolution and Robust Signal Processing*. Marcel Dekker, New York, Basel
- [10] Krabs W, Pickl SW (2003) *Analysis, Controllability and Optimization of Time-Discrete Systems and Dynamical Games*. Springer-Verlag, Berlin Heidelberg

- [11] Marple SL (1987) *Digital Spectral Analysis with Applications*. Prentice-Hall
- [12] Nagesha V, Kay S (1994) On frequency estimation with the IQML algorithm. *IEEE Trans Signal Processing* 42(9):2509–2513
- [13] de Prony BGR (1795) Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique* 1(22):24–76
- [14] Roy R, Paulraj A, Kailath T (1986) Multiple emitter location and signal parameter estimation. *IEEE Trans Acoust, Speech, Signal Process* 34(5):1340–1342
- [15] Sarachik PE (1997) *Principles of Linear Systems*. Cambridge University Press
- [16] Sarkar TK, Wicks MC, Salazar-Palma M, Bonneau RJ (2003) *Smart Antennas*. John Wiley & Sons, Hoboken, New Jersey
- [17] Schmidt R (1986) Multiple emitter location and signal parameter estimation. *IEEE Trans Antennas Propag* 34(3):276–280

Chapter 3

Optimal quadrature formulae related to solutions of initial boundary value problems¹

Abstract

An approach to the construction of optimal quadrature formulae for the case of an integrand determined by a solution of a certain initial boundary value problem is presented. Several examples of initial boundary value problems are considered.

3.1 Introduction

Let Ω be a bounded subset of \mathbb{R}^n with a piecewise smooth boundary Γ . Let Y be a certain class of real functions summable in Ω . Let us consider a linear functional $l(y)$, $y \in Y$:

$$l(y) = \int_{\Omega} y(x) dx - \sum_{k=1}^N c_k y(x_k),$$

where $c_k \in \mathbb{R}$; $x_k \in \Omega$. Let us define $d(Y, N)$ for a given class Y and a fixed number of points N :

$$d(Y, N) = \inf_{\{c_k, x_k\}} \sup_{y \in Y} |l(y)|.$$

The problem of determining the value $d(Y, N)$ is a classical problem of quadrature theory. There are many works dedicated to this problem (see [11; 7; 3; 12]). Most notably, S.L. Sobolev [11] considered the portion of this problem concerned with finding $\sup_{y \in Y} |l(y)|$ for the case of Y being a unit ball in space $H^m(\Omega)$. We consider this problem for Y defined as a set of solutions of a certain initial boundary value problem. The study of this problem is of interest for practical applications. We present several examples of estimating $\sup_{y \in Y} |l(y)|$ and finding $\inf_{\{c_k\}} \sup_{y \in Y} |l(y)|$ for a given set of quadrature points c_k . We extend the results of the previous works in this direction [9], [1] to more general classes of initial boundary value problems.

¹S.A. Avdonin, A.S. Bulanova, and D.A. Ovsyannikov, *Optimal quadrature formulae related to solutions of initial boundary value problems*, Vestnik St. Petersburg University. Series 10. Applied Mathematics, Mechanics, Control Processes (2008), no. 2.

3.2 A maximization problem in the case of a parabolic equation

Let $T > 0$, $Q = \Omega \times (0, T)$, $\Sigma = \Gamma \times (0, T)$. Let functions a_{ij} be continuous in \overline{Q} and satisfy the condition

$$\sum_{i,j=1}^n a_{ij}(x, t) \xi_i \xi_j \geq \alpha \sum_{i=1}^n \xi_i^2$$

for some $\alpha > 0$, where $x \in Q$, $\xi_i \in \mathbb{R}$ for $i = 1 \dots, n$.

In Sec. 3.2.1 we consider an initial boundary value problem controlled by the initial data; in Sec. 3.2.2 – a problem controlled by the boundary conditions.

3.2.1 Control by the initial conditions

Consider an initial boundary value problem

$$\frac{\partial y}{\partial t} = A(t)y \text{ in } Q, \quad y|_{\Sigma} = 0, \quad y|_{t=0} = v. \quad (3.1)$$

Here

$$A(t) = \sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij}(x, t) \frac{\partial y}{\partial x_j} \right) + a_0(x, t)y,$$

$a_0 \in C(\overline{Q})$, $v \in U$, U is a closed bounded convex set in space $L^2(\Omega)$.

It is known (see [4]), that for smooth enough coefficients a_{ij} , a_0 there exists a unique weak solution of problem (3.1), such that

$$y(\cdot, T) \in H_0^m(\Omega), \quad m > \frac{n}{2}. \quad (3.2)$$

By the Embedding Theorem (see, e.g. [10]), it follows that $y(\cdot, T) \in C(\overline{\Omega})$ whenever condition (3.2) is satisfied. Therefore the following functional is well defined:

$$J(v) = \left| \int_{\Omega} y(x, T) \varphi(x) dx - \sum_{k=1}^N c_k y(x_k, T) \right|^2,$$

where $\varphi \in L^2(\Omega)$, $c_k \in \mathbb{R}$, $x_k \in \Omega$. Notice that

$$|J(v)| \leq M_1 \|y(\cdot, T)\|_{C(\overline{\Omega})}^2 \leq M_2 \|v\|_{L^2(\Omega)}^2, \quad v \in L^2(\Omega), \quad M_1, M_2 > 0. \quad (3.3)$$

Our problem reduces to the problem of finding a function $u \in U$, such that $J(u) = \sup_{v \in U} J(v)$. It is more convenient to rewrite the functional $J(v)$ as

$$J(v) = |l(v)|^2, \quad (3.4)$$

where

$$l(v) = \int_{\Omega} y(x, T) f(x) dx, \quad f(x) = \varphi(x) - \sum_{k=1}^N c_k \delta(x - x_k).$$

Since $J(v)$ is a square of a linear bounded functional (see formula (3.3)), then it is weakly continuous. It follows from convexity and closedness of the set U that this set is weakly closed. Together with boundedness of the set U this implies existence of the optimal element $u \in U$ (see [2]).

To obtain optimality conditions, we define an initial boundary value problem adjoint to problem (3.1):

$$\begin{aligned} -\frac{\partial p}{\partial t} &= A(t)p \text{ in } Q, \\ p|_{\Sigma} &= 0, \quad p(x, T) = f(x) \int_{\Omega} y(x, T) f(x) dx. \end{aligned} \quad (3.5)$$

We already know that the function f determines a linear continuous functional over space $H_0^m(\Omega)$ when $m > n/2$. Therefore $p(\cdot, T) \in H^{-m}(\Omega)$. It is easy to show using the Transposition Method (see [6]) that problem (3.5) has a unique weak solution, and $p|_{t=0} \in L^2(\Omega)$.

Proposition 1. *The functional $J(v)$ achieves maximum at $u(x)$ if*

$$\int_{\Omega} p(u; x, 0)[v(x) - u(x)] dx \leq 0 \quad (3.6)$$

for all $v \in U$.

Proof. If u is an optimal element, then [2]

$$J'(u)(v - u) \leq 0, \quad v \in U. \quad (3.7)$$

The left hand side of this inequality can be expressed as

$$\begin{aligned}
J'(u)(v - u) &= J'_y(y(u))y'(u)(v - u) = J'_y(y(u))[y(v) - y(u)] \\
&= 2 \int_{\Omega} y(u; x, T)f(x)dx \int_{\Omega} [y(v; x, T) - y(u; x, T)]f(x)dx \\
&= 2 \int_{\Omega} p(u; x, T)[y(v; x, T) - y(u; x, T)]dx. \quad (3.8)
\end{aligned}$$

The following equalities hold for the solution $y(\cdot)$ of problem (3.1) and the solution $p(\cdot)$ of adjoint problem (3.5):

$$\begin{aligned}
0 &= \int_Q \left(\frac{\partial p}{\partial t} + Ap \right) y \, dx \, dt - \int_Q p \left(-\frac{\partial y}{\partial t} + Ay \right) \, dx \, dt \\
&= \int_{\Sigma} \left(\frac{\partial p}{\partial \nu_A} y - p \frac{\partial y}{\partial \nu_A} \right) \, ds \, dt + \int_{\Omega} [py]_{t=0}^T dx, \quad (3.9)
\end{aligned}$$

where

$$\frac{\partial p}{\partial \nu_A} = \sum_{i,j} a_{ij} \frac{\partial p}{\partial x_j} \nu_i, \quad \nu = (\nu_1, \nu_2, \dots, \nu_N) \text{ is the unit normal to } \Gamma.$$

Using boundary conditions $y|_{\Sigma} = p|_{\Sigma} = 0$ we obtain

$$\int_{\Omega} py|_{t=0} dx = \int_{\Omega} py|_{t=T} dx. \quad (3.10)$$

If we combine (3.10) with (3.8) we get

$$\begin{aligned}
J'(u)(v - u) &= 2 \int_{\Omega} p(u; x, 0)[y(v; x, 0) - y(u; x, 0)]dx \\
&= 2 \int_{\Omega} p(u; x, 0)[v(x) - u(x)]dx.
\end{aligned}$$

Substituting this expression in inequality (3.7), we get condition (3.6). \square

From formulas (3.4), (3.5) it follows that

$$J(v) = \int_{\Omega} p(v; x, T)y(v; x, T)dx.$$

Using equality (3.10) we obtain a representation of the functional $J(v)$ as

$$J(v) = \int_{\Omega} p(v; x, 0)v(x)dx. \quad (3.11)$$

Formula (3.11) is useful for computations and estimates. Let us consider the following example. Let U be a ball of radius ε in space $L^2(\Omega)$:

$$U = \{v \in L^2(\Omega) : \|v\|_{L^2(\Omega)} \leq \varepsilon\}.$$

Then from formula (3.11) we get an estimate

$$|J(v)| \leq \varepsilon \|p(v; \cdot, 0)\|_{L^2(\Omega)}.$$

3.2.2 Control on the boundary

Let us consider an initial boundary value problem with control on the boundary

$$\frac{\partial y}{\partial t} = A(t)y \text{ in } Q, \quad y|_{\Sigma} = v, \quad y|_{t=0} = 0. \quad (3.12)$$

Similarly, let us define a problem adjoint to (3.12):

$$\begin{aligned} -\frac{\partial p}{\partial t} &= A(t)p \text{ in } Q, \\ p|_{\Sigma} &= 0, \quad p(x, T) = f(x) \int_{\Omega} y(x, T) f(x) dx. \end{aligned} \quad (3.13)$$

Proposition 2. *The functional*

$$J(v) = \left| \int_{\Omega} y(x, T) \varphi(x) dx - \sum_{k=1}^N c_k y(x_k, T) \right|^2 \quad (3.14)$$

achieves maximum if

$$\int_{\Sigma} \frac{\partial p(u; s, t)}{\partial \nu_A} (u(s, t) - v(s, t)) ds dt \leq 0, \quad v \in U \quad (3.15)$$

Proof. As in the proof of Proposition 1, equalities (3.8) and (3.9) hold. Substituting initial and boundary conditions from (3.12) and (3.13) we get

$$\int_{\Sigma} \frac{\partial p}{\partial \nu_A} v ds dt + \int_{\Omega} py|_{t=T} dx = 0. \quad (3.16)$$

Therefore

$$J'(u)(v - u) = 2 \int_{\Sigma} \frac{\partial p(u; s, t)}{\partial \nu_A} (u(s, t) - v(s, t)) ds dt.$$

Using condition (3.7) we get

$$\int_{\Sigma} \frac{\partial p(u; s, t)}{\partial \nu_A} (u(s, t) - v(s, t)) ds dt \leq 0.$$

This completes the proof of Proposition 2 □

3.3 Minimax problem in the case of a parabolic equation

3.3.1 Control by the initial conditions

Consider problem (3.1) for $U = \{v \in L^2(\Omega) : \|v\|_{L^2(\Omega)} \leq \varepsilon\}$. As before, the functional $J(v)$, has the form

$$J(v) = \left| \int_{\Omega} y(x, T) \varphi(x) dx - \sum_{k=1}^N c_k y(x_k, T) \right|^2.$$

We show that in this case it is also possible to solve a problem of minimizing $\sup_{v \in U} J(v)$ over coefficients $c_k \in \mathbb{R}$, i.e. to find $\inf_{\{c_k\} \in \mathbb{R}^N} \sup_{v \in U} J(v)$. We use optimization methods for functionals defined on sets of solutions of initial boundary value problems (see [8]).

In addition to problem (3.1) we consider an auxiliary problem

$$\begin{aligned} -\frac{\partial z}{\partial t} &= A(t)z \text{ in } Q, \\ z|_{\Sigma} &= 0, \quad z(x, T) = \varphi(x) - \sum_{k=1}^N c_k \delta(x - x_k). \end{aligned} \quad (3.17)$$

Similarly to problem (3.5), initial boundary value problem (3.17) has a unique weak solution $z(x, t)$ and $z|_{t=0} \in L^2(\Omega)$. Clearly, identity (3.10) also holds for z , i.e.

$$\int_{\Omega} y(x, T) z(x, T) dx = \int_{\Omega} y(x, 0) z(x, 0) dx.$$

Substituting the initial conditions from problems (3.1) and (3.17) into this equality, we get

$$\int_{\Omega} y(x, T) \varphi(x) dx - \sum_{k=1}^N c_k y(x_k, T) = \int_{\Omega} v(x) z(x, 0) dx. \quad (3.18)$$

Therefore

$$\sup_{v \in U} |J(v)| = \varepsilon^2 \int_{\Omega} z^2(x, 0) dx. \quad (3.19)$$

Denote by u the element of the set U for which this maximum is achieved. Clearly,

$$u(x) = \varepsilon z(x, 0) \|z(\cdot, 0)\|_{L^2(\Omega)}^{-1}.$$

From formula (3.19) it follows that the problem of minimization of $\sup_{u \in U} J(u)$ over coefficients c_k is equivalent to a minimization problem for the functional $J_1(c) = \int_{\Omega} z^2(x, 0) dx$, where $c = \{c_k\} \in \mathbb{R}^N$, $z(x, t)$ is a solution of initial boundary value problem (3.17). Let us introduce a problem adjoint to problem (3.17):

$$\begin{aligned} \frac{\partial q}{\partial t} &= A(t)q \text{ in } Q & (3.20) \\ q|_{\Sigma} &= 0, \quad q(x, 0) = z(x, 0). \end{aligned}$$

Using the same argument as in derivation of formula (3.6), we show that the gradient $\partial J_1(c)/\partial c$ of the functional $J_1(c)$ is a vector in space \mathbb{R}^N with components $-q(c, x_k, T)$, $k = 1, 2, \dots, N$. Notice that $q(c, x, t)$ is a solution of problem (3.20), where $z(c, x, t)$ is a solution of problem (3.17) for $\{c_k\} = c$. Therefore the functional $J_1(c)$ achieves minimum if

$$q(c^*, x_k, T) = 0, \quad k = 1, 2, \dots, N, \quad (3.21)$$

where $c^* = \{c_k^*\}$ is the optimal set of coefficients.

From formulas (3.17), (3.20), (3.21) it follows

$$\begin{aligned} \int_{\Omega} z^2(c^*, x, 0) dx &= \int_{\Omega} z(c^*, x, T)q(c^*, x, T) dx \\ &= \int_{\Omega} \varphi(x)q(c^*, x, T) dx - \sum_{k=1}^N c_k^* q(c^*, x_k, T) = \int_{\Omega} \varphi(x)q(c^*, x, T) dx. \end{aligned}$$

Using formula (3.19) we obtain:

$$\inf_{\{c_k\} \in \mathbb{R}^N} \sup_{v \in U} |J(v)| = \varepsilon^2 \int_{\Omega} \varphi(x)q(c^*, x, T) dx.$$

A wider class of sets U

Consider a wider class of sets U :

$$U = \{v \in L^2(\Omega) : (Bv, v)_{L^2(\Omega)} \leq \varepsilon^2\}, \quad (3.22)$$

B is a bounded, positive-definite operator in space $L^2(\Omega)$.

From equality (3.18) it follows that

$$J(v) = \left(\int_{\Omega} v(x)z(x, 0) dx \right)^2,$$

where $z(x, t)$ is a solution of auxiliary problem (3.17). Maximum of the functional $J(v)$ on the set defined by formula (3.22) is achieved at

$$u(x) = \varepsilon B^{-1}z(x, 0) \left\| \left(\sqrt{B} \right)^{-1} z(x, 0) \right\|^{-1}$$

and is equal to

$$\sup_{v \in U} |J(v)| = \varepsilon^2 \int_{\Omega} B^{-1}z(x, 0)z(x, 0)dx.$$

In order to minimize $\sup_{u \in U} J(u)$ over the values of the coefficients c_k , we introduce a functional $J_2(c) = \int_{\Omega} B^{-1}z(x, 0)z(x, 0)dx$, $c = \{c_k\} \in \mathbb{R}^n$.

Consider an initial boundary value problem adjoint to problem (3.17):

$$\frac{\partial r}{\partial t} = A(t)r \text{ in } Q, r|_{\Sigma} = 0, r(x, 0) = B^{-1}z(x, 0). \quad (3.23)$$

It can easily be checked that

$$\int_{\Omega} rz|_{t=0} dx = \int_{\Omega} rz|_{t=T} dx.$$

Therefore

$$\begin{aligned} J_2(c) &= \int_{\Omega} B^{-1}z(x, 0)z(x, 0) = \int_{\Omega} r(x, 0)z(x, 0) \\ &= \int_{\Omega} r(x, T) \left(\varphi(x) - \sum_{k=1}^N c_k \delta(x - x_k) \right) dx = \int_{\Omega} r(x, T)\varphi(x) dx - \sum_{k=1}^N c_k r(x_k, T). \end{aligned} \quad (3.24)$$

From (3.24) it follows that $\partial J_2(c)/\partial c_j = -r(x_k, T)$. Therefore, optimality conditions for $J_2(c)$ are

$$r(c^*, x_k, T) = 0, \quad k = 1, 2, \dots, N$$

and

$$\inf_{\{c_k\} \in \mathbb{R}^N} \sup_{v \in U} |J(v)| = \varepsilon^2 \int_{\Omega} \varphi(x)r(c^*, x, T)dx,$$

where $r(c, x, t)$ is a solution of problem (3.23).

3.3.2 Control on the boundary

Using the same approach as in Sec. 3.3.1 it is possible to determine $\inf_{\{c_k\}} \sup_{v \in U} J(v)$ for the case of $y(x, t)$ being a solution of initial boundary value problem (3.12) with the control v on the boundary, $U = \{v \in L^2(\Sigma) : \|v\|_{L^2(\Sigma)} < \varepsilon\}$.

Let us consider initial boundary value problems (3.12) and (3.17). Using the same argument as in the proof of identity (3.16), we obtain

$$\int_{\Omega} y(x, T)z(x, T)dx + \int_{\Sigma} y \frac{\partial z}{\partial \nu_A} dsdt = 0.$$

Using initial and boundary conditions in (3.12) and (3.17), we get

$$\int_{\Omega} y(x, T)\varphi(x)dx - \sum_{k=1}^N c_k y(x_k, T) = \int_{\Sigma} v \frac{\partial z}{\partial \nu_A} dsdt.$$

Therefore

$$J(v) = \left| \int_{\Sigma} v(s, t) \frac{\partial z}{\partial \nu_A} dsdt \right|^2.$$

For $U = \{v \in L^2(\Sigma) : \|v\|_{L^2(\Sigma)} < \varepsilon\}$ maximum of $J(v)$ is achieved at $u(x) = \varepsilon \frac{\partial z}{\partial \nu_A} \left\| \frac{\partial z}{\partial \nu_A} \right\|_{L^2(\Sigma)}^{-1}$:

$$\sup_{v \in U} J(v) = \varepsilon^2 \int_{\Sigma} \left(\frac{\partial z}{\partial \nu_A} \right)^2 dsdt.$$

Let us minimize the functional

$$J_3(c) = \int_{\Sigma} \left(\frac{\partial z}{\partial \nu_A} \right)^2 dsdt \text{ where } c = \{c_k\} \in \mathbb{R}^N.$$

Let us introduce a problem adjoint to auxiliary problem (3.17):

$$\frac{\partial q}{\partial t} = A(t)q \text{ in } Q, q|_{\Sigma} = \frac{\partial z}{\partial \nu}, q(x, 0) = 0. \quad (3.25)$$

Then

$$\begin{aligned} J_3(c) &= \int_{\Sigma} \left(\frac{\partial z}{\partial \nu_A} \right)^2 dsdt = \int_{\Sigma} \frac{\partial z}{\partial \nu_A} q dsdt = - \int_{\Omega} z(x, T)q(x, T)dx \\ &= - \int_{\Omega} (\varphi(x) - \sum_{k=1}^N c_k \delta(x - x_k))q(x, T)dx, \end{aligned}$$

and

$$\frac{\partial J_3(c)}{\partial c_i} = q(c, x_i, T).$$

Therefore minimum of the functional $J_3(c)$ is achieved at $c = c^*$ such that

$$q(c^*, x_k, T) = 0, \quad k = 1, 2, \dots, N,$$

and

$$\inf_{\{c_k\} \in \mathbb{R}^N} \sup_{v \in U} |J(v)| = \varepsilon^2 \int_{\Omega} \varphi(x) q(c^*, x, T) dx,$$

where $q(c, x, t)$ is a solution of problem (3.25).

3.4 A maximization problem in the case of a hyperbolic equation

Our approach can be extended to hyperbolic equations, but in this case the function $f(x)$ in the definition of the functional $J(v) = \left| \int_{\Omega} f(x) y(x, T) dx \right|^2$ has to be more regular. Let us consider an initial boundary value problem for a hyperbolic partial differential equation with nonhomogeneous boundary conditions of Dirichlet type:

$$\begin{aligned} \frac{\partial^2 y}{\partial t^2} &= A(t)y \text{ in } Q, \\ y|_{\Sigma} &= v, \quad y|_{t=0} = y_t|_{t=0} = 0, \end{aligned} \quad (3.26)$$

where $v \in U$, U is a closed bounded convex set in space $L^2(\Sigma)$; the rest of notation is as in Sec. 3.2.

Initial boundary value problem (3.26) has a unique weak solution, such that $y(\cdot, T) \in L^2(\Omega)$ (see [5]). Therefore we can define the functional $J(v)$:

$$J(v) = \left| \int_{\Omega} f(x) y(x, T) dx \right|^2, \quad f \in L^2(\Omega). \quad (3.27)$$

Let us find $u \in U$ such that $J(u) = \sup_{v \in U} J(v)$.

Proposition 3. *Necessary conditions of optimality for the problem of maximization of the functional (3.27) are*

$$\int_{\Sigma} \frac{\partial p(u; s, t)}{\partial \nu_A} [v(s, t) - u(s, t)] ds dt \leq 0$$

for all $v \in L^2(\Sigma)$, where $p(v)$ is a solution of an initial boundary value problem

$$\begin{aligned} \frac{\partial^2 p}{\partial t^2} &= Ap \text{ in } Q; \\ p|_{\Sigma} &= 0, \quad p(x, T) = 0, \quad p_t(x, T) = f(x) \int_{\Omega} f(x)y(x, T)dx. \end{aligned} \quad (3.28)$$

Proof. It can be proved in the same way as in the proof of Proposition 1 that

$$J'(u)(v - u) = 2 \int_{\Omega} y(u; x, T)f(x)dx \int_{\Omega} [y(v; x, T) - y(u; x, T)]f(x)dx.$$

Using a solution of problem (3.28), the right hand side of this equality can be rewritten as

$$2 \int_{\Omega} p_t(u; x, T)[y(v; x, T) - y(u; x, T)]dx. \quad (3.29)$$

Solutions of problems (3.26) and (3.28) satisfy the following equality

$$0 = \int_{\Sigma} \left[\frac{\partial p}{\partial \nu_A} y - \frac{\partial y}{\partial \nu_A} p \right] ds dt - \int_{\Omega} [p_t y - p y_t]_{t=0}^T dx.$$

Using boundary conditions $y|_{\Sigma} = v$ and initial conditions $y|_{t=0} = y_t|_{t=0} = 0$, we get

$$\int_{\Sigma} \frac{\partial p(u)}{\partial \nu_A} (v - u) ds dt = \int_{\Omega} p_t(u; x, T)[y(v; x, T) - y(u; x, T)]dx. \quad (3.30)$$

This completes the proof of Proposition 3. \square

From formulas (3.28), (3.30) we get an expression for the functional (3.27), similar to representation (3.11):

$$J(v) = \int_{\Sigma} \frac{\partial p(v; s, t)}{\partial \nu_A} v(s, t) ds dt.$$

Thus the following inequality is true

$$|J(v)| \leq \left\| \frac{\partial p(v)}{\partial \nu_A} \right\|_{L^2(\Sigma)} \|v\|_{L^2(\Sigma)}.$$

This inequality gives a simple estimate for the functional $J(v)$.

3.5 An example of finding coefficients for a quadrature formula

Let us consider an example of finding optimal coefficients c_i for a quadrature formula

$$\int_{\Omega} y(x, T) dx \approx \sum_{k=1}^N c_k y(x_k, T)$$

in the case when $y(x, T)$ is a solution of a heat equation

$$\begin{aligned} y_t = y_{xx}, \quad y(0, t) = y(\pi, t) = 0, \quad y(x, 0) = v(x), \\ x \in [0, \pi], \quad t \geq 0, \quad v \in L^2[0, \pi], \quad \|v\| \leq \varepsilon. \end{aligned} \quad (3.31)$$

As in Sec 3.3.1 (see formulas (3.17) and (3.20)) we introduce an auxiliary problem

$$-z_t = z_{xx}, \quad z(0, t) = z(\pi, t) = 0, \quad z(x, T) = 1 - \sum_{k=1}^N c_k \delta(x - x_k), \quad (3.32)$$

and a problem adjoint to (3.32)

$$q_t = q_{xx}, \quad q(0, t) = q(\pi, t) = 0, \quad q(x, 0) = z(x, 0). \quad (3.33)$$

Solution of problem (3.32) can be represented as a Fourier series

$$z(x, t) = \sum_{n=1}^{\infty} A_n e^{-n^2(T-t)} \sin(nx), \quad (3.34)$$

where

$$A_n = \frac{2}{\pi} \int_0^{\pi} \left(1 - \sum_{k=1}^N c_k \delta(s - x_k)\right) \sin(ns) ds = \frac{2}{\pi n} (1 - \cos(\pi n)) - \frac{2}{\pi} \sum_{k=1}^N c_k \sin(nx_k).$$

Similarly, solution of problem (3.33) can be represented as:

$$q(x, t) = \sum_{n=1}^{\infty} B_n e^{-n^2 t} \sin(nx), \quad (3.35)$$

where $B_n = \frac{2}{\pi} \int_0^{\pi} z(x, 0) \sin(ns) ds = A_n e^{-n^2 T}$.

Optimality conditions for this problem were derived in Sec. 3.3.1 (see formula (3.21)). Therefore, the optimal coefficients c_i can be determined by solving a system of equations

$$q(c, x_k, T) = 0, \quad k = 1, \dots, N.$$

Using (3.34) and (3.35) we obtain:

$$\begin{aligned} q(c, x_k, T) &= \sum_{n=1}^{\infty} B_n \sin(nx_k) e^{-n^2 T} = \sum_{n=1}^{\infty} A_n \sin(nx_k) e^{-2n^2 T} \\ &= \sum_{n=1}^{\infty} \frac{2}{\pi n} (1 - \cos(\pi n)) e^{-2n^2 T} \sin(nx_k) - \sum_{n=1}^{\infty} \frac{2}{\pi} \left(\sum_{j=1}^N c_j \sin(nx_j) \right) e^{-2n^2 T} \sin(nx_k). \end{aligned}$$

Thus, optimal coefficients c_k can be found solving a linear system of equations

$$Ac = b,$$

where

$$\begin{aligned} b_i &= \sum_{n=1}^{\infty} \frac{2}{\pi n} (1 - \cos(\pi n)) e^{-2n^2 T} \sin(nx_i), \\ A_{ij} &= \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-2n^2 T} \sin(nx_i) \sin(nx_j). \end{aligned}$$

Let us consider two numerical examples of solution of this problem.

Let $T = 1$. For a uniform quadrature set of ten points on the interval $[0, \pi]$ values of the coefficients c_i are presented in the following table.

Table 3.1: Numerical Example 1

i	1	2	3	4	5	6	7	8	9	10
x_i	$\pi/11$	$2\pi/11$	$3\pi/11$	$4\pi/11$	$5\pi/11$	$6\pi/11$	$7\pi/11$	$8\pi/11$	$9\pi/11$	$10\pi/11$
c_i	0.3351	0.2611	0.2996	0.2782	0.2879	0.2879	0.2782	0.2996	0.2611	0.3351

In Table 3.1 the second row contains given values of the nodes x_i , and the third row contains computed quadrature weights c_i .

In the case of quadrature points defined by the vector

$$x = (\pi/10, 3\pi/14, \pi/3, 2\pi/5, 9\pi/17, 2\pi/3, 5\pi/7, 6\pi/7, 9\pi/10, 18\pi/19),$$

values of c_i corresponding to each node x_i are shown in Table 3.2.

Table 3.2: Numerical Example 2

i	1	2	3	4	5	6	7	8	9	10
x_i	$\pi/10$	$3\pi/14$	$\pi/3$	$2\pi/5$	$9\pi/17$	$2\pi/3$	$5\pi/7$	$6\pi/7$	$9\pi/10$	$18\pi/19$
c_i	0.3783	0.3646	0.3172	0.2416	0.4936	0.1780	0.3931	0.1444	0.6453	-0.3874

Bibliography

- [1] S. A. Avdonin and D. A. Ovsyannikov, *An approach to the construction of optimal cubature formulas*, Partial Differential Equations (1988), 153–158, (in Russian).
- [2] Jean C ea, *Lectures on optimization—theory and algorithms*, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, vol. 53, Tata Institute of Fundamental Research, Bombay, 1978.
- [3] V. F. Kuzyutin, *Error estimates for approximate integration formulae*, Leningrad, 1982, (in Russian).
- [4] Ladyzhenskaya, O. A., *The boundary value problems of mathematical physics*, Applied Mathematical Sciences, vol. 49, Springer-Verlag, New York, 1985.
- [5] J.-L. Lions, *Control of distributed singular systems*, Gauthier-Villars, Montrouge, 1985.
- [6] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications*, Springer-Verlag, New York, 1972.
- [7] I. P. Mysovskih, *Interpolation cubature formulas*, Nauka, Moscow, 1981, (in Russian).
- [8] A. G. Nakonechny, *Minimax estimation of functionals of solutions of variational equations in Hilbert Spaces*, Kiev: KGU, 1985, (in Russian).
- [9] D. A. Ovsyannikov and S. A. Avdonin, *On construction of optimal cubature formulae*, in: Mathematical methods for the control of beams (Leningrad), Leningrad. Univ., 1980, (in Russian), pp. 281–288.
- [10] V. I. Smirnov, *A course of higher mathematics. Vol. V [Integration and functional analysis]*, Pergamon Press, Oxford, 1964.

- [11] S. L. Sobolev, *Cubature formulas and modern analysis. an introduction*, Gordon and Breach Science Publishers, Montreux, 1992.
- [12] S. L. Sobolev, *Formulas of mechanical cubatures in n -dimensional space*, in: Selected works of S. L. Sobolev. Vol. I (New York), Springer, 2006, pp. 445–451.

General Conclusions

Sampling and interpolation

In chapter 1 we investigate invertibility of the convolution operator W (see formula (1.11)) proposed in [2]. Invertibility of the convolution operator W is equivalent to controllability of the corresponding dynamical system (1.1) with control supported on a union of two intervals E . So, by proving invertibility of W we prove that the system (1.1) can be made controllable by choosing an appropriate density function.

Exact controllability of the system (1.1) is equivalent to sequence $\{\lambda_n\}$ forming a sampling and interpolating set for the Paley-Wiener space L_E^2 , where λ_n^2 are eigenvalues of problem (1.3). This also proves that there are infinitely many such sequences, since there are infinitely many density functions that make the dynamical system controllable. These sequences can be found by solving the appropriate Sturm-Liouville problem (1.3).

To prove that the operator W can be made invertible by choosing the value of parameter μ , we reduce the problem to invertibility of a simpler operator. We prove that for a small enough value of the parameter μ invertibility of this operator is equivalent to invertibility of operator V (1.12), which is equal to operator W truncated to first 4 terms. Then we introduce a new operator K of the same form as V :

$$(Kf)(t) = [c_1f(t+a) + c_2f(t+a-1) + c_3f(t+b) + c_4f(t+b-1)], \quad (3.36)$$

where $t \in [0, 1]$; $a, b \in [0, 1]$; $b \leq a$; $c_1 \neq 0$ or $c_4 \neq 0$, and derive invertibility conditions for it (see theorems 4, 5 in Chapter 1). After that we proceed to prove that for small enough values of μ the coefficients of operator V satisfy conditions of theorems 4, 5 in Chapter 1, and therefore V is invertible.

Invertibility conditions for operator K is a contribution to theory of convolution operators and linear functional equations. Proof of controllability of system (1.1) with control supported on two intervals is a new result in control theory. Our proof of the existence of infinitely many sampling and interpolating sequences for two-band signals extends the knowledge in sampling and interpolation theory. Another result

concerning sampling and interpolation for two-band signals was obtained by Seip [4]. However, his method is limited to the case of two intervals.

Our approach to sampling and interpolation problem is extendable to the case of set E being any finite union of intervals. We need to prove that system (1.1) with control supported on set E can be made controllable by choosing an appropriate density function ρ . Another goal is to prove Conjecture 1 in Chapter 1, which states, that dynamical system (1.1) is controllable for any density function satisfying conditions (1.5), (1.6). First step is to try to prove it for the case of two intervals; second step is the proof for the case of any number of intervals. For the case of two intervals, we have found sufficient conditions on the density function for the system (1.1) to be controllable, but these conditions are not necessary. It is very likely that these conditions are much more restrictive than needed. If we prove that less restrictive conditions are valid, that will mean that the class of sampling and interpolating sequences is bigger than the class presented in this dissertation.

Frequency estimation

In Chapter 2 we apply a dynamical system approach to spectral estimation problem. We apply the Boundary Control method to recover signals of the form $r(t) = \sum_{n=1}^K a_n(t)e^{\lambda_n t}$, from given samples $r(0), r(1), \dots$. Here $a_n(t)$ are polynomials, λ_n are scalars. The constants λ_n and polynomials $a_n(t)$ are to be recovered. The Boundary Control method is based on connections between inverse (identification) problems and controllability of dynamical systems (see [1]). It was introduced to solve Gelfand's problem (boundary inverse problem for multidimensional wave equation), and later successfully applied to the heat, beam, Maxwell, Schrödinger equations.

We apply the Boundary Control method's framework to the frequency estimation problem by defining an auxiliary discrete-time linear dynamical system so that its input-output map is a convolution operator with a convolution kernel that has the same structure as signal $r(t)$. This system can be identified, and then the exponents and amplitudes of the signal can be found from the parameters of the system. We

show that the coefficients of the signal can be recovered by solving a generalized eigenvalue problem as in the Matrix Pencil method using the procedure in Section 2.7.

The main novelty in our approach is application of control theoretic methods to this problem. There are numerous methods for solving this problem, among them are: maximum likelihood methods, the Matrix Pencil method, MUSIC, ESPRIT. Usually the case of constant amplitudes is considered. Badeau et al. [3] develop a generalized ESPRIT algorithm for estimation of parameters of a signal modeled by the Polynomial Amplitude Complex Exponentials model. Their results are similar to ours, but they use more complex linear algebra tools. One of the advantages of our approach is that it is easily extended to the polynomial case. We also provide explicit formulas for the amplitudes $a_n(t)$.

Our signal model does not include noise, so this is one of possible directions of future work. I would also like to try to apply the Boundary Control method to spectral estimation of a signal represented by an infinite sum of exponentials, which would require introducing an auxiliary continuous-time dynamical system instead of a discrete-time system.

Approximate integration

In the last Chapter 3 we present an approach to construction of optimal quadrature formulas for classes of solutions of certain initial boundary value problems. We consider two parabolic initial value boundary problems, one with non-zero initial condition, another one with non-zero condition on the boundary. First, we investigate a maximization problem for the error functional on the space of initial or boundary conditions for both initial boundary value problems. This problem can be viewed as a problem of optimal control for initial boundary value problem. In the first case control is in the initial condition, in the second case control is on the boundary. We find optimality conditions for the initial or boundary conditions, and an error estimate in the case when the set of controls is bounded. If the set of controls is

bounded it is possible to also solve a minimax problem. We define the set of controls to be a ball of radius ϵ . For the first problem we also consider the set of controls $U = \{v \in L^2(\Omega) : (Bv, v)_{L^2(\Omega)} \leq \epsilon^2\}$, where B is a bounded, positively defined operator in space $L^2(\Omega)$. We find conditions on quadrature weights for the maximum of error functional over the set of controls to achieve its minimum. Minimization over the nodes is not considered.

For hyperbolic equations, the optimal quadrature problem does not make sense the way we stated it, since the solution is not regular enough to evaluate it at any point. We apply methods we use to solve the optimal quadrature problem to maximize a functional defined on a space of controls of a hyperbolic dynamical system. This new functional is more regular than the original quadrature error functional.

In the last section of Chapter 3 we present a simple numerical example of calculating optimal quadrature weights for a one-dimensional heat equation.

To our knowledge there is no results on this kind of problems in literature: there are results for more general classes of integrands, or for finding the solution itself instead of its integral. The interest in approximate integration of initial boundary value solutions emerged from applications in control of charged particle beams. Equations describing particle beams are much more complex than the ones that we consider, but our work is a good model example that can be extended to more complex cases. The next step in exploration of this topic will be to study optimal cubature formulas for integration of solutions of other initial boundary value problem types.

Bibliography

- [1] S.A. Avdonin and M.I. Belishev, *Boundary control and dynamical inverse problem for nonselfadjoint Sturm-Liouville operator*, Control and Cybernetics **25** (1996), 429–440.
- [2] S.A. Avdonin and W. Moran, *Sampling and interpolation of functions with multi-band spectra and controllability problems*, Optimal Control of Partial Differential Equations (K.-H. Hoffmann, G. Leugering, and Tröltzsch F., eds.), vol. 133, Birkhäuser, Basel, 1999, Internat. Ser. Numer. Math., pp. 43–51.
- [3] Roland Badeau, Bertrand David, and Gaël Richard, *High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials*, IEEE transactions on signal processing **54** (2006), no. 4, 1341–1350.
- [4] K. Seip, *A simple construction of exponential bases in L^2 of the union of several intervals*, Proc. Edinburgh Math. Soc. **38** (1995), 171–177.