Applied and Numerical Harmonic Analysis

$$\widehat{f}(\gamma)=\int f(x)e^{-2\pi i x\gamma}\,dx$$

Travis D. Andrews · Radu Balan
John J. Benedetto · Wojciech Czaja
Kasso A. Okoudjou
Editors

Excursions in Harmonic Analysis, Volume 1

The February Fourier Talks at the Norbert Wiener Center

Birkhäuser

Birkhäuser

Applied and Numerical Harmonic Analysis

For further volumes:
http://www.springer.com/series/4968

Travis D. Andrews • Radu Balan
John J. Benedetto • Wojciech Czaja
Kasso A. Okoudjou
Editors

Excursions in Harmonic Analysis, Volume 1

The February Fourier Talks at the Norbert Wiener Center

Birkhäuser

Editors
Travis D. Andrews
Norbert Wiener Center
Department of Mathematics
University of Maryland
College Park, MD, USA

Radu Balan
Norbert Wiener Center
Department of Mathematics
University of Maryland
College Park, MD, USA

John J. Benedetto
Norbert Wiener Center
Department of Mathematics
University of Maryland
College Park, MD, USA

Wojciech Czaja
Norbert Wiener Center
Department of Mathematics
University of Maryland
College Park, MD, USA

Kasso A. Okoudjou
Norbert Wiener Center
Department of Mathematics
University of Maryland
College Park, MD, USA

ISBN 978-0-8176-8375-7 ISBN 978-0-8176-8376-4 (eBook)
DOI 10.1007/978-0-8176-8376-4
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012951313

Mathematics Subject Classification (2010): 26-XX, 35-XX, 40-XX, 41-XX, 42-XX, 43-XX, 44-XX, 46-XX, 47-XX, 58-XX, 60-XX, 62-XX, 65-XX, 68-XX, 78-XX, 92-XX, 93-XX, 94-XX

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.birkhauser-science.com)

Dedicated to
Tom Grasso,
Friend and Editor Extraordinaire

ANHA Series Preface

The *Applied and Numerical Harmonic Analysis (ANHA)* book series aims to provide the engineering, mathematical, and scientific communities with significant developments in harmonic analysis, ranging from abstract harmonic analysis to basic applications. The title of the series reflects the importance of applications and numerical implementation, but richness and relevance of applications and implementation depend fundamentally on the structure and depth of theoretical underpinnings. Thus, from our point of view, the interleaving of theory and applications and their creative symbiotic evolution is axiomatic.

Harmonic analysis is a wellspring of ideas and applicability that has flourished, developed, and deepened over time within many disciplines and by means of creative cross-fertilization with diverse areas. The intricate and fundamental relationship between harmonic analysis and fields such as signal processing, partial differential equations (PDEs), and image processing is reflected in our state-of-theart *ANHA* series.

Our vision of modern harmonic analysis includes mathematical areas such as wavelet theory, Banach algebras, classical Fourier analysis, time-frequency analysis, and fractal geometry, as well as the diverse topics that impinge on them.

For example, wavelet theory can be considered an appropriate tool to deal with some basic problems in digital signal processing, speech and image processing, geophysics, pattern recognition, biomedical engineering, and turbulence. These areas implement the latest technology from sampling methods on surfaces to fast algorithms and computer vision methods. The underlying mathematics of wavelet theory depends not only on classical Fourier analysis, but also on ideas from abstract harmonic analysis, including von Neumann algebras and the affine group. This leads to a study of the Heisenberg group and its relationship to Gabor systems, and of the metaplectic group for a meaningful interaction of signal decomposition methods. The unifying influence of wavelet theory in the aforementioned topics illustrates the justification for providing a means for centralizing and disseminating information from the broader, but still focused, area of harmonic analysis. This will be a key role of ANHA. We intend to publish the scope and interaction that such a host of issues demands.

Along with our commitment to publish mathematically significant works at the frontiers of harmonic analysis, we have a comparably strong commitment to publish major advances in the following applicable topics in which harmonic analysis plays a substantial role:

Biomedical signal processing
Compressive sensing
Communications applications
Data mining/machine learning
Digital signal processing
Fast algorithms
Gabor theory and applications
Image processing
Numerical partial differential equations
Prediction theory
Radar applications
Sampling theory
Spectral estimation
Speech processing
Time-frequency and time-scale analysis
Wavelet theory

The above point of view for the ANHA book series is inspired by the history of Fourier analysis itself, whose tentacles reach into so many fields.

In the last two centuries, Fourier analysis has had a major impact on the development of mathematics, on the understanding of many engineering and scientific phenomena, and on the solution of some of the most important problems in mathematics and the sciences. Historically, Fourier series were developed in the analysis of some of the classical PDEs of mathematical physics; these series were used to solve such equations. In order to understand Fourier series and the kinds of solutions they could represent, some of the most basic notions of analysis were defined, e.g., the concept of "function". Since the coefficients of Fourier series are integrals, it is no surprise that Riemann integrals were conceived to deal with uniqueness properties of trigonometric series. Cantor's set theory was also developed because of such uniqueness questions.

A basic problem in Fourier analysis is to show how complicated phenomena, such as sound waves, can be described in terms of elementary harmonics. There are two aspects of this problem: first, to find, or even define properly, the harmonics or spectrum of a given phenomenon, e.g., the spectroscopy problem in optics; second, to determine which phenomena can be constructed from given classes of harmonics, as done, e.g., by the mechanical synthesizers in tidal analysis.

Fourier analysis is also the natural setting for many other problems in engineering, mathematics, and the sciences. For example, Wiener's Tauberian theorem in Fourier analysis not only characterizes the behavior of the prime numbers, but also provides the proper notion of spectrum for phenomena such as white light; this latter process leads to the Fourier analysis associated with correlation functions in filtering and prediction problems, and these problems, in turn, deal naturally with Hardy spaces in the theory of complex variables.

Nowadays, some of the theory of PDEs has given way to the study of Fourier integral operators. Problems in antenna theory are studied in terms of unimodular trigonometric polynomials. Applications of Fourier analysis abound in signal processing, whether with the fast Fourier transform (FFT), or filter design, or the adaptive modeling inherent in time-frequency-scale methods such as wavelet theory. The coherent states of mathematical physics are translated and modulated Fourier transforms, and these are used, in conjunction with the uncertainty principle, for dealing with signal reconstruction in communications theory. We are back to the raison d'être of the ANHA series!

University of Maryland
College Park

John J. Benedetto
Series Editor

Preface

The chapters in these two volumes have at least one (co)author who spoke at the February Fourier Talks during the period 2006–2011.

The February Fourier Talks

The February Fourier Talks (*FFT*) were initiated in 2002 as a small meeting on harmonic analysis and applications, held at the University of Maryland, College Park. Since 2006, the *FFT* has been organized by the Norbert Wiener Center in the Department of Mathematics, and it has become a major annual conference. The *FFT* brings together applied and pure harmonic analysts along with scientists and engineers from industry and government for an intense and enriching two-day meeting. The goals of the *FFT* are the following:

- To offer a forum for applied and pure harmonic analysts to present their latest cutting-edge research to scientists working not only in the academic community but also in industry and government agencies,
- To give harmonic analysts the opportunity to hear from government and industry scientists about the latest problems in need of mathematical formulation and solution,
- To provide government and industry scientists with exposure to the latest research in harmonic analysis,
- To introduce young mathematicians and scientists to applied and pure harmonic analysis,
- To build bridges between pure harmonic analysis and applications thereof.

These goals stem from our belief that many of the problems arising in engineering today are directly related to the process of making pure mathematics applicable. The Norbert Wiener Center sees the *FFT* as the ideal venue to enhance this process in a constructive and creative way. Furthermore, we believe that our vision is shared by the scientific community, as shown by the steady growth of the *FFT* over the years.

The *FFT* is formatted as a two-day single-track meeting consisting of thirty-minute talks as well as the following:

- Norbert Wiener Distinguished Lecturer series
- General interest keynote address
- Norbert Wiener Colloquium
- Graduate and postdoctoral poster session

The talks are given by experts in applied and pure harmonic analysis, including academic researchers and invited scientists from industry and government agencies.

The Norbert Wiener Distinguished Lecture caps the technical talks of the first day. It is given by a senior harmonic analyst, whose vision and depth through the years have had profound impact on our field. In contrast to the highly technical day sessions, the keynote address is aimed at a general public audience and highlights the role of mathematics, in general, and harmonic analysis, in particular. Furthermore, this address can be seen as an opportunity for practitioners in a specific area to present mathematical problems that they encounter in their work. The concluding lecture of each *FFT*, our Norbert Wiener Colloquium, features a mathematical talk by a renowned applied or pure harmonic analyst. The objective of the Norbert Wiener Colloquium is to give an overview of a particular problem or a new challenge in the field. We include here a list of speakers for these three lectures:

Distinguished lecturer

- Peter Lax
- Richard Kadison
- Elias Stein
- Ronald Coifman
- Gilbert Strang

Keynote address

- Frederick Williams
- Steven Schiff
- Peter Carr
- Barry Cipra
- William Noel
- James Coddington
- Mario Livio

Colloquium

- Christopher Heil
- Margaret Cheney
- Victor Wickerhauser
- Robert Fefferman
- Charles Fefferman
- Peter Jones

The Norbert Wiener Center

The Norbert Wiener Center for Harmonic Analysis and Applications provides a national focus for the broad area of mathematical engineering. Applied harmonic analysis and its theoretical underpinnings form the technological basis for this area. It can be confidently asserted that mathematical engineering will be to today's mathematics departments what mathematical physics was to those of a century ago. At that time, mathematical physics provided the impetus for tremendous advances within mathematics departments, with particular impact in fields such as differential equations, operator theory, and numerical analysis. Tools developed in these fields were essential in the advances of applied physics, e.g., the development of the solid-state devices which now enable our information economy.

Mathematical engineering impels the study of fundamental harmonic analysis issues in the theories and applications of topics such as signal and image processing, machine learning, data mining, waveform design, and dimension reduction into mathematics departments. The results will advance the technologies of this millennium.

The golden age of mathematical engineering is upon us. The Norbert Wiener Center reflects the importance of integrating new mathematical technologies and algorithms in the context of current industrial and academic needs and problems. The Norbert Wiener Center has three goals:

- Research activities in harmonic analysis and applications
- Education—undergraduate to postdoctoral
- Interaction within the international harmonic analysis community

We believe that educating the next generation of harmonic analysts, with a strong understanding of the foundations of the field and a grasp of the problems arising in applications, is important for a high-level and productive industrial, government, and academic workforce.

The Norbert Wiener Center web site: www.norbertwiener.umd.edu

The Structure of the Volumes

To some extent the eight parts of these two volumes are artificial placeholders for all the diverse chapters. It is an organizational convenience that reflects major areas in harmonic analysis and its applications, and it is also a means to highlight significant modern thrusts in harmonic analysis. Each of the following parts includes an introduction that describes the chapters therein:

Volume 1	Volume 2
I Sampling Theory	V Measure Theory
II Remote Sensing	VI Filtering
III Mathematics of Data Processing	VII Operator Theory
IV Applications of Data Processing	VIII Biomathematics

Acknowledgements

The Norbert Wiener Center gratefully acknowledges the indispensable support of the following groups: Australian Academy of Science, Birkhäuser, the IEEE Baltimore Section, MiMoCloud, Inc., Patton Electronics Co., Radyn, Inc., the SIAM Washington–Baltimore Section, and SR2 Group, LLC. One of the successes of the February Fourier Talks has been the dynamic participation of graduate student and postdoctoral engineers, mathematicians, and scientists. We have been fortunate to be able to provide travel and living expenses to this group due to continuing, significant grants from the National Science Foundation, which, along with the aforementioned organizations and companies, believes in and supports our vision of the FFT.

Contents

Part I
Sampling Theory

Sampling theory has a long history going at least as far back as Lagrange with explicit sampling formulas. The classical sampling theorem due to Cauchy in the 1840s is often associated with names such as Hadamard, E. T. Whittaker, J.M. Whittaker, Kotel'nikov, Wiener, Raabe, Someya, Shannon (see Chapter 1 of [1]). It is also a subject that has expanded significantly in recent years both with regard to compelling theoretical advances and modern applicability.

The first of our six chapters on sampling theory is by AKRAM ALDROUBI. Aldroubi gives an overview of the theory of nonlinear signal representations in terms of unions of subspaces. It is a theory that Aldroubi and his collaborators have spearheaded. It is formulated as a general model for data representations, which includes a variety of applications from compressive sensing and dimension reduction to motion and tracking in video. Frames, shift-invariant subspaces, and the so-called minimum subspace approximation property are key concepts in the theory, and useful algorithms are part and parcel of the presentation.

BERNHARD G. BODMANN, PETER G. CASAZZA, JESSE D. PETERSON, IHAR SMALYANAU, AND JANET C. TREMAIN provide original constructions of fusion frames with rigid geometric properties. Fusion frames can be viewed as a generalization of frames. By their definition it is tantalizing to try to invoke them in modeling environments in which data sets, e.g., communications data, overlap and where the goal is to provide efficient connectivity, e.g., good communications, over the union of the sets. Part of the authors' setting is based on Naimark's classical theorem relating frames and orthonormal bases, but their technology is both deep and beyond the original Naimark theorem. Their specific constructions are of equi-isoclinic Parseval fusion frames, motivated from a rich history and a setting for present-day applicability.

JENS G. CHRISTENSEN AND GESTUR OLAFSSON exposit their theory of the classical sampling theorem in the grand setting of commutative connected homogeneous spaces. Lie groups and homogeneous spaces are the starting point, leading to invariant differential operators on homogeneous spaces. There is a necessary and technical excursion dealing with oscillation estimates on Lie groups, as well as state-of-the-art results on the path from Gelfand pairs to Ruffino's recent topological isomorphism theorem for positive definite spherical functions. It is a virtuosic route to the definition of bandlimited functions for homogeneous spaces, and, then, to the authors' classical sampling theorem in this generality.

CHARLES FEFFERMAN began a program going back to fundamental questions, posed by Hassler Whitney in the 1930s, dealing with smooth extensions of functions. He led an effort, along with several others, solving some of these problems and addresing the comparably fundamental problems of devising best-possible and useful algorithms for interpolation of data. He and Klartay (2009) have best-possible algorithms, and Fefferman has shown the importance of constructing useful algorithms with regard to this problem. Fefferman's chapter is an important, readable introduction to Whitney's extension theorem, his problems, and to the exciting algorithmic questions that remain. It provides a marvelous setting for mathematicians of various stripes to transcend the pure power of the solutions to Whitney's problems to their relevance in current applicability and the importance of viable algorithms to effect this applicability.

Joseph D. Lakey resurrects the classical vision of Henry J. Landau's poignant encapsulation in 1983 of the fundamental issues of sampling theory. Lakey brings to bear an imposing analytic overview and also integrates the most recent topics in sampling beginning with the fundamental work of Candès, Romberg, and Tao, as well as a potpourri of deep results ranging from those of Logan, Widom, Rokhlin, and others. It is a reassuring testament to Henry Landau's magisterial understanding of the uncertainty principle and classical sampling theory.

Kabe Moen, Hrvoje Šikić, Guido Weiss, and Edward Wilson go beyond the sampling theory of the previous chapter (by Lakey) by formulating the essential ideas of classical sampling theory in terms of technological notions such as convolution idempotents, Zak transforms, frames, and principal shift-invariant spaces. This is a state-of-the-art setting for formulating and proving sampling formulas.

Reference

1. Benedetto, J.J., Ferreira, P.J.S.G. (eds.): Modern Sampling Theory. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston (2001)

Unions of Subspaces for Data Modeling and Subspace Clustering

Akram Aldroubi

Abstract Signals and data are often modeled as vectors belonging to a subspace of a Hilbert space (or Banach space)$\mathscr{H}$, e.g., bandlimited signals. However, nonlinear subsets can also serve as signal model, e.g., signals with finite rate of innovation or sparse signals. Discovering the model from data observations is a fundamental task. The goal of this chapter is to discuss the problem of learning a nonlinear model of the form $\mathscr{M} = \cup_{i=1}^{l} V_i$ from a set of observations $\mathbf{F} = \{f_1, \dots, f_m\}$, where the V_i are the unknown subspaces to be found. Learning this nonlinear model from observed data has applications to computer vision and signal processing and is connected to several areas of mathematics. In this chapter we give a brief description of the theoretical as well as the applied aspects related to this type of nonlinear modeling in terms of unions of subspaces.

Keywords Union of subspaces • Sparse signals • Subspace segmentation • Signal learning from observations • Spectral clustering • Non-linear signal models • Motion segmentation • Minimum subspace approximation property (MSAP) • Nearness to local subspaces

1 Introduction

This chapter is an overview of the problem of nonlinear signal representations in terms of unions of subspaces and its connection to other areas of mathematics and engineering. The point presented stems from recent research with my many collaborators [1–7].

A. Aldroubi (✉)
Department of Mathematics, Vanderbilt University, 1520 Stevenson Center,
Nashville, TN 37240, USA
e-mail: akram.aldroubi@vanderbilt.edu

T.D. Andrews et al. (eds.), *Excursions in Harmonic Analysis, Volume 1,*
Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-8376-4_1,

A standard model for signals, images, or multivariate data is often a shift-invariant space, such as a set of bandlimited functions, a B-spline space, or the level zero subspace of a multiresolution of $L^2(\mathbb{R}^d)$. Given a class of signals, the appropriate model can be learned from signal observations. For example, we may observe a set of signals $\mathbf{F} = \{f_1, \dots, f_m\} \subset L^2(\mathbb{R}^d)$ from a signal class of interest and then wish to use these observations to find a shift-invariant space that is consistent with these observations. Mathematically, this problem can be stated as follows. Let $\mathscr{V}_n$ be the set of all shift-invariant spaces with length at most n. That is, $V \in \mathscr{V}_n$ if

$$V = V(\boldsymbol{\Phi}) := \overline{\text{span}\{\varphi_i(\cdot - k) : \ i = 1, \dots, s, k \in \mathbb{Z}^d\}} \tag{1}$$

for some set of generators $\boldsymbol{\Phi} = \{\varphi_i \in L^2(\mathbb{R}^d) : \ i = 1, \dots, s\}$ with $s \le n$. The problem is to find a shift-invariant space that is nearest to the data $\mathbf{F}$ in the sense that

$$V^o = \text{argmin}_{V' \in \mathscr{V}_n} \sum_{i=1}^{m} \|f_i - P_{V'} f_i\|^p, \quad 0 < p < \infty, \tag{2}$$

where $P_{V'}$ is the orthogonal projection on V' and $\|\cdot\|$ denotes the standard norm in $L^2(\mathbb{R}^d)$. The choice $p = 2$ is good for noisy data; however, the choice $0 < p \le 1$ may be more advantageous when outliers are present [23]. Observe that the shift-invariant spaces in $\mathscr{V}_n$ have infinite dimensions, while there are finitely many data observations. However, the spaces in $\mathscr{V}_n$ are constrained (to be shift-invariant). Thus, if the number of generators $n < m$, it may not possible to find a space $V \in \mathscr{V}_n$ that contains $\mathbf{F}$. For $p = 2$, the existence of a minimizer V^o as in (2) that is nearest to the data $\mathbf{F}$ has been established in [1]. Moreover, a construction of a set $\{\varphi_i \in L^2(\mathbb{R}^d) : i = 1, \dots, s\}$ such that $\{\varphi_i(x - k) : \ i = 1, \dots, s, k \in \mathbb{Z}^d\}$ is a Parseval frame for V^o and a formula for the error $e(\mathbf{F}, V^o) = \sum_{i=1}^{m} \|f_i - P_{V^o} f_i\|^2$ have also been established in [1].

A general nonlinear model for signal or data representation in the finite- or infinite-dimensional cases can be stated as follows:

Problem 1 (Unions of Subspaces Model). Let $\mathscr{H}$ be a Hilbert space, $\mathbf{F} = \{f_1, \dots, f_m\}$ a finite set of vectors in $\mathscr{H}$, $\mathscr{C}$ a family of closed subspaces of $\mathscr{H}$, and $\mathscr{V}$ the set of all sequences of elements in $\mathscr{C}$ of length l (i.e., $\mathscr{V} = \mathscr{V}(l) = \big\{\{V_1, \dots, V_l\} : V_i \in \mathscr{C}, 1 \le i \le l\big\}$). Find a sequence of l subspaces $\mathbf{V}^o = \{V_1^o, \dots, V_l^o\} \in \mathscr{V}$ (if it exists) such that

$$e(\mathbf{F}, \mathbf{V}^o) = \inf\{e(\mathbf{F}, \mathbf{V}) : \mathbf{V} \in \mathscr{V}\} = \sum_{f \in F} d^2(f, EV). \tag{3}$$

The problem above is that of finding the unions of subspaces model $\mathscr{M} = \cup_{i=1}^{l} V_i$ that is nearest to a set of data $\mathbf{F} = \{f_1, \dots, f_m\}$ (see Fig. 1 for an illustration). For $l = 1$, this problem reduces to least squares problem. However, for $l > 1$ the problem becomes a nonlinear, generalized version of the least squares problem ($l = 1$), which has many applications in mathematics and engineering.

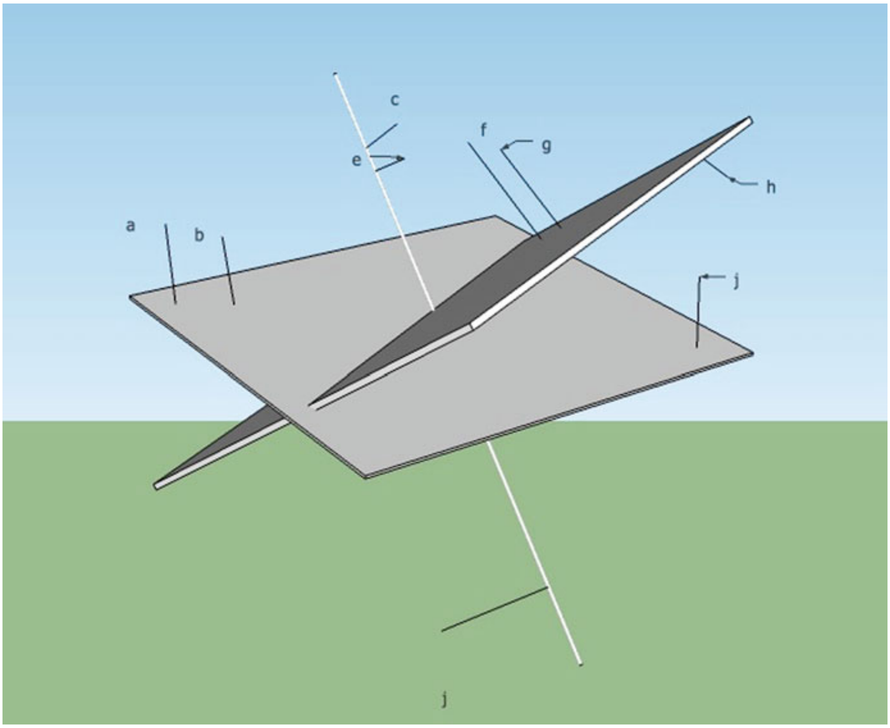

Fig. 1 An example illustrating a set of data points in $\mathbb{R}^3$ that we wish to model by a union of three subspaces, two of which are of dimension two and one of dimension one. The union drawn for this situation is clearly not the one nearest to the data points

Examples where $\mathscr{H}$ is infinite dimensional occur in signal modeling. The typical situation is $\mathscr{H} = L^2(\mathbb{R}^d)$, $\mathbf{F} \subset L^2(\mathbb{R}^d)$ is a finite set of signals, and $\mathscr{C}$ is the set of all finitely generated shift-invariant spaces of $L^2(\mathbb{R}^d)$ of length at most n [3,24,27]. The class of signals with finite rate of innovation is another example where the space $\mathscr{H}$ is infinite dimensional [26]. Applications where a union of subspaces underlies the signal model in infinite dimensions can be found in [2, 24, 26].

An important application in finite dimensions is the subspace segmentation problem in computer vision (see, e.g., [31] and the references therein). The subspace segmentation problem can be described as follows: Let $\mathscr{M} = \bigcup_{i=1}^{l} V_i$ where $\{V_i \subset \mathscr{H}\}_{i=1}^{l}$ is an unknown set of subspaces of a Hilbert space $\mathscr{H}$. Let $\mathbf{W} = \left\{w_j \in \mathscr{H}\right\}_{j=1}^{m}$ be a set of data points drawn from $\mathscr{M}$. Then,

1. Determine the number of subspaces l.
2. Determine the set of dimensions $\{d_i\}_{i=1}^{l}$.
3. Find an orthonormal basis for each subspace V_i.
4. Collect the data points belonging to the same subspace into the same cluster.

Because the data is often corrupted by noise, may have outliers, or may not be complete (e.g., there may be missing data points), a union of subspaces that is nearest to $\mathbf{W} = \left\{w_j \in \mathscr{H}\right\}_{j=1}^{m}$ (in some sense) is sought. In special cases, the number l of subspaces or the dimensions of the subspaces $\{d_i\}_{i=1}^{l}$ are known. A number of approaches have been devised to solve the problem above or some of its special cases [9, 11, 12, 15–19, 21, 22, 25, 28–35].

Other related applications include the problem of face recognition, the motion tracking problem in videos (see, e.g., [6, 13, 20, 30, 31, 33]), and the problem of segmentation and data clustering in hybrid linear models (see, e.g., [10, 23, 35] and the references therein). Compressed sensing is another related area where s-sparse signals in $\mathbb{C}^d$ can be viewed as belonging to a union of subspaces $\mathscr{M} = \cup_{i\in I} V_i$, with $\dim V_i \le s$, where s is the signals' sparsity [8].

Problem 1 raises many theoretical and practical questions, some of which will be described and discussed below. One of the theoretical questions is related to the existence of a solution. Specifically, what are the conditions on $\mathscr{C}$ that assure the existence of a solution to Problem 1? Equally important is to find search algorithms for solving Problem 1. These algorithms must be fast and must work well in noisy environments.

This overview is organized as follows. The first part discusses the conditions under which solutions to Problem 1 exist. The second part is devoted to various search algorithms and their analysis. The last part is an application to motion and tracking in video.

2 Existence of Solutions and the Minimum Subspace Approximation Property

It has been shown that, given a family of closed subspaces $\mathscr{C}$, the existence of a minimizing sequence of subspaces $\mathbf{V}^o = \{V_1^o, \ldots, V_l^o\}$ that solves Problem 1 is equivalent to the existence of a solution to the same problem but for $l = 1$ [2]:

Theorem 1. *Problem 1 has a minimizing set of subspaces for all finite sets of data and for any $l \ge 1$ if and only if it has a minimizing subspace for all finite sets of data and for $l = 1$.*

This suggests the following definition:

Definition 1. A set of closed subspaces $\mathscr{C}$ of a separable Hilbert space $\mathscr{H}$ has the minimum subspace approximation property (MSAP) if for every finite subset $\mathbf{F} \subset \mathscr{H}$, there exists an element $V \in \mathscr{C}$ that minimizes the expression

$$e(\mathbf{F}, V) = \sum_{f\in\mathbf{F}} d^2(f, V), \quad V \in \mathscr{C}. \tag{4}$$

We will say that $\mathscr{C}$ has MSAP(k) for some $k \in \mathbb{N}$ if the previous property holds for all subsets $\mathbf{F}$ of cardinality $j \leq k$.

Using this terminology, Problem 1 has a minimizing sequence of subspaces if and only if $\mathscr{C}$ satisfies the MSAP. It should be noted that MSAP($k+1$) is strictly stronger than MSAP(k). Obviously, MSAP is stronger than MSAP(k) for any $k \in \mathbb{N}$.

There are some cases for which it is known that the MSAP is satisfied. For example, if $\mathscr{H} = \mathbb{C}^d$ and $\mathscr{C} = \{V \subset \mathscr{H} : \dim V \leq s\}$, the Eckart–Young theorem [14] implies that $\mathscr{C}$ satisfies MSAP. Another example is the one described in the introduction when $\mathscr{H} = L^2(\mathbb{R}^d)$ and $\mathscr{C} = \mathscr{V}_n = \{V : V = \operatorname{span}\{\varphi_1(x-k), \ldots, \varphi_n(x-k) : k \in \mathbb{Z}^d\}\}$ is the set of all shift-invariant spaces of length at most n. For this last example, a result in [1] implies that $\mathscr{C}$ satisfies the MSAP.

The general approach for the existence of a minimizer has been recently considered in [7]. The family $\mathscr{C}$ is viewed as a set of projectors. Specifically, a subspace $V \in \mathscr{C}$ is identified with the orthogonal projector $Q = Q_V$ whose kernel is exactly V (i.e., $Q = I - P_V$ where P_V is the orthogonal projector on V). In this way, any set of closed subspaces $\mathscr{C}$ of $\mathscr{H}$ is identified with a set of projectors $\{Q \in \Pi(\mathscr{H}) : \ker(Q) \in \mathscr{C}\} \subset \Pi(\mathscr{H})$, where $\Pi(\mathscr{H}) \subset \mathscr{B}(\mathscr{H})$ denotes the set of orthogonal projectors. Using this identification, this set of projectors is denoted by $\mathscr{C}$ as well, and $e(\mathbf{F}, V)$ in (4) is expressed as

$$e(\mathbf{F}, V) = \Phi_{\mathbf{F}}(Q_V) = \sum_{f \in \mathbf{F}} \langle Q_V f, f \rangle. \tag{5}$$

The functional $\Phi_{\mathbf{F}}$ defined above is easily seen to be linear on the set $\mathscr{B}(\mathscr{H})$ of bounded linear operators on $\mathscr{H}$.

The right topology for the set of bounded linear operators $\mathscr{B}(\mathscr{H})$ in this case is the weak operator topology. Indeed, in this topology, $\Phi_{\mathbf{F}}$ is a continuous functional for each $\mathbf{F}$, and the set of projectors $\Pi(\mathscr{H})$ is pre-compact. Thus it is evident that if the set $\mathscr{C} \subset \Pi(\mathscr{H})$ is weakly closed in $\mathscr{B}(\mathscr{H})$, then $\Phi_{\mathbf{F}}$ attains its minimum (and maximum) for some $V^o \in \mathscr{C}$. This condition, however, is too strong as can be seen from this obvious example: let $\mathscr{H} = \mathbb{R}^3$ and consider the set $\mathscr{C} = \mathscr{C}_1 \cup \mathscr{C}_2$ which is the union of the plane $\mathscr{C}_1 = \operatorname{span}\{e_1, e_2\}$ and the set of lines $\mathscr{C}_2 = \cup_v \{\operatorname{span}\{v\} : v = e_3 + ce_2, \text{for some } c \in \mathbb{R}\}$. Then $\mathscr{C}$ (identified with a set of projectors as described earlier) is not closed (since $Q_{\operatorname{span}\{e_2\}} \notin \mathscr{C}$), but $\mathscr{C}$ satisfies the MSAP, since if the infimum is achieved by the missing line $\operatorname{span}\{e_2\}$, it is also achieved by the plane $\mathscr{C}_1$.

2.1 A Sufficient Geometric Condition for MSAP

Clearly, as the example above has shown, it is not necessary for $\mathscr{C}$ to be closed in order for $\mathscr{C}$ to have the MSAP even if $\mathscr{H}$ has finite dimension. It turns out that the right set to analyze is the set $\mathscr{C}^+$ consisting of $\mathscr{C}$ together with the positive semidefinite operators added to it. For example, the theorem [7] below gives a sufficient condition for MSAP in terms of the set $\mathscr{C}^+ = \mathscr{C} + \mathscr{P}^+(\mathscr{H})$ where $\mathscr{P}^+(\mathscr{H})$ is the set of positive semidefinite (self-adjoint operators). It also gives a sufficient condition for MSAP in terms of the convex hull $\mathrm{co}(\mathscr{C})$ of $\mathscr{C}$, i.e., the smallest convex set containing $\mathscr{C}$:

Theorem 2 ([7]).

$$(\mathscr{C} = \overline{\mathscr{C}}) \Rightarrow (\mathscr{C}^+ = \overline{\mathscr{C}^+}) \Rightarrow (\mathrm{co}(\mathscr{C}^+) = \mathrm{co}(\overline{\mathscr{C}^+})) \Rightarrow (\mathscr{C} \text{ satisfies MSAP}),$$

and the implications are strict in general.

The theorem above can be used to obtain another proof of the Eckart–Young theorem [14], or the fact that the set $\mathscr{C}(n) = \mathscr{V}_n = \{V : V = \overline{\mathrm{span}\{\varphi_1(\cdot - k), \ldots, \varphi_n(\cdot - k) : k \in \mathbb{Z}^d\}} \subset L^2(\mathbb{R}^d)$ of all shift-invariant spaces of length at most n satisfies MSAP as proved in [1]. It can also be used to prove that the set $\mathscr{C}_m(n) = \{V : \overline{\mathrm{span}\{\varphi_1(\cdot - k), \ldots, \varphi_n(\cdot - k) : k \in \mathbb{Z}^d\}}$ of all shift-invariant spaces of length at most n that are also $\frac{1}{m}\mathbb{Z}$ invariant satisfies MSAP as proved in [4]. Note that $\mathscr{C}_m \subset \mathscr{C}$. However, unlike the proofs in [1, 4], the use of Theorem 2 does not give constructive proofs.

2.2 *Characterization of MSAP in Finite Dimensions*

In finite dimensions, the last two implications of Theorem 2 can be reversed. Thus, the necessary and sufficient condition for MSAP is that $\mathscr{C}^+$ is closed. An equivalent characterization is that the convex hull $\mathrm{co}(\mathscr{C}^+)$ of $\mathscr{C}^+$ is equal to the convex hull $\mathrm{co}(\overline{\mathscr{C}^+})$ of its closure. A third characterization for a finite d-dimensional space $\mathscr{H}$ is that $\mathscr{C}$ satisfies the MSAP $(d-1)$ (see Definition 1). These characterizations are proved in [7], and we have:

Theorem 3. *Suppose $\mathscr{H}$ has dimension d. Then the following are equivalent:*

(i) $\mathscr{C}$ *satisfies MSAP(d-1).*
(ii) $\mathscr{C}$ *satisfies MSAP.*
(iii) $\mathscr{C}^+$ *is closed.*
(iv) $\mathrm{co}(\mathscr{C}^+) = \mathrm{co}(\overline{\mathscr{C}^+})$.

Unfortunately, the above equivalences are false in infinite dimensions. The difficulty is that although the operators of norm ≤ 1 form a compact subset for the weak operator topology, in infinite-dimensional spaces, the set of projectors is not closed in this topology. This creates most of the complications in infinite dimensions.

For example, to see why in infinite dimensions the last implication of Theorem 2 cannot be reversed, take $\mathscr{C}$ to be the set of all finite-dimensional subspaces (except the trivial vector space) of some infinite-dimensional space $\mathscr{H}$. It has MSAP since for all finite sets F in $\mathscr{H}$, one can find a finite-dimensional subspace containing it. On the other hand, the convex hull of $\mathscr{C}^+$ does not contain $\{0\}$, while the weak closure of $\mathscr{C}$ does.

An example that shows that, in infinite dimensions, the second implication of Theorem 2 is not an equivalence can be constructed as follows: let $\mathscr{H} = \ell^2$ and consider the sequence of vectors $v_n = e_1 + e_n$, which weakly converges to e_1, and the sequence $w_n = e_2 + e_{n+1}$, which weakly converges to e_2. For all $n \geq 3$, let P_n be the projector on the space spanned by v_n and w_n. One checks that the sequence $\{P_n\}$ converges weakly to $Q = (P_{E_1} + P_{E_2})/2$, where $E_1 = \text{span}\{e_1\}$ and $E_2 = \text{span}\{e_2\}$. Moreover, since $P_n = Q + (P_{E_n} + P_{E_{n+1}})/2$, where $E_n = \text{span}\{e_n\}$ and $E_{n+1} = \text{span}\{e_{n+1}\}$, we conclude that $Q < P_n$ for any n. Now define $\mathscr{C} = \{P_n,\ n \geq 3\} \cup \{P_{E_1}, P_{E_2}\}$. The closure of $\mathscr{C}$ consists of $\mathscr{C} \cup \{Q\}$. By the previous remark, Q does not belong to $\mathscr{C}^+$, so that $\mathscr{C}^+$ is not closed. But on the other hand, $Q \in \text{co}(\mathscr{C})$, hence $\overline{\mathscr{C}} \subset \text{co}(\mathscr{C})$ which implies that $\text{co}(\mathscr{C}) = \text{co}(\overline{\mathscr{C}})$. It follows that $\text{co}(\mathscr{C}^+) = \text{co}(\overline{\mathscr{C}^+})$.

2.3 Characterization of MSAP in Infinite Dimensions

For the infinite-dimensional case, the necessary and sufficient conditions are found in terms of the set of contact half-spaces. A contact half-space to a set E is a half-space containing E such that its boundary has a nontrivial intersection with E; a half-space is a closed set of the form $H_{\phi,a} = \{x \in \mathscr{B}(\mathscr{H}) : \phi(x) \geq a\}$, for some $a \in \mathbb{R}$ and ϕ an $\mathbb{R}$-linear functional on $\mathscr{B}(\mathscr{H})$. The boundary of such a half-space is the (affine) hyperplane given by $\phi(x) = a$. The set of contact half-spaces containing E is denoted by $\mathscr{T}(E)$. Using these concepts, the necessary and sufficient condition for MSAP in infinite dimensions is:

Theorem 4 ([7]). *Let $\mathscr{C}$ be a set of projectors in $\mathscr{B}(\mathscr{H})$. Then $\mathscr{C}$ has MSAP if and only if*

$$\mathscr{T}(\mathscr{C}^+) = \mathscr{T}\left(\overline{\mathscr{C}^+}\right).$$

3 Algorithms and Dimensionality Reduction

Search algorithms for finding solutions to Problem 1 are often iterative. For computational efficiency, dimensionality reduction is generally needed. Moreover, iterative algorithms often need a good initial approximation to the solution. A general, abstract algorithm of this kind is described in [2].

3.1 Algorithm for the Noise-Free Case

A fast algorithm for the noise-free case can be found using a simple reduced row echelon algorithm. Before getting into the details, we first need some definitions.

Definition 2. Let S be a linear subspace of $\mathbb{R}^D$ with dimension d. A set of data $\mathbf{F}$ drawn from $S \subset \mathbb{R}^D$ with dimension d is said to be *generic* if (i) $|\mathbf{F}| > d$, (ii) span $\mathbf{F} = S$, and (iii) every d vector from $\mathbf{F}$ forms a basis for S.

Definition 3. Matrix R is said to be the *binary reduced row echelon form* of matrix A if all non-pivot column vectors are converted to binary vectors by setting nonzero entries to one.

Theorem 5. *Let $\{S_i\}_{i=1}^k$ be a set of nontrivial linearly independent subspaces of $\mathbb{R}^D$ with corresponding dimensions $\{d_i\}_{i=1}^k$. Let $\mathbf{F} = [w_1 \cdots w_N] \in \mathbb{R}^{D\times N}$ be a matrix whose columns are drawn from $\bigcup_{i=1}^k S_i$. Assume the data is drawn from each subspace and that it is generic. Let Brref$(\mathbf{F})$ be the binary reduced row echelon form of $\mathbf{F}$. Then:*

1. *The inner product $\langle e_i, b_j\rangle$ of a pivot column e_i and a non-pivot column b_j in Brref$(\mathbf{F})$ is one, if and only if the corresponding column vectors $\{w_i, w_j\}$ in $\mathbf{F}$ belong to the same subspace S_l for some $l = 1,\ldots,k$.*
2. *Moreover, $\dim(S_l) = \|b_j\|_1$, where $\|b_j\|_1$ is the l_1-norm of b_j.*
3. *Finally, $w_p \in S_l$ if and only if $b_p = b_j$ or $\langle b_p, b_j\rangle = 1$.*

The theorem above provides an effective approach to cluster the data points. We simply find the reduced row echelon form of the data matrix W and cluster the data points according to the corresponding columns in the reduced row echelon form. For example, consider the following data matrix that contains data points drawn from the union of three linearly independent subspaces (i.e., from $\mathscr{M} = \bigcup\{V_i\}_{i=1}^3$) as column vectors.

$$W = \begin{bmatrix} 2872 & 138 & 342 & 263 & 1956 & 2016 & 1793 & 801 & 195 & 360 & 1076 & 1882 & 1918 & 2350 & 83 \\ 4041 & 249 & 467 & 516 & 129 & 288 & 2612 & 769 & 312 & 174 & 241 & 176 & 3019 & 3270 & 219 \\ 2906 & 4292 & 352 & 7240 & 2861 & 3072 & 1847 & 665 & 6968 & 646 & 1709 & 2794 & 2080 & 2366 & 1012 \\ 5803 & 1405 & 657 & 2498 & 549 & 864 & 3854 & 687 & 2158 & 390 & 629 & 628 & 4711 & 4654 & 545 \\ 5124 & 744 & 2092 & 1335 & 662 & 1056 & 2835 & 1116 & 1131 & 484 & 774 & 762 & 4867 & 4546 & 309 \\ 6701 & 3192 & 757 & 5420 & 775 & 1248 & 4502 & 578 & 5148 & 578 & 919 & 896 & 5638 & 5354 & 812 \\ 7102 & 1625 & 802 & 2862 & 888 & 1440 & 4793 & 522 & 2522 & 672 & 1064 & 1030 & 6059 & 5666 & 585 \\ 495 & 223 & 117 & 577 & 322 & 960 & 266 & 247 & 169 & 668 & 866 & 520 & 275 & 430 & 388 \\ 1184 & 2910 & 192 & 8282 & 435 & 1152 & 755 & 200 & 1482 & 762 & 1011 & 654 & 951 & 970 & 6320 \\ 2065 & 1117 & 287 & 3027 & 4040 & 4800 & 1376 & 159 & 715 & 1360 & 2920 & 4100 & 1797 & 1662 & 2172 \end{bmatrix}. \tag{6}$$

The computed (binary) reduced row echelon form of W is W_b:

$$W_b = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{7}$$