Applied and Numerical Harmonic Analysis

Holger Boche
Giuseppe Caire
Robert Calderbank
Maximilian März
Gitta Kutyniok
Rudolf Mathar
Editors

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

Compressed Sensing and its Applications

Second International MATHEON Conference 2015

Birkhäuser

Birkhäuser

Applied and Numerical Harmonic Analysis

More information about this series at http://www.springer.com/series/4968

Holger Boche • Giuseppe Caire • Robert Calderbank
Maximilian März • Gitta Kutyniok • Rudolf Mathar
Editors

Compressed Sensing and its Applications

Second International MATHEON Conference 2015

Birkhäuser

Editors
Holger Boche
Fakultät für Elektrotechnik und
Informationstechnik
Technische Universität München
Munich, Bavaria, Germany

Robert Calderbank
Department of Electrical
& Computer Engineering
Duke University
Durham, North Carolina, USA

Gitta Kutyniok
Institut für Mathematik
Technische Universität Berlin
Berlin, Germany

Giuseppe Caire
Institut für Telekommunikationssysteme
Technische Universität Berlin
Berlin, Germany

Maximilian März
Institut für Mathematik
Technische Universität Berlin
Berlin, Germany

Rudolf Mathar
Lehrstuhl und Institute für Statistik
RWTH Aachen
Aachen, Germany

ISSN 2296-5009 ISSN 2296-5017 (electronic)
Applied and Numerical Harmonic Analysis
ISBN 978-3-319-69801-4 ISBN 978-3-319-69802-1 (eBook)
https://doi.org/10.1007/978-3-319-69802-1

Library of Congress Control Number: 2017960841

Mathematics Subject Classification (2010): 94A12, 94A20, 68U10, 90C25, 15B52

Printed on acid-free paper

This book is published under the trade name Birkhäuser, www.birkhauser-science.com
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

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-the-art *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 with 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:

Antenna theory	*Prediction theory*
Biomedical signal processing	*Radar applications*
Digital signal processing	*Sampling theory*
Fast algorithms	*Spectral estimation*
Gabor theory and applications	*Speech processing*
Image processing	*Time-frequency and*
Numerical partial differential equations	*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, MD, USA

John J. Benedetto
Series Editor

Preface

The key challenge of compressed sensing arises from the task of recovering signals from a small collection of measurements, exploiting the fact that high-dimensional signals are typically governed by intrinsic low-complexity structures, for instance, being sparse in an orthonormal basis. While the reconstruction from such compressed, typically randomly selected measurements is well studied from a theoretical perspective, there also exist numerous efficient recovery algorithms exhibiting excellent practical performance and thereby making compressed sensing relevant to many different applications. In fact, from an early stage on, the field has greatly benefited from the interaction between mathematics, engineering, computer science, and physics, leading to new theoretical insights as well as significant improvements of real-world applications.

From the point of view of applied mathematics, the field makes use of tools from applied harmonic analysis, approximation theory, linear algebra, convex optimization, and probability theory, while the applications encompass many areas such as image processing, sensor networks, radar technology, quantum computing, or statistical learning, to name just a very few. Nowadays, it is fair to say that more than 10 years after its emergence, the field of compressed sensing has reached a mature state, where many of the underlying mathematical foundations are quite well understood. Therefore, some of the techniques and results are now being transferred to other related areas, leading to a broader conception of compressed sensing and opening up new possibilities for applications.

In December 2015, the editors of this volume organized the *Second International MATHEON Conference on Compressed Sensing and its Applications* at the Technische Universität Berlin. This conference was supported by the research center for *Mathematics for Key Technologies* (MATHEON), as well as the *German Research Foundation (DFG)*. It was attended by more than 150 participants from 18 different countries, and as in the first workshop of this series in 2013, experts in a variety of different research areas were present. This diverse background of participants led to a very fruitful exchange of ideas and to stimulating discussions.

This book is the second volume in the *Applied and Numerical Harmonic Analysis* book series on *Compressed Sensing and its Applications,* presenting state-of-the-art monographs on various topics in compressed sensing and related fields. It is aimed at a broad readership, reaching from graduate students to senior researchers in applied mathematics, engineering, and computer science.

This volume features contributions by two of the plenary speakers (chapters "On the Global-Local Dichotomy in Sparsity Modeling" and "Fourier Phase Retrieval: Uniqueness and Algorithms"), namely, Michael Elad (Technion—Israel Institute of Technology) and Yonina C. Eldar (Technion—Israel Institute of Technology), and by ten invited speakers (chapters "Compressed Sensing Approaches for Polynomial Approximation of High-Dimensional Functions," "Multisection in the Stochastic Block Model Using Semidefinite Programming," "Recovering Signals with Unknown Sparsity in Multiple Dictionaries," "Compressive Classification and the Rare Eclipse Problem," "Weak Phase Retrieval," "Cubatures on Grassmannians: Moments, Dimension Reduction, and Related Topics," "A Randomized Tensor Train Singular Value Decomposition," "Versatile and Scalable Cosparse Methods for Physics-Driven Inverse Problems," "Total Variation Minimization in Compressed Sensing," "Compressed Sensing in Hilbert Spaces"), namely, Ben Adcock (Simon Fraser University), Alfonso S. Bandeira (Massachusetts Institute of Technology), Peter G. Casazza (University of Missouri, Columbia), Mike E. Davies (University of Edinburgh), Martin Ehler (University of Vienna), Rémi Gribonval (INRIA Rennes), Felix Krahmer (Technische Universität München), Dustin G. Mixon (Air Force Institute of Technology), Reinhold Schneider (Technische Universität Berlin), and Philip Schniter (The Ohio State University, Columbus).

In the following, we will give a brief outline of the content of each chapter. For an introduction and a self-contained overview on compressed sensing and its major achievements, we refer the reader to chapter "A Survey of Compressed Sensing" of the first volume of this book series (*Boche, H., Calderbank, R., Kutyniok, G., and Vybiral, J. (eds.), Compressed Sensing and its Applications: MATHEON Workshop 2013. Birkhäuser Boston, 2015).*

Two of the chapters focus on phase retrieval: chapter "Fourier Phase Retrieval: Uniqueness and Algorithms" contains a detailed overview on Fourier phase retrieval and practical algorithms, whereas chapter "Weak Phase Retrieval" introduces a weaker formulation of the classical phase retrieval. Another key topic is the question how sparsity-promoting transformations are used in compressed sensing. In this realm, chapter "On the Global-Local Dichotomy in Sparsity Modeling" analyzes the gap between local and global sparsity in dictionaries, chapter "Versatile and Scalable Cosparse Methods for Physics-Driven Inverse Problems" focuses on the use of the analysis formulation in physics-driven inverse problems, chapter "Total Variation Minimization in Compressed Sensing" gives an overview over total variation minimization in compressed sensing, and chapter "Recovering Signals with Unknown Sparsity in Multiple Dictionaries" uses iterative reweighting for recovering signals with unknown sparsity in multiple dictionaries. Several chapters focus entirely on mathematical aspects, such as chapter "Compressed Sensing Approaches for Polynomial Approximation of High-Dimensional Functions" which exploits compressed sensing for approximating functions with polynomials. In chapter "Compressed Sensing in Hilbert Spaces," compressed sensing is considered in the abstract framework of Hilbert spaces, and chapter "Compressive Classification and the Rare Eclipse Problem" deals with random projections of convex sets. The other chapters study new frontiers in related areas, such as detecting community-like structures in graphs via the stochastic block model (chapter "Multisection in the Stochastic Block Model Using Semidefinite Programming"), cubatures on Grassmannians and their connection to the recovery of sparse probability measures (chapter "Cubatures on Grassmannians: Moments, Dimension Reduction, and Related Topics"), and an examination of randomized tensor train singular value decompositions (chapter "A Randomized Tensor Train Singular Value Decomposition").

We would like to thank the following current and former members of the research group "Applied Functional Analysis" at the Technische Universität Berlin without whom this conference would not have been possible: Axel Flinth, Martin Genzel, Mijail Guillemard, Anja Hedrich, Sandra Keiper, Anton Kolleck, Maximilian Leitheiser, Jackie Ma, Philipp Petersen, Friedrich Philipp, Mones Raslan, Martin Schäfer, and Yizhi Sun.

München, Germany — Holger Boche
Berlin, Germany — Giuseppe Caire
Durham, USA — Robert Calderbank
Berlin, Germany — Gitta Kutyniok
Berlin, Germany — Maximilian März
Aachen, Germany — Rudolf Mathar
July 2017

Contents

On the Global-Local Dichotomy in Sparsity Modeling

Dmitry Batenkov, Yaniv Romano, and Michael Elad

Abstract The traditional sparse modeling approach, when applied to inverse problems with large data such as images, essentially assumes a sparse model for small overlapping data patches and processes these patches as if they were independent from each other. While producing state-of-the-art results, this methodology is suboptimal, as it does not attempt to model the entire global signal in any meaningful way—a nontrivial task by itself.

In this paper we propose a way to bridge this theoretical gap by constructing a global model from the bottom-up. Given local sparsity assumptions in a dictionary, we show that the global signal representation must satisfy a constrained underdetermined system of linear equations, which forces the patches to agree on the overlaps. Furthermore, we show that the corresponding global pursuit can be solved via local operations. We investigate conditions for unique and stable recovery and provide numerical evidence corroborating the theory.

Keywords Sparse representations · Inverse problems · Convolutional sparse coding

D. Batenkov (✉)
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: batenkov@mit.edu

Y. Romano
Department of Electrical Engineering, Technion - Israel Institute of Technology, 32000 Haifa, Israel
e-mail: yromano@tx.technion.ac.il

M. Elad
Department of Computer Science, Technion - Israel Institute of Technology, 32000 Haifa, Israel
e-mail: elad@cs.technion.ac.il

H. Boche et al. (eds.), *Compressed Sensing and its Applications*,
Applied and Numerical Harmonic Analysis,
https://doi.org/10.1007/978-3-319-69802-1_1

1 Introduction

1.1 The Need for a New Local-Global Sparsity Theory

The sparse representation model [17] provides a powerful approach to various inverse problems in image and signal processing such as denoising [18, 37], deblurring [14, 57], and super-resolution [47, 56], to name a few [38]. This model assumes that a signal can be represented as a sparse linear combination of a few columns (called atoms) taken from a matrix termed dictionary. Given a signal, the sparse recovery of its representation over a dictionary is called sparse coding or pursuit (such as the orthogonal matching pursuit, OMP, or basis pursuit, BP). Due to computational and theoretical aspects, when treating high-dimensional data, various existing sparsity-inspired methods utilize local patched-based representations rather than the global ones, i.e., they divide a signal into small overlapping blocks (patches), reconstruct these patches using standard sparse recovery techniques, and subsequently average the overlapping regions [11, 17]. While this approach leads to highly efficient algorithms producing state-of-the-art results, the global signal prior remains essentially unexploited, potentially resulting in suboptimal recovery.

As an attempt to tackle this flaw, methods based on the notion of *structured sparsity* [19, 29, 30, 32, 55] started to appear; for example, in [14, 37, 47] the observation that a patch may have similar neighbors in its surroundings (often termed the self-similarity property) is injected to the pursuit, leading to improved local estimations. Another possibility to consider the dependencies between patches is to exploit the multi-scale nature of the signals [36, 40, 53]. A different direction is suggested by the expected patch log likelihood (EPLL) method [40, 52, 60], which encourages the patches of the final estimate (i.e., after the application of the averaging step) to comply with the local prior. Also, a related work [45, 46] suggests promoting the local estimations to agree on their shared content (the overlap) as a way to achieve a coherent reconstruction of the signal.

Recently, an alternative to the traditional patch-based prior was suggested in the form of the convolutional, or shift-invariant, sparse coding (CSC) model [10, 25, 27, 28, 49, 54]. Rather than dividing the image into local patches and processing each of these independently, this approach imposes a specific structure on the global dictionary—a concatenation of banded circulant matrices—and applies a global pursuit. A thorough theoretical analysis of this model was proposed very recently in [41, 42], providing a clear understanding of its success.

The empirical success of the above algorithms indicates the great potential of reducing the inherent gap that exists between the independent local processing of patches and the global nature of the signal at hand. However, a key and highly desirable part is still missing—a theory which would suggest how to modify the basic sparse model to take into account the mutual dependencies between the patches, what approximation methods to use, and how to efficiently design and learn the corresponding structured dictionary.

1.2 Content and Organization of the Paper

In this paper we propose a systematic investigation of the signals which are implicitly defined by local sparsity assumptions. A major theme in what follows is that the presence of patch overlaps reduces the number of degrees of freedom, which, in turn, has theoretical and practical implications. In particular, this allows more accurate estimates for uniqueness and stability of local sparse representations, as well as better bounds on performance of existing sparse approximation algorithms. Moreover, the global point of view allows for development of new pursuit algorithms, which consist of local operation on one hand, while also taking into account the patch overlaps on the other hand. Some aspects of the offered theory are still incomplete, and several exciting research directions emerge as well.

The paper is organized as follows. In Section 2 we develop the basic framework for signals which are patch-sparse, building the global model from the "bottom-up," and discuss some theoretical properties of the resulting model. In Section 3 we consider the questions of reconstructing the representation vector and of denoising a signal in this new framework. We describe "globalized" greedy pursuit algorithms [43] for these tasks, where the patch disagreements play a major role. We show that the frequently used local patch averaging (LPA) approach is in fact suboptimal in this case. In Section 4 and Appendix E: Generative Models for Patch-Sparse Signals, we describe several instances/classes of the local-global model in some detail, exemplifying the preceding definitions and results. The examples include piecewise constant signals, signature-type (periodic) signals, and more general bottom-up models. In Section 5 we present results of some numerical experiments, where in particular we show that one of the new globalized pursuits, inspired by the ADMM algorithm [9, 23, 24, 33], turns out to have superior performance in all the cases considered. We conclude the paper in Section 6 by discussing possible research directions.

2 Local-Global Sparsity

We start with the local sparsity assumptions for every patch and subsequently provide two complimentary characterizations of the resulting global signal space. On one hand, we show that the signals of interest admit a global "sparse-like" representation with a dictionary of convolutional type and with additional linear constraints on the representation vector. On the other hand, the signal space is in fact a union of linear subspaces, where each subspace is a kernel of a certain linear map. To complement and connect these points of view, in Appendix E: Generative Models for Patch-Sparse Signals, we show that the original local dictionary must carry a combinatorial structure, and based on this structure, we develop a generative model for patch-sparse signals. Concluding this section, we provide some theoretical analysis of the properties of the resulting model, in particular uniqueness and stability of representation. For this task, we define certain measures of the dictionary, similar to the classical spark, coherence function, and the restricted isometry property, which take the additional dictionary structure into account. In general, this additional structure implies possibly better uniqueness as well as stability to perturbations; however, it is an open question to show they are provably better in certain cases.

2.1 Preliminaries

Let $[m]$ denote the set $\{1, 2, \ldots, m\}$. If D is an $n \times m$ matrix and $S \subset [m]$ is an index set, then D_S denotes the submatrix of D consisting of the columns indexed by S.

Definition 1 (Spark of a Matrix). Given a dictionary $D \in \mathbb{R}^{n\times m}$, the *spark* of D is defined as the minimal number of columns which are linearly dependent:

$$\sigma(D) := \min\{j :\ \exists S \subset [m],\ |S| = j,\ \operatorname{rank} D_S < j\}. \tag{1}$$

Clearly $\sigma(D) \leqslant n + 1$.

Definition 2. Given a vector $\alpha \in \mathbb{R}^m$, the ℓ_0 pseudo-norm is the number of nonzero elements in α:

$$\|\alpha\|_0 := \#\left\{j :\ \alpha_j \neq 0\right\}.$$

Definition 3. Let $D \in \mathbb{R}^{n\times m}$ be a dictionary with normalized atoms. The μ_1 coherence function (Tropp's Babel function) is defined as

$$\mu_1(s) := \max_{i\in[m]} \max_{S\subset[m]\setminus\{i\},\ |S|=s} \sum_{j\in S} \left|\langle d_i, d_j\rangle\right|.$$

Definition 4. Given a dictionary D as above, the restricted isometry constant of order k is the smallest number δ_k such that

$$(1 - \delta_k)\,\|\alpha\|_2^2 \leqslant \|D\alpha\|_2^2 \leqslant (1 + \delta_k)\,\|\alpha\|_2^2$$

for every $\alpha \in \mathbb{R}^m$ with $\|\alpha\|_0 \leqslant k$.

For any matrix M, we denote by $\mathscr{R}(M)$ the column space (range) of M.

2.2 Globalized Local Model

In what follows we treat one-dimensional signals $x \in \mathbb{R}^N$ of length N, divided into $P = N$ overlapping patches of equal size n (so that the original signal is thought to be periodically extended). The other natural choice is $P = N - n + 1$, but for simplicity of derivations, we consider only the periodic case.

Let $R_1 := \left[I_{n \times n}\ \mathbf{0}\ \mathbf{0} \ldots \mathbf{0} \right] \in \mathbb{R}^{n \times N}$, and for each $i = 2, \ldots, P$, we define $R_i \in \mathbb{R}^{n \times N}$ to be the circular column shift of R_1 by $n \cdot (i-1)$ entries, i.e., this operator extracts the i-th patch from the signal in a circular fashion.

Definition 5. Given local dictionary $D \in \mathbb{R}^{n \times m}$, sparsity level $s < n$, signal length N, and the number of overlapping patches P, the *globalized local sparse* model is the set

$$\mathscr{M} = \mathscr{M}(D, s, P, N) := \left\{ x \in \mathbb{R}^N,\ R_i x = D\alpha_i,\ \|\alpha_i\|_0 \leqslant s\ \forall i = 1, \ldots, P \right\}. \tag{2}$$

This model suggests that each patch, $R_i x$ is assumed to have an s-sparse representation α_i, and this way we have characterized the global x by describing the local nature of its patches.

Next we derive a "global" characterization of $\mathscr{M}$. Starting with the equations

$$R_i x = D\alpha_i, \qquad i = 1, \ldots, P,$$

and using the equality $I_{N \times N} = \frac{1}{n} \sum_{i=1}^{P} R_i^T R_i$, we have a representation

$$x = \frac{1}{n} \sum_{i=1}^{P} R_i^T R_i x = \sum_{i=1}^{P} \left(\frac{1}{n} R_i^T D \right) \alpha_i.$$

Let the global "convolutional" dictionary D_G be defined as the horizontal concatenation of the (vertically) shifted versions of $\frac{1}{n}D$, i.e., (see Figure 1 on page 5)

$$D_G := \left[\left(\frac{1}{n} R_i^T D \right) \right]_{i=1,\ldots P} \in \mathbb{R}^{N \times mP}. \tag{3}$$

Let $\Gamma \in \mathbb{R}^{mP}$ denote the concatenation of the local sparse codes, i.e.,

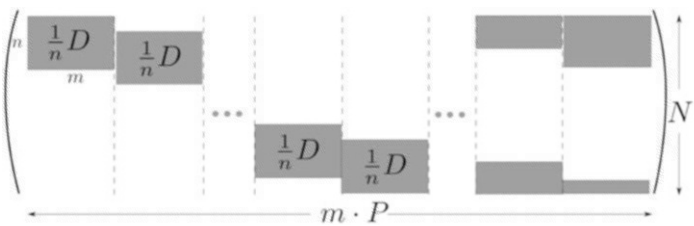

Fig. 1 The global dictionary D_G. After permuting the columns, the matrix becomes a union of circulant Toeplitz matrices, hence the term "convolutional".

$$\Gamma := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_P \end{bmatrix}.$$

Given a vector Γ as above, we will denote by $\tilde{R}_i$ the operator of extracting its i-th portion,[1], i.e., $\tilde{R}_i \Gamma \equiv \alpha_i$.

Summarizing the above developments, we have the global convolutional representation for our signal as follows:

$$x = D_G \Gamma. \tag{4}$$

Next, applying R_i to both sides of (4) and using (2), we obtain

$$D\alpha_i = R_i x = R_i D_G \Gamma. \tag{5}$$

Let $\Omega_i := R_i D_G$ denote the i-th stripe from the global convolutional dictionary D_G. Thus (5) can be rewritten as

$$\underbrace{\left[\mathbf{0} \dots \mathbf{0}\; D\; \mathbf{0} \dots \mathbf{0}\right]}_{:=Q_i} \Gamma = \Omega_i \Gamma, \tag{6}$$

or $(Q_i - \Omega_i)\,\Gamma = 0$. Since this is true for all $i = 1, \dots, P$, we have shown that the vector Γ satisfies

$$\underbrace{\begin{bmatrix} Q_1 - \Omega_1 \\ \vdots \\ Q_P - \Omega_P \end{bmatrix}}_{:=M \in \mathbb{R}^{nP \times mP}} \Gamma = 0.$$

Thus, the condition that the patches $R_i x$ agree on the overlaps is equivalent to the global representation vector Γ residing in the null-space of the matrix M.

An easy computation provides the dimension of this null-space (see proof in Appendix A: Proof of Lemma 1), or in other words the overall number of degrees of freedom of admissible Γ.

[1]Notice that while R_i extracts the i-th patch from the signal x, the operator $\tilde{R}_i$ extracts the representation α_i of $R_i x$ from Γ.

Lemma 1. *For any frame $D \in \mathbb{R}^{n\times m}$ (i.e., a full rank dictionary), we have*

$$\dim \ker M = N\,(m - n + 1)\,.$$

Note that in particular for $m = n$, we have $\dim \ker M = N$, and since in this case D is invertible, we have $R_i x = D\alpha_i$ where $\alpha_i = D^{-1}R_i x$, so that every signal admits a unique representation $x = D_G \Gamma$ with $\Gamma = \left(D^{-1}R_1 x, \dots, D^{-1}R_P x\right)^T$.

As we shall demonstrate now, the equation $M\Gamma = 0$ represents the requirement that the local sparse codes $\{\alpha_i\}$ are not independent but rather should be such that the corresponding patches $D\alpha_i$ agree on the overlaps.

Definition 6. Define the "extract from top/bottom" operators $S_T \in \mathbb{R}^{(n-1)\times n}$ and $S_B \in \mathbb{R}^{(n-1)\times n}$:

$$S_{T(op)} = \left[I_{n-1}\ \mathbf{0}\right], \quad S_{B(ottom)} = \left[\mathbf{0}\ I_{n-1}\right].$$

The following result is proved in Appendix B: Proof of Lemma 2.

Lemma 2. *Let $\Gamma = [\alpha_1, \dots, \alpha_P]^T$. Under the above definitions, the following are equivalent:*

1. $M\Gamma = 0$;
2. *For each $i = 1, \dots, P$, we have $S_B D\alpha_i = S_T D\alpha_{i+1}$.*

Definition 7. Given $\Gamma = [\alpha_1, \dots, \alpha_P]^T \in \mathbb{R}^{mP}$, the $\|\cdot\|_{0,\infty}$ pseudo-norm is defined by

$$\|\Gamma\|_{0,\infty} := \max_{i=1,\dots,P} \|\alpha_i\|_0.$$

Thus, every signal complying with the patch-sparse model, with sparsity s for each patch, admits the following representation.

Theorem 1. *Given D, s, P, and N, the globalized local sparse model (2) is equivalent to*

$$\begin{aligned}\mathcal{M} &= \left\{x \in \mathbb{R}^N : \ x = D_G\Gamma,\ M\Gamma = 0,\ \|\Gamma\|_{0,\infty} \leqslant s\right\} \\ &= \left\{x \in \mathbb{R}^N : \ x = D_G\Gamma,\ M_*\Gamma = 0,\ \|\Gamma\|_{0,\infty} \leqslant s\right\},\end{aligned} \tag{7}$$

where the matrix $M_* \in \mathbb{R}^{(n-1)P\times mP}$ is defined as

$$M_* := \begin{bmatrix} S_B D & -S_T D & & \\ & S_B D & -S_T D & \\ & & \ddots & \ddots \\ & & & \end{bmatrix}.$$

Proof. If $x \in \mathscr{M}$ (according to (2)), then by the above construction x belongs to the set defined by the RHS of (7) (let's call it $\mathscr{M}^*$ for the purposes of this proof only). In the other direction, assume that $x \in \mathscr{M}^*$. Now $R_i x = R_i D_G \Gamma = \Omega_i \Gamma$, and since $M\Gamma = 0$, we have $R_i x = Q_i \Gamma = D\tilde{R}_i \Gamma$. Denote $\alpha_i := \tilde{R}_i \Gamma$, and so we have that $R_i x = D\alpha_i$ with $\|\alpha_i\|_0 \leqslant s$, i.e., $x \in \mathscr{M}$ by definition. The second part follows from Lemma 2. □

We say that α_i is a *minimal* representation of x_i if $x_i = D\alpha_i$ such that the matrix $D_{\operatorname{supp}\alpha_i}$ has full rank—and therefore the atoms participating in the representation are linearly independent.[2]

Definition 8. Given a signal $x \in \mathscr{M}$, let us denote by $\rho(x)$ the set of all locally sparse *and minimal* representations of x:

$$\rho(x) := \left\{\Gamma \in \mathbb{R}^{mP} : \ \|\Gamma\|_{0,\infty} \leqslant s, \ x = D_G\Gamma, \ M\Gamma = 0, \ D_{\operatorname{supp}\tilde{R}_i\Gamma} \text{ is full rank}\right\}.$$

Let us now go back to the definition (2). Consider a signal $x \in \mathscr{M}$, and let $\Gamma \in \rho(x)$. Denote $S_i := \operatorname{supp}\tilde{R}_i\Gamma$. Then we have $R_i x \in \mathscr{R}(D_{S_i})$, and therefore we can write $R_i x = P_{S_i} R_i x$, where P_{S_i} is the orthogonal projection operator onto $\mathscr{R}(D_{S_i})$. In fact, since D_{S_i} is full rank, we have $P_{s_i} = D_{S_i} D_{S_i}^{\dagger}$ where $D_{S_i}^{\dagger} = \left(D_{S_i}^T D_{S_i}\right)^{-1} D_{S_i}^T$ is the Moore-Penrose pseudoinverse of D_{S_i}.

Definition 9. Given a support sequence $\mathscr{S} = (S_1, \ldots, S_P)$, define the matrix $A_{\mathscr{S}}$ as follows:

$$A_{\mathscr{S}} := \begin{bmatrix} (I_n - P_{S_1}) R_1 \\ (I_n - P_{S_2}) R_2 \\ \vdots \\ (I_n - P_{S_P}) R_P \end{bmatrix} \in \mathbb{R}^{nP \times N}.$$

The map $A_{\mathscr{S}}$ measures the local patch discrepancies, i.e., how "far" is each local patch from the range of a particular subset of the columns of D.

Definition 10. Given a model $\mathscr{M}$, denote by $\Sigma_{\mathscr{M}}$ the set of all valid supports, i.e.,

$$\begin{aligned}\Sigma_{\mathscr{M}} : = \{(S_1, \ldots, S_P) : \ \exists x \in \mathscr{M}, \ \Gamma amma \in \rho(x) \ \text{ s.t. } \forall i = 1, \ldots, P : \\ S_i = \operatorname{supp}\tilde{R}_i\Gamma\}.\end{aligned}$$

With this notation in place, it is immediate to see that the global signal model is a union of subspaces.

Theorem 2. *The global model is equivalent to the union of subspaces*

$$\mathscr{M} = \bigcup_{\mathscr{S} \in \Sigma_{\mathscr{M}}} \ker A_{\mathscr{S}}.$$

[2]Notice that α_i might be a minimal representation but not a unique one with minimal sparsity. For discussion of uniqueness, see Subsection 2.3.

Remark 1. Contrary to the well-known union of subspaces model [7, 35], the subspaces $\{\ker A_{\mathscr{S}}\}$ do not have in general a sparse joint basis, and therefore our model is distinctly different from the well-known block-sparsity model [19, 20].

An important question of interest is to estimate $\dim\ker A_{\mathscr{S}}$ for a given $\mathscr{S} \in \Sigma_{\mathscr{M}}$. One possible solution is to investigate the "global" structure of the corresponding signals (as is done in Subsection 4.1 and Subsection 4.2), while another option is to utilize information about "local connections" (Appendix E: Generative Models for Patch-Sparse Signals).

2.3 Uniqueness and Stability

Given a signal $x \in \mathscr{M}$, it has a globalized representation $\Gamma \in \rho(x)$ according to Theorem 1. When is such a representation unique, and under what conditions can it be recovered when the signal is corrupted with noise?

In other words, we study the problem

$$\min \|\Gamma\|_{0,\infty} \qquad \text{s.t. } D_G\Gamma = D_G\Gamma_0,\ M\Gamma = 0 \qquad (P_{0,\infty})$$

and its noisy version

$$\min \|\Gamma\|_{0,\infty} \qquad \text{s.t. } \|D_G\Gamma - D_G\Gamma_0\| \leqslant \varepsilon,\ M\Gamma = 0 \qquad \left(P^{\varepsilon}_{0,\infty}\right).$$

For this task, we define certain measures of the dictionary, similar to the classical spark, coherence function, and the restricted isometry property, which take the additional dictionary structure into account. In general, the additional structure implies *possibly* better uniqueness as well as stability to perturbations; however, it is an open question to show they are *provably* better in certain cases.

The key observation is that the global model $\mathscr{M}$ imposes a constraint on the allowed local supports.

Definition 11. Denote the set of allowed local supports by

$$\mathscr{T} := \{T :\ \exists\,(S_1,\dots,T,\dots,S_P) \in \Sigma_{\mathscr{M}}\}.$$

Recall the definition of the spark (1). Clearly $\sigma(D)$ can be equivalently rewritten as

$$\sigma(D) = \min\{j :\ \exists S_1, S_2 \subset [m],\ |S_1 \cup S_2| = j,\ \operatorname{rank} D_{S_1\cup S_2} < j\}. \tag{8}$$

Definition 12. The *globalized spark* $\sigma^*(D)$ is

$$\sigma^*(D) := \min\{j :\ \exists S_1, S_2 \in \mathscr{T},\ |S_1 \cup S_2| = j,\ \operatorname{rank} D_{S_1\cup S_2} < j\}. \tag{9}$$

The following proposition is immediate by comparing (8) with (9).

Proposition 1. $\sigma^*(D) \geqslant \sigma(D)$.

The globalized spark provides a uniqueness result in the spirit of [15].

Theorem 3 (Uniqueness). *Let $x \in \mathscr{M}(D,s,N,P)$. If there exists $\Gamma \in \rho(x)$ for which $\|\Gamma\|_{0,\infty} < \frac{1}{2}\sigma^*(D)$ (i.e., it is a sufficiently sparse solution of $P_{0,\infty}$), then it is the unique solution (and so $\rho(x) = \{\Gamma\}$).*

Proof. Suppose that there exists $\Gamma_0 \in \rho(x)$ which is different from Γ. Put $\Gamma_1 := \Gamma - \Gamma_0$, then $\|\Gamma_1\|_{0,\infty} < \sigma^*(D)$, while $D_G\Gamma_1 = 0$ and $M\Gamma_1 = 0$. Denote $\beta_j := \tilde{R}_j\Gamma_1$. By assumption, there exists an index i for which $\beta_i \neq 0$, but we must have $D\beta_j = 0$ for every j, and therefore $D_{\operatorname{supp}\beta_i}$ must be rank-deficient—contradicting the fact that $\|\beta_i\| < \sigma^*(D)$. □

In classical sparsity, we have the bound

$$\sigma(D) \geqslant \min\{s : \mu_1(s-1) \geqslant 1\}, \tag{10}$$

where μ_1 is given by Definition 3. In a similar fashion, the globalized spark σ^* can be bounded by an appropriate analog of "coherence"—however, computing this new coherence appears to be in general intractable.

Definition 13. Given the model $\mathscr{M}$, we define the following globalized coherence function

$$\mu_1^*(s) := \max_{S\in\mathscr{T}\cup\mathscr{T},|S|=s} \max_{j\in S} \sum_{k\in S\setminus\{j\}} \left|\langle d_j, d_k\rangle\right|,$$

where $\mathscr{T}\cup\mathscr{T} := \{S_1\cup S_2 : S_1, S_2 \in \mathscr{T}\}$.

Theorem 4. *The globalized spark σ^* can be bounded by the globalized coherence as follows*[3]*:*

$$\sigma^*(D) \geqslant \min\{s : \mu_1^*(s) \geqslant 1\}.$$

Proof. Following closely the corresponding proof in [15], assume by contradiction that

$$\sigma^*(D) < \min\{s : \mu_1^*(s) \geqslant 1\}.$$

Let $S^* \in \mathscr{T}\cup\mathscr{T}$ with $|S^*| = \sigma^*(D)$ for which D_{S^*} is rank-deficient. Then the restricted Gram matrix $G := D_{S^*}^T D_{S^*}$ must be singular. On the other hand, $\mu_1^*(|S^*|) < 1$, and so in particular

$$\max_{j\in S^*} \sum_{k\in S^*\setminus\{j\}} \left|\langle d_j, d_k\rangle\right| < 1.$$

[3]In general $\min\left\{s : \mu_1^*(s-1) \geqslant 1\right\} \neq \max\left\{s : \mu_1^*(s) < 1\right\}$ because the function μ_1^* need not be monotonic.

But that means that G is diagonally dominant and therefore $\det G \neq 0$, a contradiction. □

We see that $\mu_1^*(s+1) \leqslant \mu_1(s)$ since the outer maximization is done on a smaller set. Therefore, in general the bound of Theorem 4 appears to be sharper than (10).

A notion of globalized RIP can also be defined as follows.

Definition 14. The globalized RIP constant of order k associated to the model $\mathscr{M}$ is the smallest number $\delta_{k,\mathscr{M}}$ such that

$$(1-\delta_{k,\mathscr{M}})\,\|\alpha\|_2^2 \leqslant \|D\alpha\|_2^2 \leqslant (1+\delta_{k,\mathscr{M}})\,\|\alpha\|_2^2$$

for every $\alpha \in \mathbb{R}^m$ with $\operatorname{supp}\alpha \in \mathscr{T}$.
Immediately one can see the following (recall Definition 4).

Proposition 2. *The globalized RIP constant is upper bounded by the standard RIP constant:*

$$\delta_{k,\mathscr{M}} \leqslant \delta_k.$$

Definition 15. The generalized RIP constant of order k associated to signals of length N is the smallest number $\delta_k^{(N)}$ such that

$$\left(1-\delta_k^{(N)}\right)\|\Gamma\|_2^2 \leqslant \|D_G\Gamma\|_2^2 \leqslant \left(1+\delta_k^{(N)}\right)\|\Gamma\|_2^2$$

for every $\Gamma \in \mathbb{R}^{mN}$ satisfying $M\Gamma = 0,\ \|\Gamma\|_{0,\infty} \leqslant k$.

Proposition 3. *We have*

$$\delta_k^{(N)} \leqslant \frac{\delta_{k,\mathscr{M}} + (n-1)}{n} \leqslant \frac{\delta_k + (n-1)}{n}.$$

Proof. Obviously it is enough to show only the leftmost inequality. If $\Gamma = (\alpha_i)_{i=1}^N$ and $\|\Gamma\|_{0,\infty} \leqslant k$, this gives $\|\alpha_i\|_0 \leqslant k$ for all $i = 1,\dots,P$. Further, setting $x := D_G\Gamma$ we clearly have $\Gamma \in \rho(x)$ and so $\operatorname{supp}\Gamma \in \Sigma_{\mathscr{M}}$. Thus $\operatorname{supp}\alpha_i \in \mathscr{T}$, and therefore

$$(1-\delta_{k,\mathscr{M}})\,\|\alpha_i\|_2^2 \leqslant \|D\alpha_i\|_2^2 \leqslant (1+\delta_{k,\mathscr{M}})\,\|\alpha_i\|_2^2.$$

By Corollary 3 we know that for every Γ satisfying $M\Gamma = 0$, we have

$$\|D_G\Gamma\|_2^2 = \frac{1}{n}\sum_{i=1}^{N}\|D\alpha_i\|_2^2.$$

Now for the lower bound,

$$\|D_G\Gamma\|_2^2 \geqslant \frac{1-\delta_{k,\mathscr{M}}}{n}\sum_{i=1}^{N}\|\alpha_i\|_2^2 = \left(1-1+\frac{1-\delta_{k,\mathscr{M}}}{n}\right)\|\Gamma\|_2^2$$
$$= \left(1-\frac{\delta_{k,\mathscr{M}}+(n-1)}{n}\right)\|\Gamma\|_2^2.$$

For the upper bound,

$$\|D_G\Gamma\|_2^2 \leqslant \frac{1+\delta_{k,\mathscr{M}}}{n}\sum_{i=1}^{N}\|\alpha_i\|_2^2 < \left(1+\frac{\delta_{k,\mathscr{M}}+1}{n}\right)\|\Gamma\|_2^2$$
$$\leqslant \left(1+\frac{\delta_{k,\mathscr{M}}+(n-1)}{n}\right)\|\Gamma\|_2^2.$$

□

Theorem 5 (Uniqueness and Stability of $P_{0,\infty}$ via RIP). *Suppose that $\delta_{2s}^{(N)} < 1$, and suppose further that $x = D_G\Gamma_0$ with $\|\Gamma_0\|_{0,\infty} = s$ and $\|D_G\Gamma_0 - x\|_2 \leqslant \varepsilon$. Then every solution $\hat{\Gamma}$ of the noise-constrained $P_{0,\infty}^{\varepsilon}$ problem*

$$\hat{\Gamma} \leftarrow \arg\min_{\Gamma}\|\Gamma\|_{0,\infty} \text{ s.t. } \|D_G\Gamma - x\| \leqslant \varepsilon,\ M\Gamma = 0$$

satisfies

$$\|\hat{\Gamma}-\Gamma_0\|_2^2 \leqslant \frac{4\varepsilon^2}{1-\delta_{2s}^{(N)}}.$$

In particular, Γ_0 is the unique solution of the noiseless $P_{0,\infty}$ problem.

Proof. Immediate using the definition of the globalized RIP:

$$\|\hat{\Gamma}-\Gamma_0\|_2^2 < \frac{1}{1-\delta_{2s}^{(N)}}\|D_G\left(\hat{\Gamma}-\Gamma_0\right)\|_2^2 \leqslant \frac{1}{1-\delta_{2s}^{(N)}}\left(\|D_G\hat{\Gamma}-x\|_2+\|D_G\Gamma_0-x\|_2\right)^2$$
$$\leqslant \frac{4\varepsilon^2}{1-\delta_{2s}^{(N)}}.$$

□

3 Pursuit Algorithms

In this section we consider the problem of efficient projection onto the model $\mathscr{M}$. First we treat the "oracle" setting, i.e., when the supports of the local patches (and therefore of the global vector Γ) are known. We show that the local patch averaging