Low-overhead Communications in IoT Networks

Yuanming Shi • Jialin Dong • Jun Zhang

Low-overhead Communications in IoT Networks

Structured Signal Processing Approaches

Springer

Yuanming Shi
School of Information Science
and Technology
Shanghai Tech University
Shanghai, China

Jialin Dong
School of Information Science
and Technology
ShanghaiTech University
Shanghai, China

Jun Zhang
Department of Electronic & Information
Engineering
Hong Kong Polytechnic University
Kowloon, Hong Kong

ISBN 978-981-15-3869-8 ISBN 978-981-15-3870-4 (eBook)
https://doi.org/10.1007/978-981-15-3870-4

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The past decades have witnessed a revolution in wireless communications and networking, which has profoundly changed our daily life. Particularly, it has enabled various innovative Internet of Things (IoT) applications, e.g., smart city, healthcare, and autonomous driving and drones. The IoT architecture is established by the proliferation of low-cost and small-size mobile devices. With the explosion of IoT devices, a heavy burden is placed on the wireless access. A key characteristic of IoT data traffic is the sporadic pattern, i.e., only a portion of all the devices are active at a given time instant. In particular, in many IoT applications, devices are designed to be inactive most of the time to save energy and only be activated by external events. Thus, with massive IoT devices, it is of vital importance to manage their random access procedures, detect the active ones, and decode their data at the access point. Massive IoT connectivity has been regarded as one of the key performance requirements of 5G and beyond networks.

The emerging IoT applications have stringent demands on low-latency communications and typically transmit short packets containing both the metadata and payload. The metadata may include packet initiation and termination information, logical addresses, security and synchronization information, etc. It also contains a channel estimation sequence that facilitates channel estimation at the access point. Additionally, the metadata includes various information about the packet structure, e.g., the pilot sequences used for random access and device identification information. Considering the typical small payload size of IoT applications, it is of critical importance to reduce the size of the overhead message, e.g., identification information, pilot symbols for channel estimation, control data, etc. Such low-overhead communications also help to improve the energy efficiency of IoT devices. Recently, structured signal processing approaches have been introduced and developed to reduce the overheads for key design problems in IoT networks, such as channel estimation, device identification, and message decoding. By exploiting underlying system and problem structures, including sparsity and low rank structures, these methods can achieve significant performance gains. Chapter 1 provides more background on low-overhead communications in IoT networks and introduces general structured signal processing techniques.

This monograph shall provide an overview of four structured signal processing models, i.e., a sparse linear model, a blind demixing model, a sparse blind demixing model, and a shuffled linear regression model. Chapter 2 introduces a sparse linear model for joint activity detection and channel estimation in IoT networks with grant-free random access. A convex relaxation approach based on ℓ_p-norm minimization is firstly introduced, followed by a smoothed primal-dual first-order algorithm to solve it. For this convex relaxation approach, a trade-off between the computational cost and estimation accuracy is characterized by Proposition 2.1. The theoretical analysis of the convex relaxation approach is based on the conic integral geometry theory. This chapter only contains a brief introduction on the conic integral geometry theory. For more details, the interested reader can refer to Sect. 8.1 and other related mathematical literature enumerated in this monograph. Besides, an iterative threshold algorithm, namely approximate message passing (AMP), is introduced in Chap. 2, followed by the performance analysis based on the state evolution technique.

Blind demixing is introduced in Chap. 3, which facilitates joint data decoding and channel estimation without explicit pilot sequences. After presenting the basic convex relaxation approach for solving the blind demixing problem, we introduce three nonconvex approaches: the regularized Wirtinger flow, the regularization-free Wirtinger flow, and a Riemannian optimization algorithm. Theorems 3.1 and 3.2 provide the theoretical analysis of the convex relaxation approach and regularized Wirtinger flow, respectively. Furthermore, Theorem 3.3 presents the theoretical guarantees of the Wirtinger flow with the spectral initialization, which provides readers an easy access to well-round results. Readers who are interested in the intrinsic mechanism of the theoretical analysis can refer to Sect. 8.3 for more discussions. The theoretical analysis of the Wirtinger flow via random initialization is further provided in Sect. 8.4. Additionally, the basic concepts of Riemannian manifold optimization are presented in Sect. 8.5, which provide sufficient background for related algorithms in Chaps. 3 and 4. The extension of blind demixing, i.e., sparse blind demixing, is introduced in Chap. 4, which further takes device activity into consideration. The sparse blind demixing formulation is able to jointly consider device activity detection, data decoding, and channel estimation, for which three approaches are presented: a convex relaxation approach, a difference-of-convex-functions approach, and smoothed Riemannian optimization.

A further step to reduce the overhead is to remove the device identification information from the metadata. To support the joint data decoding and device identification, shuffled linear regression is introduced in Chap. 5. We first present maximum likelihood estimation (MLE) based approaches for solving the shuffled linear regression problem. Theorems 5.1 and 5.2 provide the statistical properties of the MLE, and both an upper bound and a lower bound on the probability of error of the permutation matrix estimator are introduced. To solve the MLE problem, two algorithms are presented: one is based on sorting, and the other algorithm returns an approximate solution to the MLE problem. Next, theoretical analysis of the shuffled linear regression problem based on the algebraic–geometric theory is presented. Based on the analysis, an algebraically initialized expectation–maximization algorithm is introduced to solve the shuffled linear regression problem, which enjoys better algorithmic performance than previous works. To give a comprehensive introduction of the algebraic–geometric theory, besides the concepts mentioned in this chapter, we introduce several related definitions on the algebraic–geometric theory in Sect. 8.7, including the geometric characterization of dimension, algebraic characterization of dimension, homogenization, and regular sequences.

Furthermore, Chap. 6 provides some cutting-edge learning augmented techniques for structured signal processing on the aspects of structured signal model design (e.g., structured signal processing under a generative prior) and algorithm design (e.g., deep-learning-based algorithm). We begin with compressed sensing under a generative prior, and other structure signal processing techniques under a generative model are worth further investigating, e.g., blind deconvolution. We next consider the joint design of measurement matrix and sparse support recovery for the sparse linear model (e.g., compressed sensing). Some basic methods are firstly presented, i.e., sample scheduling and sensing matrix optimization, and then learning augmented techniques are introduced. Additionally, for estimating the sparse linear model, several deep-learning-based AMP methods are introduced in this chapter: learned AMP, learned Vector-AMP, and learned ISTA for group row sparsity. In Chap. 7, we summarize the book and discuss some potential extensions of the area of interest. Tables 7.1 and 7.2 list the main theorems, propositions, and algorithms presented in this monograph.

The monograph is not only suitable for beginners in structured signal processing for applications in IoT networks but also helpful to experienced researchers who intend to work in-depth on the theoretical analysis of structured signals. For beginners, the background of both low-overhead communications and structured signal processing in Chap. 1 is helpful, and the problem formulation section in each chapter may be referred for further details with respect to each model. Tables 1.1, 7.1, and 7.2 provide quick references for the main results. Readers who are more interested in the intrinsic mechanism of the theoretical analysis of the specific models can refer to Chap. 8.

Low-overhead communications supported by structured signal processing approaches have received significant attention in recent years. The main motivation of this monograph is to provide an overview of the major structured signal processing models, along with their applications in low-overhead communications in IoT networks. Practical algorithms, via both convex and nonconvex optimization approaches, and theoretical analysis, using various mathematical tools, will be introduced. While the structured signal models concerned in this monograph have certain limitations, we hope the presented results will galvanize researchers into investigating this influential and intriguing area.

Acknowledgements

We express our gratitude to the support of National Nature Science Foundation of China under Grant 61601290. We also thank Xinyu Bian for proofreading an early version of the manuscript and Prof. Liang Liu for providing codes for part of the simulations in Sect. 2.4.

Shanghai, China Yuanming Shi
Shanghai, China Jialin Dong
Kowloon, Hong Kong Jun Zhang

Contents

Mathematical Notations

- The set of real numbers is denoted by $\mathbb{R}$ and the set of positive real numbers is denoted by $\mathbb{R}_+$. The set of complex numbers is denoted by $\mathbb{C}$. Denote $\mathbb{S}_+$ as the set of Hermitian positive semidefinite matrices. Moreover, $\mathbb{N}$ represents the set of natural numbers.
- The boldface and lowercase alphabets, e.g., $\boldsymbol{x}$, $\boldsymbol{y}$, denote vectors. The zero vector is denoted by $\mathbf{0}$. A vector $\boldsymbol{x} \in \mathbb{R}^d$ is in the column format. The transpose of a vector is denoted by $\boldsymbol{x}^\top$. The complex conjugate of $\boldsymbol{x}$ is represented as $\bar{\boldsymbol{x}}$. The conjugate transpose of a vector is denoted by $\boldsymbol{x}^{\mathsf{H}}$ or $\boldsymbol{x}^*$. x_i denotes the i-th coordinate of a vector $\boldsymbol{x}$.
- For a complex vector $\boldsymbol{x}$ or a complex matrix $\boldsymbol{X}$, the real parts of them are represented by $\Re\{\boldsymbol{x}\}$ and $\Re\{\boldsymbol{X}\}$, respectively. Likewise, the imaginary parts are denoted as $\Im\{\boldsymbol{x}\}$ and $\Im\{\boldsymbol{X}\}$.
- The boldface and uppercase alphabets, e.g., $\boldsymbol{A}$, $\boldsymbol{B}$, denote matrices. A_{ij} denotes the element at the i-th row and the j-th column.
- The support function of a vector $\boldsymbol{x}$ is denoted as

$$\operatorname{supp}(\boldsymbol{x}) := \{i : x_i \neq 0\}.$$

  A vector $\boldsymbol{x}$ such that $|\operatorname{supp}(\boldsymbol{x})| \leq s$ is defined as s-sparse.
- For a vector $\boldsymbol{x} \in \mathbb{R}^d$ or $\boldsymbol{x} \in \mathbb{C}^d$, its ℓ_p-norm is given by

$$\|\boldsymbol{x}\|_p = \sum_{i=1}^{d} |x_i|^p.$$

  In certain cases, we define ℓ_0-norm as $\|\boldsymbol{x}\|_0 := |\operatorname{supp}(\boldsymbol{x})|$.
- For a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ or $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, the Frobenius norm of $\boldsymbol{A}$ is defined as

$$\|\boldsymbol{A}\|_F := \sqrt{\sum_{i,j} |\boldsymbol{A}_{ij}|^2} = \sqrt{\sum_i \sigma_i(\boldsymbol{A})^2},$$

where $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_{\min\{m,n\}}(A)$ denote its singular values. The nuclear norm of $\boldsymbol{A}$ is denoted as $\|\boldsymbol{A}\|_* := \sum_i \sigma_i(\boldsymbol{A})$. The spectral norm of a matrix $\boldsymbol{A}$ is denoted as

$$\|\boldsymbol{A}\| := \max_i \sigma_i(\boldsymbol{A}).$$

- The cardinality of a set $\mathscr{S}$ is denoted by $|\mathscr{S}|$.
- Random variables or events are denoted as uppercase letters, i.e., X, Y, E.
- The indicator function of an event E is denoted by $y = \mathbb{I}(E)$, where $y = 1$ if the event E is true, otherwise $y = 0$.
- Throughout this book, $f(n) = \mathcal{O}(g(n))$ or $f(n) \lesssim g(n)$ denotes that there exists a constant $c > 0$ such that $|f(n)| \leq c|g(n)|$, whereas $f(n) = \Omega(g(n))$ or $f(n) \gtrsim g(n)$ means that there exists a constant $c > 0$ such that $|f(n)| \geq c|g(n)|$. $f(n) \gg g(n)$ denotes that there exists some sufficiently large constant $c > 0$ such that $|f(n)| \geq c|g(n)|$. In addition, the notation $f(n) \asymp g(n)$ means that there exist constants $c_1, c_2 > 0$ such that $c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|$.
- For a general cone $C \subset \mathbb{R}^d$, the *polar cone* C° is the set of outward normals of C:

$$C^\circ := \left\{\boldsymbol{u} \in \mathbb{R}^d : \langle \boldsymbol{u},\ \boldsymbol{x}\rangle \leq 0 \quad \text{for all } \boldsymbol{x} \in C\right\}.$$

The polar cone C° is always closed and convex.