Compressed Sensing with Applications in Wireless Networks

Other titles in Foundations and Trends® in Signal Processing

A Brief Introduction to Machine Learning for Engineers
Osvaldo Simeone
ISBN: 978-1-68083-472-7

A Survey on the Low-Dimensional-Model-based Electromagnetic Imaging
Lianlin Li, Martin Hurtado, Feng Xu, Bing Chen Zhang, Tian Jin, Tie Jun Xui, Marija Nikolic Stevanovic and Arye Nehorai
ISBN: 978-1-68083-462-8

Synchronization and Localization in Wireless Networks
Bernhard Etzlinger and Henk Wymeersch
ISBN: 978-1-68083-434-5

Compressed Sensing with Applications in Wireless Networks

Markus Leinonen
Centre for Wireless Communications
University of Oulu
markus.leinonen@oulu.fi

Marian Codreanu
Department of Science and Technology
Linköping University
marian.codreanu@liu.se

Georgios B. Giannakis
Department of Electrical and Computer Engineering
University of Minnesota
georgios@umn.edu

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Signal Processing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

M. Leinonen, M. Codreanu and G. B. Giannakis. *Compressed Sensing with Applications in Wireless Networks*. Foundations and Trends® in Signal Processing, vol. 13, no. 1-2, pp. 1–282, 2019.

ISBN: 978-1-68083-647-9

Foundations and Trends® in Signal Processing
Volume 13, Issue 1-2, 2019

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multirate signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing
  - Classification and detection
  - Estimation and regression
  - Tree-structured methods

Information for Librarians

Foundations and Trends® in Signal Processing, 2019, Volume 13, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Contents

Compressed Sensing with Applications in Wireless Networks

Markus Leinonen[1], Marian Codreanu[2] and Georgios Giannakis[3]

[1] *Centre for Wireless Communications, University of Oulu; markus.leinonen@oulu.fi*

[2] *Department of Science and Technology, Linköping University; marian.codreanu@liu.se*

[3] *Department of Electrical and Computer Engineering, University of Minnesota; georgios@umn.edu*

ABSTRACT

Sparsity is an attribute present in a myriad of natural signals and systems, occurring either inherently or after a suitable projection. Such signals with lots of zeros possess minimal degrees of freedom and are thus attractive from an implementation perspective in wireless networks. While sparsity has appeared for decades in various mathematical fields, the emergence of compressed sensing (CS) – the joint sampling and compression paradigm – in 2006 gave rise to plethora of novel communication designs that can efficiently exploit sparsity. In this monograph, we review several CS frameworks where sparsity is exploited to improve the quality of signal reconstruction/detection while reducing the use of radio and energy resources by decreasing, e.g., the sampling rate, transmission rate, and number of computations. The first part focuses on several advanced CS signal reconstruction techniques along with wireless applications. The second part deals with efficient data gathering and lossy compression techniques in wireless sensor networks. Finally, the third part addresses CS-driven designs for spectrum sensing and multi-user detection for cognitive and wireless communications.

Markus Leinonen, Marian Codreanu and Georgios Giannakis (2019), Compressed Sensing with Applications in Wireless Networks, Foundations and Trends® in Signal Processing: Vol. 13, No. 1-2, pp 1–282. DOI: 10.1561/2000000107.

1

Introduction

Efficient data acquisition and signal reconstruction methods used for various purposes in wireless communication systems have been under extensive research and development for a while. One overarching design objective is to improve the quality of reconstruction, classification, detection etc. while reducing the use of radio and energy resources by decreasing, e.g., the sampling rate, transmission rate, and number of computations. Due to proliferation of the number of deployed wireless devices and the amount of data, implementing advanced signal processing techniques to achieve such objectives becomes of particular interest for emerging multi-user communication systems.

One fundamental feature of a desired information signal that can provide a solution to reach the aforementioned goal is signal *sparsity*. To this end, the key tool for exploiting sparsity is the modern theory of *compressed sensing/compressive sampling* (CS). Accordingly, this monograph reviews several CS techniques to utilize the sparsity of an underlying signal in data gathering, signal reconstruction and detection tasks in wireless networks. In Section 1, we first give a general overview of the different communication frameworks and applications addressed in the monograph. As a continuation, Section 1 elaborates signal sparsity and discusses how sparsity may be present or be-

come available in the considered frameworks. The main idea of CS is detailed in Section 1. Finally, the outline of the monograph is given in Section 1.

Communication Frameworks and Applications

Cognitive Radio Communications: It has been recognized that the licensed radio frequency (RF) spectrum is often severely under-utilized depending on the time and location of communication, in spite of the evident scarcity of the spectral resources due to the growing use of wireless devices [105]. *Cognitive radios (CRs)* aim to mitigate this issue by opportunistically utilizing the unused licensed spectrum through spectrum sensing and dynamic spectrum access in multi-user communication systems.
In particular, *RF cartography* is an instrumental concept for CR tasks [160]. Based on the measurements collected by spatially distributed CR sensors, RF cartography constructs the maps over the space, time, and frequency, portraying the RF landscape in which the CR network is deployed. Notable RF maps that have been proposed include the power spectral density maps, which acquire the ambient interference power distribution, revealing the crowded regions that CR transceivers need to avoid [21]; and the channel gain maps, which capture the channel gains between any two points in space, allowing CR networks to perform accurate spectrum sensing and aggressive spatial reuse [161]. By modelling the channel gains as the tomographic accumulations of an underlying spatial loss field (SLF), the technique captures the attenuation in the signal strength due to the obstacles in the propagation path.
Multi-User Detection: Multiuser detection (MUD) algorithms play a major role for mitigating multi-access interference present in code-division multiple access (CDMA) systems; see e.g., [288] and references therein. These well-appreciated MUD algorithms simultaneously detect the transmitted symbols of all active user terminals. Conventional techniques require knowledge of which terminals are active and exploit no possible user (in)activity. In the design of practical CDMA systems, one is also interested in saving bandwidth and power resources. Such savings are possible by reducing the size of the required spreading gain, which in turn reduces latency and energy consumption. This may be enabled by a low activity factor (probability of each user being active) in which case the system can be designed for spreading gains smaller than the number of candidate users.

Sensor Monitoring: In near future, there will be a burgeoning demand for the deployment of low-power smart sensors, especially to serve the myriad of diverse *internet of things* (IoT) applications including environmental, industrial, healthcare, and military monitoring tasks [316, 119, 213]. Accordingly, *wireless sensor networks* (WSNs) consisting of battery-powered sensors will be a key technology in creating the ubiquitous networked world and smart cities under the IoT framework. IoT opens a new era of intelligent networking, where collaborative sensors sense their environment with no human intervention, enabling to, e.g., automate an underlying process, improve the system performance, and reduce the maintenance costs. By 2020, the number of IoT devices is anticipated to reach hundreds of billions and the IoT market to become on the order of trillions of dollars, with a major portion on healthcare applications [119, 213].

In a typical monitoring task, geographically distributed sensors measure a correlated information source, encode the observations separately, and communicate the information to a sink for joint reconstruction of the source signals. As the sensors have limited batteries, which are often non-rechargeable or irreplaceable, it is crucial to minimize the energy consumption to prolong the lifespan of a WSN. The main contributors to sensors' energy consumption are wireless communications [239], and in certain applications, also the sensing/sampling part [6]. Consequently, it is crucial to minimize the amount of information (i.e., the number of data packets or bits) that must be communicated from each sensor to the sink in order to satisfy given application requirements. Accordingly, an energy-efficient sensor acquires a small number of data samples of a physical phenomenon (e.g., temperature, humidity, or light intensity), and encodes and communicates them at a minimum rate to the sink to reconstruct the information signal with, e.g., a given fidelity or maximum allowed delay. This engenders the need for energy-efficient distributed data gathering techniques that preserve autonomous operation of sensors and have a simple infrastructure with low battery consumption and computational complexity.

Signal Sparsity in the Considered Applications

Sparsity is an attribute present in a plethora of natural signals and systems, occurring either naturally or after projecting them over appropriate bases. A

signal is said to be *sparse* if it has many zero-valued elements or can be represented by few non-zero coefficients under a proper transformation. Similarly, a signal is termed *compressible*, if the energy of transform coefficients is concentrated in a small set of elements. Focusing on sparse signal structures is reasonable not only because nature itself is parsimonious but also because processing and simple models with minimal degrees of freedom are attractive from an implementation perspective. Whereas sparsity is by default an attribute defined for a single signal, it can be extended to characterize joint sparse structures; see, e.g., the regression problems dealing with *group sparsity* [314], and *joint sparsity models* introduced in [19].

Exploitation of sparsity is critical in a wide range of communication tasks. Regarding the frameworks addressed in this monograph, utilization of sparse signals has been investigated in the following applications.

- **WSNs:** Sparse signals are encountered in diverse WSN applications in, e.g., environmental monitoring [19, 238], source localization [201], and biomedical sensing [83]. For instance, universal transformations suitable for revealing the underlying sparsity of many smooth/piecewise smooth signals include the discrete Fourier, cosine, and wavelet transform (DFT, DCT, DWT), respectively [202, 15, 36, 37]. In particular, the efficacy of the DWT matrices in sparsifying signals of several natural phenomena such as temperature, humidity, and light has been especially reported in, e.g., [197, 37]. Sparsity in the sensed data allows to reduce the computations of simple sensor devices, and most importantly, to significantly reduce sensors' energy consumption for communicating the data to a fusion center [175, 176, 178]. One particular direction of great interest is so-called *quantized CS* [129, 273, 333, 151, 180, 184, 182, 181] – a lossy compression setup where the CS measurements are converted into finite-rate bit sequences, and the aim is to design efficient quantization-aware CS reconstruction algorithms and analyze their rate-distortion performance. Application requirements may demand as low quantization rate as one bit per measurement sample, referred as *1-bit CS* [152].

- **Cellular Networks:** Sparsity has been utilized in, e.g., estimation of wireless multipath channels [71, 275], estimation of parameters of com-

munication systems [53, 23, 9, 109], and sparse sphere decoding [277]. As for MUD algorithms, the sparsity arises because the active terminals are unknown and the activity factor is low in a typical system [322]. Sparsity can be exploited to either relax or judiciously search over subsets of the alphabet of the desired symbol vector so that the resultant MUD algorithms trade off optimality in detection error performance with computational complexity. Moreover, source localization based on the DoA estimation typically involves sparsity in the angular domain under certain radio propagation features [118, 201, 228].

- **CR Applications:** Distributed spectrum sensing for CR communications is a crucial task and has been addressed in, e.g., [20, 10, 212, 261]. In spectrum sensing, the sparsity manifests itself in two forms: 1) narrow-band nature of transmit power spectral density relative to the broad range of usable spectrum, and 2) sparsely located active radios in the operational space [20]. This type of *compressive* wideband power spectrum estimation allows to recover an unknown power spectrum of a wide-sense stationary signal from samples obtained at a sub-Nyquist rate [10], even if the samples were coarsely quantized sensors' measurements [212].

  Another emergent topic in CR networks is RF cartography. In RF cartography, SLF may have a low-rank structure potentially corrupted by sparse outliers [171]. Such a model is particularly appealing for urban and indoor propagation scenarios, where regular placement of buildings and walls renders a scene inherently of low rank, while sparse outliers can pick up the artifacts that do not conform to the low-rank model. Earlier works on sparsity-leveraging cartography include network anomaly monitoring in [205].

Applications from other research fields which deal with sparse signals include variable selection in linear regression models for diabetes [278], image compression [46], signal decomposition using overcomplete bases [61], and more.

The Concept of Compressive Sampling

Sampling Followed by Compression: The key principle underlying the data sampling methods and analog-to-digital conversion in modern consumer devices is the Nyquist-Shannon sampling theorem – the celebrated result of the seminal works by Nyquist [227] and Shannon [259]. The theorem states that if the sampling rate of a signal is at least twice its maximum frequency component, the signal can be perfectly reconstructed. This threshold rate is called the Nyquist rate. In a resource-limited digital sensor, the acquisition of signal samples is typically followed by data compression which aims to encode the information with fewer bits. Consequently, a substantial portion of expensively acquired data is eventually discarded prior to storage or transmission. Fortunately, if a signal has certain additional features, perfect reconstruction may be possible even below the Nyquist rate. Namely, the inefficiency caused by the separate sampling and compression may be alleviated by *sub-Nyquist sampling* – an unorthodox paradigm violating the conventional sampling notion.

Compressive Sampling: A feature that enables sub-Nyquist sampling is the sparsity/compressibility of a signal, discussed in Section 1. While sparsity has been exploited for a while in numerical linear algebra, statistics, and signal processing, renewed interest emerged in recent years because sparsity plays an instrumental role in modern *compressive sampling/compressed sensing* (CS) theory and applications [43, 87, 44, 146, 46, 38, 94, 17, 49]. CS is a joint sampling and compression paradigm which enables a sparse/compressible length-N signal to be accurately reconstructed from its $M < N$ (random) linear measurements. This engenders the sub-Nyquist sampling interpretation of CS [217]: instead of sampling at a rate proportional to the signal bandwidth, the sampling rate in CS is dictated by the signal's "information content" [218]. The primary asset of CS is its simple and universal encoding since most computational work load is shifted to the decoder [94]. As a rough comparison, computational complexity at the encoder for CS scales as MN (at most for a dense measurement matrix), whereas for a standard compression method like fast Fourier transform it scales as $N\log N$ [112, Appendix C.1]. While for high-dimensional signals $MN > N\log N$, the use of sparse measurements matrices drastically reduces the computational and memory requirements for CS [28]. The other benefits of CS include robustness to measurement/quantization noise, resiliency to packet losses, security via pseudorandom projections, and

the gradual improvement of reconstruction accuracy from increased number of measurements [94]. All above benefits are especially beneficial for low-power sensor applications.

A Historical Note: Prior to the CS era launched in 2006, there has already been lots of research interest to tackle signal processing tasks involving sparse signals. This becomes also evident in Table 1 of Section 2 where many popular present-day CS reconstruction algorithms trace back to early 1990's. Indeed, in a diversity of engineering applications, one encounters solving ill-posed/underdetermined inverse problems, i.e., problems where the number of available measurements is smaller than the dimension of the signal/model to be estimated. Luckily, in many such practical situations, the encountered models have structural constraints similar to sparsity, i.e., they can be described by only a few degrees of freedom relative to their ambient dimension or as linear combinations of a few basic building blocks.

One cornerstone of the emerging area of CS is the notion of *variable selection* (VS) associated with sparse linear regression [278]. VS is a combinatorially complex task closely related (but not identical) to the well-known *model order selection* problem tackled through Akaike's information [2], Bayesian information [258], and risk inflation [111] criteria. A typically convex function of the model fitting error is penalized with the ℓ_0-norm of the unknown vector which equals the number of nonzero entries, and thus accounts for model complexity (degrees of freedom). To bypass the non-convexity of the ℓ_0-norm, VS and CS approaches replace it with convex penalty terms (e.g., the ℓ_1-norm) that capture sparsity but also lead to computationally efficient solvers. One another line of work preceding the CS era is the concept "sampling signals with *finite rate of innovation*" introduced in [289], which generalizes the classic sampling theorem of bandlimited signals with sinc kernels. The rate of innovation is a number that describes a finite number of degrees of freedom (i.e., sparsity) per unit of time for certain classes of signals.

Outline of the Monograph

This monograph addresses several CS techniques to utilize sparsity of an underlying signal in data gathering, signal reconstruction and detection tasks in wireless networks. In this monograph, the "sub-Nyquist feature" of CS refers to measuring an underlying (continuous-time) analog source via dimensionality

reducing projections, which can be represented by discrete-time measurements of form (1) (see also Fig. 1). Another line of work is analog-to-digital compression, where analog signals are encoded into bits via a combined sub-Nyquist sampling and quantization process [162]. A hardware implementation of a sub-Nyquist sampling system was presented in [215], and a unified Xampling framework was introduced in [216].

This monograph is organized as follows. We first introduce fundamentals of CS and give an overview about common reconstruction techniques of sparse signals from compressed measurements in Section 2. After that, the monograph is split into three parts as follows:

- The first part focuses on several advanced CS signal reconstruction techniques along with wireless applications. Accordingly, Section 3 develops sparsity-aware recursive algorithms for estimating and tracking sparse and (possibly time-varying) signals. Section 4 devises regularized total least-squares algorithms under sparsity constraints for a perturbed CS signal model, along with applying them to directions-of-arrival estimation.

- The second part deals with efficient data gathering and lossy compression techniques in wireless sensor networks. Compressed acquisition of streaming correlated data in WSNs is presented in Section 5. Section 6 and Section 7 consider CS signal acquisition setups under quantization of measurements. Accordingly, Section 6 devises an efficient quantized CS algorithm for distributed source coding of correlated sparse sources in WSNs. Rate-distortion performance of a quantized CS setup is investigated in Section 7, including both information theoretic analyses and the design of several types of practical quantized CS algorithms.

- The third part addresses CS-driven designs for spectrum sensing and multi-user detection for cognitive and wireless communications. Section 8 addresses channel gain cartography for CR networks under a limited number of available measurements. Section 9 proposes efficient sparsity-utilizing MUD algorithms in CDMA systems.

Finally, the monograph is concluded with the summary in Section 10.

Part III

Sparsity-Enabled Cognitive and Cellular Communications

8

Channel Gain Cartography for Cognitive Radios Leveraging Low Rank and Sparsity

This section focuses on channel gain cartography which aims at inferring the channel gains between two arbitrary points in space based on the measurements (samples) of the gains collected by a set of radios deployed in the area. Channel gain maps are useful for various sensing and resource allocation tasks essential for the operation of cognitive radio (CR) networks. In the considered framework, the channel gains are modeled as the tomographic accumulations of an underlying spatial loss field (SLF), which captures the attenuation in the signal strength due to the obstacles in the propagation path. In order to estimate the map accurately with a relatively small number of measurements, the SLF is postulated to have a low-rank structure possibly with sparse deviations. Efficient batch and online algorithms are developed for the resulting map reconstruction problem. Comprehensive tests with both synthetic and real datasets corroborate that the algorithms can accurately reveal the structure of the propagation medium and produce the desired channel gain maps.

Related Works

RF cartography is an instrumental concept for CR tasks [160], motivated by the under-utilization of the licensed RF spectrum [105]. Based on the

measurements collected by spatially distributed CR sensors, RF cartography constructs the maps over the space, time, and frequency, portraying the RF landscape in which the CR network is deployed. Notable RF maps that have been proposed include the power spectral density (PSD) maps, which acquire the ambient interference power distribution, revealing the crowded regions that CR transceivers need to avoid [21]; and the channel gain (CG) maps, which capture the channel gains between any two points in space, allowing CR networks to perform accurate spectrum sensing and aggressive spatial reuse [161].

Prior works on channel gain cartography capitalized on experimentally validated notion of a spatial loss field (SLF) [1], which expresses the shadow fading over an arbitrary link as the weighted integral of the underlying attenuation that the RF propagation experiences due to the blocking objects in the path. Linear interpolation techniques such as kriging were employed to estimate the shadow fading based on spatially correlated measurements, and the spatio-temporal dynamics were tracked using Kalman filtering approaches [161, 79]. It is worth noting that SLF reconstruction is tantamount to the radio tomographic imaging (RTI), useful in a wide range of applications, from locating survivors in rescue operations to environmental monitoring [297, 298, 141]. The method in [297] captures the variation of the propagation medium by taking SLF differences at consecutive time slots into consideration. To cope with multipath fading in a cluttered environment, multiple channel measurements were utilized to enhance localization accuracy in [155]. However, the methods in [297, 155] do not reveal static objects in the imaging area. In contrast, a method to track moving objects using a dynamic SLF model, as well as identifying the static ones, was reported in [141]. Exploiting the sparse occupancy of the monitored area by the target objects, sparsity-leveraging algorithms for constructing obstacle maps were developed [220, 158, 219]. This work adopts a related data model, but mainly focuses on the channel gain map construction for CR applications.

Although more sophisticated methodologies for channel modeling do exist [282, 305], the computational cost and requirements on various structural/geometric prior information may hinder their use in CR applications. On the other hand, the SLF model has been experimentally validated [1], as well as in our work through a real tomographic imaging example. The proposed

approach in this section provides a computationally efficient solution, by capitalizing on the inherent structure of measurement data, rather than relying heavily on the physics of RF propagation.

System Model and Problem Statement

Notation. Bold uppercase (lowercase) letters denote matrices (column vectors). Calligraphic letters are used for sets; $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\mathbf{0}_n$ denotes an $n \times 1$ vector of all zeros, and $\mathbf{0}_{n\times n}$ an $n \times n$ matrix of all zeros. Operators $(\cdot)^{\mathcal{T}}$, $\mathrm{tr}(\cdot)$, and $\sigma_i(\cdot)$ represent the transposition, trace, and the i-th largest singular value of a matrix, respectively; $|\cdot|$ is used for the cardinality of a set, and the magnitude of a scalar. $\mathbf{R} \succeq \mathbf{0}$ signifies that $\mathbf{R}$ is positive semidefinite. The ℓ_1-norm of $\mathbf{X} \in \mathbb{R}^{n\times n}$ is $||\mathbf{X}||_1 := \sum_{i,j=1}^{n} |X_{ij}|$. The ℓ_∞-norm of $\mathbf{X} \in \mathbb{R}^{n\times n}$ is represented by $||\mathbf{X}||_\infty := \max\{|X_{ij}| : i, j = 1, \ldots, n\}$. For two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$, the matrix inner product $\langle \mathbf{X}, \mathbf{Y} \rangle := \mathrm{tr}(\mathbf{X}^{\mathcal{T}}\mathbf{Y})$. The Frobenius norm of matrix $\mathbf{Y}$ is $||\mathbf{Y}||_F := \sqrt{\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{\mathcal{T}})}$. The spectral norm of $\mathbf{Y}$ is $||\mathbf{Y}|| := \max_{||\mathbf{x}||_2=1} ||\mathbf{Y}\mathbf{x}||_2$, and $||\mathbf{Y}||_* := \sum_i \sigma_i(\mathbf{Y})$ is the nuclear norm of $\mathbf{Y}$. For a function $h : \mathbb{R}^{m\times n} \to \mathbb{R}$, the directional derivative of h at $\mathbf{X} \in \mathbb{R}^{m\times n}$ along a direction $\mathbf{D} \in \mathbb{R}^{m\times n}$ is denoted as $h'(\mathbf{X}; \mathbf{D}) := \lim_{t\to 0+}[h(\mathbf{X} + t\mathbf{D}) - h(\mathbf{X})]/t$.

Consider a set of N CRs deployed over a geographical area represented by a two-dimensional plane $\mathcal{A} \subset \mathbb{R}^2$. Let $\mathbf{x}_n^{(t)} \in \mathcal{A}$ denote the position of CR $n \in \{1, 2, \ldots, N\}$ at time t. By exchanging pilot sequences, the CR nodes can estimate the channel gains among them. A typical channel gain between nodes n and n' can be modeled as the product of pathloss, shadowing, and small-scale fading. By averaging out the effect of the small-scale fading, the (averaged) channel gain measurement over a link (n, n') at time t, denoted by $G(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$, can be represented (in dB) as

$$G(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) = G_0 - \gamma 10 \log_{10} ||\mathbf{x}_n^{(t)} - \mathbf{x}_{n'}^{(t)}|| + s(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) \tag{152}$$

where G_0 is the path gain at unit distance; $||\mathbf{x}_n^{(t)} - \mathbf{x}_{n'}^{(t)}||$ is the distance between nodes n and n'; γ is the pathloss exponent; and $s(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$ is the attenuation due to the shadow fading. By subtracting the known pathloss component in (152), the noisy shadowing measurement

$$\check{s}(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) = s(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) + \epsilon(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) \tag{153}$$

is obtained, where $\epsilon(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$ denotes the measurement noise. Let $\mathcal{M}^{(t)}$ be the set of links, for which channel gain measurements are made at time t, and collect those measurements in vector $\check{\mathbf{s}}^{(t)} \in \mathbb{R}^{|\mathcal{M}^{(t)}|}$. The goal of channel gain cartography is to predict the channel gain between arbitrary points $\mathbf{x}, \mathbf{x}' \in \mathcal{A}$ at time t, based on the known nodal positions $\{\mathbf{x}_n^{(t)}\}$ and the channel gain measurements collected up to time t, that is, $\{\check{\mathbf{s}}^{(\tau)}\}_{\tau=1}^{t}$ [161, 79].

In order to achieve this interpolation, the structure of shadow fading experienced by co-located radio links will be leveraged. To this end, a variety of correlation models for shadow fading have been proposed [1, 139, 137]. In particular, the models in [161, 1, 298, 141] rely on the so-termed *spatial loss field* (SLF), which captures the attenuation due to obstacles in the line-of-sight propagation.

Let $f : \mathcal{A} \to \mathbb{R}$ denote the SLF, which captures the attenuation at location $\tilde{\mathbf{x}} \in \mathcal{A}$, and $w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}})$ is the weight function modeling the influence of the SLF at $\tilde{\mathbf{x}}$ to the shadowing experienced by link $\mathbf{x}$–$\mathbf{x}'$. Then, $s(\mathbf{x}, \mathbf{x}')$ is expressed as [142]

$$s(\mathbf{x}, \mathbf{x}') = \int_{\mathcal{A}} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}. \tag{154}$$

The *normalized ellipse model* is often used for the weight function, with w taking the form [297]

$$w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) := \begin{cases} 1/\sqrt{d(\mathbf{x}, \mathbf{x}')}, & \text{if } d(\mathbf{x}, \tilde{\mathbf{x}}) + d(\mathbf{x}', \tilde{\mathbf{x}}) \\ & < d(\mathbf{x}, \mathbf{x}') + \delta \\ 0, & \text{otherwise} \end{cases} \tag{155}$$

where $d(\mathbf{x}, \mathbf{x}') := \|\mathbf{x} - \mathbf{x}'\|$ is the distance between positions $\mathbf{x}$ and $\mathbf{x}'$, and $\delta > 0$ is a tunable parameter. The value of δ is commonly set to half the wavelength to assign non-zero weights only within the first Fresnel zone. The integral in (154) can be approximated by

$$s(\mathbf{x}, \mathbf{x}') \simeq \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}_{i,j}) f(\tilde{\mathbf{x}}_{i,j}) \tag{156}$$

where $\{\tilde{\mathbf{x}}_{i,j}\}_{i=1,j=1}^{N_x,N_y}$ are the pre-specified grid points over $\mathcal{A}$. Let matrix $\mathbf{F} \in \mathbb{R}^{N_x \times N_y}$ denote the SLF, sampled by the N_x-by-N_y grid. Similarly, the weight matrix $\mathbf{W}_{\mathbf{x}\mathbf{x}'}$ corresponding to link $\mathbf{x}$–$\mathbf{x}'$ is constructed. The shadow fading

over link $\mathbf{x}$–$\mathbf{x}'$ in (156) can then be expressed as a linear projection of the SLF given by

$$s(\mathbf{x}, \mathbf{x}') \simeq \langle \mathbf{W}_{\mathbf{x}\mathbf{x}'}, \mathbf{F} \rangle = \mathrm{tr}(\mathbf{W}_{\mathbf{x}\mathbf{x}'}^{\mathcal{T}} \mathbf{F}). \tag{157}$$

The goal is to form an estimate $\hat{\mathbf{F}}^{(t)}$ of $\mathbf{F}^{(t)}$ at time t, based on $\{\mathbf{x}_n^{(t)}\}$ and $\{\check{\mathbf{s}}^{(\tau)}\}_{\tau=1}^{t}$. Once $\hat{\mathbf{F}}^{(t)}$ is obtained, the shadowing and the overall channel gain across any link $\mathbf{x}$–$\mathbf{x}'$ at time t can be estimated via (157) and (152) as

$$\hat{s}(\mathbf{x}^{(t)}, \mathbf{x}'^{(t)}) = \langle \mathbf{W}_{\mathbf{x}\mathbf{x}'}^{(t)}, \hat{\mathbf{F}}^{(t)} \rangle \tag{158}$$

$$\hat{G}(\mathbf{x}^{(t)}, \mathbf{x}'^{(t)}) = G_0 - \gamma 10 \log_{10} ||\mathbf{x}^{(t)} - \mathbf{x}'^{(t)}|| + \hat{s}(\mathbf{x}^{(t)}, \mathbf{x}'^{(t)}). \tag{159}$$

The number of unknown $\mathbf{F}^{(t)}$ entries is less than $N_x N_y$, while the number of measurements is $\mathcal{O}(tN^2)$, provided that the SLF remains invariant for t slots. If the number of entries to be estimated in $\mathbf{F}^{(t)}$ is larger than the number of measurements, the problem is underdetermined and cannot be solved uniquely. To overcome this and further improve the performance even in the overdetermined cases, a priori knowledge on the structure of $\mathbf{F}^{(t)}$ will be exploited next to regularize the problem.

Channel Gain Prediction Using Low Rank and Sparsity

Problem Formulation

The low-rank plus sparse structure has been advocated in various problems in machine learning and signal processing [40, 102, 207]. Low-rank matrices are effective in capturing slow variation or regular patterns, and sparsity is instrumental for incorporating robustness against outliers. Inspired by these, we postulate that $\mathbf{F}$ has a low-rank-plus-sparse structure as

$$\mathbf{F} = \mathbf{L} + \mathbf{E} \tag{160}$$

where matrix $\mathbf{L}$ is low-rank, and $\mathbf{E}$ is sparse. This model is particularly attractive in urban or indoor scenarios where the obstacles often possess regular patterns, while the sparse term can capture irregularities that do not conform to the low-rank model.

Redefine $\mathbf{W}^{(t)}_{nn'} := \mathbf{W}^{(t)}_{\mathbf{x}_n\mathbf{x}_{n'}}$ and $\check{s}^{(t)}_{nn'} := \check{s}(\mathbf{x}^{(t)}_n, \mathbf{x}^{(t)}_{n'})$ for brevity.[21] Toward estimating $\mathbf{F}^{(t)}$ that obeys (160), consider the cost

$$c^{(t)}(\mathbf{L},\mathbf{E}) := \frac{1}{2}\sum_{(n,n')\in\mathcal{M}^{(t)}} \left(\langle \mathbf{W}^{(t)}_{nn'}, \mathbf{L}+\mathbf{E}\rangle - \check{s}^{(t)}_{nn'}\right)^2 \tag{161}$$

which fits the shadowing measurements to the model. Then, with T denoting the total number of time slots taking measurements, we adopt the following optimization criterion

$$\text{(P1)} \quad \min_{\mathbf{L},\mathbf{E}\in\mathbb{R}^{N_x\times N_y}} \sum_{\tau=1}^{T} \beta^{T-\tau}\left[c^{(\tau)}(\mathbf{L},\mathbf{E}) + \lambda||\mathbf{L}||_* + \mu||\mathbf{E}||_1\right] \tag{162}$$

where $\beta \in (0,1]$ is the forgetting factor that can be optionally put in to weigh the recent observations more heavily. The nuclear norm regularization term promotes a low-rank $\mathbf{L}$, while the ℓ_1-norm encourages sparsity in $\mathbf{E}$. Parameters λ and μ are appropriately chosen to control the effect of these regularizers. Conditions for exact recovery through a related convex formulation in the absence of measurement noise can be found in [307].

Problem (162) is convex, and can be tackled using existing efficient solvers, such as the interior-point method. Once the optimal $\hat{\mathbf{L}}$ and $\hat{\mathbf{E}}$ are found, the desired $\hat{\mathbf{F}}$ is obtained as $\hat{\mathbf{F}} = \hat{\mathbf{L}} + \hat{\mathbf{E}}$. However, the general-purpose optimization packages tend to scale poorly as the problem size grows. Specialized algorithms developed for related problems often employ costly SVD operations iteratively [307]. Furthermore, such an algorithm might not be amenable for an online implementation. Building on [206] and [245], an efficient solution is proposed next with reduced complexity.

[21] Prompted by [155], the benefit of multi-channel diversity for RTI may be incorporated in the presented framework. Suppose K channels $\mathcal{K}^{(t)}_{nn'}$ are available to sensors n and n' at time t, and let $\acute{s}^{(t)}_{nn',k}$ denote the noisy measurement including fading over link $\mathbf{x}_n$–$\mathbf{x}_{n'}$ at t in channel $k \in \mathcal{K}^{(t)}_{nn'}$. Construct a new measurement as $\bar{s}^{(t)}_{nn'} = \phi(\acute{s}^{(t)}_{nn',1}, \acute{s}^{(t)}_{nn',2}, \ldots, \acute{s}^{(t)}_{nn',K})$, where $\phi(\cdot)$ is a channel selection function [155]. By replacing $\check{s}^{(t)}_{nn'}$ in (161) with $\bar{s}^{(t)}_{nn'}$, the multiple channel measurements can be incorporated without altering the method. However, the dynamic channel availability and multi-channel measurements will increase algorithm complexity. On the other hand, it is not clear whether such a multi-channel approach can be adopted for estimating *any channel gain* over multiple frequency bands.

Efficient Batch Solution

Without loss of generality, consider replacing $\mathbf{L}$ with the low-rank product $\mathbf{P}\mathbf{Q}^{\mathcal{T}}$, where $\mathbf{P} \in \mathbb{R}^{N_x \times \rho}$ and $\mathbf{Q} \in \mathbb{R}^{N_y \times \rho}$, and ρ is a pre-specified overestimate of the rank of $\mathbf{L}$. It is known that (e.g., [245])

$$||\mathbf{L}||_* = \min_{\mathbf{P},\mathbf{Q}} \frac{1}{2}\left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2\right) \quad \text{subject to } \mathbf{L} = \mathbf{P}\mathbf{Q}^{\mathcal{T}}. \tag{163}$$

Thus, a natural re-formulation of (162) is (see also [206])

$$\text{(P2)} \quad \min_{\mathbf{P},\mathbf{Q},\mathbf{E}} f(\mathbf{P},\mathbf{Q},\mathbf{E}) := \tag{164}$$
$$\sum_{\tau=1}^{T} \beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}},\mathbf{E}) + \frac{\lambda}{2}\left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2\right) + \mu||\mathbf{E}||_1\right].$$

Instead of seeking the $N_x N_y$ entries of $\mathbf{L}$, the factorization approach (164) entails only $(N_x + N_y)\rho$ unknowns, thus reducing complexity and memory requirements significantly when $\rho \ll \min\{N_x, N_y\}$. Furthermore, adoption of the separable Frobenius norm regularizer in (P2) comes with no loss of optimality as asserted in the following lemma.

Lemma 3: *If* $\{\hat{\mathbf{L}}, \hat{\mathbf{E}}\}$ *minimize (P1) and we choose* $\rho \geq$ *rank*$(\hat{\mathbf{L}})$*, then, (P2) is equivalent to (P1) at the minimum.*

Proof: It is clear that the minimum of (P1) is no larger than that of

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{E}} \sum_{\tau=1}^{T} \beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}},\mathbf{E}) + \lambda||\mathbf{P}\mathbf{Q}^{\mathcal{T}}||_* + \mu||\mathbf{E}||_1\right] \tag{165}$$

since the search space is reduced by the re-parametrization $\mathbf{L} = \mathbf{P}\mathbf{Q}^{\mathcal{T}}$ with $\rho \leq \min\{N_x, N_y\}$. Now (163) implies that the minimum of (165) is no larger than that of (P2). However, the inequality is tight since the objectives of (P1) and (P2) are identical for $\mathbf{E} := \hat{\mathbf{E}}$, $\mathbf{P} := \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^{1/2}$, and $\mathbf{Q} := \hat{\mathbf{V}}\hat{\mathbf{\Sigma}}^{1/2}$, where $\hat{\mathbf{L}} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^{\mathcal{T}}$ is the SVD. Consequently, (P1) and (P2) have identical costs at the minimum. ■

Although (P1) is a convex optimization problem, (P2) is not. Thus, in general, one can obtain only a *locally* optimal solution of (P2), which may not be the globally optimal solution of (P1). Interestingly, under appropriate conditions,

global optimality can be guaranteed for the local optima of (P2), as claimed in the following proposition.

Proposition 7: *If* $\{\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}\}$ *is a stationary point of (P2),* $\bar{\beta} := \sum_{\tau=1}^{T} \beta^{T-\tau}$, *and* $||\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}})|| \leq \lambda\bar{\beta}$ *with*

$$\tilde{f}(\hat{\mathbf{L}}, \hat{\mathbf{E}}) := \sum_{\tau=1}^{T} \beta^{T-\tau} \left[\sum_{(n,n')\in\mathcal{M}^{(\tau)}} \left(\langle \mathbf{W}_{nn'}^{(\tau)}, \hat{\mathbf{L}} + \hat{\mathbf{E}} \rangle - \check{s}_{nn'}^{(\tau)} \right) \mathbf{W}_{nn'}^{(\tau)} \right] \tag{166}$$

then $\{\hat{\mathbf{L}} := \bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \hat{\mathbf{E}} := \bar{\mathbf{E}}\}$ *is a globally optimal solution to (P1).*

Proof: See Appendix G.

A stationary point of (P2) can be obtained through a block coordinate-descent (BCD) algorithm, where the optimization is performed in a cyclic fashion over one of $\{\mathbf{E}, \mathbf{P}, \mathbf{Q}\}$ with the remaining two variables fixed. In fact, since the term $\mu||\mathbf{E}||_1$ is separable in the individual entries as well, the cyclic update can be stretched all the way up to the individual entries of $\mathbf{E}$ without affecting convergence [280]. The proposed solver entails an iterative procedure comprising three steps per iteration $k = 1, 2, \ldots$

[S1] Update E:

$$\mathbf{E}[k+1] = \arg\min_{\mathbf{E}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{P}[k]\mathbf{Q}^{\mathcal{T}}[k], \mathbf{E}) + \mu||\mathbf{E}||_1 \right]$$

[S2] Update P:

$$\mathbf{P}[k+1] = \arg\min_{\mathbf{P}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}}[k], \mathbf{E}[k+1]) + \frac{\lambda}{2}||\mathbf{P}||_F^2 \right]$$

[S3] Update Q:

$$\mathbf{Q}[k+1] = \arg\min_{\mathbf{Q}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{P}[k+1]\mathbf{Q}^{\mathcal{T}}, \mathbf{E}[k+1]) + \frac{\lambda}{2}||\mathbf{Q}||_F^2 \right].$$

To update each block variable, the cost in (P2) is minimized while fixing the other block variables to their up-to-date iterates.

To detail the update rules, let $\boldsymbol{\mathcal{W}}^{(t)} \in \mathbb{R}^{N_x N_y \times |\mathcal{M}(t)|}$ be a matrix with columns equal to $\text{vec}\left(\mathbf{W}_{nn'}^{(t)}\right)$ for $(n, n') \in \mathcal{M}(t)$, where $\text{vec}(\cdot)$ produces a column vector by stacking the columns of a matrix one below the other ($\text{unvec}(\cdot)$ denotes the reverse process). Define $\boldsymbol{\mathcal{W}} := [\sqrt{\beta^{T-1}}\boldsymbol{\mathcal{W}}^{(1)} \ldots \sqrt{\beta^{0}}\boldsymbol{\mathcal{W}}^{(T)}]$, $\check{\mathbf{s}} := [\sqrt{\beta^{T-1}}\check{\mathbf{s}}^{(1)\mathcal{T}} \ldots \sqrt{\beta^{0}}\check{\mathbf{s}}^{(T)\mathcal{T}}]^{\mathcal{T}}$, and $\mathbf{e} := \text{vec}(\mathbf{E})$. Then, one can write $\sum_{\tau=1}^{T} \beta^{T-\tau} c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}}, \mathbf{E}) = \|\boldsymbol{\mathcal{W}}^{\mathcal{T}} \text{vec}(\mathbf{P}\mathbf{Q}^{\mathcal{T}} + \mathbf{E}) - \check{\mathbf{s}}\|_2^2$. Let e_l denote the l-th entry of $\mathbf{e}$, and $\mathbf{e}_{-l}$ represent the replica of $\mathbf{e}$ without its

l-th entry. Similarly, let $\boldsymbol{\omega}_l^{\mathcal{T}}$ denote the l-th row of the matrix $\boldsymbol{\mathcal{W}}$, and $\boldsymbol{\mathcal{W}}_{-l}$ denote the matrix $\boldsymbol{\mathcal{W}}$ with its l-th row removed. The soft-thresholding function soft_th$(\cdot;\mu)$ is defined as

$$\text{soft\_th}(x;\mu) := \text{sgn}(x)\max\{0, |x| - \mu\}. \tag{167}$$

Minimization in **[S1]** proceeds sequentially over the individual entries of $\mathbf{e}$. At iteration k, each entry is updated via

$$e_l[k+1] = \arg\min_{e_l} \frac{1}{2}||e_l\boldsymbol{\omega}_l - \check{\check{\mathbf{s}}}||_2^2 + \mu\bar{\beta}|e_l|,\ l = 1, \ldots, N_xN_y \tag{168}$$

where $\check{\check{\mathbf{s}}}_l[k] := \check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^{\mathcal{T}}\text{vec}(\mathbf{P}[k]\mathbf{Q}^{\mathcal{T}}[k]) - \boldsymbol{\mathcal{W}}_{-l}^{\mathcal{T}}\mathbf{e}_{-l}$. The closed-form solution for e_l is obtained as

$$e_l[k+1] = \frac{\text{soft\_th}(\boldsymbol{\omega}_l^{\mathcal{T}}\check{\check{\mathbf{s}}}_l[k]; \mu\bar{\beta})}{||\boldsymbol{\omega}_l||_2^2}. \tag{169}$$

Matrices $\mathbf{P}$ and $\mathbf{Q}$ are similarly updated over their rows through **[S2]** and **[S3]**. Let $\mathbf{p}_i$ be the i-th row of $\mathbf{P}$, transposed to a column vector; i.e., $\mathbf{P} := [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N_x}]^{\mathcal{T}}$. Define $\tilde{\boldsymbol{\mathcal{W}}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)| \times N_y}$ to be the matrix whose rows are the i-th rows of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n,n') \in \mathcal{M}(t)}$ denoted as $\tilde{\mathbf{w}}_{nn',i}^{(t)\mathcal{T}}$, and $\tilde{\mathbf{s}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|}$ a vector with entries equal to

$$\tilde{s}_{nn',i}^{(t)} := \check{s}_{nn'}^{(t)} - \langle \mathbf{W}_{nn'}^{(t)}, \mathbf{E}[k+1] \rangle - \sum_{j \neq i}^{N_x} \tilde{\mathbf{w}}_{nn',j}^{(t)\mathcal{T}} \mathbf{Q}[k]\mathbf{p}_j \tag{170}$$

for $(n, n') \in \mathcal{M}(t)$. Define also $\tilde{\boldsymbol{\mathcal{W}}}_i := [\sqrt{\beta^{T-1}}\tilde{\boldsymbol{\mathcal{W}}}_i^{(1)\mathcal{T}} \ldots \sqrt{\beta^0}\tilde{\boldsymbol{\mathcal{W}}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$ and $\tilde{\mathbf{s}}_i := [\sqrt{\beta^{T-1}}\tilde{\mathbf{s}}_i^{(1)\mathcal{T}} \ldots \sqrt{\beta^0}\tilde{\mathbf{s}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$. Then, $\mathbf{p}_i$ is updated by solving a ridge-regression problem as

$$\mathbf{p}_i[k+1] = \arg\min_{\mathbf{p}_i} \left[\frac{1}{2}||\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k]\mathbf{p}_i - \tilde{\mathbf{s}}_i||_2^2 + \frac{\lambda\bar{\beta}}{2}||\mathbf{p}_i||_2^2 \right]$$

whose solution is given in closed form by

$$\mathbf{p}_i[k+1] = \left[\mathbf{Q}^{\mathcal{T}}[k]\tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1}\mathbf{Q}^{\mathcal{T}}[k]\tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\tilde{\mathbf{s}}_i \tag{171}$$

which involves matrix inversion of dimension only ρ-by-ρ. Likewise, let $\mathbf{q}_i$ denote the i-th row of $\mathbf{Q}$, transposed to a column vector; i.e., $\mathbf{Q} :=$

$[\mathbf{q}_1, \ldots, \mathbf{q}_{N_y}]^{\mathcal{T}}$. Define also $\breve{\boldsymbol{\mathcal{W}}}_i := [\sqrt{\beta^{T-1}}\breve{\boldsymbol{\mathcal{W}}}_i^{(1)\mathcal{T}} \ldots \sqrt{\beta^0}\breve{\boldsymbol{\mathcal{W}}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$ and $\breve{\mathbf{s}}_i := [\sqrt{\beta^{T-1}}\breve{\mathbf{s}}_i^{(1)\mathcal{T}} \ldots \sqrt{\beta^0}\breve{\mathbf{s}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$, where $\breve{\boldsymbol{\mathcal{W}}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)| \times N_x}$ is the matrix whose rows are the transpositions of the i-th columns of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n,n')\in\mathcal{M}(t)}$, denoted as $\breve{\mathbf{w}}_{nn',i}^{(t)}$, and $\breve{\mathbf{s}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|}$ has entries

$$\breve{s}_{nn',i}^{(t)} := \breve{s}_{nn'}^{(t)} - \langle \mathbf{W}_{nn'}^{(t)}, \mathbf{E}[k+1] \rangle - \sum_{j \neq i}^{N_y} \breve{\mathbf{w}}_{nn',j}^{(t)\mathcal{T}} \mathbf{P}[k+1]\mathbf{q}_j \tag{172}$$

for $(n, n') \in \mathcal{M}(t)$. The update for $\mathbf{q}_i$ is then given by solving another ridge regression problem to obtain

$$\mathbf{q}_i[k+1] = \arg\min_{\mathbf{q}_i} \left[\frac{1}{2} ||\breve{\boldsymbol{\mathcal{W}}}_i \mathbf{P}[k+1]\mathbf{q}_i - \breve{\mathbf{s}}_i||_2^2 + \frac{\lambda\bar{\beta}}{2} ||\mathbf{q}_i||_2^2 \right]$$

whose solution is given also in closed form by

$$\mathbf{q}_i[k+1] = \left[\mathbf{P}^{\mathcal{T}}[k+1]\breve{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \breve{\boldsymbol{\mathcal{W}}}_i \mathbf{P}[k+1] + \lambda\bar{\beta}\mathbf{I}_\rho \right]^{-1} \mathbf{P}^{\mathcal{T}}[k+1]\breve{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \breve{\mathbf{s}}_i \tag{173}$$

which again involves matrix inversion of dimension ρ-by-ρ. The overall algorithm is tabulated in Table 5.

Although the proposed batch algorithm exhibits low computational and memory requirements, it is not suitable for online processing, since (164) must be re-solved every time a new set of measurements arrive, incurring major computational burden. Thus, the development of an online recursive algorithm is well motivated.

Online Algorithm

Stochastic Approximation Approach

In practice, it is often the case that a new set of data becomes available *sequentially* in time. Then, it is desirable to have an algorithm that can process the newly acquired data incrementally and refine the previous estimates, rather than re-computing the batch solution, which may incur prohibitively growing computational burden. Furthermore, when the channel is time-varying due to, e.g., mobile obstacles, online algorithms can readily track such variations.

Stochastic approximation (SA) is an appealing strategy for deriving online algorithms [267, 169]. Moreover, techniques involving minimizing majorized

Table 5: Batch solver of (P2)

Initialize $\mathbf{E}[1] := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}[1]$ and $\mathbf{Q}[1]$ at random.
For $k = 1, 2, \ldots$
  [S1] *Update* $\mathbf{E}$*:*
  Set $\mathbf{e} = \text{vec}(\mathbf{E}[k])$
  For $l = 1, 2, \ldots, N_x N_y$
    Set $\check{\mathbf{s}}_l[k] := \check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^{\mathcal{T}} \text{vec}(\mathbf{P}[k]\mathbf{Q}^{\mathcal{T}}[k]) - \boldsymbol{\mathcal{W}}_{-l}^{\mathcal{T}} \mathbf{e}_{-l}$
    $e_l[k+1] = \text{soft\_th}(\boldsymbol{\omega}_l^{\mathcal{T}} \check{\mathbf{s}}_l[k]; \mu\bar{\beta}) / ||\boldsymbol{\omega}_l||_2^2$
  Next l
  Set $\mathbf{E}[k+1] = \text{unvec}(\mathbf{e}[k+1])$
  [S2] *Update* $\mathbf{P}$*:*
  For $i = 1, 2, \ldots, N_x$
    Set $\tilde{\boldsymbol{\mathcal{W}}}_i$ and $\tilde{\mathbf{s}}_i$
    $\mathbf{p}_i[k+1] = \left[\mathbf{Q}^{\mathcal{T}}[k] \tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \tilde{\boldsymbol{\mathcal{W}}}_i \mathbf{Q}[k] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1} (\mathbf{Q}^{\mathcal{T}}[k] \tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \tilde{\mathbf{s}}_i)$
  Next i
  $\mathbf{P}[k+1] = [\mathbf{p}_1[k+1], \mathbf{p}_2[k+1], \ldots, \mathbf{p}_{N_x}[k+1]]^{\mathcal{T}}$
  [S3] *Update* $\mathbf{Q}$*:*
  For $i = 1, 2, \ldots, N_y$
    Set $\check{\boldsymbol{\mathcal{W}}}_i$ and $\check{\mathbf{s}}_i$
    $\mathbf{q}_i[k+1] = \left[\mathbf{P}^{\mathcal{T}}[k+1] \check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \check{\boldsymbol{\mathcal{W}}}_i \mathbf{P}[k+1] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1} \times (\mathbf{P}^{\mathcal{T}}[k+1] \check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \check{\mathbf{s}}_i)$
  Next i
  $\mathbf{Q}[k+1] = [\mathbf{q}_1[k+1], \mathbf{q}_2[k+1], \ldots, \mathbf{q}_{N_y}[k+1]]^{\mathcal{T}}$
Next k

surrogate functions were developed to handle nonconvex cost functions in online settings [206, 208, 199, 242]. An online algorithm to solve a dictionary learning problem was proposed in [199]. A stochastic gradient descent algorithm was derived for subspace tracking and anomaly detection in [206]. Next, an online algorithm for the CPCP problem is developed. The proposed approach employs quadratic surrogate functions with diagonal weighting to capture disparate curvatures in the directions of different block variables.

For simplicity, let the number of measurements per time slot t be constant $M := |\mathcal{M}^{(t)}|$ for all t. Define $\mathbf{X} := (\mathbf{P}, \mathbf{Q}, \mathbf{E}) \in \mathcal{X} \subset \mathcal{X}' := \mathbb{R}^{(N_x \times \rho)} \times \mathbb{R}^{(N_y \times \rho)} \times \mathbb{R}^{(N_x \times N_y)}$, where $\mathcal{X}$ is a compact convex set, and $\mathcal{X}'$ a bounded open set, and $\boldsymbol{\xi}^{(t)} := [\{\check{s}_m^{(t)}\}_{m=1}^M, \{\mathbf{W}_m^{(t)}\}_{m=1}^M] \in \Xi$, where Ξ is assumed to be bounded. Define with slight abuse of notation

$$g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) = g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)}) := \frac{1}{2} \sum_{m=1}^{M} \left(\langle \mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^{\mathcal{T}} + \mathbf{E} \rangle - \check{s}_m^{(t)} \right)^2 \tag{174}$$

$$g_2(\mathbf{X}) = g_2(\mathbf{P}, \mathbf{Q}, \mathbf{E}) := \frac{\lambda}{2} \left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2 \right) + \mu ||\mathbf{E}||_1. \tag{175}$$

A quadratic surrogate function for $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ is then constructed as

$$\begin{aligned} \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) &:= g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \\ &+ \langle \mathbf{P} - \mathbf{P}^{(t-1)}, \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{P}}^{(t)}}{2} \|\mathbf{P} - \mathbf{P}^{(t-1)}\|_F^2 \\ &+ \langle \mathbf{Q} - \mathbf{Q}^{(t-1)}, \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{Q}}^{(t)}}{2} \|\mathbf{Q} - \mathbf{Q}^{(t-1)}\|_F^2 \\ &+ \langle \mathbf{E} - \mathbf{E}^{(t-1)}, \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{E}}^{(t)}}{2} \|\mathbf{E} - \mathbf{E}^{(t-1)}\|_F^2 \end{aligned} \tag{176}$$

where $\eta_{\mathbf{P}}^{(t)}$, $\eta_{\mathbf{Q}}^{(t)}$, and $\eta_{\mathbf{E}}^{(t)}$ are positive constants, and with $\tilde{\tilde{f}}_m^{(t)}(\mathbf{P}, \mathbf{Q}, \mathbf{E}) := \langle \mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^{\mathcal{T}} + \mathbf{E} \rangle - \check{s}_m^{(t)}$ it can be readily verified that

$$\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \tag{177}$$

$$\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)\mathcal{T}} \mathbf{P}^{(t-1)} \tag{178}$$

$$\nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)}. \tag{179}$$

Let us focus on the case without the forgetting factor, i.e., $\beta = 1$. A convergent SA algorithm for (P2) is obtained by considering the following surrogate

problem

$$\text{(P3)} \quad \min_{\mathbf{X}} \frac{1}{t}\sum_{\tau=1}^{t}\left[\breve{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X})\right]. \tag{180}$$

In fact, solving (P3) yields a stochastic gradient descent (SGD) algorithm. In particular, since variables $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$ can be separately optimized in (P3), the proposed algorithm updates the variables in parallel in each time slot t as

$$\mathbf{P}^{(t)} = \arg\min_{\mathbf{P}} \sum_{\tau=1}^{t}\Bigg[\langle \mathbf{P} - \mathbf{P}^{(\tau-1)}, \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)})\rangle + \frac{\eta_{\mathbf{P}}^{(\tau)}}{2}||\mathbf{P} - \mathbf{P}^{(\tau-1)}||_F^2 + \frac{\lambda}{2}||\mathbf{P}||_F^2\Bigg] \tag{181}$$

$$\mathbf{Q}^{(t)} = \arg\min_{\mathbf{Q}} \sum_{\tau=1}^{t}\Bigg[\langle \mathbf{Q} - \mathbf{Q}^{(\tau-1)}, \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)})\rangle + \frac{\eta_{\mathbf{Q}}^{(\tau)}}{2}||\mathbf{Q} - \mathbf{Q}^{(\tau-1)}||_F^2 + \frac{\lambda}{2}||\mathbf{Q}||_F^2\Bigg] \tag{182}$$

$$\mathbf{E}^{(t)} = \arg\min_{\mathbf{E}} \sum_{\tau=1}^{t}\Bigg[\langle \mathbf{E} - \mathbf{E}^{(\tau-1)}, \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)})\rangle + \frac{\eta_{\mathbf{E}}^{(\tau)}}{2}||\mathbf{E} - \mathbf{E}^{(\tau-1)}||_F^2 + \mu||\mathbf{E}||_1\Bigg]. \tag{183}$$

By checking the first-order optimality conditions, and defining $\bar{\eta}_{\mathbf{P}}^{(t)} := \sum_{\tau=1}^{t}\eta_{\mathbf{P}}^{(\tau)}$ and $\bar{\eta}_{\mathbf{Q}}^{(t)} := \sum_{\tau=1}^{t}\eta_{\mathbf{Q}}^{(\tau)}$, the update rules for $\mathbf{P}$ and $\mathbf{Q}$ are obtained as

$$\mathbf{P}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \lambda t}\sum_{\tau=1}^{t}\left[\eta_{\mathbf{P}}^{(\tau)}\mathbf{P}^{(\tau-1)} - \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)})\right] \tag{184}$$

$$\mathbf{Q}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \lambda t}\sum_{\tau=1}^{t}\left[\eta_{\mathbf{Q}}^{(\tau)}\mathbf{Q}^{(\tau-1)} - \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)})\right] \tag{185}$$

which can be written in recursive forms as

$$\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \lambda t}\left(\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda\mathbf{P}^{(t-1)}\right) \tag{186}$$

$$\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \lambda t}\left(\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda\mathbf{Q}^{(t-1)}\right). \tag{187}$$

Table 6: Online SGD solver of (P2)

1: **Initialize** $\mathbf{E}^{(0)} := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}^{(0)}$ and $\mathbf{Q}^{(0)}$ at random.
2: **For** $t = 1, 2, \ldots$
3: **Set** $L_{\mathbf{P}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \right\|_F^2$,
$L_{\mathbf{Q}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)\mathcal{T}} \mathbf{P}^{(t-1)} \right\|_F^2$,
$L_{\mathbf{E}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)} \right\|_F^2$, and $L_{\min} = \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$.
4: **Set** $\eta_{\mathbf{P}}^{(t)} \geq \frac{L_{\mathbf{P}}^2}{L_{\min}}$, $\eta_{\mathbf{Q}}^{(t)} \geq \frac{L_{\mathbf{Q}}^2}{L_{\min}}$, and $\eta_{\mathbf{E}}^{(t)} \geq \frac{L_{\mathbf{E}}^2}{L_{\min}}$.
5: **Set** $\bar{\eta}_{\mathbf{P}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{P}}^{(\tau)}$, $\bar{\eta}_{\mathbf{Q}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{Q}}^{(\tau)}$, and $\bar{\eta}_{\mathbf{E}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{E}}^{(\tau)}$.
6: $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \lambda t} \left(\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda \mathbf{P}^{(t-1)} \right)$
7: $\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \lambda t} \left(\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda \mathbf{Q}^{(t-1)} \right)$
8: $\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} + \bar{\eta}_{\mathbf{E}}^{(t-1)} \mathbf{Z}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \right]$
9: $\mathbf{E}^{(t)} = \text{soft\_th}(\mathbf{Z}^{(t)}; \mu t / \bar{\eta}_{\mathbf{E}}^{(t)})$
10: **Next** t

Due to the non-smoothness of $||\mathbf{E}||_1$, the update for $\mathbf{E}$ proceeds in two steps. First, an auxiliary variable $\mathbf{Z}^{(t)}$ is introduced, which is computed as

$$\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\sum_{k=1}^{t} \eta_{\mathbf{E}}^{(k)} \mathbf{E}^{(k-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(k-1)}, \boldsymbol{\xi}^{(k)}) \right]. \tag{188}$$

Again defining $\bar{\eta}_{\mathbf{E}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{E}}^{(\tau)}$, matrix $\mathbf{Z}^{(t)}$ can be obtained recursively as

$$\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} + \bar{\eta}_{\mathbf{E}}^{(t-1)} \mathbf{Z}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \right]. \tag{189}$$

Then, $\mathbf{E}^{(t)}$ is updated as

$$\mathbf{E}^{(t)} = \text{soft\_th}(\mathbf{Z}^{(t)}; \mu t / \bar{\eta}_{\mathbf{E}}^{(t)}). \tag{190}$$

The overall online algorithm is listed in Table 6.

Remark 1 (Computational complexity). In the batch algorithm of Table 5, the complexity orders for computing the updates for each of $\mathbf{p}_i$ and $\mathbf{q}_i$ are $\mathcal{O}(N_y M T)$ and $\mathcal{O}(N_x M T)$, respectively, due to the computation of $\tilde{\boldsymbol{\mathcal{W}}}^{\mathcal{T}} \tilde{\mathbf{s}}_i$

and $\check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}} \check{\mathbf{s}}_i$. Thus, the complexity orders for updating $\mathbf{P}$ and $\mathbf{Q}$ per iteration k are both $\mathcal{O}(N_x N_y M T)$. The update of e_l incurs complexity $\mathcal{O}(MT)$ for computing $\boldsymbol{\omega}_l^{\mathcal{T}} \check{\mathbf{s}}_l$. Thus, the complexity order for updating $\mathbf{E}$ per iteration k is $\mathcal{O}(N_x N_y M T)$. Accordingly, the overall per-iteration complexity of the batch algorithm becomes $\mathcal{O}(N_x N_y M T)$. On the other hand, the complexity of the online algorithm in Table 6 is dominated by the gradient computations, which require $\mathcal{O}(\rho N_x N_y M)$. Since ρ is smaller than N_x and N_y, and the per-iteration complexity does not grow with T, the online algorithm has a much more affordable complexity than its batch counterpart, and it is scalable for large network scenarios.

Convergence

The iterates $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ generated from the algorithm in Table 6 converge to a stationary point of (P2), as asserted in the following proposition. First define

$$C_t(\mathbf{X}) := \frac{1}{t} \sum_{\tau=1}^{t} \left[g_1(\mathbf{X}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right] \tag{191}$$

$$\check{C}_t(\mathbf{X}) := \frac{1}{t} \sum_{\tau=1}^{t} \left[\check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right] \tag{192}$$

$$C(\mathbf{X}) := \mathbb{E}_{\boldsymbol{\xi}} \left[g_1(\mathbf{X}, \boldsymbol{\xi}) + g_2(\mathbf{X}) \right]. \tag{193}$$

Note that $C_t(\mathbf{X})$ is essentially identical to the cost of (P2). Furthermore, the minimizer of $C_t(\mathbf{X})$ approaches that of $C(\mathbf{X})$ when $t \to \infty$, provided $\boldsymbol{\xi}$ obeys the law of large numbers, which is clearly the case when e.g., $\{\boldsymbol{\xi}^{(t)}\}$ is i.i.d. Assume that $\nabla_{\mathbf{P}} g_1(\cdot, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi})$, $\nabla_{\mathbf{Q}}(\mathbf{P}, \cdot, \mathbf{E}, \boldsymbol{\xi})$ and $\nabla_{\mathbf{E}}(\mathbf{P}, \mathbf{Q}, \cdot, \boldsymbol{\xi})$ are Lipschitz with respect to $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$, respectively, with constants $L_{\mathbf{P}}$, $L_{\mathbf{Q}}$, and $L_{\mathbf{E}}$, respectively (which will be shown in Appendix H). Furthermore, let $\bar{\alpha}_i^{(t)} := (\sum_{\tau=1}^{t} (\eta_i^{(\tau)} + \lambda))^{-1}$ for $i \in \{\mathbf{P}, \mathbf{Q}\}$, and $\bar{\alpha}_{\mathbf{E}}^{(t)} := (\bar{\eta}_{\mathbf{E}}^{(t)})^{-1}$ denote step sizes.

Proposition 8: *If* (a1) $\{\boldsymbol{\xi}^{(t)}\}_{t=1}^{\infty}$ *is an independent and identically distributed (i.i.d) random sequence;* (a2) $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ *are in a compact set* $\mathcal{X}$*;* (a3) Ξ *is bounded;* (a4) *For* $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$, $\bar{\eta}_i^{(t)} \geq ct$ $\forall t$ *for some* $c \geq 0$*; and* (a5) $c' \geq \eta_i^{(t)} \geq L_i^2 / L_{\min}$ $\forall t$ *for some* $c' > 0$ *and* $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$, *then the iterates* $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ *generated by the algorithm in Table 6 converge to*

the set of stationary points of (P2) *with* $\beta = 1$, *i.e.,*

$$\lim_{t\to\infty} \inf_{\bar{\mathbf{X}}\in\bar{\mathcal{X}}} \|\mathbf{X}^{(t)} - \bar{\mathbf{X}}\|_F = 0 \qquad \text{a.s.} \tag{194}$$

where $\bar{\mathcal{X}}$ *is the set of stationary points of* $C(\mathbf{X})$.
Proof: See Appendix H.

Numerical Tests

Performance of the presented batch and online algorithms is assessed through numerical tests using both synthetic and real datasets. A few existing methods are also tested for comparison. The ridge-regularized least-squares (LS) scheme estimates the SLF as $\text{vec}(\hat{\mathbf{F}}) = (\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^{\mathcal{T}} + \omega\mathbf{C}_{\mathbf{f}}^{-1})^{-1}\boldsymbol{\mathcal{W}}\check{\mathbf{s}}$, where $\mathbf{C}_{\mathbf{f}}$ is the spatial covariance matrix of the SLF, and ω is a regularization parameter [297, 155, 142]. The total variation (TV)-regularized LS scheme in [236] is also tested, which solves $\min_{\mathbf{f}} \|\check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^{\mathcal{T}}\mathbf{f}\|_2^2 + \omega(\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y}|f_{i+1,j} - f_{i,j}| + \sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1}|f_{i,j+1} - f_{i,j}|)$ where $\mathbf{f} := \text{vec}(\mathbf{F})$ and $f_{i,j}$ corresponds to the (i,j)-th element of $\mathbf{F}$. Finally, the LASSO estimator is obtained by solving (P1) with $\lambda = 0$.

Test with Synthetic Data

Random tomographic measurements were taken by sensors deployed uniformly over $\mathcal{A} := [0.5, 40.5] \times [0.5, 40.5]$, from which the SLF with $N_x = N_y = 40$ was reconstructed. Per-time slot, 10 measurements were taken, corrupted by zero-mean white Gaussian noise with variance $\sigma^2 = 0.1$. The regularization parameters were set to $\lambda = 0.05$ and $\mu = 0.01$ through cross-validation by minimizing the normalized error $\|\hat{\mathbf{F}} - \mathbf{F}_0\|_F/\|\mathbf{F}_0\|_F$, where $\mathbf{F}_0$ is the ground-truth SLF depicted in Fig. 36. Other parameters were set to $\rho = 13$, $\beta = 1$, and $\delta = 0.06$; while $\mathbf{C}_{\mathbf{f}} = \mathbf{I}_{N_xN_y}$ and $\omega = 0.13$ were used for the ridge-regularized LS.
To validate the batch algorithm in Table 5, two cases were tested. In the first case, the measurements were generated for $T = 130$ time slots using $N = 52$ sensors, while in the second case, $T = 260$ and $N = 73$ were used. As a comparison, the accelerated proximal gradient (APG) algorithm was also derived for (P1) [187]. Note that the APG requires the costly SVD operation of an N_x-by-N_y matrix per iteration, while only the inversion of a ρ-by-ρ

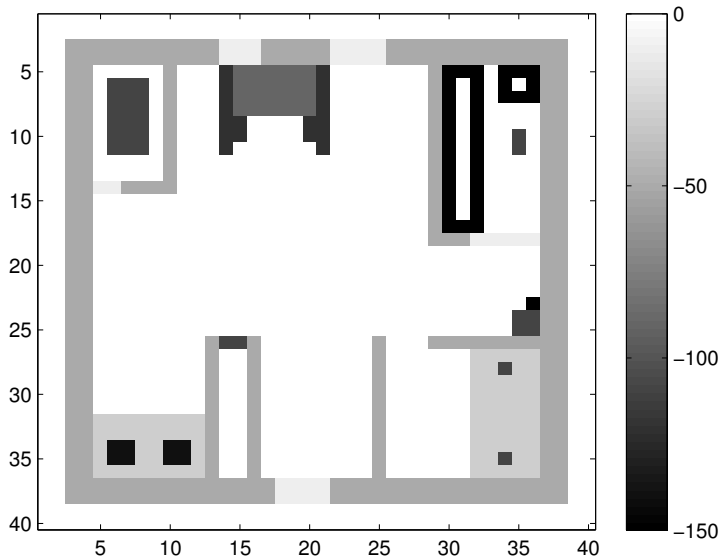
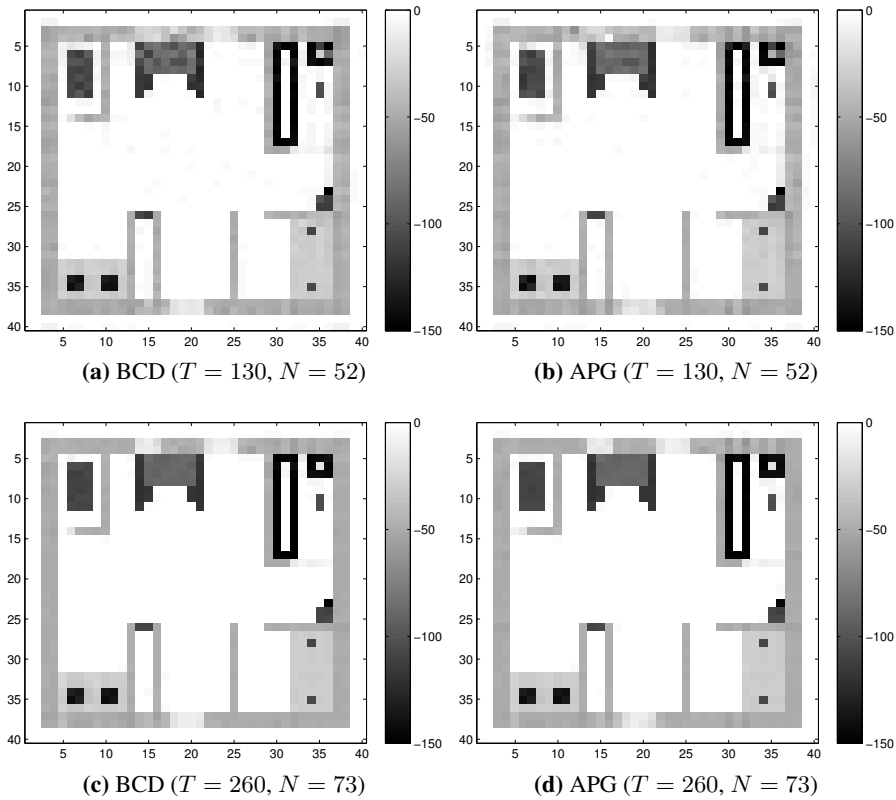


Figure 36: True SLF.

matrix is necessary in the proposed BCD algorithm. Fig. 37 shows the SLFs reconstructed by APG and BCD algorithms for the two cases. Apparently, the reconstructed SLFs capture well the features of the ground-truth SLF in Fig. 36. Note that (P2) is underdetermined when $T = 130$ since the total number of unknowns in (P2) is $2,640$ while the total number of measurements is only $1,300$. This verifies that the channel gain maps can be accurately interpolated with a small number of measurements by leveraging the attributes of the low rank and sparsity. Fig. 38a shows the convergence of the BCD and APG algorithms. The cost of (P2) from the BCD algorithm converges to that of (P1) from APG after $k = 550$ iterations, showing that the performance of solving (P1) directly is achievable by the proposed algorithm solving (P2) instead. This can also be corroborated from the reconstructed SLFs in Fig. 37 as well.

Table 7 lists the reconstruction error when $T = 130$ and the per-iteration complexity of the batch algorithms. It is seen that the proposed method outperforms benchmark algorithms in terms of the reconstruction error. Note that the ridge-regularized LS has a one-shot (non-iterative) complexity of



(a) BCD ($T = 130$, $N = 52$) **(b)** APG ($T = 130$, $N = 52$)

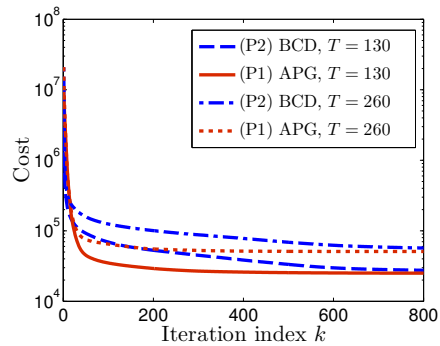
(c) BCD ($T = 260$, $N = 73$) **(d)** APG ($T = 260$, $N = 73$)

Figure 37: SLFs reconstructed by the batch algorithms.

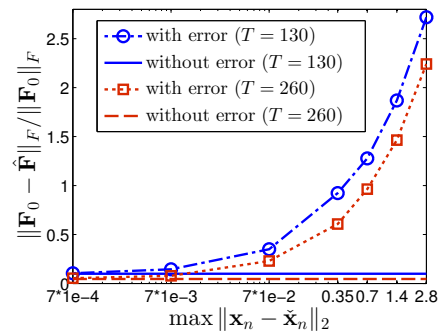
$\mathcal{O}((N_x N_y)^3)$, but its reconstruction capability is worse than the proposed algorithm as the true SLF is not smooth.

To test robustness of the proposed algorithm against imprecise CR location estimates, the reconstruction error versus the maximum sensor location error is depicted in Fig. 38b. To reconstruct $\mathbf{F}$ matrix, $\boldsymbol{\mathcal{W}}$ was computed via a set of erroneous sensor locations $\breve{\mathbf{x}}_n^{(t)}$ obtained by adding uniformly random perturbations to true locations $\mathbf{x}_n^{(t)}$. It is seen that the SLF could be accurately reconstructed when the location error was small.

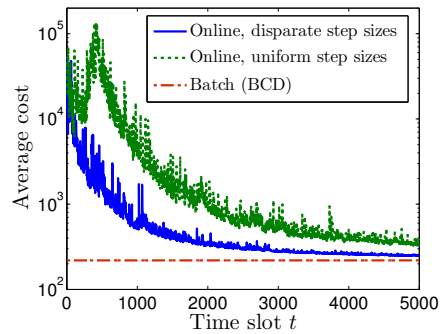
The numerical tests for the online algorithm were carried out with the same parameter setting as the batch experiments with $N = 317$. Fig. 38c depicts the evolution of the average cost in (191) for two sets of values for $(\bar{\eta}_{\mathbf{P}}^{(t)}, \bar{\eta}_{\mathbf{Q}}^{(t)}, \bar{\eta}_{\mathbf{E}}^{(t)})$.



(a)



(b)



(c)

Figure 38: SLF reconstruction using the batch and online algorithms. (a) Cost versus iterations (batch). (b) Reconstruction error versus CR location error (batch). (c) Average cost over time slots (online).

Table 7: Reconstruction error at $T = 130$ and computational complexity per iteration.

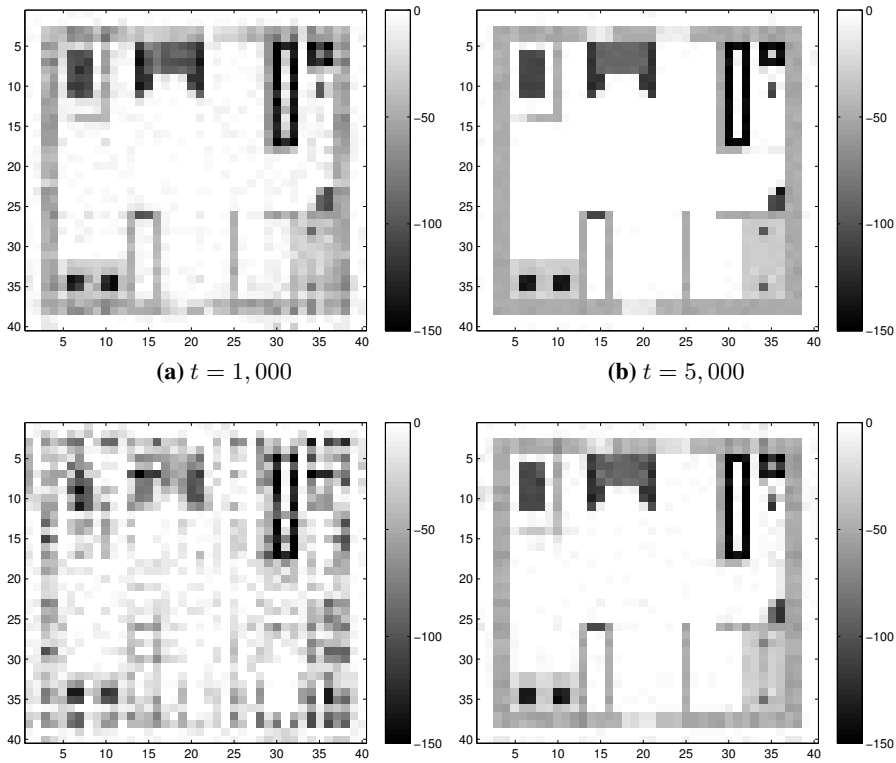
Algorithm	Proposed (BCD)	Ridge-regularized LS	Total variation (ADMM)	LASSO
$\|\mathbf{F}_0 - \hat{\mathbf{F}}\|_F / \|\mathbf{F}_0\|_F$	0.1064	0.1796	0.1196	0.1828
Complexity per iteration	$\mathcal{O}(N_x N_y M T)$	N/A	$\mathcal{O}((N_x N_y)^3 + (N_x N_y)^2 M T)$	$\mathcal{O}(N_x N_y M T)$

The green dotted curve corresponds to using $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = \bar{\eta}_{\mathbf{E}}^{(t)} = 300$, while the blue solid curve is for $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 300$, and $\bar{\eta}_{\mathbf{E}}^{(t)} = 10$. It can be seen that the uniform step sizes for all variables result in convergence rate that is slower than that with the disparate step sizes. Fig. 39 shows the SLFs reconstructed via the online algorithm at $t = 1,000$ and $t = 5,000$ using the two choices of step sizes. It can be seen that for a given time slot t, flexibly choosing the step sizes yields much more accurate reconstruction. As far as reconstruction error, the online algorithm with disparate step sizes yields 6.3×10^{-2} at $t = 5,000$, while its batch counterpart has 2.4×10^{-2}. Although slightly less accurate SLF is obtained by the online algorithm, it comes with greater computational efficiency.

To assess the tracking ability of the online algorithm, the slow channel variation was simulated. The measurements were generated using the SLF in Fig. 36 with three additional objects slowly moving in the rate of unit pixel width per 70 time slots. Fig. 40 depicts instances of the true and reconstructed SLFs at $t = 2,400$ and $t = 3,200$, respectively, obtained by the online algorithm. The moving objects are marked by the red circles. It is seen that the reconstructed SLFs correctly capture the moving objects, while the stationary objects are estimated more clearly as t increases.

Test with Real Data

To validate the performance of the proposed framework for SLF and channel gain map estimation in realistic scenarios, real received signal strength (RSS) measurements were also processed. The data were collected by a set of $N = 20$ sensors deployed in the perimeter of a square-shaped testbed as shown in Fig. 41, where the crosses indicate the sensor positions. Data collection was performed in two steps [142]. First, free-space measurements were taken to



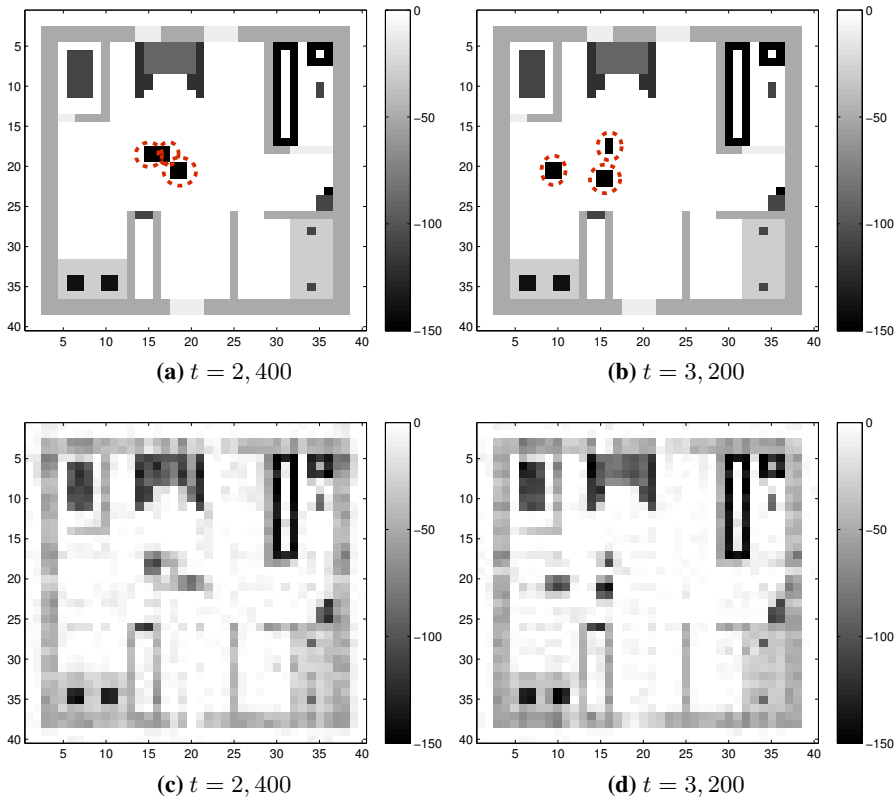
(a) $t = 1,000$ **(b)** $t = 5,000$

(c) $t = 1,000$ **(d)** $t = 5,000$

Figure 39: SLFs reconstructed by the online algorithm. (a) and (b) correspond to using $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 300$ and $\bar{\eta}_{\mathbf{E}}^{(t)} = 10$. (c) and (d) use $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = \bar{\eta}_{\mathbf{E}}^{(t)} = 300$.

obtain estimates of the path gain G_0 and the pathloss exponent γ via least-squares. The estimated γ was approximately 2, and G_0 was found to be 75. Then, tomographic measurements were formed with the artificial structure shown in Fig. 41. For the both measurements, 100 measurements were taken per time slot, in the 2.425 GHz frequency band, across 24 time slots. The shadowing measurements were obtained by subtracting the estimated pathloss from the RSS measurements.

The SLFs of size $N_x = N_y = 61$ were reconstructed by the proposed batch algorithm. The regularization parameters were set to $\lambda = 4.5$ and $\mu = 3.44$, which were determined by cross-validation. The parameter δ in (155) was set



(a) $t = 2,400$ **(b)** $t = 3,200$

(c) $t = 2,400$ **(d)** $t = 3,200$

Figure 40: (a) and (b) are true SLFs; (c) and (d) show reconstructed SLFs at different time slots.

to 0.2 feet to capture the non-zero weights within the first Fresnel zone, and $\rho = 10$ and $\beta = 1$ were used.

For comparison, the ridge-regularized LS estimator was also tested. To construct $\mathbf{C_f}$, the exponential decay model in [1] was used, which models the covariance between points $\mathbf{x}$ and $\mathbf{x}'$ as $\mathbf{C_f}(\mathbf{x}, \mathbf{x}') = \sigma_s^2 e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\kappa}}$, where σ_s^2 and $\kappa > 0$ are model parameters. In our tests, $\sigma_s^2 = \kappa = 1$, and $\omega = 79.9$ were used.

The SLF, shadow fading map, and channel gain map reconstructed by the proposed BCD algorithm are depicted in Fig. 42. The shadow fading and channel gain maps portray the gains in dB between any point in the map and the

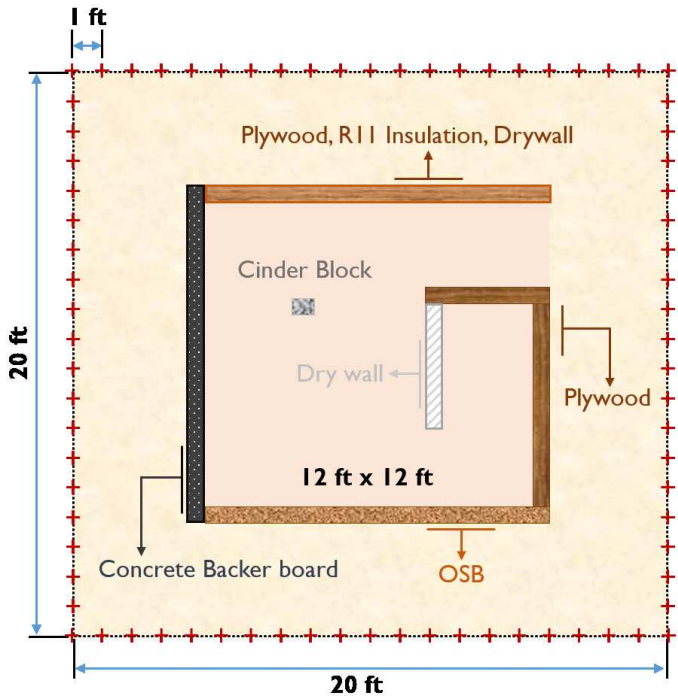
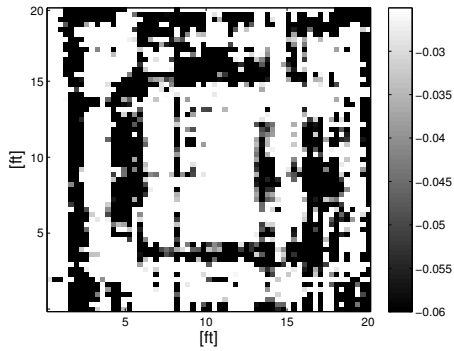


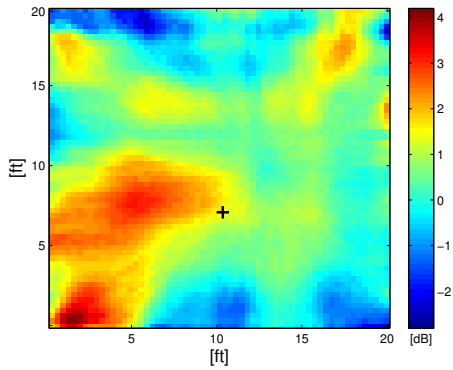
Figure 41: Configuration of the testbed.

fixed CR location at $(10.2, 7.2)$ (marked by the cross). Fig. 43 shows the results from the ridge-regularized LS estimation. It can be seen from Fig. 42(a) and Fig. 43(b) that the proposed low-rank plus sparse model produces a somewhat sharper SLF image than the ridge-regularized LS approach. Although the latter yields a smooth SLF image, it produces more artifacts near the isolated block and the boundary of the SLF. Such artifacts may lead to less accurate shadowing and channel gain maps. For instance, Fig. 42(b) and Fig. 43(b) both show that the shadow fading is stronger as more building material is crossed in the communication path. However, somewhat strong attenuations are observed near the cinder block location and the interior of the oriented strand board (OSB) wall only in Fig. 43(b), which seems anomalous.

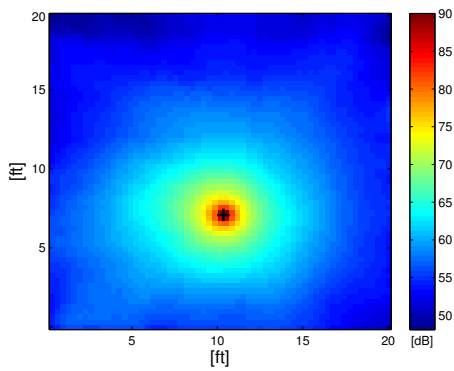
The online algorithm was also tested with the real data. Parameters $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 620$ and $\bar{\eta}_{\mathbf{E}}^{(t)} = 200$ were selected, and 6×10^5 measurements were



(a) SLF

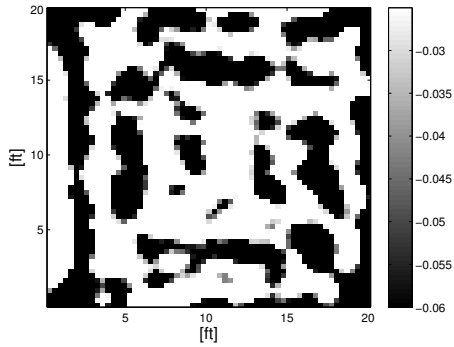


(b) Shadow fading map

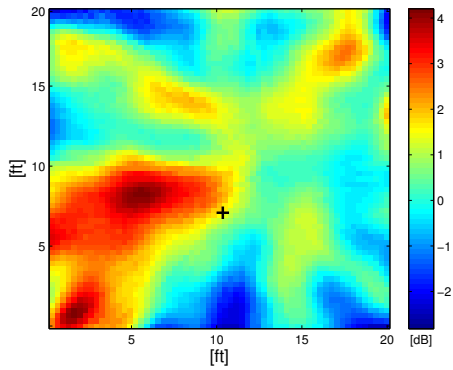


(c) Channel gain map

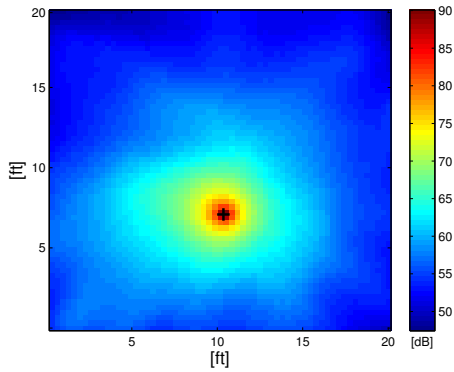
Figure 42: Reconstructions by the proposed batch algorithm.



(a) SLF

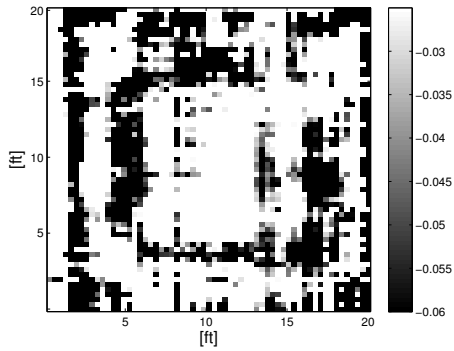


(b) Shadow fading map

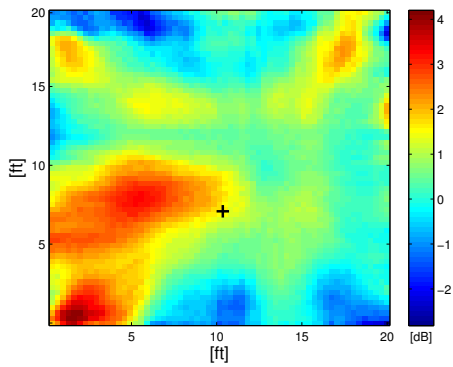


(c) Channel gain map

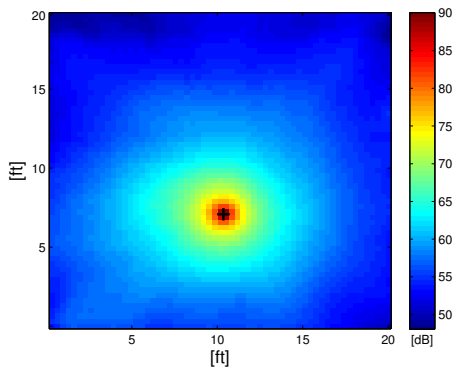
Figure 43: Reconstructions by the ridge-regularized LS.



(a) SLF

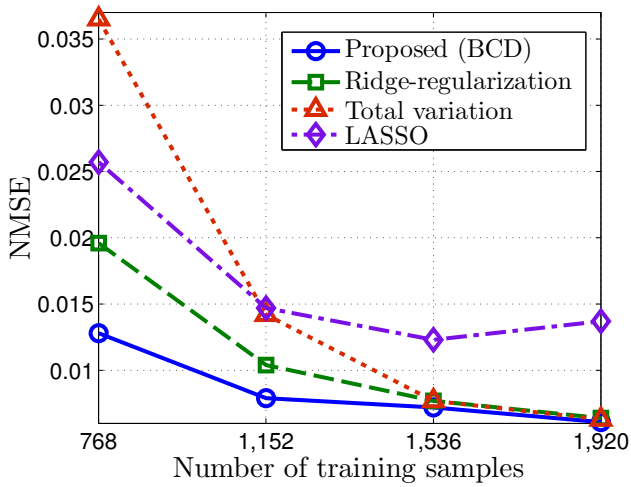


(b) Shadow fading map

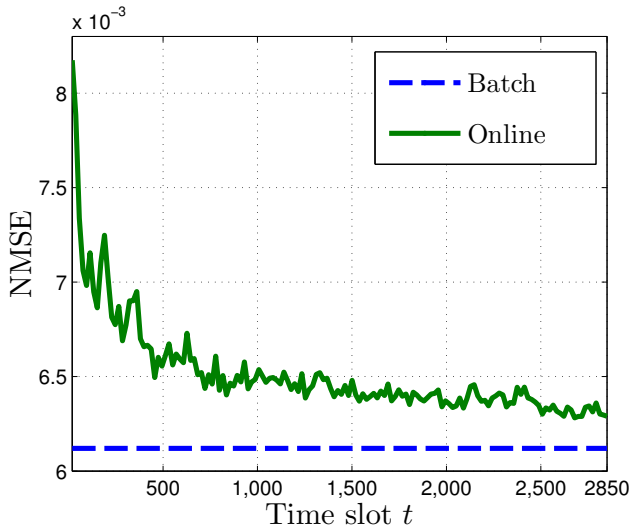


(c) Channel gain map

Figure 44: Reconstructions by the proposed online algorithm.



(a)



(b)

Figure 45: NMSE of channel gain prediction by (a) the batch; and (b) online algorithms.

uniformly drawn from the original dataset with replacement to demonstrate the asymptotic performance. Fig. 44 depicts the reconstructed SLF, shadow fading and channel gain maps obtained from the online algorithm. It can be seen

that the SLF shown in Fig. 44a is close to that depicted in Fig. 42a. Similar observations can be made for the shadow fading and channel gain maps as well. Thus, the online algorithm is a viable alternative to the batch algorithm with reduced computational complexity, and affordable memory requirement. Channel gain estimation performance of the proposed algorithms was assessed via 5-fold cross-validation. Let $\check{\mathbf{g}}_{\text{test}}$ and $\hat{\mathbf{g}}_{\text{test}}$ denote RSS measurement vectors in the test set and its estimate, respectively. Prediction performance is measured by the normalized mean-square error (NMSE) $\|\check{\mathbf{g}}_{\text{test}} - \hat{\mathbf{g}}_{\text{test}}\|^2/\|\check{\mathbf{g}}_{\text{test}}\|^2$. Fig. 45a displays the NMSE of batch algorithms with 480 test samples versus the number of training samples. It is shown that the proposed algorithm outperforms competing alternatives, particularly when a small number of training samples are available, validating the usefulness of the proposed model. The online algorithm was also tested with 2.85×10^5 measurements uniformly drawn from $1,920$ training samples with replacement. Fig. 45b depicts the evolution of the NMSE measured on 480 test samples at every t. It is observed that the online algorithm attains the batch performance as t increases.

Conclusions

A low-rank plus sparse matrix model was proposed for channel gain cartography, which is instrumental for various CR spectrum sensing and resource allocation tasks. The channel gains were modeled as the sum of the distance-based pathloss and the tomographic accumulation of shadowing due to the underlying SLF. The SLF was postulated to have a low-rank structure corrupted by sparse outliers. Efficient batch and online algorithms were developed by leveraging a bifactor-based characterization of the matrix nuclear norm. The algorithms enjoy low computational complexity and a reduced memory requirement, without sacrificing the optimality, with provable convergence properties. Tests with both synthetic and real measurement datasets corroborated the claims and showed that the algorithms could accurately reveal the structure of the propagation medium.

9

Exploiting Sparse User Activity in Multiuser Detection

This section focuses on efficient sparsity-utilizing multiuser detectors in code-division multiple access (CDMA) systems. Relying on the fact that the number of *active* users in CDMA systems is often much lower than the spreading gain, the presented approach fruitfully exploits this *a priori* information to improve performance of multiuser detectors. A low-activity factor manifests itself in a sparse symbol vector with entries drawn from a finite alphabet that is augmented by the zero symbol to capture user inactivity. The non-equiprobable symbols of the augmented alphabet motivate a sparsity-exploiting maximum *a posteriori* probability (S-MAP) criterion, which is shown to yield a cost comprising the ℓ_2 least-squares error penalized by the p-th norm of the wanted symbol vector ($p = 0, 1, 2$). Related optimization problems appear in variable selection (shrinkage) schemes developed for linear regression, as well as in the emerging field of CS. The contribution of this work to such *sparse* CDMA systems is a gamut of sparsity-exploiting multiuser detectors trading off performance for complexity requirements. From the vantage point of CS and the Lasso spectrum of applications, the contribution amounts to sparsity-exploiting algorithms when the entries of the wanted signal vector adhere to finite-alphabet constraints.

Related Works

Multiuser detection (MUD) algorithms play a major role for mitigating multi-access interference present in CDMA systems; see e.g., [288] and references therein. These well-appreciated MUD algorithms simultaneously detect the transmitted symbols of all active user terminals. However, they require knowledge of which terminals are active, and exploit no possible user (in)activity. In this work, MUD algorithms are developed when the active terminals are unknown and the activity factor (probability of each user being active) is low – a typical scenario in tactical or commercial CDMA systems deployed. The inactivity per user can be naturally incorporated by augmenting the underlying alphabet with an extra zero constellation point. Low activity thus implies a sparse symbol vector to be detected. With non-equiprobable symbols in the augmented alphabet, the optimal sparsity-embracing MUD naturally suggests a maximum *a posteriori* (MAP) criterion for detection. Sparse sphere decoding has been considered in, e.g., [277]. Sparsity has been used for estimating parameters of communication systems in, e.g., [53, 23, 9, 109], but not for multiuser detection.

Modeling and Problem Statement

Consider the uplink of a CDMA system with K user terminals and spreading gain N. Assume first that $N \geq K$. The under-determined case ($N < K$) will also be addressed later on in Section 9. Suppose the system has a relatively low activity factor, which analytically means that each terminal is active with probability (w.p.) $p_a < 1/2$ per symbol, and the events "*active*" are independent across symbols and across users. The case of correlated (in)activity of users across symbols will be dealt with in Section 9. Let $b_k \in \mathcal{A}$ denote the symbol, drawn from a finite alphabet by the k-th user, when active; otherwise, $b_k = 0$. Incorporating possible (in)activity per user is equivalent to having b_k take values from an *augmented* alphabet $\mathcal{A}_a := \mathcal{A} \bigcup \{0\}$.

The access point (AP) receives the superimposed modulated (quasi-) synchronous signature waveforms through (possibly frequency-selective) fading channels in the presence of additive white Gaussian noise (AWGN); and projects on the orthonormal space spanned by the aggregate waveforms to obtain the received chip samples collected in the $N \times 1$ vector $\mathbf{y}$. With the

$K \times 1$ vector $\mathbf{b}$ containing the symbols of all (active and inactive) users, the canonical input-output relationship is, see e.g., [288, Sec. 2.9]

$$\mathbf{y} = \mathbf{H}\mathbf{b} + \mathbf{w} \tag{195}$$

where $\mathbf{H}$ is an $N \times K$ matrix capturing transmit-receive filters, spreading sequences, channel impulse responses, and timing offsets; and the $N \times 1$ vector $\mathbf{w}$ is the AWGN. Without loss of generality (w.l.o.g.), $\mathbf{w}$ can be scaled to have unit variance. Note that (195) holds for quasi-synchronous systems too, where relative user asynchronism is bounded to a few chips per symbol, provided that: either i) user transmissions include guard bands to eliminate inter-symbol interference (ISI); or ii) the received vector $\mathbf{y}$ collects only the chips of each user that belong to the part of the common symbol interval under consideration.

The low activity factor implies that $\mathbf{b}$ is a sparse vector. However, the AP is neither aware of the positions nor the number of zero entries in $\mathbf{b}$. In order to perform multiuser detection (MUD) needed to determine the optimal $\hat{\mathbf{b}}$, the AP must account for the augmented alphabet $\mathcal{A}_a$, i.e., consider all the possible candidate vectors $\mathbf{b} \in \mathcal{A}_a^K$. This way, the MUD also determines the k-th user's activity captured by the extra constellation point $b_k = 0$. Supposing that the AP has the channel matrix $\mathbf{H}$ available (e.g., via training), the goal of this paper is to detect the optimal $\hat{\mathbf{b}}$ given the received vector $\mathbf{y}$ by exploiting the sparsity of active users.

To motivate this sparsity-exploiting MUD setup in the CDMA context, consider a set of terminals wishing to link with a common AP. Suppose that the AP acquires the full matrix $\mathbf{H}$ (with all terminals active) during a training phase. Those channels may include either non-dispersive or multipath fading, and are assumed invariant during the coherence time, which is typically larger compared to the symbol period. Each terminal accesses the channel randomly, and the AP receives the superposition of signals from the active terminals only. The AP is interested in determining both the active terminals and the symbols transmitted.

Another scenario where $\mathbf{H}$ is known and sparsity-exploiting MUD is well motivated, entails an unmanned aerial vehicle (UAV) collecting information from a ground wireless sensor network (WSN) placed over a grid, as depicted in Fig. 46. As the UAV flies over the grid of sensors, it collects the signals from a random subset of them. If the channel fading is predominantly affected by

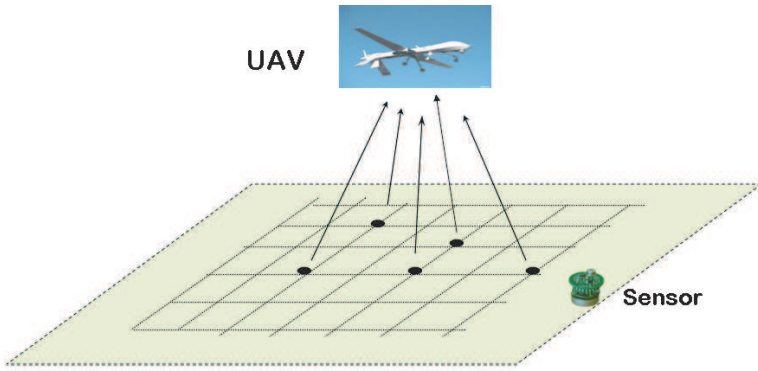


Figure 46: Wireless sensors access a UAV.

path loss, the UAV can acquire $\mathbf{H}$ based on the relative AP-to-sensor distances. Again, the UAV faces the problem of determining both identities of active sensors, and the symbols each active sensor transmits.

With this problem setup in mind, we will develop different MUD strategies, which account for the low activity factor. First, we will look at the maximum *a posteriori* probability (MAP) optimum MUD that exploits the sparsity present.

Sparsity-Exploiting MAP Detector

The goal is to detect $\mathbf{b}$ in (195), given a prescribed activity factor, the received vector $\mathbf{y}$, and the matrix $\mathbf{H}$. Recall though that the low activity factor leads to a sparse $\mathbf{b}$, i.e., each entry b_k is more likely to take the value 0 from the alphabet. Because entries $\{b_k\}_{k=1}^K$ are non-equiprobable, the optimal detector in the sense of minimizing the detection error rate is the MAP one.

Aiming at a sparsity-aware MAP criterion, consider first the prior probability for $\mathbf{b}$. For simplicity in exposition, suppose for now that each terminal transmits binary symbols when active, i.e., $\mathcal{A} = \{\pm 1\}$. (It will become clear later on that all sparsity-cognizant MUD schemes are applicable to finite alphabets of general constellations, not necessarily binary.) If b_k takes values from $\{-1, 0, 1\}$, with corresponding probabilities $\{p_a/2, 1 - p_a, p_a/2\}$, and since each entry b_k is independent from $b_{k'}$ for $k \neq k'$, the prior probability for $\mathbf{b}$

can be expressed as

$$\Pr(\mathbf{b}) = \prod_{k=1}^{K} \Pr(b_k) = (1-p_a)^{K-\|\mathbf{b}\|_0}(p_a/2)^{\|\mathbf{b}\|_0} \tag{196}$$

where $\|\mathbf{b}\|_0$ denotes the ℓ_0 (pseudo) norm that is by definition equal to the number of non-zero entries in the vector $\mathbf{b}$. Upon taking logarithms, (196) yields

$$\ln \Pr(\mathbf{b}) = -\lambda\|\mathbf{b}\|_0 + K\ln(1-p_a) \tag{197}$$

where

$$\lambda := \ln \frac{1-p_a}{p_a/2}. \tag{198}$$

Since low activity factor means $p_a < 1/2$, it follows readily from (198) that $\lambda > 0$. With the prior distribution of $\mathbf{b}$ in (197), the sparsity-aware MAP (S-MAP) detector is

$$\begin{aligned}\hat{\mathbf{b}}^{\text{MAP}} &= \arg\max_{\mathbf{b}\in\mathcal{A}_a^K} \Pr(\mathbf{b}|\mathbf{y}) = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} -\ln p(\mathbf{y}|\mathbf{b}) - \ln\Pr(\mathbf{b}) \\ &= \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_0 \end{aligned} \tag{199}$$

where the last equality follows from (197) and the Gaussianity of $\mathbf{w}$. Hence, the S-MAP detection task amounts to finding the vector in the constraint set $\mathcal{A}_a^K$, which minimizes the cost in (199).

Two remarks are now in order.

Remark 4: *(General constellations).* Beyond binary alphabets, it is easy to see why the S-MAP detector in (199) applies to more general constellations, including pulse amplitude modulation (PAM), phase-shift keying (PSK), and quadrature amplitude modulation (QAM). Specifically, for general M-ary constellations with $M \geq 2$ it suffices to adjust accordingly the prior probability as a function of $\|\mathbf{b}\|_0$ in (196). This will render the S-MAP MUD in (199) applicable to general M-ary constellations, provided that λ in (198) is replaced by $\lambda := \ln \frac{1-p_a}{p_a/M}$.

Remark 5: *(Scale λ as a function of p_a).* The definition in (198) reveals the explicit relationship of λ with the activity factor p_a. Different from CS and VS approaches, where λ is a tuning parameter often chosen with cross-validation techniques as a function of the data size N and K, here it is directly coupled with the user activity factor p_a. Such a coupling carries over even when users

have distinct activity factors. Specifically, if the user k is active w.p. $p_{a,k}$, then the term $\lambda\|\mathbf{b}\|_0 := \lambda\sum_{k=1}^K |b_k|$ in (199) should change to $\sum_{k=1}^K \lambda_k|b_k|$, where λ_k is defined as in (198) with $p_{a,k}$ substituting p_a. This user-specific regularization will be used in Section 9 to adaptively estimate user activity factors on-the-fly, and thus enable sparsity-aware MUD which accounts for correlated user (in)activity across the time slots.

With the variable b_k only taking values from $\{\pm 1, 0\}$, it holds for $p \geq 1$ that

$$\|\mathbf{b}\|_0 = \sum_{k=1}^K |b_k|^p = \|\mathbf{b}\|_p^p, \quad \forall\, \mathbf{b} \in \mathcal{A}_a^K. \tag{200}$$

Hence, the S-MAP detector (199) for binary transmissions is equivalent to

$$\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_p^p\,, \quad \forall\, p \geq 1. \tag{201}$$

Notice that the equivalence between (199) and (201) is based on the norm equivalence in (200), which holds only for constant modulus constellations. Although the cost in (201) will turn out to be of interest on its own, it is not an S-MAP detector for non-constant modulus constellations.

Interestingly, since low-activity factor implies a positive λ, the problem (201) entails a convex cost function, compared to the non-convex one in (199). In lieu of the finite-alphabet constraint, the criterion of (201) consists of the least-squares (LS) cost regularized by the ℓ_p norm, which in the context of linear regression has been adopted to mitigate data overfitting. In contrast, the LS estimator only considers the goodness-of-fit, and thus tends to overfit the data. Shrinking the LS estimator, by penalizing its size through the ℓ_p norm, typically outperforms LS in practice. For example, the Lasso adopts the ℓ_1 norm through which it effects sparsity[278]. In the MUD context for CDMA systems with low-activity factor, the vector $\mathbf{b}$ has a sparse structure, which motivates well this regularizing strategy. What is distinct and interesting here is that this penalty-augmented LS approach under finite-alphabet constraints emerges naturally as the logarithm of the prior in the S-MAP detector.

However, the finite-alphabet constraint renders the solution of (201) combinatorially complex. For general $\mathbf{H}$ and $\mathbf{y}$, the solution of (201) requires exhaustive search over all the 3^K feasible points, with the complexity growing exponentially in the problem dimension K. Likewise, for general M-ary alphabets the complexity incurred by (199) is $\mathcal{O}((M+1)^K)$. On the other

hand, many (sub-) optimal alternatives are available in the MUD literature; see e.g., [288, Ch. 5-7]. Similarly here, we will develop different (sub-) optimal algorithms to trade off complexity for probability of error performance in sparsity-exploiting MUD alternatives.

Since the exponential complexity of MUD stems from the finite-alphabet constraint, one reduced-complexity approach is to solve the unconstrained convex problem, and then quantize the resultant soft decision to the nearest point in the alphabet. This approach includes the sub-optimal linear MUD algorithms (decorrelating and minimum mean-square error (MMSE) detectors). Another approach is to search over (possibly a subset of) the alphabet lattice directly as in the decision-directed detectors or the sphere decoders [97]. Likewise, it is possible to devise (sub-) optimal algorithms for solving the S-MAP MUD problem (201) along these two categories. First, we will present the sparsity-exploiting MUD algorithms by relaxing the finite-alphabet constraint.

Relaxed S-MAP Detectors

In addition to offering an S-MAP detector for constant modulus constellations, the cost in (201) is convex. Thus, by relaxing the combinatorial constraint, the optimization problem (201) can be solved efficiently by capitalizing on convexity. As mentioned earlier, this problem is similar to the penalty-augmented LS criterion that is used for VS in linear regression, where the choice of p is important for controlling the shrinking effect, that is the degree of sparsity in the solution. Next, we will develop detectors for two choices of p, and compare them in terms of complexity and performance.

Linear Ridge MUD

The choice $p = 2$ is a popular one in statistics, well-known as *Ridge regression*. Its popularity is mainly due to the fact that it can regularize the LS solution while retaining its closed-form expression as a linear function of the data $\mathbf{y}$. A relaxed detection algorithm for S-MAP MUD can be developed accordingly with $p = 2$, what we term Ridge detector (RD). Ignoring the finite-alphabet

constraint, the optimal solution of (201) for $p = 2$ takes a linear form

$$\begin{aligned}\mathbf{b}^{\mathrm{RD}} &= \arg\min_{\mathbf{b}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_2^2 \\ &= (\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}. \end{aligned} \tag{202}$$

In addition to its simplicity, and different from LS, the existence of the inverse in (202) is ensured even for ill-posed or under-determined problems; i.e., when $\mathbf{H}$ is rank deficient or fat ($N < K$). Notice that $\mathbf{b}^{\mathrm{RD}}$ takes a form similar to the linear MMSE multiuser detector, with the parameter λ replacing the noise variance, and connecting the activity factor with the degree of regularization applied to the matrix $\mathbf{H}^T\mathbf{H}$.

Upon quantizing each entry of the soft decision $\mathbf{b}^{\mathrm{RD}}$ with the operator

$$\mathcal{Q}_\theta(x) := \mathrm{sign}(x)\mathbb{1}(|x| \geq \theta) \tag{203}$$

where $\theta > 0$, $\mathrm{sign}(x) = 1(-1)$ with $x > (<)0$, and $\mathbb{1}$ denoting the indicator function, the hard RD is

$$\hat{\mathbf{b}}^{\mathrm{RD}} = \mathcal{Q}_\theta(\mathbf{b}^{\mathrm{RD}}) = \mathcal{Q}_\theta\left((\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}\right). \tag{204}$$

Because the detector in (202) is linear, it is possible to express linearly its soft output $\mathbf{b}^{\mathrm{RD}}$ with respect to (w.r.t.) the input symbol vector $\mathbf{b}$. Based on this relationship, one can subsequently derive the symbol error rate (SER) of the hard detected symbols in $\hat{\mathbf{b}}^{\mathrm{RD}}$ as a function of the quantization threshold θ. These steps will be followed next to obtain the performance of the RD.

Performance Analysis

Letting $\check{\mathbf{b}}$ denote the vector transmitted, and substituting $\mathbf{y} = \mathbf{H}\check{\mathbf{b}} + \mathbf{w}$ into (202) yields

$$\mathbf{b}^{\mathrm{RD}} = (\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-1}\mathbf{H}^T(\mathbf{H}\check{\mathbf{b}} + \mathbf{w}) = \mathbf{G}\check{\mathbf{b}} + \mathbf{w}' \tag{205}$$

where $\mathbf{G} := \mathbf{I} - 2\lambda(\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-1}$, and the colored noise $\mathbf{w}' := (\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-1}\mathbf{H}^T\mathbf{w}$ is zero-mean Gaussian with covariance matrix $\mathbf{\Sigma}_{w'} := E\{\mathbf{w}'(\mathbf{w}')^T\} = (\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-2}\mathbf{H}^T\mathbf{H}$.

It follows readily from (205) that the k-th entry of $\mathbf{b}^{\mathrm{RD}}$ satisfies

$$b_k^{\mathrm{RD}} = G_{kk}\check{b}_k + \sum_{\ell \neq k} G_{k\ell}\check{b}_\ell + w'_k \tag{206}$$

where $G_{k\ell}$ and w'_k are the (k,ℓ)-th and k-th entries of $\mathbf{G}$ and $\mathbf{w}'$, respectively. The last two terms in the right-hand side of (206) capture the multiuser interference-plus-noise effect, which has variance

$$\sigma_k^2 = \text{var}\left\{\sum_{\ell\neq k} G_{k\ell}\check{b}_\ell + w'_k\right\} = \sum_{\ell\neq k} G_{k\ell}^2 p_a + \Sigma_{w',kk} \tag{207}$$

where $\Sigma_{w',kk}$ denotes the (k,k)-th entry of $\mathbf{\Sigma}_{w'}$.

With the interference-plus-noise term being approximately Gaussian distributed, deciphering $\check{b}_k$ from (206) amounts to detecting a ternary deterministic signal in the presence of zero-mean, Gaussian noise of known variance. Hence, the symbol error rate (SER) for the k-th entry using the quantization rule in (204) entry-wise can be analytically obtained as

$$P_{e,k}^{\text{RD}} = 2(1-p_a)Q\left(\frac{\theta}{\sigma_k}\right) + p_a Q\left(\frac{G_{kk}-\theta}{\sigma_k}\right) \tag{208}$$

where $Q(\mu) := (1/\sqrt{2\pi})\int_\mu^\infty \exp(-\nu^2/2)d\nu$ denotes the Gaussian tail function.

The SER in (208) is a convex function of the threshold θ. Thus, taking the first-order derivative w.r.t. θ and setting it equal to zero yields the optimal threshold for the k-th entry as

$$\hat{\theta}_k = \frac{G_{kk}}{2} + \frac{\sigma_k^2}{G_{kk}}\lambda\,. \tag{209}$$

The corresponding minimum SER becomes [cf. (208) and (209)]

$$\hat{P}_{e,k}^{\text{RD}} = 2(1-p_a)Q\left(\frac{G_{kk}}{2\sigma_k} + \frac{\lambda\sigma_k}{G_{kk}}\right) + p_a Q\left(\frac{G_{kk}}{2\sigma_k} - \frac{\lambda\sigma_k}{G_{kk}}\right). \tag{210}$$

As the CDMA system signal-to-noise ratio (SNR) goes to infinity, asymptotically we have $\mathbf{G}\to\mathbf{I}$ and $\mathbf{\Sigma}_{w'}\to\mathbf{0}$, so the optimal threshold in (209) approaches 0.5. The numerical tests in Section 9 will also confirm that selecting $\theta_k = 0.5$ approaches the minimum SER $\hat{P}_{e,k}^{\text{RD}}$ over the range of SNR values encountered in most practical settings.

The clear advantage of RD-MUD is its simplicity as a linear detector. However, using the ℓ_2 norm for regularization, the RD-MUD inherently introduces a Gaussian prior for the unconstrained symbol vector and is thus not affecting sparsity in $\mathbf{b}^{\text{RD}}$; see also [278]. This renders the performance of RD dependent

on the quantization threshold θ – a fact also corroborated by the simulations in Section 9. These considerations motivate the ensuing development of an alternative relaxed S-MAP algorithm, which accounts for the sparsity present in $\mathbf{b}$.

Lasso-Based MUD

Another popular regression method is the Lasso one, which regularizes the LS cost with the ℓ_1 norm. In the Bayesian formulation, regularization with the ℓ_1 norm corresponds to adopting a Laplacian prior for $\mathbf{b}$ [278]. The nice feature of Lasso-based regression is that it ensures sparsity in the resultant estimates. The degree of sparsity depends on the value of λ, which is selected here using the *a priori* information available on the activity factor [cf. (198)]. The optimal solution of (201) for $p = 1$ without the finite-alphabet constraint yields the Lasso detector (LD) as

$$\mathbf{b}^{\mathrm{LD}} = \arg\min_{\mathbf{b}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1. \tag{211}$$

While a closed-form solution is impossible for general $\mathbf{H}$, the minimization in (211) is a quadratic programming (QP) problem that can be readily accomplished using available QP solvers, such as SeDuMi [271]. Upon slicing the solution in (211), we obtain the detection result as

$$\hat{\mathbf{b}}^{\mathrm{LD}} = \mathcal{Q}_\theta(\mathbf{b}^{\mathrm{LD}}). \tag{212}$$

The larger λ is, the more sparse $\mathbf{b}^{\mathrm{LD}}$ becomes [cf. (198)]. This is intuitively reasonable, because λ is inversely proportional to the activity factor p_a. Since the Lasso approach (211) yields sparse estimates systematically, and can be obtained via QP solvers in polynomial time, LD is a competitive MUD alternative. Lack of a closed-form solution prevents analytical evaluation of the SER, which will be tested using simulations in Section IV.

Remark 6: So far, we assumed $\mathcal{A} = \{\pm 1\}$ to ensure equivalence of the ℓ_p-norm regularized S-MAP detector (201) with the more general one in (199). However, the sub-optimal algorithms of this section ignore the finite-alphabet constraint, and just rely on the convexity of the cost function in (201) to offer MUD schemes that can be implemented efficiently, either in linear closed-form or through quadratic programming. In fact, starting from (201) and forgoing its equivalence with (199), the RD and LD relaxations of (201) apply also

for general M-ary alphabets for any $M > 2$. Of course, the quantization thresholds required for slicing the soft symbol estimates in order to obtain hard symbol estimates must be modified in accordance with the corresponding constellation. For non-constant modulus transmissions, the cost in (201) favors low-energy (close to the origin) constellation points, but this effect is mitigated by the judicious selection of the quantization thresholds.

Note that forgoing the equivalence of (201) with (199) is less of an issue for RD and LD because the major limitation of these simple relaxation-based algorithms is their sub-optimality, which emerges because they do not account for the finite-alphabet symbol constraints. Next, MUD algorithms are developed to minimize the S-MAP cost while adhering to the constraint in (199) explicitly.

S-MAP Detectors with Lattice Search

User symbols in this section are drawn from an M-ary PAM alphabet $\mathcal{A} = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$, with M even. Consider also reformulating the S-MAP problem in (199) using the QR decomposition of the matrix $\mathbf{H}$ (assumed here to be square or tall with full column rank) as $\mathbf{H} = \mathbf{QR}$, where $\mathbf{R}$ is a $K \times K$ upper triangular matrix, and $\mathbf{Q}$ is an $N \times K$ unitary matrix. Substituting this QR decomposition into (199), and left multiplying with the unitary $\mathbf{Q}$ inside the LS cost, the S-MAP detector becomes: $\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^k} \frac{1}{2}\left\|\mathbf{Q}^T\mathbf{y} - \mathbf{Q}^T(\mathbf{QR})\mathbf{b}\right\|_2^2 + \lambda\|\mathbf{b}\|_0 = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\mathbf{y}' - \mathbf{Rb}\|_2^2 + \lambda\|\mathbf{b}\|_0$, where $\mathbf{y}' := \mathbf{Q}^T\mathbf{y}$; or, after using the definitions of the norms,

$$\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \sum_{k=1}^{K} \left\{ \frac{1}{2}\left(y'_k - \sum_{\ell=k}^{K} R_{k\ell} b_\ell \right)^2 + \lambda |b_k|_0 \right\} . \tag{213}$$

Although the optimal solution of (213) still incurs exponential complexity, the upper triangular form of $\mathbf{R}$ enables decomposition of (213) into sub-problems involving only scalar variables. As it will be seen later in Section 9, the S-MAP problem accepts a neat closed-form solution in the scalar case ($K = 1$). This is instrumental for the development of efficient (near-) optimal algorithms searching over the finite-alphabet induced lattice. One such sub-optimal MUD algorithm is described next.

Algorithm 16 (DDD): Input $\mathbf{y}'$, $\mathbf{R}$, λ, and M even. Output $\hat{\mathbf{b}}^{\mathrm{DDD}}$.

1: **for** $k = K, K-1, \ldots, 1$ **do**
2: (Back substitution) Compute the unconstrained LS solution $b_k^{\mathrm{LS}} := (y'_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_\ell^{\mathrm{DDD}})/R_{kk}$.
3: (Quantize to $\mathcal{A}$) Set $\hat{b}_k^{\mathrm{DDD}} := \lfloor b_k^{\mathrm{LS}} \rceil$.
4: (Compare with 0) If $2b_k^{\mathrm{LS}}\lfloor b_k^{\mathrm{LS}} \rceil - \lfloor b_k^{\mathrm{LS}} \rceil^2 - 2\lambda/R_{kk}^2 \leq 0$, then set $\hat{b}_k^{\mathrm{DDD}} := 0$.
5: **end for**

Sparsity-Exploiting Decision-Directed MUD

Close look at (213) reveals that once the estimates $\{\hat{b}_\ell\}_{\ell=k+1}^{K}$ are available, the optimal $\hat{b}_k$ can be obtained by minimizing the cost corresponding to the k-th summand of (213). This leads to the per-symbol optimal decision-directed detector (DDD), which following related schemes in different contexts, could be also called *successive interference cancellation* or *decision-feedback decoding*, see e.g., [123, Sec. 9.4] and [288, Ch. 7]. The main difference here is that $\mathbf{b}$ is sparse.

The DDD algorithm relies on back substitution to decompose the overall S-MAP cost into K sub-costs each dependent on a single scalar variable and accepting a closed-form solution. Specifically, supposing the symbols $\{\hat{b}_\ell^{\mathrm{DDD}}\}_{\ell=k+1}^{K}$ have been already detected, the DDD algorithm detects the k-th symbol as

$$\hat{b}_k^{\mathrm{DDD}} = \arg\min_{b_k \in \mathcal{A}_a} \frac{1}{2}\left(y'_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_\ell^{\mathrm{DDD}} - R_{kk}b_k\right)^2 + \lambda|b_k|_0 \,. \quad (214)$$

This minimization problem entails only one scalar variable taking one of $(M+1)$ possible points in $\mathcal{A}_a$. Thus, the minimum is found after comparing the costs corresponding to these $(M+1)$ values. Appendix I proves that this optimal solution can be found in closed form as

$$\hat{b}_k^{\mathrm{DDD}} = \lfloor b_k^{\mathrm{LS}} \rceil \mathbb{1}\left(2b_k^{\mathrm{LS}}\lfloor b_k^{\mathrm{LS}} \rceil - \lfloor b_k^{\mathrm{LS}} \rceil^2 - 2\lambda/R_{kk}^2 > 0\right) \quad (215)$$

where $b_k^{\mathrm{LS}} := (y'_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_\ell^{\mathrm{DDD}})/R_{kk}$, and $\lfloor \cdot \rceil$ quantizes to the nearest point in $\mathcal{A}$. The simple implementation steps are tabulated as Algorithm 16.

When $\mathbf{R}$ is diagonal, Algorithm 1 yields the optimal S-MAP detection result; i.e., $\hat{\mathbf{b}}^{\text{MAP}} = \hat{\mathbf{b}}^{\text{DDD}}$ for this case. However, since the DDD detects symbols sequentially, it is prone to error propagation, especially at low SNR values. The error propagation can be mitigated by preprocessing and ordering methods [288, Ch. 7]. Also similar to all related detectors that rely on back substitution, error performance of the sparse DDD will be analyzed assuming there is no error propagation. For the special case of $M = 2$, Appendix J shows that under this assumption, the SER for (215) becomes

$$P_{e,k}^{\text{DDD}} = 2(1-p_a)Q\left(\frac{|R_{kk}|}{2}+\frac{\lambda}{|R_{kk}|}\right)+p_aQ\left(\frac{|R_{kk}|}{2}-\frac{\lambda}{|R_{kk}|}\right) . \quad (216)$$

For a general M-ary constellation, it is also possible to approximate the SER using the union bound.

Because it accounts for the finite-alphabet constraint, sparse DDD outperforms the relaxed detectors of the previous section – a fact that will be confirmed also by simulations. However, the error propagation emerging at medium-low SNR degrades the sparse DDD performance when compared to the optimal but computationally complex S-MAP detector. As a compromise, a branch-and-bound type of MUD algorithm is developed next, to attain (near-) optimal performance by exploiting the finite-alphabet and sparsity constraints, at the price of increased complexity compared to DDD.

Sparsity-Exploiting Sphere Decoding-based MUD

Sphere decoding (SD) algorithms have been widely used for maximum-likelihood (ML) demodulation of multiuser and/or multiple-input multiple-output (MIMO) transmissions. Given a PAM or QAM alphabet, SD yields (near-) ML performance at polynomial average complexity; see e.g., [123, Sec. 5.2]. However, different from the ML-optimal SD that minimizes an LS cost, the S-MAP problem (213) entails also a regularization term to account for sparsity in $\mathbf{b}$. Although the resultant algorithm will be termed sparse sphere decoder (SSD), it searches in fact within an "ℓ_0-*norm regularized* sphere," which is not a sphere but a hyper-solid.

The goal is to find the unknown $K \times 1$ vector $\mathbf{b} \in \mathcal{A}_a^K$, which minimizes the distance metric [cf. (213)]

$$D_1^K(\mathbf{b}) := \sum_{k=1}^{K} \left\{ \frac{1}{2} \left(y'_k - \sum_{\ell=k}^{K} R_{k\ell} b_\ell \right)^2 + \lambda |b_k|_0 \right\}. \tag{217}$$

For a large enough threshold τ, candidate vectors (and thus the minimizer of D_1^K too) satisfy[22]

$$D_1^K(\mathbf{b}) < \tau \tag{218}$$

that specifies a hyper-solid inside which the wanted minimizer must lie. Define now [cf. (215)]

$$\rho_k := \left(y'_k - \sum_{\ell=k+1}^{K} R_{k\ell} b_\ell \right) \Big/ R_{kk}, \quad k = K, K-1 \cdots 1 \tag{219}$$

where $\rho_K := y'_K / R_{KK}$. Note that ρ_k depends on $\{b_\ell\}_{\ell=k+1}^K$.
Using (219), the hyper-solid in (218) can be expressed as

$$D_1^K(\mathbf{b}) = \sum_{k=1}^{K} \left\{ \frac{R_{kk}^2}{2} (\rho_k - b_k)^2 + \lambda |b_k|_0 \right\} < \tau \tag{220}$$

or, in a more compact form as $D_1^K(\mathbf{b}) = \sum_{k=1}^K d_k(b_k) < \tau$, where $d_k(b_k) := (R_{kk}^2/2)(\rho_k - b_k)^2 + \lambda |b_k|_0$. In addition to the overall metric D_1^K assessing a candidate vector $\mathbf{b} \in \mathcal{A}_a^K$, as well as the per entry metric d_k for each candidate symbol $b_k \in \mathcal{A}_a$, it will be useful to define the *accumulated* metric $D_k^K := \sum_{\ell=k}^K d_\ell(b_\ell)$ corresponding to the $K-k+1$ candidate symbols from entry K down to entry k.
Per entry k, which subsequently will be referred to as level k, eq. (220) implies a set of inequalities:

$$\text{Level } k: \quad d_k(b_k) < \tau - D_{k+1}^K, \quad \text{for } k = K, K-1 \ldots 1, \tag{221}$$

with $D_{K+1}^K := 0$. SSD relies on the Schnorr-Euchner (SE) enumeration, see e.g., [80], properly adapted here to account for the ℓ_0-norm regularization. SE capitalizes on the inequalities (221) to search efficiently over all possible

[22] At initialization, τ is set equal to ∞ so that (218) is always satisfied.

vectors $\mathbf{b}$ with entries belonging to $\mathcal{A}_a$. Any candidate $\mathbf{b} \in \mathcal{A}_a^K$ obeying the K inequalities in (221) for a given τ, will be termed *admissible*. Threshold τ is reduced after each admissible $\mathbf{b}$ is identified, as will be detailed soon. The SE-based SSD amounts to a *depth-first* tree search, which seeks and checks candidate vectors starting from entry (level) K and working downwards to entry 1 per candidate vector.

At level K, SE search chooses $b_K = \lfloor \rho_K \rceil \mathbb{1}(2\rho_K \lfloor \rho_K \rceil - \lfloor \rho_K \rceil^2 - 2\lambda/R_{KK}^2 > 0)$, which we know from (215) is the constellation point yielding the smallest d_K. If this choice of b_K does not satisfy the inequality (221) with $k = K$, no other constellation point will satisfy it either, and the minimizer of D_1^K in (217) must lie outside[23] the hyper-solid postulated by (218). If this choice of b_K satisfies (221), SE proceeds to level $K-1$ in which (219) is used first with $k = K-1$ to find ρ_{K-1} that depends on the chosen b_K from level K; subsequently, b_{K-1} is selected as $b_{K-1} = \lfloor \rho_{K-1} \rceil \mathbb{1}(2\rho_{K-1} \lfloor \rho_{K-1} \rceil - \lfloor \rho_{K-1} \rceil^2 - 2\lambda/R_{K-1,K-1}^2 > 0)$. If this choice of b_{K-1} does not satisfy (221) with $k = K-1$, then we move back to level K, and select b_K equal to the constellation point yielding the *second* smallest d_K, and so on; otherwise, we proceed to level $K-2$. Continuing this procedure down to level 1, yields the first candidate vector $\hat{\mathbf{b}}$, which is deemed admissible since it has entries belonging to $\mathcal{A}_a$ and also satisfying (218). This candidate is stored, and the threshold is updated to $\tau := D_1^K(\hat{\mathbf{b}})$.

Then, the search proceeds looking for a better candidate. Now at level 1, we move up to level 2 and choose b_2 equal to the constellation point yielding the *second* smallest cost d_2. If this b_2 satisfies (221) at level 2 with the current τ, we move down to level 1 to update the value of b_1 (note that b_2 has just been updated and $\{b_\ell\}_{\ell=3}^K$ are equal to the corresponding entries in $\hat{\mathbf{b}}$). If (221) at level 2 is not satisfied with the current τ, we move up to level 3 to update the value of b_3, and so on.

Finally, when it fails to find any other admissible candidate satisfying (221), the search stops, and the latest admissible candidate $\hat{\mathbf{b}}$ is the optimal $\hat{\mathbf{b}}^{\mathrm{MAP}}$ solution sought. With $\tau = \infty$, the first found admissible candidate $\hat{\mathbf{b}}$ is the $\hat{\mathbf{b}}^{\mathrm{DDD}}$ solution of Section V-A.

Before summarizing the SSD steps, it is prudent to elaborate on the *ordered* enumeration of the constellation points *per level*, which in fact constitutes

[23] This will never happen with $\tau = \infty$ in (218).

the main difference of SSD relative to SD. In lieu of the 0 constellation point and the ℓ_0 norm, SE in SD enumerates the PAM symbols per level k in the order of increasing cost as: $\{b_k, b_k + 2\Delta_k, b_k - 2\Delta_k, b_k + 4\Delta_k, b_k - 4\Delta_k, \cdots\} \bigcap \mathcal{A}$, with $b_k := \lfloor \rho_k \rceil$ and $\Delta_k := \text{sign}(\rho_k - b_k)$. (With $\lfloor \rho_k \rceil$ yielding the smallest d_k, if $\Delta_k = 1$, then $\lfloor \rho_k \rceil + 2$ yields the second smallest d_k and $\lfloor \rho_k \rceil - 2$ the third; and the other way around, if $\Delta_k = -1$.) SD effects such an ordered enumeration by alternately updating $b_k = b_k + 2\Delta_k$ and $\Delta_k = -\Delta_k - \text{sign}(\Delta_k)$, [80]. To demonstrate how SSD further accounts for the ℓ_0-norm regularization and the augmented alphabet of S-MAP which includes 0, let $b_k^{(i)} \in \mathcal{A}_a$ denote the symbol for level k incurring the i-th smallest ($i = 1, 2, \ldots, M+1$) cost d_k. If $b_k^{(i)} \in \mathcal{A}$, then $b_k^{(i+1)}$ will be either 0 or $b_k^{(i)} + 2\Delta_k$. If $d_k(0) < d_k(b_k^{(i)} + 2\Delta_k)$, then the next symbol in the ordered enumeration should be $b_k^{(i+1)} = 0$, and an auxiliary variable $b_k^{(c)}$ is used to cache the subsequent symbol in the order as $b_k^{(i+2)} = b_k^{(i)} + 2\Delta_k$. With $b_k^{(i+1)} = 0$, the auxiliary variable allows the wanted $b_k^{(i+2)}$ at the next enumeration step to be retrieved from $b_k^{(c)}$.

Similar to SD, the ordered enumeration pursued by SSD per level implies a corresponding order in considering all $\mathbf{b} \in \mathcal{A}_a^K$, which leads to a repetition-free and exhaustive search of all admissible candidate vectors. At the same time, the hyper-solid postulated by (218) shrinks as τ decreases, until no other admissible vector can be found. This guarantees that the SSD outputs the vector with the smallest D_1^K, and thus the optimal solution to (213). The SSD algorithm can be summarized in the following six steps 1–6 tabulated as Algorithm 17.

Remark 7: SSD inherits all the attractive features of SD [80]. Specifically, during the search one basically needs to store D_k^K per level k. Its *in place update* for each b_k candidate implies that SSD *memory requirements* are only linear in K. In addition, the *computational efficiency* of SSD (relative to that of ML which is $\mathcal{O}((M+1)^K)$) stems from four factors: (i) the DDD solution provides an admissible initialization reducing the search space at the outset; (ii) the recursive search enabled by the QR decomposition gives rise to the causally dependent inequalities (221), which restrict admissible candidate entries to choices that even decrease over successive depth-first passes of the search; (iii) ordering per level increases the likelihood of finding "near-optimal admissible" candidates early, which means quick and sizeable shrinkage of

Algorithm 17 (SSD): Input $\tau = \infty$, $\mathbf{y}'$, $\mathbf{R}$, λ, and M even. Output the solution $\hat{\mathbf{b}}^{\text{MAP}} := \hat{\mathbf{b}}^{\text{SSD}}$ to (213).

1: (Initialization) Set $k := K$, $D_{K+1}^K := 0$.

2: Compute ρ_k as in (219), set $b_k := \lfloor \rho_k \rceil$, $b_k^{(c)} := 0$, $\Delta_k := \text{sign}(\rho_k - b_k)$.
If $2\rho_k b_k - b_k^2 - 2\lambda/R_{kk}^2 < 0$, then
// `symbol 0 yields smaller` d_k `than` $\lfloor \rho_k \rceil$
Set $b_k^{(c)} := b_k$, and $b_k := 0$.
End if and go to Step 3.

3: **If** $d_k(b_k) := (R_{kk}^2/2)(\rho_k - b_k)^2 + \lambda |b_k|_0 \geq \tau - D_{K+1}^K$, then
go to Step 4. // `outside hyper-solid in` (218)
Else if $|b_k| > M - 1$, then
go to Step 6. // `inside hyper-solid in` (218), `but outside` $\mathcal{A}_a$
Else if $k > 1$, then
compute $D_k^K := D_{K+1}^K + d_k(b_k)$; set $k := k - 1$, and go to Step 2.
// `go the next level (deeper in the tree)`
Else go to Step 5. // $k = 1$, `at the tree's bottom`
End if

4: **If** $k = K$, then terminate
Else set $k := k + 1$, go to Step 6.
End if

5: (An admissible $\mathbf{b}$ is found)
Set $\tau := D_2^K + d_1(b_1)$, $\hat{\mathbf{b}}^{\text{SSD}} := \mathbf{b}$, and $k := k + 1$; then go to Step 6.

6: (Enumeration at level k proceeds to the candidate symbol next in the order)
If $b_k = 0$, then
Retrieve the next (based on cost d_k ordering) symbol $b_k := b_k^{(c)}$, and set $b_k^{(c)} := FLAG$.
Else set $b_k := b_k + 2\Delta_k$, and $\Delta_k := -\Delta_k - \text{sign}(\Delta_k)$.
If $b_k^{(c)} \neq FLAG$ and $2\rho_k b_k - b_k^2 - 2\lambda/R_{kk}^2 < 0$, then
// `0 yields smaller` d_k `than` b_k
Set $b_k^{(c)} := b_k$, and $b_k := 0$.
End if
End if and go to Step 3.

the hyper-solid, and thus fast convergence to the S-MAP optimal solution; and (iv) metrics involved in the search can be efficiently reused since children of the same level in the tree share the already computed accumulated metric of the "partial path" from this level to the root.

Compared to other sub-optimal detection schemes presented in previous sections, the SSD algorithm can return the S-MAP optimal solution possibly at exponential complexity, unless one stops the search at the affordable complexity – case in which the solution is only ensured to be near-optimal. Fortunately, at medium-high SNR, both SD and SSD return the optimal solution at average complexity which is polynomial (typically cubic). Moreover, SSD can be generalized to provide symbol-by-symbol soft output with approximate *a posteriori* probabilities, as is the case with the SD; see e.g., [123, Chapter 5].

Generalizations of S-MAP Detectors

Up to now, four sparsity-exploiting MUD algorithms have been developed to solve the integer program associated with the linear model in (195). The present section will present interesting generalizations to account for correlated user (in)activity across symbols, and under-determined CDMA systems.

Exploiting User (In)Activity Across Symbols

Sparsity-aware detectors for the linear model in (195) were so far developed on a symbol-by-symbol basis, which does not account for the fact that user (in)activity typically persists across multiple symbols. To this end, user activity across time can for instance be thought of as a Markov chain with two states (active-inactive). Once a user terminal starts transmitting to the AP, it becomes more likely to stay active for the next symbol slot too; and likewise, inactive once it stops transmitting. In this model, the state transition probability from either one state to the other is relatively much smaller than that of staying unchanged, and this manifests itself to the said dependence of user (in)activities across time.

Admittedly, MUD schemes accounting for this dependence must process the aggregation of data vectors $\mathbf{y}$ obeying (195) across slots. With N_s denoting the number of slots, the number of unknowns (KN_s) can grow prohibitively with N_s. One approach to cope with this "curse of dimensionality" is via

dynamic programming, which can take advantage of the fact that correlation is only present between two consecutive slots; see e.g., [288, Sec. 4.2]. However, for M-ary alphabets, the resultant sequential detector requires evaluating per symbol slot the path weights of all $(M+1)^K$ possible symbol vectors. This high computational burden is impractical for real-time implementations.

The proposed alternative to bypass this challenge stems from the observation that for a given slot t the influence of all previous slots $\{t'\}_{t'=0}^{t-1}$ on the S-MAP detection rule is reflected only in the prior probability of each user being active at time t; i.e., user k's current (and time-varying) activity factor $p_{a,k}(t)$. The natural means to capture this influence online is to track each user's activity factor using the recursive LS (RLS) estimator [254, Ch. 12], based on activity factors from previous slots; that is,

$$\hat{p}_{a,k}(t) = \arg\min_{p} \sum_{t'=0}^{t-1} \beta_k^{t,t'} (p - |\hat{b}_k(t')|_0)^2, \quad t = 1, \ldots \tag{222}$$

where $\hat{b}_k(t')$ denotes user k's detected symbol at time t', and the so-called "forgetting-factor" $\beta_k^{t,t'}$ describes the effective memory (data windowing). A popular choice is the exponentially decaying window, for which $\beta_k^{t,t'} := \beta_k^{t-t'}$ for some $0 \ll \beta_k < 1$. Accordingly, (222) can expressed in closed form, recursively as

$$\begin{aligned}\hat{p}_{a,k}(t) &= \frac{1-\beta_k}{\beta_k} \left(\sum_{t'=0}^{t-1} \beta_k^{t-t'} |\hat{b}_k(t')|_0 \right) / (1-\beta_k^t) \\ &= \frac{\beta_k - \beta_k^t}{1-\beta_k^t} \hat{p}_{a,k}(t-1) + \frac{1-\beta_k}{1-\beta_k^t} |\hat{b}_k(t-1)|_0, \ t = 1, \ldots \end{aligned} \tag{223}$$

where the last equality comes from back substitution of $\hat{p}_{a,k}(t-1)$. The choice of β_k critically depends on the (in)activity correlation between consecutive slots. In the extreme case where user (in)activities across slots are independent, the infinite-memory window ($\beta_k = 1$) is optimal, and (223) reduces to the simple online time-averaging estimate $\hat{p}_{a,k}(t) = \frac{1}{t} \sum_{t'=0}^{t-1} |\hat{b}_k(t')|_0$.

Adapting the user activity factors allows one to weigh entries of ℓ_0-norm regularization which in turn affects the prior probability in the S-MAP detector (199) through the coefficient $\lambda_k(t)$ corresponding to $\hat{p}_{a,k}(t)$ (cf. Remark 5). Note that when the correlation across time is strong, it is possible that $\hat{p}_{a,k}(t)$ can approach 1, case in which $\lambda_k(t)$ is not guaranteed to stay positive. This will

cause problems to the relaxed S-MAP detectors of Section 9, as those schemes rely on the convexity of the cost function in (201). However, the DDD and SSD algorithms will remain operational, because they rely on enumeration per symbol (in DDD) or per group of symbols within a sphere (in SSD). Note also that with $\lambda < 0$ the regularization term in the minimization of (213) is non-positive; hence, $\lambda < 0$ encourages searching over the non-zero constellation points (and thus discourages sparsity), whereas $\lambda > 0$ promotes sparsity.

Under-Determined CDMA Systems

Minimal-size spreading sequences, even smaller than the number of users, is well motivated for bandwidth and power savings. Without finite-alphabet constraints on the wanted vector, results available in the CS literature guarantee recovery of sparse signals from a few observations; see e.g., [46, 49] and references therein. Specifically, [46] shows that if the vector of interest is sparse (or compressible) over a known basis, then it is possible to reconstruct it with very high accuracy from a small number of random linear projections at least in the ideal noise-free case. For non-ideal observations corrupted with unknown noise of bounded perturbation, [49] provides an upper bound on the reconstruction error, which is proportional to the noise variance for a sufficiently sparse signal. However, CS theory pertains to sparse analog-valued signals. Moreover, the noise considered in a practical communication system is typically Gaussian, or generally drawn from a distribution having possibly unbounded support. Therefore, existing results from the CS literature do not carry over to the present context.

Nevertheless, it is still interesting to consider extensions of all the (sub-)optimal sparsity-exploiting MUD algorithms to an under-determined CDMA system with $N < K$, where the observation matrix $\mathbf{H}$ becomes fat. Consider first the two types of relaxed S-MAP detectors. The RD-MUD in (204) clearly works when $N < K$, since the $2\lambda\mathbf{I}$ term inside the inversion renders the overall matrix full rank. However, since the RD is a linear detector, it is expected to lose identifiability in the under-determined case, similar to the MMSE detectors for the sparsity-agnostic MUD schemes. Similarly, the LD problem (211) is also solvable for a fat $\mathbf{H}$ matrix, as the Lasso problem in CS. However, neither of them accounts for the augmented finite-alphabet constraint present in the original S-MAP problem (201).

The S-MAP detectors with lattice search are challenging to implement when $N < K$. The main obstacle is the QR decomposition of the fat matrix $\mathbf{H}$, which yields the upper triangular matrix $\mathbf{R}$ of the same dimension $N \times K$. Instead of a single unknown symbol, now the sparse DDD must optimize over the last $(N - K + 1)$ symbols in $\mathbf{b}$. However, apart from exhaustive search there is no low-complexity method to solve the aforementioned problem involving $(N - K + 1)$ variables, because sub-optimal alternatives introduce severe error propagation.

The same problem appears also with the SSD. To tackle the under-determined case, the generalized SD in [80] fixes the last $(N - K)$ symbols of $\mathbf{b}$ and relies on the standard SSD to detect the remaining K symbols that minimize a cost similar to the one in (217). Repeating this search for every choice of the last $(N - K)$ symbols, yields eventually the overall optimum vector. The complexity of the latter is exponential in $(N - K)$, regardless of the SNR. Recently, an alternative SD approach to avoid this exponential complexity has been developed for the under-determined case [75]. This algorithm takes advantage of the fact that for constant-modulus constellations the usual LS cost can be modified without affecting optimality, by adding the ℓ_2-norm $\mathbf{b}^T\mathbf{b}$ constant for every vector in the alphabet. This extra term allows one to obtain an equivalent full-rank system on which the standard SD algorithm can be applied. This efficient method can be readily extended to handle non-constant modulus constellations.

Interestingly, for our S-MAP detectors in (201) with lattice search of binary transmitted symbols, the norm term needed for regularization comes naturally from the Bernoulli prior. Specifically, with $p = 2$ the reformulated S-MAP detectors in (201) can be equivalently written as

$$\begin{aligned}\hat{\mathbf{b}}^{\text{MAP}} &= \arg \min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}[\mathbf{b}^T(\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})\mathbf{b} - 2\mathbf{y}^T\mathbf{H}\mathbf{b}] \\ &= \arg \min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\tilde{\mathbf{y}}' - \tilde{\mathbf{R}}\mathbf{b}\|_2^2 \end{aligned} \tag{224}$$

where $\tilde{\mathbf{R}}$ is the full rank $K \times K$ upper triangular matrix such that $\tilde{\mathbf{R}}^T\tilde{\mathbf{R}} = \mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I}$, and $\tilde{\mathbf{y}}' := \tilde{\mathbf{R}}^{-T}\mathbf{H}^T\mathbf{y}$. Utilizing the metric of (224), the back substitution of DDD and the lattice point search of SSD can be implemented easily. Hence, these S-MAP detectors can be readily extended to under-determined systems. In this way, all the (sub-)optimal S-MAP detectors are applicable

with less observations than unknowns in a CDMA system with low activity factor, but their SER performance will certainly be affected. In Section 9, we will provide simulated performance comparisons of the different optimal and sub-optimal S-MAP algorithms proposed, for a variable number of observations.

Group Lassoing Block Activity

The last generalization considered pertains to user (in)activity in a (quasi-)synchronous block fashion, where during a block of N_s symbol slots, user k remains (in)active independently from other users and across blocks. Concatenate the K user symbols across N_s time slots in the $K \times N_s$ matrix $\mathbf{B} := [\mathbf{b}(1) \dots \mathbf{b}(N_s)]$, where $\mathbf{b}(t)$ collects the symbols of all K users at slot t, and likewise for the receive-data matrix $\mathbf{Y}$ as well as the noise matrix $\mathbf{W}$, both of size $N \times N_s$. With these definitions, the counterpart of (195) for this block model is $\mathbf{Y} = \mathbf{HB} + \mathbf{W}$. Letting the $N_s \times 1$ vector $\breve{\mathbf{b}}_k := [b_k(1), \dots, b_k(N_s)]^T$ collect the N_s symbols of user k, it is useful to consider it drawn from an augmented (due to possible inactivity) block alphabet $\mathcal{A}_{a,N_s} := \mathcal{A}^{N_s} \bigcup \{\mathbf{0}_{N_s}\}$. Assuming again binary transmissions, the S-MAP *block* detector will now yield

$$\begin{aligned}\hat{\mathbf{B}}^{\text{MAP}} &= \arg \min_{\breve{\mathbf{b}}_k \in \mathcal{A}_{a,N_s}} \frac{1}{2}\|\mathbf{Y} - \mathbf{HB}\|_F^2 + \sum_{k=1}^{K} \frac{\lambda_b}{\sqrt{N_s}} \|\breve{\mathbf{b}}_k\|_0 \\ &= \arg \min_{\breve{\mathbf{b}}_k \in \mathcal{A}_{a,N_s}} \frac{1}{2}\|\mathbf{Y} - \mathbf{HB}\|_F^2 + \sum_{k=1}^{K} \lambda_b \|\breve{\mathbf{b}}_k\|_2 \end{aligned} \tag{225}$$

where $\lambda_b := \frac{1}{\sqrt{N_s}} \ln\left(\frac{1-p_a}{p_a/2^{N_s}}\right)$.

Similar to (201), the convex reformulation of the cost in (225) will lead to what is referred to in statistics as *Group Lasso* [315], which effects group sparsity on a block of symbols ($\breve{\mathbf{b}}_k$ in our case). This Group-Lasso based formulation is particularly handy for under-determined CDMA systems. In fact, the unconstrained version of (225) can be solved first to unveil the nonzero rows (i.e., the support) of $\mathbf{B}$, with improved reliability as N_s increases. Subsequently, standard sparsity-agnostic MUD schemes can be run on the estimated set of active users. Note that such a two-step approach works in the

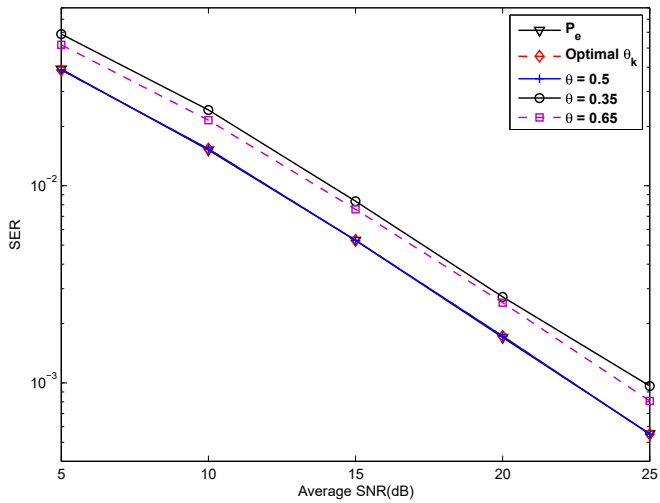


Figure 47: SER vs. average SNR (in dB) for RD-MUD with $N = 32$ and $K = 20$ and different quantization threshold θ's.

under-determined case, and also reduces number of symbols to be detected per time slot.

Simulations

We simulate a CDMA system with $K = 20$ users, each with activity factor $p_a = 0.3$. Random sequences with length $N = 32$ are used for spreading. We consider non-dispersive independent Rayleigh fading channels between AP and users, where the channel gain g_k of the k-th user is Rayleigh distributed with variance $E[g_k^2] = \sigma^2$. Thus, the average system SNR is set to be σ^2 since the AWGN $\mathbf{w}$ has unit variance.

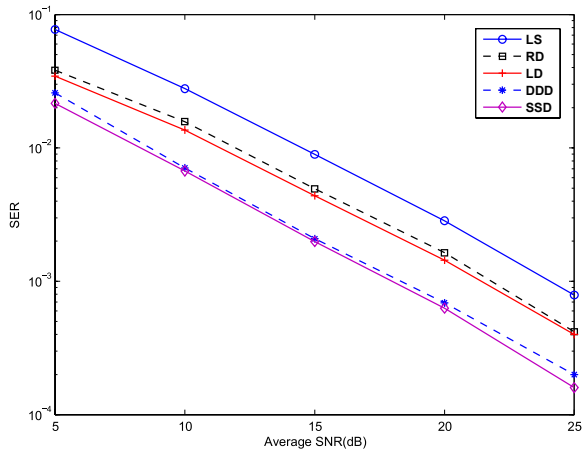
Test Case 1 (Quantization thresholds for RD). First, we test the RD scheme with different quantization thresholds θ in (203). The optimum threshold for the k-th symbol is obtained as in (209) per channel realization $\mathbf{H}$. The resulting SER is compared for four choices of θ: $0.5,\ 0.35,\ 0.65$, and $\hat{\theta}_k$. The theoretical minimum SER $\hat{P}_{e,k}^{\mathrm{RD}}$ in (210) using the optimum θ is also added for comparison. Fig. 47 shows that the SER curve with $\theta = 0.5$ comes very close to the one of the optimal $\hat{\theta}_k$, and thus constitutes a near-optimal choice

in practice. Moreover, those two curves also coincide with the analytical SER formulation corresponding to $\hat{P}_{e,k}^{\mathrm{RD}}$, thus corroborating the closed-form expression in (210).

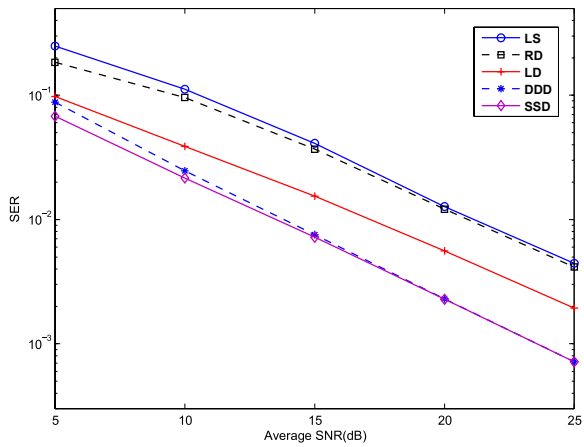
Test Case 2 (S-MAP MUD algorithms). Next, the RD, LD, DDD, and SSD MUD algorithms are all tested for both BPSK and 4-PAM constellations, and their SER performance is compared. For LD, the quadratic program in (211) is solved using the general-purpose SeDuMi toolbox [271]. The quantization rule chooses the nearest point in $\mathcal{A}_a$ for both RD and LD. For comparison, we also include the ordinary LS detector, which corresponds to the RD solution in (204) with $\lambda = 0$.

Fig. 48(a) shows that the LS detector exhibits the worst performance. This is intuitive since it neither exploits the finite alphabet nor the sparsity present in $\mathbf{b}$. The SSD exhibits the best performance at the price of highest complexity. The LD outperforms the RD algorithm, as predicted. It is interesting to observe that even at low SNR region the DDD algorithm is surprisingly competitive, especially in view of its low complexity that grows only linearly in the number of symbols K. The diversity orders for those detectors are basically the same. This is reasonable since independent Rayleigh fading channels between AP and users were simulated here. The corresponding curves for 4-PAM depicted in Fig. 48(b) follow the same trend. However, compared to Fig. 48(a), the RD algorithm degrades noticeably as its SER approaches the LS one. This is because choices of quantization thresholds become more influential as the constellation size increases. As expected, the LD exhibits resilience to this influence. The DDD and SSD algorithms have almost identical performance in high-SNR region.

Test Case 3 ((In)activity across symbols). In this case, the user (in)activity is correlated across time slots. We model this random (in)activity process as a two-state (active-inactive) stationary Markov chain. The state transition matrix is $\mathbf{P} = [a\ (1-a);\ b\ (1-b)]$, where a is uniformly distributed over $[0.8\ 0.85]$, and b over $[0.05\ 0.1]$, for each user. For this model, the expected number of successive active slots is $1/(1-a)$, and $1/b$ for the inactive ones. Also, the limiting probability for the "active" state becomes $b/(1-a+b)$, taking values from the interval $[0.2\ 0.4]$. Note that the activity factor over time is still quite low. We use the RLS approach to estimate $\hat{p}_{a,k}(t)$ as in (223) using $\beta = 0.5$, and test both the DDD and SSD algorithms in solving the resultant S-MAP



(a)



(b)

Figure 48: SER vs. average SNR (in dB) of sparsity-exploiting MUD algorithms with $N = 32$ and $K = 20$ for (a) BPSK, and (b) 4-PAM alphabets.

detection problem. The empirical SER is plotted in Fig. 49 across time for different SNR values. Clearly, the proposed scheme is effective in tracking

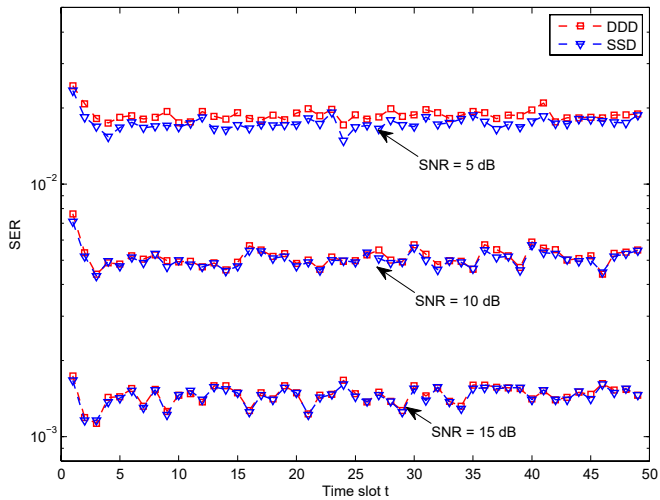


Figure 49: SER vs. time t of sparsity-exploiting MUD algorithms with RLS estimation of the activity factors.

the evolving time-correlated activity. For the same SNR value, it yields SER performance similar to the independent case in Fig. 48(a).

Test Case 4 (Under-determined CDMA systems). We also test the S-MAP MUD algorithms for under-determined systems, by varying N from 32 to 16 and 8. The results are depicted in Fig. 50. Since the RD is a simple linear detector, it is expected that once $N < K$, it will lose identifiability, and exhibits a considerably flat SER curve. At the same time, DDD still enjoys almost full diversity with a moderate choice of $N = 16$. Being the optimum detector, the SSD collects the full diversity even if $N = 8$; however, the other two kinds of detectors exhibit flat SER curves, as expected.

The Group Lasso scheme for recovering block activity is also included for the under-determined case. Fig. 51 illustrates the activity recovery error rate for different values of N and N_s. The number of observations N affects the diversity order, while the block size N_s influences the recovery accuracy.

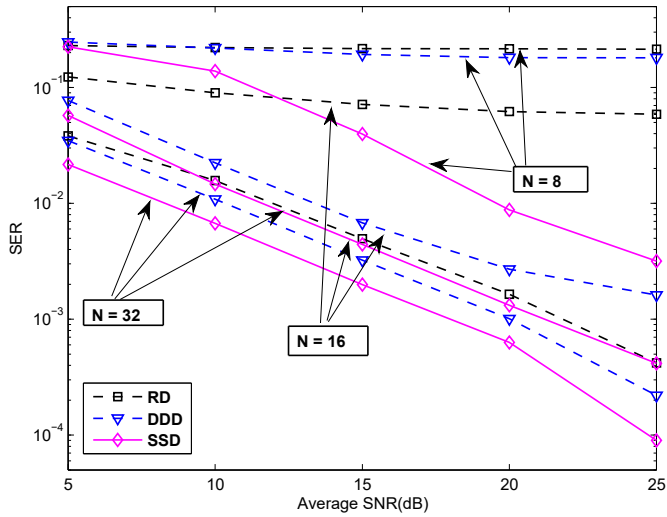


Figure 50: SER vs. average SNR (in dB) of sparsity-exploiting MUD algorithms with $N = 32, 16,$ or 8 and $K = 20$.

Conclusions

The MUD problem of sparse symbol vectors emerging with CDMA systems having low-activity factor was considered. Viewing user inactivity as augmenting the underlying alphabet, the *a priori* available sparsity information was exploited in the optimal S-MAP detector. The exponential complexity associated with the S-MAP detector was reduced by resorting to (sub-) optimal algorithms. Relaxed S-MAP detectors (RD and LD) come with low complexity but sacrifice optimality, because they ignore the finite-alphabet constraint. The second kind of detectors (DDD and SSD), searches over a subset of the alphabet and exhibits improved performance at increased complexity. The performance was analyzed for the RD and DDD algorithms, and closed-form expressions were derived for the corresponding symbol error rates.

S-MAP detectors were further generalized to deal with correlated user (in)activity across symbols by recursively estimating each user's activity factor online; and also with under-determined (a.k.a. over-saturated CDMA) settings emerging when the spreading gain is smaller than the potential number of

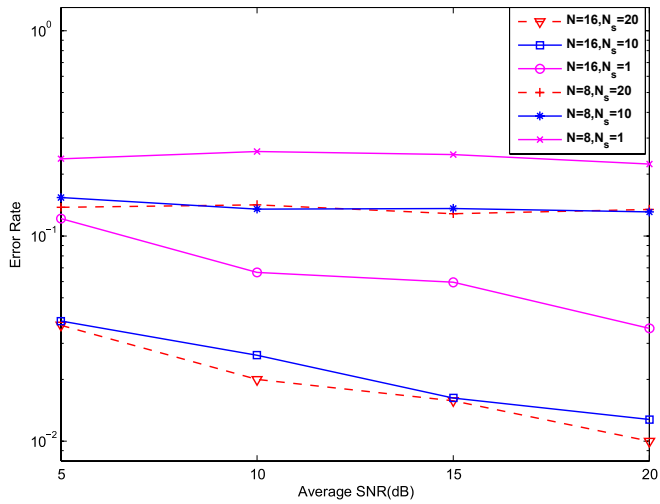


Figure 51: Error rate for block activity vs. average SNR (in dB) of Group Lasso algorithm with $N = 16, 8$ and $N_s = 20, 10, 1$.

users. Coping with the latter becomes possible through regularization with the ℓ_0 norm prior, or, with the Group Lasso-based recovery of the active user set. The numerical tests corroborated our analytical findings and quantified the relative performance of the developed sparsity-exploiting MUD algorithms.

10

Summary

As it was manifested, sparsity is present – either originally or after projecting the signal onto an appropriate basis – in a plethora of natural signals and practical communication systems. Accordingly, this monograph reviewed various CS frameworks where the signal sparsity is used as the key attribute to ameliorate the signal reconstruction/detection performance while conserving the radio and energy resources by reducing the sampling rate, transmission rate, and computational load. In the first part, advanced CS techniques based on least-squares and Lasso were presented for sparsity-aware recursive real-time estimation and signal reconstruction in a perturbed sensing setup. The second part was dedicated to CS based environmental monitoring and lossy compression techniques in wireless sensor networks. The third part addressed spectrum sensing in cognitive radio communications and multi-user detection in wireless networks. The numerical experiments of each framework demonstrated that exploiting an inherent sparsity of the underlying signal provides significant system performance gains. Thus, the principles of the presented sparse signal processing techniques are worth taking into consideration when striving for high performance in an application involving signal sparsity.

Acknowledgements

The work of M. Leinonen and M. Codreanu has been financially supported in part by Infotech Oulu, the Academy of Finland (grant 323698), and Academy of Finland 6Genesis Flagship (grant 318927). M. Codreanu would like to acknowledge the support of the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 793402 (COMPRESS NETS). G. B. Giannakis' work was supported in part by NSF grants NSF 1508993, 1514056, 1711471, and 1901134.

Appendices

A

Proof of Proposition 4

Define the vector $\boldsymbol{\chi}_k := \widehat{\mathbf{x}}_N^{\mathrm{OCD}}$ for $N = kP$, that is, the one containing the iterates at the end of the kth cycle when each variable has been updated k times. The proof that $\boldsymbol{\chi}_k$ converges to $\mathbf{x}_o$ as $k \to \infty$ will proceed in five stages. In the first one, Algorithm 1 is put in the form of a noisy vector-matrix difference equation. The second and third stages prove that the corresponding discrete-time dynamical system is exponentially stable, and that the sequence $\{\boldsymbol{\chi}_k\}_{k=1}^{\infty}$ is bounded. In the fourth stage, convergence to a limit point $\boldsymbol{\chi}_\infty$ is proved. The proof concludes by showing that $\boldsymbol{\chi}_\infty = \mathbf{x}_o$.

Dynamical System

Let $\bar{\mathbf{R}}_k$ denote the matrix with entries $\bar{R}_k(p,q) := R_{kP+p}(p,q)/(kP+p)$, and $\bar{\mathbf{r}}_k$ the vector with entries $\bar{r}_k(p) := r_{kP+p}(p)/(kP+p)$. Conditions (r1) and (r2) guarantee that $\bar{\mathbf{R}}_k \underset{k\to\infty}{\longrightarrow} \mathbf{R}_\infty$ and $\bar{\mathbf{r}}_k \underset{k\to\infty}{\longrightarrow} \mathbf{r}_\infty$ w.p. 1. Consider the decomposition $\bar{\mathbf{R}}_k = \bar{\mathbf{D}}_k + \bar{\mathbf{L}}_k + \bar{\mathbf{U}}_k$ where $\bar{\mathbf{D}}_k$ is diagonal, and $\bar{\mathbf{L}}_k$ ($\bar{\mathbf{U}}_k$) is strictly lower (upper) triangular. Observe that $\bar{\mathbf{L}}_k \neq \bar{\mathbf{U}}_k^T$.
Dividing the cost function of the problem in (17) by $kP+p$ yields

$$\chi_{k+1}(p) = \arg\min_{\chi} \left[\frac{1}{2} \left(\frac{R_{kP+p}(p,p)}{kP+p} \right) \chi^2 - \left(\frac{r_{kP+p,p}}{kP+p} \right) \chi + \left(\frac{\lambda}{kP+p} \right) |\chi| \right]. \tag{226}$$

The solution of this scalar minimization problem can be obtained in two steps. First, solve the differentiable linear-quadratic part of (226) using the auxiliary vector $\mathbf{z}_{k+1}$ to obtain

$$z_{k+1}(p) = \arg\min_{z} \left[\frac{1}{2} \left(\frac{R_{kP+p}(p,p)}{kP+p} \right) z^2 - \left(\frac{r_{kP+p}(p)}{kP+p} - \sum_{q<p} \frac{R_{kP+p}(p,q)}{kP+p} \chi_{k+1}(q) - \sum_{q>p} \frac{R_{kP+p}(p,q)}{kP+p} \chi_k(q) \right) z \right] \tag{227}$$

where $r_{kP+p,p}$ was expanded according to its definition in (21) and divided in two sums: the one already updated in the cycle $k+1$ ($q<p$), and the second one updated in the kth cycle ($q>p$). The second step to solve (226) is to pass $z_{k+1}(p)$ through the soft-threshold operator

$$\chi_{k+1}(p) = \operatorname{sgn}\left(z_{k+1}(p)\right)\left[|z_{k+1}(p)| - \frac{\lambda_{kP+p}}{kP+p}\right]_+. \tag{228}$$

Using the decomposition of $\bar{\mathbf{R}}_k$, (227) can be rewritten as

$$z_{k+1}(p) = \arg\min_z \frac{1}{2}\bar{D}_k(p,p)\, z^2 - \left(\bar{r}_k(p) - \sum_{q<p}\bar{L}_k(p,q)\chi_{k+1}(q) - \sum_{q>p}\bar{U}_k(p,q)\chi_k(q)\right) z$$

whose solution is obtained (after equating the derivative to zero) as

$$\bar{D}_k(p,p)\, z_{k+1}(p) = \bar{r}_k(p) - \sum_{q<p}\bar{L}_k(p,q)\chi_{k+1}(q) - \sum_{q>p}\bar{U}_k(p,q)\chi_k(q). \tag{229}$$

Concatenating the latter with $p=1,\ldots,P$ yields the matrix-vector difference equation

$$\bar{\mathbf{D}}_k \boldsymbol{z}_{k+1} = \bar{\mathbf{r}}_k - \bar{\mathbf{L}}_k\boldsymbol{\chi}_{k+1} - \bar{\mathbf{U}}_k\boldsymbol{\chi}_k. \tag{230}$$

The soft-thresholding operation in (228) can be accounted for by defining the error vector $\boldsymbol{\epsilon}_k := \boldsymbol{\chi}_{k+1} - \mathbf{z}_{k+1}$ in which case (230) can be re-written as

$$\bar{\mathbf{D}}_k(\boldsymbol{\chi}_{k+1} - \boldsymbol{\epsilon}_k) = \bar{\mathbf{r}}_k - \bar{\mathbf{L}}_k\boldsymbol{\chi}_{k+1} - \bar{\mathbf{U}}_k\boldsymbol{\chi}_k. \tag{231}$$

Assuming that there exists a $k^\star$ such that $\bar{\mathbf{D}}_k + \bar{\mathbf{L}}_k$ is invertible for each $k > k^\star$, (231) can be written as

$$\boldsymbol{\chi}_{k+1} = \bar{\mathbf{G}}_k\boldsymbol{\chi}_k + (\bar{\mathbf{D}}_k + \bar{\mathbf{L}}_k)^{-1}\bar{\mathbf{r}}_k + (\bar{\mathbf{D}}_k + \bar{\mathbf{L}}_k)^{-1}\bar{\mathbf{D}}_k\boldsymbol{\epsilon}_k \tag{232}$$

with $\bar{\mathbf{G}}_k := (\bar{\mathbf{D}}_k + \bar{\mathbf{L}}_k)^{-1}\bar{\mathbf{U}}_k$. The key point to be used subsequently is that (228) guarantees that the entries of $\boldsymbol{\epsilon}_k$ are bounded by a vanishing sequence. Specifically, it holds that

$$|\epsilon_k(p)| \le \frac{\lambda_{kP+p}}{kP+p}, \text{ for } p=1,\ldots,P \tag{233}$$

since the input-output variables of the soft-threshold operator $\chi = \operatorname{sgn}(\mathrm{z})[|\mathrm{z}| - \lambda]_+$ obey $|\chi - z| \le \lambda$.

Exponential Stability

Let $\bar{\mathbf{G}}(l:k) := \prod_{i=l}^{k} \bar{\mathbf{G}}_i$ in (232) denote the product of the transition matrices $\bar{\mathbf{G}}_k$. The goal of this stage is to prove that $||\bar{\mathbf{G}}(l:k)|| \leq c\rho^{k-l+1}$, with $\rho < 1$. The convergence of $\bar{\mathbf{R}}_k$ to $\mathbf{R}_\infty$ implies convergence of $\bar{\mathbf{G}}_k$ to $\mathbf{G}_\infty := -(\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}\mathbf{U}_\infty$, where $\mathbf{D}_\infty$, $\mathbf{L}_\infty$ and $\mathbf{U}_\infty$ are the diagonal, lower triangular, and upper triangular parts of $\mathbf{R}_\infty$, respectively. Since $\mathbf{R}_\infty$ is positive definite, the spectral radius of $\mathbf{G}_\infty$ is strictly less than one, i.e., $\varrho(G_\infty) < 1$ [127, p. 512]. Furthermore, for every $\delta > 0$ there exists a $c(\delta)$ constant w.r.t. k, such that $||\mathbf{G}_\infty^k|| < c(\delta)[\varrho(\mathbf{G}_\infty) + \delta]^k$ [127, p. 336]. Then, by selecting $\delta = (1 - \varrho(\mathbf{G}_\infty))/2$, and defining $\rho_\infty := (1 + \varrho(\mathbf{G}_\infty))/2$, it holds that

$$||\mathbf{G}_\infty^k|| < c\rho_\infty^k, \qquad \text{with } \rho_\infty < 1. \tag{234}$$

Upon defining $\tilde{\mathbf{G}}_k := \bar{\mathbf{G}}_k - \mathbf{G}_\infty$, the following recursion is obtained

$$\begin{aligned}
\bar{\mathbf{G}}(1:k) &= \bar{\mathbf{G}}_k\bar{\mathbf{G}}(1:k-1) \\
&= \mathbf{G}_\infty\bar{\mathbf{G}}(1:k-1) + \tilde{\mathbf{G}}_k\bar{\mathbf{G}}(1:k-1) \\
&= \mathbf{G}_\infty^k + \sum_{i=1}^{k} \mathbf{G}_\infty^{k-i}\tilde{\mathbf{G}}_i\bar{\mathbf{G}}(1:i-1), \quad \bar{\mathbf{G}}(1:0) := \mathbf{I}.
\end{aligned}$$

Using (234), the latter can be bounded as

$$||\bar{\mathbf{G}}(1:k)|| \leq c\rho_\infty^k + c\sum_{i=1}^{k} \rho_\infty^{k-i}||\tilde{\mathbf{G}}_i||||\bar{\mathbf{G}}(1:i-1)||$$

which after multiplying both sides by ρ_∞^{-k} yields

$$\rho_\infty^{-k}||\bar{\mathbf{G}}(1:k)|| \leq c + \sum_{i=1}^{k} c\rho_\infty^{-1}||\tilde{\mathbf{G}}_i||||\bar{\mathbf{G}}(1:i-1)||\rho_\infty^{-(i-1)} \tag{235}$$

and allows one to apply the discrete Bellman-Gronwall lemma (see e.g., [269, p. 315]).

Lemma 5 (Bellman-Gronwall)**.** *If $c, \xi_k, h_k \geq 0$ $\forall k$ satisfy the recursive inequality*

$$\xi_k \leq c + \sum_{i=1}^{k} h_{i-1}\xi_{i-1} \tag{236}$$

then ξ_k obeys the non-recursive inequality

$$\xi_k \leq c \prod_{i=1}^{k} (1 + h_{i-1}). \tag{237}$$

For $\xi_k = \rho_\infty^{-k} ||\bar{\mathbf{G}}(1:k)||$ and $h_k = c\rho_\infty^{-1} ||\tilde{\mathbf{G}}_{k+1}||$, (235) takes the form of (236), so that (237) holds and (after multiplying both sides by ρ_∞^k) results in

$$||\bar{\mathbf{G}}(1:k)|| \leq \rho_\infty^k c \prod_{i=1}^{k} \left(1 + c\rho_\infty^{-1} ||\tilde{\mathbf{G}}_i||\right) = c \prod_{i=1}^{k} \left(\rho_\infty + c||\tilde{\mathbf{G}}_i||\right). \tag{238}$$

Raising both sides of (238) to the power of $1/k$ and applying the geometric-arithmetic mean inequality, it follows that

$$||\bar{\mathbf{G}}(1:k)||^{1/k} \leq c^{1/k} \frac{1}{k} \sum_{i=1}^{k} \left(\rho_\infty + c||\tilde{\mathbf{G}}_i||\right)$$

which is readily rewritten as

$$||\bar{\mathbf{G}}(1:k)|| \leq c \left(\rho_\infty + c\frac{1}{k} \sum_{i=1}^{k} ||\tilde{\mathbf{G}}_i||\right)^k.$$

Since $\bar{\mathbf{G}}_k \underset{k\to\infty}{\longrightarrow} \mathbf{G}_\infty$ and $||\tilde{\mathbf{G}}_k|| \underset{k\to\infty}{\longrightarrow} \mathbf{0}$ w.p. 1, for every $\delta > 0$ there exists an integer k_0 such that if $k \geq k_0$, then $\frac{1}{k}\sum_{i=1}^{k} ||\tilde{\mathbf{G}}_i|| \leq \delta$ w.p 1. Thus, if δ is selected as $(1 - \rho_\infty)/(2c)$, and ρ as $(1 + \rho_\infty)/2$, the following bound is obtained

$$||\bar{\mathbf{G}}(1:k)|| \leq c\rho^k, \qquad \rho < 1,\ k \geq k_0, \qquad \text{w.p. 1.}$$

It is clear, by inspection, that the proof so far carries over even if the product of transition matrices starts at $l > 1$; that is

$$||\bar{\mathbf{G}}(l:k)|| \leq c\rho^{k-l+1}, \qquad \rho < 1,\ k \geq k_0, \qquad \text{w.p. 1.} \tag{239}$$

Certainly, c, ρ, and k_0 do not depend on l. However, k_0 does depend on the realization of the random sequence $\bar{\mathbf{G}}_k$, and its existence and finiteness are guaranteed w.p. 1.

Boundedness

Define $\mathbf{v}_k := (\bar{\mathbf{D}}_k + \bar{\mathbf{L}}_k)^{-1}(\bar{\mathbf{r}}_k + \bar{\mathbf{D}}_k\boldsymbol{\epsilon}_k)$, and rewrite (232) as

$$\boldsymbol{\chi}_{k+1} = \bar{\mathbf{G}}_k\boldsymbol{\chi}_k + \mathbf{v}_k.$$

Using back substitution, $\boldsymbol{\chi}_{k+1}$ can then be expressed as

$$\boldsymbol{\chi}_{k+1} = \bar{\mathbf{G}}(k_0 : k)\boldsymbol{\chi}_{k_0} + \sum_{i=1}^{k-k_0} \bar{\mathbf{G}}(k-i+1 : k)\mathbf{v}_{k-i} + \mathbf{v}_k.$$

Since $\bar{\mathbf{D}}_k$, $\bar{\mathbf{L}}_k$, $\bar{\mathbf{r}}_k$, and $\boldsymbol{\epsilon}_k$ converge, the random sequence $\mathbf{v}_k$ converges too w.p. 1; hence, it can be stochastically bounded by a random variable v; that is, $||\mathbf{v}_k|| < v,\ \forall k$, w.p. 1. This, combined with the exponential stability ensured by (239), guarantees that the realizations of the random sequence $\boldsymbol{\chi}_k$ are (stochastically) bounded; thus

$$\begin{aligned} ||\boldsymbol{\chi}_{k+1}|| &\leq c\rho^{-k+k_0-1}||\boldsymbol{\chi}_{k_0}|| + \sum_{i=1}^{k-k_0} c\rho^{-i}v + v \\ &\leq c||\boldsymbol{\chi}_{k_0}|| + cv\left(\frac{1}{1-\rho}+1\right), \quad \text{w.p. 1.} \end{aligned} \tag{240}$$

Convergence

Define the error $\tilde{\mathbf{D}}_k := \bar{\mathbf{D}}_k - \mathbf{D}_\infty$, and similarly $\tilde{\mathbf{L}}_k := \bar{\mathbf{L}}_k - \mathbf{L}_\infty$, $\tilde{\mathbf{U}}_k := \bar{\mathbf{U}}_k - \mathbf{U}_\infty$, and $\tilde{\mathbf{r}}_k := \bar{\mathbf{r}}_k - \mathbf{r}_\infty$. Using these new variables, (231) can be rewritten in error form as

$$(\mathbf{D}_\infty + \tilde{\mathbf{D}}_k)(\boldsymbol{\chi}_{k+1} - \boldsymbol{\epsilon}_k) = (\mathbf{r}_\infty + \tilde{\mathbf{r}}_k) - (\mathbf{L}_\infty + \tilde{\mathbf{L}}_k)\boldsymbol{\chi}_{k+1} - (\mathbf{U}_\infty + \tilde{\mathbf{U}}_k)\boldsymbol{\chi}_k \tag{241}$$

and, after regrouping terms, as

$$\begin{aligned} \boldsymbol{\chi}_{k+1} &= -(\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}\mathbf{U}_\infty\boldsymbol{\chi}_k + (\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1} \\ &\times\Big(\mathbf{r}_\infty + \tilde{\mathbf{r}}_k - (\tilde{\mathbf{D}}_k + \tilde{\mathbf{L}}_k)\boldsymbol{\chi}_{k+1} + (\mathbf{D}_\infty + \tilde{\mathbf{D}}_k)\boldsymbol{\epsilon}_k - \tilde{\mathbf{U}}_k\boldsymbol{\chi}_k\Big). \end{aligned} \tag{242}$$

Equation (242) describes an exponentially stable linear time-invariant system with transition matrix $-(\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}\mathbf{U}_\infty$, and input $\mathbf{u}_k := (\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}[\mathbf{r}_\infty + \tilde{\mathbf{r}}_k - (\tilde{\mathbf{D}}_k + \tilde{\mathbf{L}}_k)\boldsymbol{\chi}_{k+1} + (\mathbf{D}_\infty + \tilde{\mathbf{D}}_k)\boldsymbol{\epsilon}_k - \tilde{\mathbf{U}}_k\boldsymbol{\chi}_k]$. The input can be divided into its limiting point $\mathbf{u}_\infty := (\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}\mathbf{r}_\infty$, and the error

$\tilde{\mathbf{u}}_k := (\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}[\tilde{\mathbf{r}}_k - (\tilde{\mathbf{D}}_k + \tilde{\mathbf{L}}_k)\boldsymbol{\chi}_{k+1} + (\mathbf{D}_\infty + \tilde{\mathbf{D}}_k)\boldsymbol{\epsilon}_k - \tilde{\mathbf{U}}_k\boldsymbol{\chi}_k]$. As $k \to \infty$, the vector $\tilde{\mathbf{u}}_k$ goes to zero almost surely because the sequence $\boldsymbol{\chi}_k$ is bounded, and the error $\tilde{\mathbf{r}}_k$, $\tilde{\mathbf{D}}_k$, $\tilde{\mathbf{L}}_k$ and $\tilde{\mathbf{U}}_k$ as well as $\boldsymbol{\epsilon}_k$, all go to zero w.p. 1. With this notation and recalling the definition $\mathbf{G}_\infty := -(\mathbf{D}_\infty + \mathbf{L}_\infty)^{-1}\mathbf{U}_\infty$, (242) is rewritten as

$$\boldsymbol{\chi}_{k+1} = \mathbf{G}_\infty\boldsymbol{\chi}_k + \mathbf{u}_\infty + \tilde{\mathbf{u}}_k$$

and back-substituting again the new expression for $\boldsymbol{\chi}_{k+1}$ yields

$$\boldsymbol{\chi}_{k+1} = \mathbf{G}_\infty^k\boldsymbol{\chi}_1 + \sum_{i=1}^{k}\mathbf{G}_\infty^{k-i}\mathbf{u}_\infty + \sum_{i=1}^{k}\mathbf{G}_\infty^{k-i}\tilde{\mathbf{u}}_i. \tag{243}$$

Convergence of this recursion will be established by showing that the first and third terms in the right-hand side vanish as $k \to \infty$, while the surviving one corresponds to a stable geometric series. Given that $\exists\ c > 0, \rho_\infty < 1$ such that $\forall n\ ||\boldsymbol{G}_\infty^n|| \leq c\rho_\infty^n$, convergence of the first term to zero follows readily from (234). The third term represents the limiting output value of a multiple input-multiple output stable linear time-invariant system with vanishing input; that is $\lim_{i\to\infty}\tilde{\boldsymbol{u}}_i = 0$. As $\forall k\ ||\boldsymbol{G}_\infty^k|| \leq c\rho_\infty^k$, (234) implies that it is possible to bound the sum under consideration as

$$\left\|\sum_{i=1}^{k}\mathbf{G}_\infty^{k-i}\tilde{\mathbf{u}}_i\right\| \leq c\sum_{i=1}^{k}\rho_\infty^{k-i}||\tilde{\mathbf{u}}_i|| \ . \tag{244}$$

Since $\lim_{i\to\infty}\tilde{\mathbf{u}}_i = 0$, it holds by the definition of the limit that for any $\epsilon > 0\ \exists\ N \in \mathbb{N}$ so that $||\tilde{\mathbf{u}}_i|| \leq \epsilon, \forall\ i \geq N$. Using the latter along with (244), it follows that for $k \geq N$

$$\begin{aligned}\left\|\sum_{i=1}^{k}\mathbf{G}_\infty^{k-i}\tilde{\mathbf{u}}_i\right\| &\leq c\sum_{i=1}^{N-1}\rho_\infty^{k-i}||\tilde{\mathbf{u}}_i|| + c\epsilon\sum_{i=N}^{k}\rho_\infty^{k-i} \\ &= c\rho_\infty^{k-N}\sum_{i=1}^{N-1}\rho_\infty^{N-i}||\tilde{\mathbf{u}}_i|| + c\epsilon\sum_{i=N}^{k}\rho_\infty^{k-i} \ . \end{aligned} \tag{245}$$

Because $\sum_{i=1}^{N-1}\rho_\infty^{N-i}||\tilde{\mathbf{u}}_i||$ does not depend on k, the limit of the first summand in (245) goes to zero; hence,

$$\lim_{k\to\infty}\left\|\sum_{i=1}^{k}\mathbf{G}_\infty^{k-i}\tilde{\mathbf{u}}_i\right\| \leq c\epsilon/(1-\rho_\infty) \ .$$

The last inequality holds $\forall \epsilon > 0$; thus,

$$\lim_{k\to\infty} \left\| \sum_{i=1}^{k} \mathbf{G}_{\infty}^{k-i} \tilde{\mathbf{u}}_i \right\| = 0$$

which establishes convergence to zero of the third sum in the right-hand side of (243).

Limit Point

Once convergence is established, it is possible to take the limit as $k \to \infty$ in (241) to obtain

$$\mathbf{D}_{\infty}(\boldsymbol{\chi}_{\infty} - \boldsymbol{\epsilon}_{\infty}) = \mathbf{r}_{\infty} - \mathbf{L}_{\infty}\boldsymbol{\chi}_{\infty} - \mathbf{U}_{\infty}\boldsymbol{\chi}_{\infty}, \qquad \text{w.p. 1.} \tag{246}$$

Recalling that $||\boldsymbol{\epsilon}_{\infty}|| \leq \lim_{N\to\infty} \frac{\lambda_N}{N} = 0$, (246) reduces to

$$(\mathbf{D}_{\infty} + \mathbf{L}_{\infty} + \mathbf{U}_{\infty})\boldsymbol{\chi}_{\infty} = \mathbf{r}_{\infty}, \qquad \text{w.p. 1} \tag{247}$$

and since $\mathbf{D}_{\infty} + \mathbf{L}_{\infty} + \mathbf{U}_{\infty} = \mathbf{R}_{\infty}$, it holds that

$$\boldsymbol{\chi}_{\infty} = (\mathbf{R}_{\infty})^{-1}\mathbf{r}_{\infty} = \mathbf{x}_o, \qquad \text{w.p. 1} \tag{248}$$

which concludes the proof.

B

Construction of Matrix Ψ' in (76)

The matrix $\boldsymbol{\Psi}'$ in (76) can be expressed as

$$\begin{aligned}\boldsymbol{\Psi}' &= \boldsymbol{\Psi}'_{\mathrm{T}} \otimes \boldsymbol{\Psi}_{\mathrm{S}} \\ &\overset{(a)}{=} (\mathbf{I}'_{\mathrm{T}} \boldsymbol{\Psi}_{\mathrm{T}}) \otimes \boldsymbol{\Psi}_{\mathrm{S}} \\ &\overset{(b)}{=} (\mathbf{I}'_{\mathrm{T}} \boldsymbol{\Psi}_{\mathrm{T}}) \otimes (\mathbf{I}_N \boldsymbol{\Psi}_{\mathrm{S}}) \\ &\overset{(c)}{=} (\mathbf{I}'_{\mathrm{T}} \otimes \mathbf{I}_N)(\boldsymbol{\Psi}_{\mathrm{T}} \otimes \boldsymbol{\Psi}_{\mathrm{S}}) \\ &\overset{(d)}{=} (\mathbf{I}'_{\mathrm{T}} \otimes \mathbf{I}_N)\boldsymbol{\Psi} \\ &\overset{(e)}{=} [\boldsymbol{\psi}_1 \ \cdots \ \boldsymbol{\psi}_{N(W-1)}]^{\mathrm{T}}\end{aligned} \tag{249}$$

where (a) follows from defining a binary sparse matrix $\mathbf{I}'_{\mathrm{T}} \triangleq [\mathbf{I}_{W-1} \ \mathbf{0}_{(W-1)\times 1}] \in \mathbb{B}^{(W-1)\times W}$ that performs a row selection of $\boldsymbol{\Psi}_{\mathrm{T}}$ as $\boldsymbol{\Psi}'_{\mathrm{T}} = \mathbf{I}'_{\mathrm{T}} \boldsymbol{\Psi}_{\mathrm{T}}$; (b) follows because $\boldsymbol{\Psi}_{\mathrm{S}} = \mathbf{I}_N \boldsymbol{\Psi}_{\mathrm{S}}$; (c) follows from $\mathbf{AB} \otimes \mathbf{CD} = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D})$; (d) follows from (69); (e) follows from performing a row selection of $\boldsymbol{\Psi}$ via a binary sparse matrix $\mathbf{I}'_{\mathrm{T}} \otimes \mathbf{I}_N = [\mathbf{I}_{N(W-1)} \ \mathbf{0}_{N(W-1)\times N}] \in \mathbb{B}^{N(W-1)\times NW}$ as $(\mathbf{I}'_{\mathrm{T}} \otimes \mathbf{I}_N)\boldsymbol{\Psi} = [\boldsymbol{\psi}_1 \ \cdots \ \boldsymbol{\psi}_{N(W-1)}]^{\mathrm{T}}$, which is the desired result.

C

MMSE Orthogonality Principle for (125)

In order to show the equality $\mathbb{E}\big[\|\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s + \underline{\mathbf{Z}}_s - \underline{\hat{\mathbf{X}}}_s\|_2^2\big] = \mathbb{E}\big[\|\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s\|_2^2\big] + \mathbb{E}\big[\|\underline{\mathbf{Z}}_s - \underline{\hat{\mathbf{X}}}_s\|_2^2\big]$ in step (a) of (125), we show that the associated cross-term $\mathbb{E}\big[(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)^{\mathrm{T}}(\underline{\mathbf{Z}}_s - \underline{\hat{\mathbf{X}}}_s)\big]$ is zero. This term can be written as

$$\mathbb{E}\big[(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)^{\mathrm{T}}(\underline{\mathbf{Z}}_s - \underline{\hat{\mathbf{X}}}_s)\big] = \mathbb{E}\big[\underline{\mathbf{Z}}_s^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)\big] - \mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)\big]. \quad (250)$$

By the law of total expectation, the first term of (250) can be written as

$$\begin{aligned}
\mathbb{E}\big[\underline{\mathbf{Z}}_s^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)\big] &= \mathbb{E}\Big\{\mathbb{E}\big[\underline{\mathbf{Z}}_s^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)|\mathbf{Y}_s\big]\Big\} \\
&= \mathbb{E}\Big\{\underline{\mathbf{Z}}_s^{\mathrm{T}}(\mathbb{E}[\underline{\mathbf{X}}_s|\mathbf{Y}_s] - \underline{\mathbf{Z}}_s)\Big\} \\
&\overset{(a)}{=} \mathbb{E}\Big\{\underline{\mathbf{Z}}_s^{\mathrm{T}}(\underline{\mathbf{Z}}_s - \underline{\mathbf{Z}}_s)\Big\} \\
&= 0
\end{aligned} \quad (251)$$

where (a) follows from (123). Similarly, the second term of (250) can be written as

$$\begin{aligned}
\mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)\big] &= \mathbb{E}\Big(\mathbb{E}\Big\{\mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)|\underline{\mathbf{X}}_s, \mathbf{Y}_s\big]|\mathbf{Y}_s\Big\}\Big) \\
&= \mathbb{E}\Big(\mathbb{E}\Big\{\mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s|\underline{\mathbf{X}}_s, \mathbf{Y}_s\big]^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)|\mathbf{Y}_s\Big\}\Big) \\
&\overset{(a)}{=} \mathbb{E}\Big(\mathbb{E}\Big\{\mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s|\mathbf{Y}_s\big]^{\mathrm{T}}(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)|\mathbf{Y}_s\Big\}\Big) \\
&= \mathbb{E}\Big(\mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s|\mathbf{Y}_s\big]^{\mathrm{T}}(\mathbb{E}\{\underline{\mathbf{X}}_s|\mathbf{Y}_s\} - \underline{\mathbf{Z}}_s)\Big) \\
&\overset{(b)}{=} \mathbb{E}\Big(\mathbb{E}\big[\underline{\hat{\mathbf{X}}}_s|\mathbf{Y}_s\big]^{\mathrm{T}}(\underline{\mathbf{Z}}_s - \underline{\mathbf{Z}}_s)\Big) \\
&= 0
\end{aligned} \quad (252)$$

where (a) follows because $\underline{\mathbf{X}}_s \to \mathbf{Y}_s \to \underline{\hat{\mathbf{X}}}_s$ forms a Markov chain, and (b) follows from (123). By (251) and (252), $\mathbb{E}\big[(\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s)^{\mathrm{T}}(\underline{\mathbf{Z}}_s - \underline{\hat{\mathbf{X}}}_s)\big] = 0$.

D

MMSE Estimation Error in (126)

The average MMSE estimation error in (125) can be expressed as

$$\begin{aligned} D_{\mathbf{Z}|\mathbf{b}_s} &= N^{-1}\mathbb{E}\big[\|\underline{\mathbf{X}}_s - \underline{\mathbf{Z}}_s\|_2^2\big] \\ &\overset{(a)}{=} N^{-1}\mathbb{E}\big[\|\mathbf{X}_s - \mathbf{Z}_s\|_2^2\big] \\ &= N^{-1}\mathrm{Tr}\left\{\mathbb{E}\big[(\mathbf{X}_s - \mathbf{Z}_s)(\mathbf{X}_s - \mathbf{Z}_s)^{\mathrm{T}}\big]\right\} \\ &= N^{-1}\mathrm{Tr}\left\{\mathbb{E}\big[\mathbf{X}_s\mathbf{X}_s^{\mathrm{T}} - \mathbf{X}_s\mathbf{Z}_s^{\mathrm{T}} - \mathbf{Z}_s\mathbf{X}_s^{\mathrm{T}} + \mathbf{Z}_s\mathbf{Z}_s^{\mathrm{T}}\big]\right\} \\ &= N^{-1}\mathrm{Tr}\left\{\boldsymbol{\Sigma}_{\mathbf{X}_s} - \boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Z}_s} - \boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Z}_s}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathbf{Z}_s}\right\} \end{aligned} \tag{253}$$

where (a) follows from removing the zero parts of $\underline{\mathbf{X}}_s$ and $\underline{\mathbf{Z}}_s$ (see (121) and (124)), and the cross-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Z}_s} \in \mathbb{R}^{K\times K}$ is

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Z}_s} &= \mathbb{E}[\mathbf{X}_s(\mathbf{F}_s\mathbf{Y}_s)^{\mathrm{T}}] = \boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Y}_s}\mathbf{F}_s^{\mathrm{T}} = \boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Y}_s}\boldsymbol{\Sigma}_{\mathbf{Y}_s}^{-1}\boldsymbol{\Sigma}_{\mathbf{X}_s\mathbf{Y}_s}^{\mathrm{T}} \\ &= \boldsymbol{\Sigma}_{\mathbf{Z}_s}. \end{aligned} \tag{254}$$

Substituting (254) into (253) results in $D_{\mathbf{Z}|\mathbf{b}_s} = N^{-1}\mathrm{Tr}\big(\boldsymbol{\Sigma}_{\mathbf{X}_s} - \boldsymbol{\Sigma}_{\mathbf{Z}_s}\big)$.

E

The Proof of Theorem 1

Using the equivalence $R^{\text{rem}}_{\mathbf{X}|\mathbf{b}_s}(D_s) = R^{\text{dir}}_{\mathbf{Z}|\mathbf{b}_s}(D'_s)$ of Proposition 6, and $D'_s = D_s - D_{\mathbf{Z}|\mathbf{b}_s} \geq 0$ in (129), the conditional remote RDF in (119) can be recast as

$$R^{\text{rem}}_{\mathbf{X}|\mathbf{B}}(D) = \min_{\substack{\sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s)[D'_s + D_{\mathbf{Z}|\mathbf{b}_s}] = D \\ D'_s \geq 0,\, s=1,\ldots,\binom{N}{K}}} \sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s) R^{\text{dir}}_{\mathbf{Z}|\mathbf{b}_s}(D'_s) \tag{255}$$

with optimization variables D'_s, $s = 1, \ldots, \binom{N}{K}$. Let $D'_{s,k} \geq 0$, $k = 1, \ldots, K$, $s = 1, \ldots, \binom{N}{K}$, be auxiliary variables for (255). Inserting the new variables with constraint $\sum_{k=1} D'_{s,k} = D'_s$, and substituting $D_{\mathbf{Z}|\mathbf{B}} = \sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s) D_{\mathbf{Z}|\mathbf{b}_s}$ in (133) and the expression of $R^{\text{dir}}_{\mathbf{Z}|\mathbf{b}_s}(D'_s)$ in (131), $R^{\text{rem}}_{\mathbf{X}|\mathbf{B}}(D)$ in (255) can be equivalently expressed as

$$R^{\text{rem}}_{\mathbf{X}|\mathbf{B}}(D) = \min_{\substack{\sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s) D'_s = D - D_{\mathbf{Z}|\mathbf{B}} \\ \sum_{k=1}^{K} D'_{s,k} = D'_s,\, s=1,\ldots,\binom{N}{K} \\ D'_s \geq 0,\, s=1,\ldots,\binom{N}{K} \\ D'_{s,k} \geq 0,\, k=1,\ldots,K,\, s=1,\ldots,\binom{N}{K}}} N^{-1} \sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s) \sum_{k=1}^{K} \max\left\{0, \frac{1}{2}\log\frac{\lambda_{s,k}}{D'_{s,k}}\right\} \tag{256}$$

with optimization variables D'_s, and $D'_{s,k}$, $k = 1, \ldots, K$, $s = 1, \ldots, \binom{N}{K}$. Finally, eliminating the variables D'_s, $s = 1, \ldots, \binom{N}{K}$, by substituting the second set of equality constraints into the first one yields the expression for $R^{\text{rem}}_{\mathbf{X}|\mathbf{B}}(D)$ in (134).

Remark E.1. *A valid distortion is $D \geq D_{\mathbf{Z}|\mathbf{B}} \geq 0$ because the distortion criterion must be non-negative.*

Remark E.2. *For all distortion criteria $D \geq \frac{1}{N} \sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s) \text{Tr}(\mathbf{\Sigma}_{\mathbf{X}_s})$, $R^{\text{rem}}_{\mathbf{X}|\mathbf{B}}(D) = 0$; if the encoder sends no information (i.e., $R = 0$),*

then the decoder can set $\hat{\mathbf{X}} = \mathbf{0}_N$*, resulting in an admissible distortion because* $\mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] = \frac{1}{N}\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2] = \frac{1}{N}\mathbb{E}[\|\mathbf{X}\|_2^2] = \frac{1}{N}\sum_{s=1}^{\binom{N}{K}} p(\mathbf{b}_s)\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{X}_s}) \leq D$.

Combining the above derivations with Remarks E.1 and E.2, $R^{\mathrm{rem}}_{\mathbf{X}|\mathbf{B}}(D)$ has the characterization of Theorem 1, which concludes the proof. ■

F

The Conditional Direct RDF

The conditional direct RDF $R^{\text{dir}}_{\mathbf{X}|\mathbf{B}}(D)$ determines the minimum achievable rate R for distortion $D \geq 0$ in the compression scheme, where the encoder observes $\mathbf{X}$ directly (i.e., without CS), and $\mathbf{B}$ is available as SI at the encoder and decoder. It is defined as (cf. (118))

$$R^{\text{dir}}_{\mathbf{X}|\mathbf{B}}(D) = \min_{\{f(\hat{\mathbf{x}}|\mathbf{x},\mathbf{b}_s)\}_{s=1}^{|\mathcal{B}|}:\mathbb{E}[d(\mathbf{X},\hat{\mathbf{X}})]\leq D} \frac{1}{N} I(\mathbf{X};\hat{\mathbf{X}}|\mathbf{B}) \tag{257a}$$

where the optimization is over the conditional PDFs $f(\hat{\mathbf{x}}|\mathbf{x},\mathbf{b}_s)$, $s = 1,\ldots,|\mathcal{B}|$. The conditional mutual information between $\mathbf{X}$ and $\hat{\mathbf{X}}$ given $\mathbf{B}$ is

$$\begin{aligned} I(\mathbf{X};\hat{\mathbf{X}}|\mathbf{B}) &= \textstyle\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) I(\mathbf{X};\hat{\mathbf{X}}|\mathbf{B}=\mathbf{b}_s) \\ &= \textstyle\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) \int_{\mathbf{x}}\int_{\hat{\mathbf{x}}} f(\mathbf{x}|\mathbf{b}_s) f(\hat{\mathbf{x}}|\mathbf{x},\mathbf{b}_s) \log \frac{f(\hat{\mathbf{x}}|\mathbf{x},\mathbf{b}_s)}{f(\hat{\mathbf{x}}|\mathbf{b}_s)} \mathrm{d}\mathbf{x}\mathrm{d}\hat{\mathbf{x}} \end{aligned} \tag{257b}$$

and the average MSE distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ is

$$\begin{aligned} \mathbb{E}\big[d(\mathbf{X},\hat{\mathbf{X}})\big] &= \textstyle\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) \mathbb{E}\big[d(\mathbf{X},\hat{\mathbf{X}})|\mathbf{B}=\mathbf{b}_s\big] \\ &= \textstyle\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) \int_{\mathbf{x}}\int_{\hat{\mathbf{x}}} f(\mathbf{x}|\mathbf{b}_s) f(\hat{\mathbf{x}}|\mathbf{x},\mathbf{b}_s) d(\mathbf{x},\hat{\mathbf{x}}) \mathrm{d}\mathbf{x}\mathrm{d}\hat{\mathbf{x}}. \end{aligned} \tag{257c}$$

Following the steps analogous to those for $R^{\text{rem}}_{\mathbf{X}|\mathbf{B}}(D)$, the conditional direct RDF is given as (cf. (134))

$$R^{\text{dir}}_{\mathbf{X}|\mathbf{B}}(D) = \min_{\substack{\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) \sum_{k=1}^{K} D_{s,k} = D \\ D_{s,k}\geq 0,\, k=1,\ldots,K,\, s=1,\ldots,|\mathcal{B}|}} N^{-1} \textstyle\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) \sum_{k=1}^{K} \max\left\{0, \frac{1}{2}\log\frac{\bar{\lambda}_{s,k}}{D_{s,k}}\right\} \tag{258}$$

where $D_{s,k}$, $k = 1,\ldots,K$, $s = 1,\ldots,|\mathcal{B}|$, are the optimization variables; $\bar{\lambda}_{s,1} \geq \ldots \geq \bar{\lambda}_{s,K} > 0$ are the eigenvalues of covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}_s} = \bar{\mathbf{Q}}_s \bar{\boldsymbol{\Lambda}}_s \bar{\mathbf{Q}}_s^{\mathrm{T}}$, where the columns of $\bar{\mathbf{Q}}_s \in \mathbb{R}^{K\times K}$ are the eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{X}_s} \in \mathbb{S}^K_{++}$, and $\bar{\boldsymbol{\Lambda}}_s \triangleq \operatorname{diag}(\bar{\lambda}_{s,1},\ldots,\bar{\lambda}_{s,K})$.

G

Proof of Proposition 7

A stationary point $\bar{\mathbf{P}}, \bar{\mathbf{Q}}$ and $\bar{\mathbf{E}}$ of (P2) must satisfy the following first-order optimality conditions [34]

$$\mathbf{0}_{N_x \times N_y} \in \partial_{\mathbf{E}} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \tag{259}$$
$$\left\{ \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}}) \; + \mu\bar{\beta} \left[\operatorname{sgn}(\bar{\mathbf{E}}) + \tilde{\mathbf{E}} \right] \middle| \bar{\mathbf{E}} \odot \tilde{\mathbf{E}} = \mathbf{0}, \; \|\tilde{\mathbf{E}}\|_\infty \leq 1 \right\}$$
$$\nabla_{\mathbf{P}} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}})\bar{\mathbf{Q}} + \lambda\bar{\beta}\bar{\mathbf{P}} = \mathbf{0}_{N_x \times \rho} \tag{260}$$
$$\nabla_{\mathbf{Q}^{\mathcal{T}}} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \bar{\mathbf{P}}^{\mathcal{T}} \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}}) + \lambda\bar{\beta}\bar{\mathbf{Q}}^{\mathcal{T}} = \mathbf{0}_{\rho \times N_y} \tag{261}$$

where $\odot$ denotes the element-wise (Hadamard) product. Through post-multiplying (260) by $\bar{\mathbf{P}}^{\mathcal{T}}$ and pre-multiplying (261) by $\bar{\mathbf{Q}}$, one can see that

$$\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}}) = -\mu\bar{\beta}(\operatorname{sgn}(\bar{\mathbf{E}}) + \tilde{\mathbf{E}})$$
$$\operatorname{tr}\left(\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}})\bar{\mathbf{Q}}\bar{\mathbf{P}}^{\mathcal{T}} \right) = -\lambda\bar{\beta}\operatorname{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}}) = -\lambda\bar{\beta}\operatorname{tr}(\bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}). \tag{262}$$

Define now $\kappa(\mathbf{R}_1, \mathbf{R}_2) := \frac{1}{2}\left(\operatorname{tr}(\mathbf{R}_1) + \operatorname{tr}(\mathbf{R}_2)\right)$, and consider the following *convex* problem

$$\text{(P4)} \quad \min_{\substack{\mathbf{L}, \mathbf{E} \in \mathbb{R}^{N_x \times N_y}, \\ \mathbf{R}_1 \in \mathbb{R}^{N_x \times N_x}, \\ \mathbf{R}_2 \in \mathbb{R}^{N_y \times N_y}}} \sum_{\tau=1}^{T} \beta^{T-\tau} c^{(\tau)}(\mathbf{L}, \mathbf{E}) + \lambda\bar{\beta}\, \kappa(\mathbf{R}_1, \mathbf{R}_2) + \mu\bar{\beta}\, ||\mathbf{E}||_1$$
$$\text{subject to } \mathbf{R} := \begin{pmatrix} \mathbf{R}_1 & \mathbf{L} \\ \mathbf{L}^{\mathcal{T}} & \mathbf{R}_2 \end{pmatrix} \succeq \mathbf{0} \tag{263}$$

which is *equivalent* to (P1). Equivalence can be easily inferred by minimizing (P4) with respect to $\{\mathbf{R}_1, \mathbf{R}_2\}$ and noting an alternative characterization of the nuclear norm given by [245]

$$\|\mathbf{L}\|_* = \min_{\mathbf{R}_1, \mathbf{R}_2} \kappa(\mathbf{R}_1, \mathbf{R}_2)$$
$$\text{subject to } \mathbf{R} \succeq \mathbf{0}. \tag{264}$$

In what follows, the optimality conditions of the conic program (P4) are explored. Introducing a Lagrange multiplier matrix $\mathbf{M} \in \mathbb{R}^{(N_x+N_y)\times(N_x+N_y)}$ associated with the conic constraint in (263), the Lagrangian is first formed as

$$\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \sum_{\tau=1}^{T} \beta^{T-\tau} c^{(\tau)}(\mathbf{L},\mathbf{E}) + \lambda\bar{\beta}\,\kappa(\mathbf{R}_1,\mathbf{R}_2) + \mu\bar{\beta}\,||\mathbf{E}||_1 - \langle \mathbf{M},\mathbf{R}\rangle. \tag{265}$$

For notational convenience, partition $\mathbf{M}$ as

$$\mathbf{M} := \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_4 & \mathbf{M}_3 \end{pmatrix} \tag{266}$$

in accordance with the block structure of $\mathbf{R}$ in (263), where $\mathbf{M}_1 \in \mathbb{R}^{N_x\times N_x}$ and $\mathbf{M}_3 \in \mathbb{R}^{N_y\times N_y}$. The optimal solution to (P4) must satisfy: (i) the stationarity conditions

$$\nabla_{\mathbf{L}}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \tilde{f}(\mathbf{L},\mathbf{E}) - \mathbf{M}_2 - \mathbf{M}_4^{\mathcal{T}} = \mathbf{0} \tag{267}$$

$$\mathbf{0} \in \partial_{\mathbf{E}}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \left\{ \tilde{f}(\mathbf{L},\mathbf{E}) + \mu\bar{\beta}\left[\operatorname{sgn}(\mathbf{E}) + \tilde{\mathbf{E}}\right] \,\middle|\, \mathbf{E}\odot\tilde{\mathbf{E}} = \mathbf{0},\ \|\tilde{\mathbf{E}}\|_\infty \le 1 \right\} \tag{268}$$

$$\nabla_{\mathbf{R}_1}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_x} - \mathbf{M}_1 = \mathbf{0} \tag{269}$$

$$\nabla_{\mathbf{R}_2}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_y} - \mathbf{M}_3 = \mathbf{0} \tag{270}$$

(ii) complementary slackness condition $\langle \mathbf{M},\mathbf{R}\rangle = 0$; (iii) primal feasibility $\mathbf{R} \succeq \mathbf{0}$; and (iv) dual feasibility $\mathbf{M} \succeq \mathbf{0}$.

Using the stationary point $\bar{\mathbf{P}}$, $\bar{\mathbf{Q}}$ and $\bar{\mathbf{E}}$ of (P2), construct a candidate solution for (P4) as $\hat{\mathbf{L}} := \bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}$, $\hat{\mathbf{E}} := \bar{\mathbf{E}}$, $\hat{\mathbf{R}}_1 := \bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}}$, and $\hat{\mathbf{R}}_2 := \bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}$, as well as $\hat{\mathbf{M}}_1 := \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_x}$, $\hat{\mathbf{M}}_2 := \frac{1}{2}\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})$, $\hat{\mathbf{M}}_3 := \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_y}$, and $\hat{\mathbf{M}}_4 := \hat{\mathbf{M}}_2^{\mathcal{T}}$. After substituting these into (267)–(270), it can be readily verified that condition (i) holds. Condition (ii) also holds since

$$\begin{aligned}\langle \hat{\mathbf{M}},\hat{\mathbf{R}}\rangle &= \langle \hat{\mathbf{M}}_1,\hat{\mathbf{R}}_1\rangle + \langle \hat{\mathbf{M}}_2,\hat{\mathbf{L}}\rangle + \langle \hat{\mathbf{M}}_3,\hat{\mathbf{R}}_2\rangle + \langle \hat{\mathbf{M}}_4,\hat{\mathbf{L}}^{\mathcal{T}}\rangle \\ &= \frac{\lambda\bar{\beta}}{2}\operatorname{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}} + \bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}) + \operatorname{tr}\left(\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})\bar{\mathbf{Q}}\bar{\mathbf{P}}^{\mathcal{T}}\right) \\ &= \mathbf{0}\end{aligned} \tag{271}$$

where the last equality follows from (262). Condition (iii) is met since $\mathbf{R}$ can be rewritten as

$$\mathbf{R} = \begin{pmatrix} \bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}} & \bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}} \\ \bar{\mathbf{Q}}\bar{\mathbf{P}}^{\mathcal{T}} & \bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}} \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{P}} \\ \bar{\mathbf{Q}} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{P}} \\ \bar{\mathbf{Q}} \end{pmatrix}^{\mathcal{T}} \succeq \mathbf{0}. \tag{272}$$

For (iv), according to the Schur complement condition for positive semidefinite matrices, $\mathbf{M} \succeq \mathbf{0}$ holds if and only if

$$\hat{\mathbf{M}}_3 - \hat{\mathbf{M}}_4 \hat{\mathbf{M}}_1^{-1} \hat{\mathbf{M}}_2 \succeq \mathbf{0} \tag{273}$$

which is equivalent to $\lambda_{\max}(\hat{\mathbf{M}}_2^{\mathcal{T}} \hat{\mathbf{M}}_2) \leq (\lambda\bar{\beta}/2)^2$, or $||\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}, \bar{\mathbf{E}})|| \leq \lambda\bar{\beta}$. ■

H

Proof of Proposition 8

The proof uses the technique similar to the one employed in [199], where the convergence of online algorithms for optimizing objectives involving non-convex bilinear terms and sparse matrices was established in the context of dictionary learning.

In order to proceed with the proof, three lemmata are first established. The first lemma concerns some properties of $g(\mathbf{X}, \boldsymbol{\xi}^{(t)}) := g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$, and $\breve{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) := \breve{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$.

Lemma 4: *If the assumptions* (a1)–(a5) *in Proposition 8 hold, then*

(p1) $\breve{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ *majorizes* $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$*, i.e.,* $\breve{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) \; \forall \; \mathbf{X} \in \mathcal{X}'$*;*

(p2) $\breve{g}_1$ *is locally tight, i.e.,* $\breve{g}_1(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$*;*

(p3) $\nabla \breve{g}_1(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$*;*

(p4) $\breve{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) := \breve{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$ *is uniformly strongly convex in* $\mathbf{X}$*, i.e.,* $\forall (\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \in \mathcal{X} \times \mathcal{X} \times \Xi$*, it holds that*

$$\breve{g}(\mathbf{X} + \mathbf{D}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) - \breve{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq \breve{g}'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) + \frac{\zeta}{2} ||\mathbf{D}||_F^2$$

where $\zeta > 0$ *is a constant and* $\breve{g}'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D})$ *is a directional derivative of* $\breve{g}$ *at* $\mathbf{X}$ *along the direction* $\mathbf{D}$*;*

(p5) g_1 *and* $\breve{g}_1$*, their derivatives, and their Hessians are uniformly bounded;*

(p6) g_2 *and its directional derivative* g_2' *are uniformly bounded; and*

(p7) *there exists* $\bar{g} \in \mathbb{R}$ *such that* $|\breve{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})| \leq \bar{g}$*.*

Proof: For (p1), let us first show that $\nabla_{\mathbf{P}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$, $\nabla_{\mathbf{Q}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$, and $\nabla_{\mathbf{E}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ are Lipschitz continuous for $\mathbf{X} := (\mathbf{P},\mathbf{Q},\mathbf{E}) \in \mathcal{X}'$ and $\boldsymbol{\xi}^{(t)} \in \Xi$. For arbitrary $\mathbf{X}_1 := (\mathbf{P}_1,\mathbf{Q}_1,\mathbf{E}_1)$, $\mathbf{X}_2 := (\mathbf{P}_2,\mathbf{Q}_2,\mathbf{E}_2) \in \mathcal{X}'$, the variation of ∇g_1 in (177) can be bounded as

$$\begin{aligned}
&\|\nabla_{\mathbf{P}} g_1(\mathbf{P}_1,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)}) - \nabla_{\mathbf{P}} g_1(\mathbf{P}_2,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})\|_F \\
&\qquad = \left\| \sum_{m=1}^{M} \langle \mathbf{W}_m^{(t)}, (\mathbf{P}_1-\mathbf{P}_2)\mathbf{Q}^{\mathcal{T}} \rangle \mathbf{W}_m^{(t)} \mathbf{Q} \right\|_F \\
&\qquad \overset{\text{(i1)}}{\leq} \sum_{m=1}^{M} |\langle \mathbf{W}_m^{(t)}, (\mathbf{P}_1-\mathbf{P}_2)\mathbf{Q}^{\mathcal{T}} \rangle| \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F \\
&\qquad \overset{\text{(i2)}}{\leq} \sum_{m=1}^{M} \|\mathbf{P}_1-\mathbf{P}_2\|_F \, \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F^2
\end{aligned}$$

where (i1) and (i2) are due to the triangle and Cauchy-Schwarz inequalities, respectively. Since $\mathcal{X}'$ and Ξ are assumed to be bounded, $\sum_{m=1}^{M} \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F^2$ is bounded. Therefore, there exists a positive constant $L_{\mathbf{P}}$ such that

$$\|\nabla_{\mathbf{P}} g_1(\mathbf{P}_1,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)}) - \nabla_{\mathbf{P}} g_1(\mathbf{P}_2,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})\|_F \leq L_{\mathbf{P}} \|\mathbf{P}_1-\mathbf{P}_2\|_F \tag{274}$$

meaning that $\nabla_{\mathbf{P}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ is Lipschitz continuous with constant $L_{\mathbf{P}}$. Similar arguments hold for $\nabla_{\mathbf{Q}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ and $\nabla_{\mathbf{E}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ as well, with Lipschitz constants $L_{\mathbf{Q}}$ and $L_{\mathbf{E}}$, respectively. Then, upon defining $\|\mathbf{X}\|_\Delta := \sqrt{L_{\mathbf{P}}^2\|\mathbf{P}\|_F^2 + L_{\mathbf{Q}}^2\|\mathbf{Q}\|_F^2 + L_{\mathbf{E}}^2\|\mathbf{E}\|_F^2}$, it is easy to verify

$$\|\nabla g_1(\mathbf{X}_1,\boldsymbol{\xi}^{(t)}) - \nabla g_1(\mathbf{X}_2,\boldsymbol{\xi}^{(t)})\|_F \leq \|\mathbf{X}_1-\mathbf{X}_2\|_\Delta. \tag{275}$$

On the other hand, the proof of the Descent Lemma [29] can be adopted to show

$$\begin{aligned}
&g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) - g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \\
&\leq \langle \mathbf{X}-\mathbf{X}^{(t-1)}, \nabla g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \rangle + \int_0^1 \|\mathbf{X}-\mathbf{X}^{(t-1)}\|_F \\
&\times \|\nabla g_1(\mathbf{X}^{(t-1)} + \alpha(\mathbf{X}-\mathbf{X}^{(t-1)}),\boldsymbol{\xi}^{(t)}) \\
&\qquad\qquad - \nabla g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})\|_F d\alpha.
\end{aligned} \tag{276}$$

Note that

$$\|\mathbf{X}\|_F \leq \frac{1}{L_{\min}} \|\mathbf{X}\|_\Delta \tag{277}$$

where $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$. Then, substitution of (275) into (276) with $\mathbf{X}_1 = \mathbf{X}^{(t-1)} + \alpha(\mathbf{X} - \mathbf{X}^{(t-1)})$ and $\mathbf{X}_2 = \mathbf{X}^{(t-1)}$ yields

$$g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \langle \mathbf{X} - \mathbf{X}^{(t-1)}, \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})\rangle + \frac{1}{2L_{\min}}\|\mathbf{X} - \mathbf{X}^{(t-1)}\|_{\Delta}^2 \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) \tag{278}$$

which completes the proof by the construction of $\check{g}_1$, provided that $\eta_i^{(t)} \geq L_i^2/L_{\min}$ for all $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$.

To show (p2) and (p3), let us first denote

$$\nabla g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) = (\nabla_{\mathbf{P}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}), \nabla_{\mathbf{Q}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}), \nabla_{\mathbf{E}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})) \tag{279}$$

$$\begin{aligned}\nabla \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = (&\nabla_{\mathbf{P}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{P}}^{(t)}(\mathbf{P} - \mathbf{P}^{(t-1)}),\\ &\nabla_{\mathbf{Q}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{Q}}^{(t)}(\mathbf{Q} - \mathbf{Q}^{(t-1)}),\\ &\nabla_{\mathbf{E}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{E}}^{(t)}(\mathbf{E} - \mathbf{E}^{(t-1)})).\end{aligned} \tag{280}$$

Then, it suffices to evaluate $\check{g}_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ and $\nabla \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ at $\mathbf{X}^{(t-1)}$ to see that (p2) and (p3) hold.

To show (p4), let us first find $\check{g}_1'$ and g_2'. Along a direction $\mathbf{D} := (\mathbf{D}_{\mathbf{P}}, \mathbf{D}_{\mathbf{Q}}, \mathbf{D}_{\mathbf{E}}) \in \mathcal{X}'$, it holds that $\check{g}_1'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) = \langle \nabla \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}), \mathbf{D}\rangle$ since $\check{g}_1$ is differentiable. Similarly, $g_2'(\mathbf{X}; \mathbf{D}) = \lambda(\langle \mathbf{P}, \mathbf{D}_{\mathbf{P}}\rangle + \langle \mathbf{Q}, \mathbf{D}_{\mathbf{Q}}\rangle) + \mu h'(\mathbf{E}; \mathbf{D}_{\mathbf{E}})$ where $h(\mathbf{E}) := \|\mathbf{E}\|_1$, $\mathbf{d}_{\mathbf{E}} := \mathrm{vec}(\mathbf{D}_{\mathbf{E}})$ with its l-th entry being $d_{\mathbf{E},l}$, and

$$\begin{aligned} h'(\mathbf{E}; \mathbf{D}_{\mathbf{E}}) &:= \lim_{t \to 0+} \frac{h(\mathbf{E} + t\mathbf{D}_{\mathbf{E}}) - h(\mathbf{E})}{t}\\ &= \lim_{t \to 0+} \frac{\sum_{l, e_l \neq 0}(|e_l + t d_{\mathbf{E},l}| - |e_l|) + \sum_{l, e_l = 0} |t d_{\mathbf{E},l}|}{t}\\ &= \sum_{l, e_l \neq 0} \mathrm{sgn}(e_l) d_{\mathbf{E},l} + \sum_{l, e_l = 0} |d_{\mathbf{E},l}|. \end{aligned} \tag{281}$$

On the other hand, the variation of $\check{g}$ can be written as

$$\begin{aligned}&\check{g}(\mathbf{X} + \mathbf{D}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})\\ &= \check{g}_1'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) + \sum_{i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}} \frac{\eta_i^{(t)}}{2}\|\mathbf{D}_i\|_F^2 + g_2(\mathbf{X} + \mathbf{D}) - g_2(\mathbf{X}).\end{aligned} \tag{282}$$

Note that $\sum_i \frac{\eta_i^{(t)}}{2}\|\mathbf{D}_i\|_F^2 \geq \frac{L_{\min}}{2}||\mathbf{D}||_F^2$ since $\eta_i^{(t)} \geq L_i^2/L_{\min}$ by algorithmic construction. Furthermore, $g_2(\mathbf{X}+\mathbf{D}) - g_2(\mathbf{X}) \geq g_2'(\mathbf{X};\mathbf{D})$ since g_2 is convex [226]. Then, the variation of $\check{g}$ in (282) can be lower-bounded as

$$\begin{aligned}\check{g}(\mathbf{X}+\mathbf{D},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \\ \geq \check{g}'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) + \frac{L_{\min}}{2}||\mathbf{D}||_F^2 \end{aligned} \tag{283}$$

where $\check{g}'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) = \check{g}_1'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) + g_2'(\mathbf{X};\mathbf{D})$. Therefore, $(\mathrm{p4})$ holds for a positive constant $\zeta \leq L_{\min}$.

By the compactness of $\mathcal{X}$ and boundedness of Ξ by $(\mathrm{a3})$, $(\mathrm{p5})$ is automatically satisfied since g_1 and $\check{g}_1$ are continuously twice differentiable in $\mathbf{X}$ [242]. In addition, one can easily show $(\mathrm{p6})$ since g_2 and g_2' are also uniformly bounded by the compactness of $\mathcal{X}$.

Let K_1 and K_2 denote constants where $|\check{g}_1| \leq K_1$ and $|g_2| \leq K_2$, respectively, by $(\mathrm{p5})$ and $(\mathrm{p6})$. Then, $(\mathrm{p7})$ readily follows since

$$\begin{aligned}|\check{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})| &= |\check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})| \\ &\leq |\check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})| + |g_2(\mathbf{X})| \\ &\leq K_1 + K_2 =: \bar{g}. \blacksquare \end{aligned} \tag{284}$$

The next lemma asserts that a distance between two subsequent estimates asymptotically goes to zero, which will be used to show $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, almost surely.

Lemma 5: *If* $(\mathrm{a2})$–$(\mathrm{a5})$ *hold, then* $||\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}||_F = \mathcal{O}(1/t)$.

Proof: See [242, Lemma 2]. A proof of Lemma 5 is omitted to avoid duplication of the proof of [242, Lemma 2]. Hence, it suffices to mention that Lemma 4 guarantees the formulation of the proposed work satisfying the general assumptions on the formulation in [242]. ■

Lemma 5 does not guarantee convergence of the iterates to the stationary point of (P2). However, the final lemma asserts that the overestimated cost sequence converges to the cost of (P2), almost surely. Before proceeding to the next

lemma, let us first define

$$C_{1,t}(\mathbf{X}) := \frac{1}{t}\sum_{\tau=1}^{t} g_1(\mathbf{X}, \boldsymbol{\xi}^{(\tau)}) \tag{285}$$

$$\check{C}_{1,t}(\mathbf{X}) := \frac{1}{t}\sum_{\tau=1}^{t} \check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \tag{286}$$

and $C_2(\mathbf{X}) := g_2(\mathbf{X})$. Note also that $\check{C}_t(\mathbf{X}) - C_t(\mathbf{X}) = \check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X})$.

Lemma 6: *If* (a1)–(a5) *hold,* $\check{C}_t(\mathbf{X}^{(t)})$ *converges almost surely, and* $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, *almost surely.*

Proof: See [242, Lemma 1]. A proof of Lemma 6 is omitted to avoid duplication of the proof of [242, Lemma 1]. Instead, a sketch of the proof is following. It is firstly shown that the sequence $\{\check{C}_t(\mathbf{X}^{(t)})\}_{t=1}^{\infty}$ follows a quasi-martingale process and converges almost surely. Then, the lemma on positive converging sums (see [199, Lemma 8]) and Lemma 3 are used to claim that $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, almost surely. ■

The last step of the proof for Proposition 8 is inspired by [242]. Based on Lemma 6, it will be shown that the sequence $\{\nabla\check{C}_{1,t}(\mathbf{X}^{(t)}) - \nabla C_{1,t}(\mathbf{X}^{(t)})\}_{t=1}^{\infty}$ goes to zero, almost surely. Together with C_2', it follows that $\lim_{t\to\infty} C_t'(\mathbf{X}^{(t)}; \mathbf{D}) \geq 0\ \forall \mathbf{D}$, a.s. by algorithmic construction, implying convergence of a sequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ to the set of stationary points of $C(\mathbf{X})$.

By the compactness of $\mathcal{X}$, it is always possible to find a convergent subsequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ to a limit point $\bar{\mathbf{X}} \in \mathcal{X}$. Then, by the strong law of large numbers [116] under (a1) and equicontinuity of a family of functions $\{C_{1,t}(\cdot)\}_{t=1}^{\infty}$ due to the uniform boundedness of ∇g_1 in (p5) [35], upon restricting to the subsequence, one can have

$$\lim_{t\to\infty} C_{1,t}(\mathbf{X}^{(t)}) = \mathbb{E}_{\boldsymbol{\xi}}[g_1(\bar{\mathbf{X}}, \boldsymbol{\xi})] =: C_1(\mathbf{X}). \tag{287}$$

Similarly, a family of functions $\{\check{C}_{1,t}(\cdot)\}_{t=1}^{\infty}$ is equicontinuous due to the uniform boundedness of $\nabla\check{g}_1$ in (p5). Furthermore, $\{\check{C}_{1,t}(\cdot)\}_{t=1}^{\infty}$ is pointwisely bounded by (a1)–(a3). Thus, Arzelá-Ascoli theorem (see [35, Cor. 2.5] and [96]) implies that there exists a uniformly continuous function $\check{C}_1(\mathbf{X})$ such that $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}) = \check{C}_1(\mathbf{X})\ \forall\ \mathbf{X} \in \mathcal{X}$ and after restricting to the subsequence

$$\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) = \check{C}_1(\bar{\mathbf{X}}). \tag{288}$$

Furthermore, since $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ as in (p1), it follows that

$$\check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X}) \geq 0 \; \forall \mathbf{X}. \tag{289}$$

By letting $t \to \infty$ on (289) and combining Lemma 6 with (287) and (288), one deduces that

$$\check{C}_1(\bar{\mathbf{X}}) - C_1(\bar{\mathbf{X}}) = 0, \text{ a.s.} \tag{290}$$

meaning that $\check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X})$ takes its minimum at $\bar{\mathbf{X}}$ and

$$\nabla\check{C}_1(\bar{\mathbf{X}}) - \nabla C_1(\bar{\mathbf{X}}) = 0, \text{ a.s.} \tag{291}$$

by the first-order optimality condition.
On the other hand, the fact that $\mathbf{X}^{(t)}$ minimizes $\check{C}_t(\mathbf{X})$ by algorithmic construction and g_2' exists (so does C_2'), yields

$$\check{C}_{1,t}(\mathbf{X}^{(t)}) + C_2(\mathbf{X}^{(t)}) \leq \check{C}_{1,t}(\mathbf{X}) + C_2(\mathbf{X}) \quad \forall \mathbf{X} \in \mathcal{X} \tag{292}$$

and $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) + C_2(\mathbf{X}^{(t)}) \leq \lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}) + C_2(\mathbf{X}) \; \forall \mathbf{X} \in \mathcal{X}$, which implies

$$\lim_{t\to\infty} \langle \nabla\check{C}_{1,t}(\mathbf{X}^{(t)}), \mathbf{D}\rangle + C_2'(\mathbf{X}^{(t)}; \mathbf{D}) \geq 0 \quad \forall \mathbf{D}. \tag{293}$$

Using the result in (291), (293) can be re-written as $\langle \nabla C_1(\bar{\mathbf{X}}), \mathbf{D}\rangle + C_2'(\bar{\mathbf{X}}; \mathbf{D}) \geq 0 \; \forall \mathbf{D}$, a.s. or

$$C'(\bar{\mathbf{X}}; \mathbf{D}) \geq 0 \quad \forall \mathbf{D}, \text{ a.s.} \tag{294}$$

Thus, the subsequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ asymptotically coincides with the set of stationary points of $C(\mathbf{X})$. ■

I

Proof of (215)

To solve the optimization problem (214), consider first the unconstrained solution to the LS part of the cost, which can be written as

$$b_k^{\mathrm{LS}} := \left(y'_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_\ell^{\mathrm{DDD}}\right)/R_{kk}. \tag{295}$$

The detected symbol in (214) can be equivalently expressed as

$$\begin{aligned}\hat{b}_k^{\mathrm{DDD}} &= \arg\min_{b_k\in\mathcal{A}_a} f(b_k),\\ f(b_k) &:= \left(b_k^{\mathrm{LS}} - b_k\right)^2 + (2\lambda/R_{kk}^2)|b_k|_0.\end{aligned} \tag{296}$$

The solution $\hat{b}_k^{\mathrm{DDD}}$ can be obtained by comparing $f(0)$ with $\min_{b_k\in\mathcal{A}} f(b_k)$. Specifically, as the cost $f(b_k)$ is quadratic for $b_k \in \mathcal{A}$, the minimum is achieved at $f(\lfloor b_k^{\mathrm{LS}}\rceil)$, by quantizing b_k^{LS} to the nearest point in $\mathcal{A}$. Thus, $\hat{b}_k^{\mathrm{DDD}} = 0$ only if $f(0) \le f(\lfloor b_k^{\mathrm{LS}}\rceil)$, or equivalently, after using the definition of $f(\cdot)$, if $2b_k^{\mathrm{LS}}\lfloor b_k^{\mathrm{LS}}\rceil - \lfloor b_k^{\mathrm{LS}}\rceil^2 - 2\lambda/R_{kk}^2 \le 0$. This completes the proof of (215).

J

Proof of (216)

When $M = 2$, the DDD solution (215) reduces to

$$\hat{b}_k^{\text{DDD}} = \text{sign}(b_k^{\text{LS}})\mathbb{1}(2|b_k^{\text{LS}}| - 1 - 2\lambda/R_{kk}^2 > 0) \tag{297}$$

due to the fact that $\lfloor b_k^{\text{LS}} \rceil = \text{sign}(b_k^{\text{LS}}) \in \{\pm 1\}$.
Recalling that $\check{\mathbf{b}}$ denotes the transmitted vector, and substituting $\mathbf{y} = \mathbf{H}\check{\mathbf{b}} + \mathbf{w}$ yields

$$\mathbf{y}' := \mathbf{Q}^T\mathbf{y} = \mathbf{R}\check{\mathbf{b}} + \mathbf{u} \tag{298}$$

where $\mathbf{u} := \mathbf{Q}^T\mathbf{w}$ is zero-mean Gaussian with identity covariance matrix. Supposing that there is no error propagation, the b_k^{LS} term in (297) becomes [cf. (295)]

$$\begin{aligned} b_k^{\text{LS}} &= \left(\sum_{\ell=k}^{K} R_{k\ell}\check{b}_\ell + u_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_k^{\text{DDD}}\right) \Big/ R_{kk} \\ &= \check{b}_\ell + u_k/R_{kk} \end{aligned} \tag{299}$$

where u_k denotes the k-th entry of $\mathbf{u}$.
To analyze the error probability for the detector in (297) consider the following three cases.
a) $\check{b}_k = 0$ is sent: under this case, $b_k^{\text{LS}} = u_k/R_{kk}$ and a detection error emerges when $\hat{b}_k^{\text{DDD}} = 1$ or -1. With the closed-form DDD detector (297) in mind, such an error occurs only if both $b_k^{\text{LS}} \neq 0$ and $2|b_k^{\text{LS}}| - 2\lambda/R_{kk}^2 - 1 > 0$ hold. The first case corresponds to $u_k \neq 0$ and the second one is equivalent to $|u_k| > |R_{kk}|/2 + \lambda/|R_{kk}|$, which is included in the event $u_k \neq 0$. Hence, to evaluate the error probability it suffices to consider only the case $|u_k| > |R_{kk}|/2 + \lambda/|R_{kk}|$.
b) $\check{b}_k = 1$ is sent: following the analysis in a), an error occurs if $\text{sign}(R_{kk})u_k \leq -|R_{kk}|/2 + \lambda/|R_{kk}|$.

c) $\check{b}_k = -1$ is sent: following the analysis in a), an error occurs if $\text{sign}(R_{kk})u_k \geq |R_{kk}|/2 - \lambda/|R_{kk}|$.
Given that the Gaussian distributed u_k has variance 1, the overall SER for the k-th entry b_k becomes

$$\begin{aligned} P_{e,k}^{\text{DDD}} &= \textstyle\sum_{i=0,\pm 1} P(\text{error}|\check{b}_k = i) P(\check{b}_k = i) \\ &= 2(1-p_a) Q\left(\frac{|R_{kk}|}{2} + \frac{\lambda}{|R_{kk}|}\right) \\ &\quad + p_a Q\left(\frac{|R_{kk}|}{2} - \frac{\lambda}{|R_{kk}|}\right) . \end{aligned} \tag{300}$$

References

[1] Agrawal, P. and N. Patwari. 2009. "Correlated Link Shadow Fading in Multi-Hop Wireless Networks". *IEEE Trans. Wireless Commun.* 8(9): 4024–4036.

[2] Akaike, H. 1974. "A New Look at the Statistical Model Identification". *IEEE Trans. Automat. Contr.*: 716–723.

[3] Akçakaya, M. and V. Tarokh. 2010. "Shannon-Theoretic Limits on Noisy Compressive Sampling". *IEEE Trans. Inform. Theory*. 56(1): 492–504.

[4] Akyildiz, I. F., W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. "Wireless sensor networks: a survey". *Computer Networks*. 38: 393–422.

[5] Akyildiz, I. F., M. C. Vuran, and Ö. B. Akan. 2004. "On Exploiting Spatial and Temporal Correlation in Wireless Sensor Networks". In: *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*. Cambridge, UK. 71–80.

[6] Anastasi, G., M. Conti, M. D. Francesco, and A. Passarella. 2009. "Energy Conservation in Wireless Sensor Networks: A Survey". *Ad Hoc Netw.* 7(3): 537–568.

[7] Angelosante, D., J. A. Bazerque, and G. B. Giannakis. 2010a. "Online Adaptive Estimation of Sparse Signals: Where RLS Meets the ℓ_1-Norm". *IEEE Trans. Signal Processing*. 58(7): 3436–3447.

[8] Angelosante, D. and G. B. Giannakis. 2009. "RLS-Weighted Lasso for Adaptive Estimation of Sparse Signals". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Taipei, Taiwan.

[9] Angelosante, D., E. Grossi, G. B. Giannakis, and M. Lops. 2010b. "Sparsity-Aware Estimation of CDMA System Parameters". *EURASIP J. Advances Signal Processing*. 2010(June). Article ID 417981.

[10] Arianandа, D. D. and G. Leus. 2012. "Compressive Wideband Power Spectrum Estimation". *IEEE Trans. Signal Processing*. 60(9): 4775–4789.

[11] Arimoto, S. 1972. "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels". *IEEE Trans. Inform. Theory*. 18(1): 14–20.

[12] Asif, M. S. and J. Romberg. 2013. "Fast and Accurate Algorithms for Re-Weighted l_1-Norm Minimization". *IEEE Trans. Signal Processing*. 61(23): 5905–5916.

[13] Asif, M. S. and J. Romberg. 2014. "Sparse Recovery of Streaming Signals Using ℓ_1-Homotopy". *IEEE Trans. Signal Processing*. 62(16): 4209–4223.

[14] Bajwa, W., J. Haupt, A. Sayeed, and R. Nowak. 2007. "Joint Source-Channel Communication for Distributed Estimation in Sensor Networks". *IEEE Trans. Inform. Theory*. 53(10): 3629–3653.

[15] Baraniuk, R. G., V. Cevher, M. F. Duarte, and C. Hegde. 2010. "Model-Based Compressive Sensing". *IEEE Trans. Inform. Theory*. 56(4): 1982–2001.

[16] Baraniuk, R., M. Davenport, R. DeVore, and M. Wakin. 2008. "A Simple Proof of the Restricted Isometry Property for Random Matrices". *Constructive Approximation*. 28(3): 253–263.

[17] Baraniuk, R. 2007. "Compressive Sensing [Lecture Notes]". *IEEE Signal Processing Mag.* 24(4): 118–121.

[18] Baron, D., S. Sarvotham, and R. G. Baraniuk. 2010. "Bayesian Compressive Sensing Via Belief Propagation". *IEEE Trans. Signal Processing*. 58(1): 269–280.

[19] Baron, D., M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk. 2005. "Distributed Compressed Sensing". *Tech. rep.* No. ECE-0612. Rice University, Houston, TX: Dept. Elect. Comput. Eng. URL: http://webee.technion.ac.il/people/drorb/pdf/DCS112005.pdf.

[20] Bazerque, J. A. and G. B. Giannakis. 2010. "Distributed Spectrum Sensing for Cognitive Radio Networks by Exploiting Sparsity". *IEEE Trans. Signal Processing*. 58(3): 1847–1862.

[21] Bazerque, J. A., G. Mateos, and G. B. Giannakis. 2011. "Group-Lasso on Splines for spectrum Cartography". *IEEE Trans. Signal Processing*. 59(10): 4648–4663.

[22] Beck, A., A. Ben-Tal, and M. Teboulle. 2006. "Finding a Global Optimal Solution for a Quadratically Constrained Fractional Quadratic Problem with Applications to the Regularized Total Least Squares". *SIAM J. on Matrix Anal. and Appl.* 28(2): 425–445.

[23] Berger, C. R., S. Zhou, J. C. Preisig, and P. Willett. 2010. "Sparse Channel Estimation for Multicarrier Underwater Acoustic Communication: From Subspace Methods to Compressed Sensing". *IEEE Trans. Signal Processing*. 58(3): 1708–1721.

[24] Berger, J., V. D. Oliveira, and B. Sanso. 2000. "Objective Bayesian Analysis of Spatially Correlated Data". *Journal of the American Statistical Association*. 96(Dec.): 1361–1374.

[25] Berger, T. 1971. *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. *Prentice-Hall Series in Information and System Sciences*. Prentice Hall. 352.

[26] Berger, T. and J. Gibson. 1998. "Lossy Source Coding". *IEEE Trans. Inform. Theory*. 44(6): 2693–2723.

[27] Berinde, R., A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. 2008. "Combining Geometry and Combinatorics: A Unified Approach to Sparse Signal Recovery". In: *Proc. Annual Allerton Conf. Commun., Contr., Computing*. Illinois, USA. 798–805.

[28] Berinde, R. and P. Indyk. 2008. "Sparse Recovery Using Sparse Random Matrices". *Tech. rep.* MIT.

[29] Bertsekas, D. P. 1999. *Nonlinear Programming*. 2nd. Belmont, Massachusetts, USA: Athena Scientific.

[30] Blahut, R. 1972. "Computation of Channel Capacity and Rate-Distortion Functions". *IEEE Trans. Inform. Theory*. 18(4): 460–473.

[31] Blumensath, T. and M. E. Davies. 2008. "Iterative Thresholding for Sparse Approximations". *Journal of Fourier Analysis and Applications*. 14(5): 629–654.

[32] Boufounos, P. T. 2012. "Universal Rate-Efficient Scalar Quantization". *IEEE Trans. Inform. Theory*. 58(3): 1861–1872.

[33] Boyd, S. 2008. "EE363: Linear Dynamical Systems [lecture material 2008 – 2009], Stanford University". http://stanford.edu/class/ee363/index.html.

[34] Boyd, S. and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge, UK: Cambridge University Press.

[35] Brown, R. F. 2004. *A Topological Introduction to Nonlinear Analysis*. Boston, MA: Birkhäuser.

[36] Caione, C., D. Brunelli, and L. Benini. 2012. "Distributed Compressive Sampling for Lifetime Optimization in Dense Wireless Sensor Networks". *IEEE Trans. Ind. Informat.* 8(1): 30–40.

[37] Caione, C., D. Brunelli, and L. Benini. 2014. "Compressive Sensing Optimization for Signal Ensembles in WSNs". *IEEE Trans. Ind. Informat.* 10(1): 382–392.

[38] Candés, E. J. 2006. "Compressive Sampling". In: *Proceedings of the International Congress of Mathematicians*. Vol. 3. Madrid, Spain. 1433–1452.

[39] Candés, E. J. 2008. "The Restricted Isometry Property and Its Implications for Compressed Sensing". *C.R. Acad. Sci. Paris, Ser. I.* 346(9-10): 589–592.

[40] Candès, E. J., X. Li, Y. Ma, and J. Wright. 2011. "Robust Principal Component Analysis?" *J. of the ACM*. 58(3, article no. 11).

[41] Candés, E. J. and J. Romberg. 2005. "l_1-MAGIC: Recovery of Sparse Signals via Convex Programming". http://users.ece.gatech.edu/~justin/l1magic/.

[42] Candés, E. J. and J. Romberg. 2007. "Sparsity and Incoherence in Compressive Sampling". *Inverse Problems*. 23(3): 969–985.

[43] Candés, E. J., J. Romberg, and T. Tao. 2006a. "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information". *IEEE Trans. Inform. Theory*. 52(2): 489–509.

[44] Candés, E. J., J. Romberg, and T. Tao. 2006b. "Stable Signal Recovery from Incomplete and Inaccurate Measurements". *Communications on Pure and Applied Mathematics*. 59(8): 1207–1223.

[45] Candés, E. J. and T. Tao. 2005. "Decoding by Linear Programming". *IEEE Trans. Inform. Theory*. 51(12): 4203–4215.

[46] Candés, E. J. and T. Tao. 2006. "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Trans. Inform. Theory*. 52(12): 5406–5425.

[47] Candés, E. J. and T. Tao. 2007. "The Dantzig Selector: Statistical Estimation When p Is Much Larger than n". *The Annals of Statistics*. 35(6): 2313–2351.

[48] Candés, E. J., M. Wakin, and S. Boyd. 2008. "Enhancing Sparsity by Reweighted l_1 Minimization". *Journal of Fourier Analysis and Applications*. 14(5-6): 877–905.

[49] Candés, E. J. and M. Wakin. 2008. "An Introduction To Compressive Sampling". *IEEE Signal Processing Mag.* 25(2): 21–30.

[50] Cardinal, J. and G. V. Assche. 2002. "Joint Entropy-Constrained Multiterminal Quantization". In: *Proc. IEEE Int. Symp. Inform. Theory*. Lausanne, Switzerland. 63.

[51] Carmi, A. Y., L. S. Mihaylova, and S. J. Godsill. 2014. *Compressed Sensing & Sparse Filtering*. first. *Signals and Communication Technology*. Springer-Verlag Berlin Heidelberg. 502.

[52] Carrillo, R. E. and K. E. Barner. 2009. "Iteratively Re-Weighted Least Squares for Sparse Signal Reconstruction from Noisy Measurements". In: *Proc. Conf. Inform. Sciences Syst. (CISS)*. Baltimore, MD. 448–453.

[53] Cetin, M. and B. M. Sadler. 2005. "Semi-Blind Sparse Channel Estimation with Constant Modulus Symbols". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Philadelphia, PN.

[54] Cevher, V., M. F. Duarte, and R. G. Baraniuk. 2008. "Distributed Target Localization via Spatial Sparsity". In: *Proc. European Sign. Proc. Conf.* Lausanne, Switzerland.

[55] Chae, D. H., P. Sadeghi, and R. A. Kennedy. 2010. "Effects of Basis-Mismatch in Compressive Sampling of Continuous Sinusoidal Signals". In: *Proc. of 2nd Intl. Conf. on Future Computer and Communication*. Vol. 2. Wuhan, China. 739–743.

[56] Chandrasekaran, V., B. Recht, P. A. Parrilo, and A. S. Willsky. 2012. "The Convex Geometry of Linear Inverse Problems". *Found. Comput. Math.* 12(6): 805–849.

[57] Chang, C. 2010. "On the Rate Distortion Function of Bernoulli Gaussian Sequences". In: *Proc. IEEE Int. Symp. Inform. Theory*. Austin, TX. 66–70.

[58] Chang, C.-I. and L. D. Davisson. 1988. "On Calculating the Capacity of an Infinite-Input Finite (Infinite)-Output Channel". *IEEE Trans. Inform. Theory*. 34(5): 1004–1010.

[59] Chartrand, R. 2007. "Exact Reconstruction of Sparse Signals via Nonconvex Minimization". *IEEE Signal Processing Lett.* 14(10): 707–710.

[60] Chartrand, R. and Y. Wotao. 2008. "Iteratively Reweighted Algorithms for Compressive Sensing". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Las Vegas, NV. 3869–3872.

[61] Chen, S., D. Donoho, and M. Saunders. 1998. "Atomic Decomposition by Basis Pursuit". *SIAM Journal on Scientific Computing*. 20(1): 33–61.

[62] Cheng, S., V. Stankovic, and Z. Xiong. 2005. "Computing the Channel Capacity and Rate-Distortion Function with Two-Sided State Information". *IEEE Trans. Inform. Theory*. 51(12): 4418–4425.

[63] Chepuri, S. P. and G. Leus. 2016. "Sparse Sensing for Statistical Inference". *Found. Trends Signal Proc.* 9(3–4): 233–368.

[64] Chi, Y., A. Pezeshki, L. Scharf, and R. Calderbank. 2010. "Sensitivity to Basis Mismatch in Compressed Sensing". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Dallas, TX.

[65] Chiang, M. and S. Boyd. 2004. "Geometric Programming Duals of Channel Capacity and Rate Distortion". *IEEE Trans. Inform. Theory*. 50(2): 245–258.

[66] Chou, P., T. Lookabaugh, and R. Gray. 1989. "Entropy-Constrained Vector Quantization". *IEEE Trans. Acoust., Speech, Signal Processing*. 37(1): 31–42.

[67] Coluccia, G., S. Kuiteing, A. Abrardo, M. Barni, and E. Magli. 2012. "Progressive Compressed Sensing and Reconstruction of Multidimensional Signals Using Hybrid Transform/Prediction Sparsity Model". *IEEE J. Emerg. Select. Topics Circuits Syst.* 2(3): 340–352.

[68] Coluccia, G., E. Magli, A. Roumy, and V. Toto-Zarasoa. 2011. "Lossy Compression of Distributed Sparse Sources: A Practical Scheme". In: *Proc. European Sign. Proc. Conf.* Barcelona, Spain. 422–426.

[69] Coluccia, G., A. Roumy, and E. Magli. 2014. "Operational Rate-Distortion Performance of Single-Source and Distributed Compressed Sensing". *IEEE Trans. Commun.* 62(6): 2022–2033.

[70] Cotter, S. F. and B. D. Rao. 2002a. "Application Of Total Least Squares (TLS) to the Design of Sparse Signal Representation Dictionaries". In: *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* Pacific Grove, CA.

[71] Cotter, S. F. and B. D. Rao. 2002b. "Sparse Channel Estimation via Matching Pursuit with Application to Equalization". *IEEE Trans. Commun.* 50(3): 374–377.

[72] Cotter, S. F., B. D. Rao, K. Engan, and K. Kreutz-Delgado. 2005. "Sparse Solutions to Linear Inverse Problems with Multiple Measurement Vectors". *IEEE Trans. Signal Processing*. 53(7): 2477–2488.

[73] Cover, T. and M. Chiang. 2002. "Duality Between Channel Capacity and Rate Distortion with Two-Sided State Information". *IEEE Trans. Inform. Theory*. 48(6): 1629–1638.

[74] Cover, T. and J. Thomas. 2006. *Elements of Information Theory*. second. New York, USA: John Wiley.

[75] Cui, T. and C. Tellambura. 2005. "An Efficient Generalized Sphere Decoder for Rank-Deficient MIMO Systems". *IEEE Commun. Lett.* 9(May): 423–425.

[76] D. M. Sima, S. Van Huffel, and G. H. Golub. 2004. "Regularized Total Least Squares based on Quadratic Eigenvalue Problem Solvers". *BIT Numerical Mathematics*. 44(4): 793–812.

[77] Dai, W. and O. Milenkovic. 2009. "Subspace Pursuit for Compressive Sensing Signal Reconstruction". *IEEE Trans. Inform. Theory*. 55(5): 2230–2249.

[78] Dai, W. and O. Milenkovic. 2011. "Information Theoretical and Algorithmic Approaches to Quantized Compressive Sensing". *IEEE Trans. Commun.* 59(7): 1857–1866.

[79] Dall'Anese, E., S.-J. Kim, and G. B. Giannakis. 2011. "Channel Gain Map Tracking via Distributed Kriging". *IEEE Trans. Veh. Technol.* 60(3): 1205–1211.

[80] Damen, M. O., H. E. Gamal, and G. Caire. 2003. "On Maximum-Likelihood Detection and the Search for the Closest Lattice Point". *IEEE Trans. Inform. Theory*. 49(Oct.): 2389–2402.

[81] Davenport, M. A., M. F. Duarte, Y. C. Eldar, and G. Kutyniok. 2012. "Introduction to Compressed Sensing". In: *Compressed Sensing: Theory and Applications*. Ed. by Y. C. Eldar and G. Kutyniok. Cambridge University Press. Chap. 1. 1–64.

[82] Dinkelbach, W. 1967. "On Nonlinear Fractional Programming". *Management Science*. 13(7): 492–498.

[83] Dixon, A. M. R., E. G. Allstot, D. Gangopadhyay, and D. J. Allstot. 2012. "Compressed Sensing System Considerations for ECG and EMG Wireless Biosensors". *IEEE Trans. Biomed. Circ. Syst.* 6(2): 156–166.

[84] Do, T., L. Gan, N. Nguyen, and T. Tran. 2012. "Fast and Efficient Compressive Sensing Using Structurally Random Matrices". *IEEE Trans. Signal Processing*. 60(1): 139–154.

[85] Dobrushin, R. and B. Tsybakov. 1962. "Information Transmission with Additional Noise". *IRE Trans. Inform. Theory*. 8(5): 293–304.

[86] Dong, Y., S. Chang, and L. Carin. 2005. "Rate-Distortion Bound for Joint Compression and Classification with Application to Multiaspect Scattering". *IEEE Sensors J.* 5(3): 481–492.

[87] Donoho, D. L. 2006. "Compressed Sensing". *IEEE Trans. Inform. Theory*. 52(4): 1289–1306.

[88] Donoho, D. L., M. Elad, and V. N. Temlyakov. 2006a. "Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise". *IEEE Trans. Inform. Theory*. 52(1): 6–18.

[89] Donoho, D. L., A. Maleki, and A. Montanari. 2009. "Message-Passing Algorithms for Compressed Sensing". *Proceedings of the National Academy of Sciences of the USA*. 106(45): 18914–18919.

[90] Donoho, D., A. Maleki, and M. Shahram. 2006b. "Wavelab 850: Wavelet Toolbox for Matlab". http://www-stat.stanford.edu/~wavelab.

[91] Dragotti, P. L. and M. Gastpar. 2009. *Distributed Source Coding: Theory, Algorithms and Applications*. Academic Press.

[92] Duarte, M. and R. Baraniuk. 2010. "Kronecker Product Matrices for Compressive Sensing". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Dallas, TX. 3650–3653.

[93] Duarte, M. and R. Baraniuk. 2012. "Kronecker Compressive Sensing". *IEEE Trans. Image Processing*. 21(2): 494–504.

[94] Duarte, M., M. Wakin, D. Baron, and R. Baraniuk. 2006. "Universal Distributed Sensing Via Random Projections". In: *Proc. IEEE Int. Symp. on Inform. Proc. in Sensor Networks*. Nashville, Tennessee, USA. 177–185.

[95] Duarte, M., M. Wakin, D. Baron, S. Sarvotham, and R. Baraniuk. 2013. "Measurement Bounds for Sparse Signal Ensembles via Graphical Models". *IEEE Trans. Inform. Theory*. 59(7): 4280–4289.

[96] Dunford, N. and J. T. Schwartz. 1958. *Linear Operators. Part 1: General Theory*. Newyork, NY: Interscience publishers.

[97] E. Viterbo and J. Boutros. 1999. "A Universal Lattice Code Decoder for Fading Channels". *IEEE Trans. Inform. Theory*. 45(July): 1639–1642.

[98] Elad, M. 2010. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. 1st ed. Springer-Verlag New York.

[99] Elad, M. and I. Yavneh. 2009. "A Plurality of Sparse Representations Is Better Than the Sparsest One Alone". *IEEE Trans. Inform. Theory*. 55(10): 4701–4714.

[100] Eldar, Y. C. and G. Kutyniok. 2012. *Compressed Sensing: Theory and Applications*. Cambridge University Press.

[101] Eldar, Y. C. and M. Mishali. 2009. "Robust Recovery of Signals From a Structured Union of Subspaces". *IEEE Trans. Inform. Theory*. 55(11): 5302–5316.

[102] Eriksson, A. and A. van den Hengel. 2010. "Efficient Computation of Robust Low-Rank Matrix Approximations in the Presence of Missing Data Using the L_1 Norm". In: *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*. San Francisco, CA. 771–778.

[103] Fan, J. and R. Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties". *Journal of the American Stat. Assoc.* 96(456): 1348–1360.

[104] Fazel, F., M. Fazel, and M. Stojanovic. 2011. "Random Access Compressed Sensing for Energy-Efficient Underwater Sensor Networks". *IEEE J. Select. Areas Commun.* 29(8): 1660–1670.

[105] FCC. 2002. "Spectrum Policty Task Force". *Tech. rep.* No. ET Docket 02-135.

[106] Finamore, W. and W. Pearlman. 1980. "Optimal Encoding of Discrete-Time Continuous-Amplitude Memoryless Sources with Finite Output Alphabets". *IEEE Trans. Inform. Theory*. 26(2): 144–155.

[107] Fine, T. 1964. "Properties of An Optimum Digital System and Applications". *IEEE Trans. Inform. Theory*. 10(4): 287–296.

[108] Fleming, M., Q. Zhao, and M. Effros. 2004. "Network Vector Quantization". *IEEE Trans. Inform. Theory*. 50(8): 1584–1604.

[109] Fletcher, A. K., S. Rangan, and V. K. Goyal. 2009. "A Sparsity Detection Framework for On-Off Random Access Channels". In: *Proc. IEEE Int. Symp. Inform. Theory*. Seoul, South Korea. 169–173.

[110] Flynn, T. and R. Gray. 1987. "Encoding of Correlated Observations". *IEEE Trans. Inform. Theory*. 33(6): 773–787.

[111] Foster, D. P. and E. I. George. 1994. "The Risk Inflation Criterion for Multiple Regression". *Annals of Statistics*. 22(4): 1947–1975.

[112] Foucart, S. and H. Rauhut. 2013. *A Mathematical Introduction to Compressive Sensing*. *Applied and Numerical Harmonic Analysis*. Springer New York.

[113] Fraysse, A., B. Pesquet-Popescu, and J.-C. Pesquet. 2009. "On the Uniform Quantization of a Class of Sparse Sources". *IEEE Trans. Inform. Theory*. 55(7): 3243–3263.

[114] Freris, N. M., O. Öçal, and M. Vetterli. 2013. "Compressed Sensing of Streaming Data". In: *Proc. Annual Allerton Conf. Commun., Contr., Computing*. Monticello, IL, USA. 1242–1249.

[115] Friedman, J., T. Hastie, H. Höfling, and R. Tibshirani. 2007. "Pathwise Coordinate Optimization". *Annals of Applied Statistics*. 1(Dec.): 302–332.

[116] Fristedt, B. E. and L. F. Gray. 1997. *A Modern Approach to Probability Theory*. Boston, MA: Birkhäuser.

[117] Fuchs, J.-J. 1999. "Multipath Time-Delay Detection and Estimation". *IEEE Trans. Signal Processing*. 47(1): 237–243.

[118] Fuchs, J.-J. 2001. "On the Application of the Global Matched Filter to DOA Estimation with Uniform Circular Arrays". *IEEE Trans. Signal Processing*. 49(4): 702–709.

[119] Al-Fuqaha, A., M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. 2015. "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications". *IEEE Commun. Surveys & Tutorials*. 17(4): 2347–2376.

[120] Gallager, R. 2013. *Stochastic Processes: Theory for Applications*. New York, USA: Cambridge University Press. 536.

[121] Gamal, A. E. and Y.-H. Kim. 2012. *Network Information Theory*. New York, USA: Cambridge University Press.

[122] Gersho, A. 1982. "On The Structure of Vector Quantizers". *IEEE Trans. Inform. Theory*. 28(2): 157–166.

[123] Giannakis, G. B., Z. Liu, X. Ma, and S. Zhou. 2007. *Space-Time Coding for Broadband Wireless Communications*. John Wiley & Sons, Inc.

[124] Gibson, J. 2013. *Information Theory and Rate Distortion Theory for Communications and Compression*. 1st ed. *Synthesis Lectures on Communications*. Morgan & Claypool.

[125] Gish, H. and J. Pierce. 1968. "Asymptotically Efficient Quantizing". *IEEE Trans. Inform. Theory*. 14(5): 676–683.

[126] Goblirsch, D. 1989. "Quantization Systems for Hidden Markov Sources". *PhD thesis*. College Park, MD, USA: The Institute for Systems Research, Univ. Maryland. URL: http://drum.lib.umd.edu/handle/1903/4942.

[127] Golub, G. and C. Loan. 1996. *Matrix Computations*. Baltimore: Johns Hopkins University Press.

[128] Gorodnitsky, I. F. and B. D. Rao. 1997. "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-Weighted Minimum Norm Algorithm". *IEEE Trans. Signal Processing*. 45(3): 600–616.

[129] Goyal, V., A. Fletcher, and S. Rangan. 2008. "Compressive Sampling and Lossy Compression". *IEEE Signal Processing Mag.* 25(2): 48–56.

[130] Grant, M. and S. Boyd. 2012. "CVX: Matlab Software for Disciplined Convex Programming, version 1.22". http://cvxr.com/cvx.

[131] Gray, R. 1972. "Conditional Rate-Distortion Theory". *Tech. rep.* Stanford Electron. Lab, Stanford University. 17.

[132] Gray, R. 1973. "A New Class of Lower Bounds to Information Rates of Stationary Sources via Conditional Rate-Distortion Functions". *IEEE Trans. Inform. Theory*. 19(4): 480–489.

[133] Gray, R. 1984. "Vector Quantization". *IEEE Trans. Acoust., Speech, Signal Processing*. 1(2): 4–29.

[134] Gray, R. 1990. *Source Coding Theory*. Vol. 83. *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers.

[135] Gray, R. and D. Neuhoff. 1998. "Quantization". *IEEE Trans. Inform. Theory*. 44(6): 2325–2383.

[136] Gray, R., D. Neuhoff, and J. Omura. 1975. "Process Definitions of Distortion-Rate Functions and Source Coding Theorems". *IEEE Trans. Inform. Theory*. 21(5): 524–532.

[137] Graziosi, F. and F. Santucci. 2002. "A General Correlation Model for Shadow Fading in Mobile Radio Systems". *IEEE Commun. Lett.* 6(3): 102–104.

[138] Gu, Y., Y. Chen, and A. O. Hero. 2009. "Sparse LMS for System Identification". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Taipei, Taiwan.

[139] Gudmundson, M. 1991. "Correlation Model for Shadow Fading in Mobile Radio Systems". *Elec. Lett.* 27(23): 2145–2146.

[140] Guler, B., E. MolavianJazi, and A. Yener. 2015. "Remote Source Coding with Two-Sided Information". In: *Proc. IEEE Int. Symp. Inform. Theory*. 2176–2180.

[141] Hamilton, B. R., X. Ma, R. J. Baxley, and S. M. Matechik. 2012. "Radio Frequency Tomography in Mobile Networks". In: *Proc. of the IEEE Stat. Sig. Proc. Workshop*. Ann Arbor, MI. 508–511.

[142] Hamilton, B. R., X. Ma, R. J. Baxley, and S. M. Matechik. 2014. "Propagation Modeling for Radio Frequency Tomography in Wireless Networks". *IEEE J. Select. Topics Signal Processing*. 8(1): 55–65.

[143] Hansen, P. C. 1994. "Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems". *Numerical Algorithms*. 6(1): 1–35.

[144] Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, And Prediction*. Second. New York: Springer.

[145] Haupt, J., W. Bajwa, M. Rabbat, and R. Nowak. 2008. "Compressed Sensing for Networked Data: A Different Approach to Decentralized Compression". *IEEE Signal Processing Mag.* 25(2): 92–101.

[146] Haupt, J. and R. Nowak. 2006. "Signal Reconstruction From Noisy Random Projections". *IEEE Trans. Inform. Theory*. 52(9): 4036–4048.

[147] Haupt, J. and R. Nowak. 2012. "Adaptive Sensing for Sparse Recovery". In: *Compressive Sensing: Theory and Applications*. Ed. by Y. C. Eldar and G. Kutyniok. Cambridge University Press. Chap. 6. 269–304.

[148] Herman, M. A. and T. Strohmer. 2010. "General Deviants: an Analysis of Perturbations in Compressive Sensing". *IEEE J. Select. Topics Signal Processing*. 4(Apr.): 342–349.

[149] Huffman, D. 1952. "A Method for the Construction of Minimum-Redundancy Codes". *Proc. IRE*. 40(9): 1098–1101.

[150] I. G. Akrotirianakis and C. A. Floudas. 2004. "Computational Experience with a New Class of Convex Underestimators: Box-Constrained NLP Problems". *Journal of Global Optimization*. 29(3): 249–264.

[151] Jacques, L., D. Hammond, and J. Fadili. 2011. "Dequantizing Compressed Sensing: When Oversampling and Non-Gaussian Constraints Combine". *IEEE Trans. Inform. Theory*. 57(1): 559–571.

[152] Jacques, L., J. N. Laska, P. T. Boufounos, and R. G. Baraniuk. 2013. "Robust 1-Bit Compressive Sensing via Binary Stable Embeddings of Sparse Vectors". *IEEE Trans. Inform. Theory*. 59(4): 2082–2102.

[153] Jansson, M., A. L. Swindlehurst, and B. Ottersten. 1998. "Weighted Subspace Fitting for General Array Error Models". *IEEE Trans. Signal Processing*. 46(9): 2484–2498.

[154] Ji, S., Y. Xue, and L. Carin. 2008. "Bayesian Compressive Sensing". *IEEE Trans. Signal Processing*. 56(6): 2346–2356.

[155] Kaltiokallio, O., M. Bocca, and N. Patwari. 2012. "Enhancing the Accuracy of Radio Tomographic Imaging Using Channel Diversity". In: *Proc. of the 9th IEEE Int. Conf. on Mobile Ad hoc and Sensor Syst.* Las Vegas, NV. 254–262.

[156] Kamilov, U. S., S. Rangan, A. K. Fletcher, and M. Unser. 2014. "Approximate Message Passing With Consistent Parameter Estimation and Applications to Sparse Learning". *IEEE Trans. Inform. Theory*. 60(5): 2969–2985.

[157] Kamilov, U., V. Goyal, and S. Rangan. 2012. "Message-Passing De-Quantization With Applications to Compressed Sensing". *IEEE Trans. Signal Processing*. 60(12): 6270–6281.

[158] Kanso, M. A. and M. G. Rabbat. 2009. "Compressed RF Tomography for Wireless Sensor Networks: Centralized And Decentralized Approaches". In: *Int. Conf. Distributed Comput. Sensor Syst.* Springer. 173–186.

[159] Khajehnejad, M., W. Xu, A. Avestimehr, and B. Hassibi. 2011. "Analyzing Weighted ℓ_1 Minimization for Sparse Recovery with Nonuniform Sparse Models". *IEEE Trans. Signal Processing*. 59(5): 1985–2001.

[160] Kim, S.-J., E. Dall'Anese, J. A. Bazerque, K. Rajawat, and G. B. Giannakis. 2014. "Advances in Spectrum Sensing and Cross-Layer Design for Cognitive Radio Networks". In: *Academic Press Library in Signal Processing: Volume 2 Communications and Radar Signal Processing R. Chellapa and S. Theodoridis, Eds.* Academic Press. Chap. 9. 471–502.

[161] Kim, S.-J., E. Dall'Anese, and G. B. Giannakis. 2011. "Cooperative Spectrum Sensing for Cognitive Radios Using Kriged Kalman Filtering". *IEEE J. Select. Topics Signal Processing*. 5(1): 24–36.

[162] Kipnis, A., Y. C. Eldar, and A. J. Goldsmith. 2018a. "Analog-to-Digital Compression: A New Paradigm for Converting Signals to Bits". *IEEE Signal Processing Mag.* 35(3): 16–39.

[163] Kipnis, A., A. J. Goldsmith, T. Weissman, and Y. C. Eldar. 2016. "Distortion Rate Function of Sub-Nyquist Sampled Gaussian Sources". *IEEE Trans. Inform. Theory*. 62(1): 401–429.

[164] Kipnis, A., G. Reeves, and Y. C. Eldar. 2018b. "Single Letter Formulas for Quantized Compressed Sensing with Gaussian Codebooks". In: *Proc. IEEE Int. Symp. Inform. Theory*. Vail, CO, USA. 71–75.

[165] Kipnis, A., G. Reeves, Y. Eldar, and A. Goldsmith. 2017. "Compressed Sensing Under Optimal Quantization". In: *Proc. IEEE Int. Symp. Inform. Theory*. Aachen, Germany. 2148–2152.

[166] Kipnis, A., S. Rini, and A. J. Goldsmith. 2015. "The Indirect Rate-Distortion Function of a Binary i.i.d Source". In: *Proc. IEEE Inform. Theory Workshop*. Jeju Island, Korea. 352–356.

[167] Knight, K. and W. J. Fu. 2000. "Asymptotics for Lasso-Type Estimators". *Annals of Statistics*. 28: 1356–1378.

[168] Ku, G., J. Ren, and J. M. Walsh. 2015. "Computing the Rate Distortion Region for the CEO Problem With Independent Sources". *IEEE Trans. Signal Processing*. 63(3): 567–575.

[169] Kushner, H. and G. G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY: Springer-Verlag.

[170] Lalos, A. S., A. Antonopoulos, E. Kartsakli, M. D. Renzo, S. Tennina, L. Alonso, and C. Verikoukis. 2015. "RLNC-Aided Cooperative Compressed Sensing for Energy Efficient Vital Signal Telemonitoring". *IEEE Trans. Wireless Commun.* 14(7): 3685–3699.

[171] Lee, D., S. Kim, and G. B. Giannakis. 2017. "Channel Gain Cartography for Cognitive Radios Leveraging Low Rank and Sparsity". *IEEE Trans. Wireless Commun.* 16(9): 5953–5966.

[172] Lee, S., S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega. 2009a. "Compressed Sensing and Routing in Sensor Networks". *Tech. rep.* USC.

[173] Lee, S., S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega. 2009b. "Spatially-Localized Compressed Sensing and Routing in Multi-hop Sensor Networks". In: *Proceedings of the 3rd International Conference on GeoSensor Networks*. Oxford, UK. 11–20.

[174] Leiner, B. and R. Gray. 1974. "Rate-Distortion Theory for Ergodic Sources with Side Information (Corresp.)" *IEEE Trans. Inform. Theory*. 20(5): 672–675.

[175] Leinonen, M., M. Codreanu, and M. Juntti. 2013. "Distributed Correlated Data Gathering in Wireless Sensor Networks via Compressed Sensing". In: *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* Pacific Grove, CA. 418–422.

[176] Leinonen, M., M. Codreanu, and M. Juntti. 2014. "Compressed Acquisition and Progressive Reconstruction of Multi-Dimensional Correlated Data in Wireless Sensor Networks". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Florence, Italy. 6449–6453.

[177] Leinonen, M., M. Codreanu, and M. Juntti. 2015a. "Channel-Robust Compressed Sensing via Vector Pre-Quantization in Wireless Sensor Networks". In: *Proc. IEEE Global Conf. on Signal and Inform. Proc.* Orlando, Florida, USA. 383–387.

[178] Leinonen, M., M. Codreanu, and M. Juntti. 2015b. "Sequential Compressed Sensing With Progressive Signal Reconstruction in Wireless Sensor Networks". *IEEE Trans. Wireless Commun.* 14(3): 1622–1635.

[179] Leinonen, M., M. Codreanu, and M. Juntti. 2016a. "Distributed Variable-Rate Quantized Compressed Sensing in Wireless Sensor Networks". In: *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.,* invited paper. Edinburgh, UK. 1–5.

[180] Leinonen, M., M. Codreanu, and M. Juntti. 2018a. "Distributed Distortion-Rate Optimized Compressed Sensing in Wireless Sensor Networks". *IEEE Trans. Commun.* 66(4): 1609–1623.

[181] Leinonen, M., M. Codreanu, and M. Juntti. 2019a. "Practical Compression Methods for Quantized Compressed Sensing". In: *Proc. IEEE Int. Conf. on Comp. Commun.* Paris, France. 5.

[182] Leinonen, M., M. Codreanu, and M. Juntti. 2019b. "Signal Reconstruction Performance under Quantized Noisy Compressed Sensing". In: *Proc. IEEE Data Compr. Conf.* Snowbird, Utah. 586.

[183] Leinonen, M., M. Codreanu, M. Juntti, and G. Kramer. 2016b. "Rate-Distortion Lower Bound for Compressed Sensing via Conditional Remote Source Coding". In: *Proc. IEEE Inform. Theory Workshop*. Cambridge, UK. 275–279.

[184] Leinonen, M., M. Codreanu, M. Juntti, and G. Kramer. 2018b. "Rate-Distortion Performance of Lossy Compressed Sensing of Sparse Sources". *IEEE Trans. Commun.* 66(10): 4498–4512.

[185] Levin, D. A., Y. Peres, and W. L. Wilmer. 2008. *Markov Chains and Mixing Times*. 1st ed. Providence, Rhode Island: American Mathematical Society.

[186] Li, S., L. Xu, and X. Wang. 2013. "Compressed Sensing Signal and Data Acquisition in Wireless Sensor Networks and Internet of Things". *IEEE Trans. Ind. Informat.* 9(4): 2177–2186.

[187] Lin, Z., A. Ganesh, J. Right, L. Wu, M. Chen, and Y. Ma. 2009. "Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix". *UIUC Technical Report UILU-ENG-09-2214*: 1–18.

[188] Linde, Y., A. Buzo, and R. Gray. 1980. "An Algorithm for Vector Quantizer Design". *IEEE Trans. Commun.* 28(1): 84–95.

[189] Linder, T., R. Zamir, and K. Zeger. 2000. "On Source Coding with Side-Information-Dependent Distortion Measures". *IEEE Trans. Inform. Theory*. 46(7): 2697–2704.

[190] Ling, Q. and Z. Tian. 2010. "Decentralized Sparse Signal Recovery for Compressive Sleeping Wireless Sensor Networks". *IEEE Trans. Signal Processing*. 58(7): 3816–3827.

[191] Liu, Z., S. Cheng, A. D. Liveris, and Z. Xiong. 2006. "Slepian-Wolf Coded Nested Lattice Quantization for Wyner-Ziv Coding: High-Rate Performance Analysis and Code Design". *IEEE Trans. Inform. Theory*. 52(10): 4358–4379.

[192] Lloyd, S. 1982. "Least Squares Quantization in PCM". *IEEE Trans. Inform. Theory*. 28(2): 129–137.

[193] Löfberg, J. 2006. "Yalmip: Software for Solving Convex (and Non-convex) Optimization Problems". In: *Proc. Amer. Contr. Conf.* Minneapolis, MN.

[194] Lookabaugh, T. and R. Gray. 1989. "High-Resolution Quantization Theory and The Vector Quantizer Advantage". *IEEE Trans. Inform. Theory*. 35(5): 1020–1033.

[195] Lu, W. and N. Vaswani. 2012. "Regularized Modified BPDN for Noisy Sparse Reconstruction With Partial Erroneous Support and Signal Value Knowledge". *IEEE Trans. Signal Processing*. 60(1): 182–196.

[196] Luo, C., F. Wu, J. Sun, and C. Chen. 2009. "Compressive Data Gathering for Large-Scale Wireless Sensor Networks". In: *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*. Beijing, China. 145–156.

[197] Luo, C., F. Wu, J. Sun, and C. Chen. 2010. "Efficient Measurement Generation and Pervasive Sparsity for Compressive Data Gathering". *IEEE Trans. Wireless Commun.* 9(12): 3728–3738.

[198] M. Godavarti, A. O. H. I. 2005. "Partial Update LMS Algorithms". *IEEE Trans. Signal Processing*. 53(7): 2382–2399.

[199] Mairal, J., F. Bach, J. Ponce, and G. Sapiro. 2010. "Online Learning for Matrix Factorization and Sparse Coding". *J. Machine Learning Res.* 11(Jan.): 19–60.

[200] Malioutov, D. M., S. R. Sanghavi, and A. S. Willsky. 2010. "Sequential Compressed Sensing". *IEEE J. Select. Topics Signal Processing*. 4(2): 435–444.

[201] Malioutov, D., M. Çetin, and A. S. Willsky. 2005. "A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays". *IEEE Trans. Signal Processing*. 53(8): 3010–3022.

[202] Mallat, S. 2008. *A Wavelet Tour of Signal Processing*. 3rd. Academic Press.

[203] Mallat, S. G. and Z. Zhang. 1993. "Matching Pursuits with Time-Frequency Dictionaries". *IEEE Trans. Signal Processing*. 41(12): 3397–3415.

[204] Mamaghanian, H., N. Khaled, D. Atienza, and P. Vandergheynst. 2011. "Compressed Sensing for Real-Time Energy-Efficient ECG Compression on Wireless Body Sensor Nodes". *IEEE Trans. Biomed. Eng.* 58(9): 2456–2466.

[205] Mardani, M., G. Mateos, and G. B. Giannakis. 2013a. "Dynamic Anomalography: Tracking Network Anomalies Via Sparsity and Low Rank". *IEEE J. Select. Topics Signal Processing*. 7(1): 50–66.

[206] Mardani, M., G. Mateos, and G. B. Giannakis. 2013b. "Dynamic Anomalography: Tracking Network Anomalies via Sparsity and Low Rank". *IEEE J. Select. Topics Signal Processing*. 7(1): 50–66.

[207] Mardani, M., G. Mateos, and G. B. Giannakis. 2013c. "Recovery of Low-Rank Plus Compressed Sparse Matrices with Application to Unveiling Traffic Anomalies". *IEEE Trans. Inform. Theory*. 59(8): 5186–5205.

[208] Mardani, M., G. Mateos, and G. B. Giannakis. 2015. "Subspace Learning and Imputation for Streaming Big Data Matrices and Tensors". *IEEE Trans. Signal Processing*. 63(10): 2663–2677.

[209] Markovsky, I. and S. Van Huffel. 2007. "Overview of Total Least-Squares Methods". *Signal Processing*. 87(10): 2283–2302.

[210] Masiero, R., G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi. 2009. "Data Acquisition Through Joint Compressive Sensing and Principal Component Analysis". In: *Proc. IEEE Global Telecommun. Conf.* Honolulu, HI. 1–6.

[211] Max, J. 1960. "Quantizing for Minimum Distortion". *Proc. IRE*. 6(1): 7–12.

[212] Mehanna, O. and N. D. Sidiropoulos. 2013. "Frugal Sensing: Wideband Power Spectrum Sensing From Few Bits". *IEEE Trans. Signal Processing*. 61(10): 2693–2703.

[213] Mehmood, Y., F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani. 2017. "Internet-of-Things-Based Smart Cities: Recent Advances and Challenges". *IEEE Commun. Mag.* 55(9): 16–24.

[214] Miosso, C., R. von Borries, M. Argàez, L. Velazquez, C. Quintero, and C. Potes. 2009. "Compressive Sensing Reconstruction with Prior Information by Iteratively Reweighted Least-Squares". *IEEE Trans. Signal Processing*. 57(6): 2424–2431.

[215] Mishali, M., Y. C. Eldar, O. Dounaevsky, and E. Shoshan. 2011a. "Xampling: Analog to Digital at Sub-Nyquist Rates". *IET Circ. Dev. Syst.* 5(1): 8–20.

[216] Mishali, M., Y. C. Eldar, and A. J. Elron. 2011b. "Xampling: Signal Acquisition and Processing in Union of Subspaces". *IEEE Trans. Signal Processing*. 59(10): 4719–4734.

[217] Mishali, M. and Y. Eldar. 2011. "Sub-Nyquist Sampling". *IEEE Signal Processing Mag.* 28(6): 98–124.

[218] Mishali, M. and Y. Eldar. 2012. "Xampling: Compressed Sensing of Analog Signals". In: *Compressive Sensing: Theory and Applications*. Ed. by Y. C. Eldar and G. Kutyniok. Cambridge University Press. Chap. 3. 88–147.

[219] Moshtaghpour, A., M. A. Akhaee, and M. Attarifar. 2016. "Obstacle Mapping in Wireless Sensor Networks via Minimum Number Of Measurements". *IET Sig. Proc.* 10(3): 237–246.

[220] Mostofi, Y. 2011. "Compressive Cooperative Sensing and Mapping in Mobile Networks". 10(12): 1769–1784.

[221] Mota, J. F. C., N. Deligiannis, and M. R. D. Rodrigues. 2017. "Compressed Sensing With Prior Information: Strategies, Geometry, and Bounds". *IEEE Trans. Inform. Theory*. 63(7): 4472–4496.

[222] Msechu, E. J. and G. B. Giannakis. 2012. "Sensor-Centric Data Reduction for Estimation With WSNs via Censoring and Quantization". *IEEE Trans. Signal Processing*. 60(1): 400–414.

[223] Needell, D. 2009. "Noisy Signal Recovery via Iterative Reweighted L1-Minimization". In: *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* Pacific Grove, CA. 113–117.

[224] Needell, D. and J. A. Tropp. 2009. "CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples". *Applied and Computational Harmonic Analysis*. 26(3): 301–321.

[225] Nestares, O., D. J. Fleet, and D. J. Heeger. 2000. "Likelihood Functions and Confidence Bounds for Total-Least-Squares Problems". In: *Proc. of Conf. on Computer Vision and Pattern Recognition*. Hilton Head Island, SC.

[226] Nesterov, Y. 2004. *Introductory Lectures on Convex optimization : A Basic Course*. Boston, MA: Kluwer Academic.

[227] Nyquist, H. 1928. "Certain Topics in Telegraph Transmission Theory". *Trans. Amer. Inst. Elec. Eng.* 47(2): 617–644.

[228] Önhon, N. Ö. and M. Çetin. 2009. "A Nonquadratic Regularization-Based Technique for Joint SAR Imaging and Model Error Correction". In: *Proc. of SPIE*. Vol. 7337.

[229] Oppenheim, A. and G. Verghese. 2015. *Signals, Systems and Inference*. NJ, USA: Prentice Hall. 608.

[230] Palzer, L. and R. Timo. 2016. "A Lower Bound for the Rate-Distortion Function of Spike Sources that is Asymptotically Tight". In: *Proc. IEEE Inform. Theory Workshop*. Cambridge, UK. 1–5.

[231] Pati, Y. C., R. Rezaiifar, and P. S. Krishnaprasad. 1993. "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition". In: *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* Pacific Grove, CA. 40–44.

[232] Picard, R. R. and R. D. Cook. 1984. "Cross-Validation of Regression Models". *Journal of the American Stat. Assoc.* 79(387): 575–583.

[233] Pradhan, S. S., J. Kusuma, and K. Ramchandran. 2002. "Distributed Compression in a Dense Microsensor Network". *IEEE Signal Processing Mag.* 19(2): 51–60.

[234] Pradhan, S. S. and K. Ramchandran. 2003. "Distributed Source Coding Using Syndromes (DISCUS): Design and Construction". *IEEE Trans. Inform. Theory*. 49(3): 626–643.

[235] Protter, M., I. Yavneh, and M. Elad. 2010. "Closed-Form MMSE Estimation for Signal Denoising Under Sparse Representation Modeling Over a Unitary Dictionary". *IEEE Trans. Signal Processing*. 58(7): 3471–3484.

[236] Qin, Z., D. Goldfarb, and S. Ma. 2015. "An Alternating Direction Method For Total Variation Denoising". *Optim. Method Softw.* 30(3): 594–615.

[237] Quer, G., R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi. 2009. "On the Interplay Between Routing and Signal Representation for Compressive Sensing in Wireless Sensor Networks". In: *Information Theory and Applications Workshop*. San Diego, CA. 206–215.

[238] Quer, G., R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi. 2012. "Sensing, Compression, and Recovery for WSNs: Sparse Signal Modeling and Monitoring Framework". *IEEE Trans. Wireless Commun.* 11(10): 3447–3461.

[239] Raghunathan, V., C. Schurgers, S. Park, and M. B. Srivastava. 2002. "Energy-Aware Wireless Microsensor Networks". *IEEE Signal Processing Mag.* 19(2): 40–50.

[240] Rani, M., S. B. Dhok, and R. B. Deshmukh. 2018. "A Systematic Review of Compressive Sensing: Concepts, Implementations and Applications". *IEEE Acc.* 6: 4875–4894.

[241] Rasmussen, C. and C. Williams. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

[242] Razaviyayn, M., M. Sanjabi, and Z. Luo. 2016. "A Stochastic Successive Minimization Method for Nonsmooth Nonconvex Optimization with Applications to Transceiver Design in Wireless Communication Networks". *Math. Program.* 157(2): 515–545.

[243] Rebollo-Monedero, D. 2007. "Quantization and Transforms for Distributed Source Coding". *PhD thesis*. Dept. Elect. Eng., Stanford University.

[244] Rebollo-Monedero, D., R. Zhang, and B. Girod. 2003. "Design of Optimal Quantizers for Distributed Source Coding". In: *Proc. IEEE Data Compr. Conf.* Snowbird, UT, USA. 13–22.

[245] Recht, B., M. Fazel, and P. A. Parrilo. 2010. "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization". *SIAM Rev.* 52(3): 471–501.

[246] Reeves, G. and M. Gastpar. 2012. "The Sampling Rate-Distortion Tradeoff for Sparsity Pattern Recovery in Compressed Sensing". *IEEE Trans. Inform. Theory*. 58(5): 3065–3092.

[247] Rissanen, J. and G. Langdon. 1979. "Arithmetic Coding". *IBM Journal of Research and Development*. 23(2): 149–162.

[248] Rivenson, Y. and A. Stern. 2009. "Compressed Imaging With a Separable Sensing Operator". *IEEE Signal Processing Lett.* 16(6): 449–452.

[249] Rose, K. 1994. "A Mapping Approach to Rate-Distortion Computation and Analysis". *IEEE Trans. Inform. Theory*. 40(6): 1939–1952.

[250] Rose, K., E. Gurewitz, and G. C. Fox. 1992. "Vector Quantization by Deterministic Annealing". *IEEE Trans. Inform. Theory*. 38(4): 1249–1257.

[251] Rose, K. and D. Miller. 1993. "A Deterministic Annealing Algorithm for Entropy-Constrained Vector Quantizer Design". In: *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* Pacific Grove, CA. 1651–1655.

[252] Rudin, L. I., S. Osher, and E. Fatemi. 1992. "Nonlinear Total Variation Based Noise Removal Algorithms". *Phys. D: Nonlin. Phenom.* 60(1): 259–268.

[253] Sakrison, D. 1968. "Source Encoding in the Presence of Random Disturbance (corresp.)" *IEEE Trans. Inform. Theory*. 14(1): 165–167.

[254] Sayed, A. H. 2003. *Fundamentals of Adaptive Filtering*. John Wiley & Sons.

[255] Scharf, L. L. and B. Friedlander. 1994. "Matched Subspace Detectors". *IEEE Trans. Signal Processing*. 42(8): 2146–2157.

[256] Schizas, I. D., G. B. Giannakis, and N. Jindal. 2008. "Distortion-Rate Bounds for Distributed Estimation Using Wireless Sensor Networks". *EURASIP J. Advances Signal Processing*. 1: 12. article ID: 748605.

[257] Schniter, P., L. C. Potter, and J. Ziniel. 2008. "Fast Bayesian Matching Pursuit". In: *Proc. Inform. Theory and Appl. Workshop*. San Diego, CA, USA. 326–333.

[258] Schwarz, G. 1978. "Estimating the Dimension of a Model". *Annals of Statistics*. 6(2): 461–464.

[259] Shannon, C. E. 1949. "Communication in the Presence of Noise". *Proc. IRE*. 37(1): 10–21.

[260] Shannon, C. 1959. "Coding Theorems for a Discrete Source With a Fidelity Criterion". *IRE Nat. Conv. Rec*. 7: 142–163.

[261] Sharma, S. K., E. Lagunas, S. Chatzinotas, and B. Ottersten. 2016. "Application of Compressive Sensing in Cognitive Radio Communications: A Survey". *IEEE Commun. Surveys & Tutorials*. 18(3): 1838–1860.

[262] Shirazinia, A. 2014. "Source and Channel Coding for Compressed Sensing and Control". *PhD thesis*. KTH Royal Institute of Technology. URL: http://kth.diva-portal.org/smash/get/diva2:709596/FULLTEXT01.pdf.

[263] Shirazinia, A., S. Chatterjee, and M. Skoglund. 2013. "Analysis-by-Synthesis Quantization for Compressed Sensing Measurements". *IEEE Trans. Signal Processing*. 61(22): 5789–5800.

[264] Shirazinia, A., S. Chatterjee, and M. Skoglund. 2014a. "Distributed Quantization for Compressed Sensing". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Florence, Italy. 6439–6443.

[265] Shirazinia, A., S. Chatterjee, and M. Skoglund. 2014b. "Joint Source-Channel Vector Quantization for Compressed Sensing". *IEEE Trans. Signal Processing*. 62(14): 3667–3681.

[266] Shirazinia, A., S. Chatterjee, and M. Skoglund. 2015. "Distributed Quantization for Measurement of Correlated Sparse Sources over Noisy Channels". available at http://arxiv.org/abs/1404.7640v2.

[267] Slavakis, K., S.-J. Kim, G. Mateos, and G. B. Giannakis. 2014. "Stochastic Approximation vis-a-vis Online Learning for Big Data Analytics". *IEEE Signal Processing Mag.* 31(Nov.).

[268] Slepian, D. and J. Wolf. 1973. "Noiseless Coding of Correlated Information Sources". *IEEE Trans. Inform. Theory*. 19(4): 471–480.

[269] Solo, V. and X. Kong. 1995. *Adaptive Signal Processing Algorithms: Stability and Performance*. Upper Saddle River, NJ: Prentice-Hall.

[270] Stoica, P. and P. Babu. 2012. "SPICE and LIKES: Two Hyperparameter-Free Methods for Sparse-Parameter Estimation". *Signal Processing, Elsevier*. 92(7): 1580–1590.

[271] Sturm, J. F. 1999. "Using SeDuMi 1.02, a MATLAB Toolbox for Optimization over Symmetric Cones". *Optimization Methods Software*. 11–12(Aug.): 625–653.

[272] Subramaniam, A. D. and B. D. Rao. 2003. "PDF Optimized Parametric Vector Quantization of Speech Line Spectral Frequencies". *IEEE Trans. Speech Audio Processing*. 11(2): 130–142.

[273] Sun, J. and V. Goyal. 2009. "Optimal Quantization of Random Measurements in Compressed Sensing". In: *Proc. IEEE Int. Symp. Inform. Theory*. Seoul, Korea. 6–10.

[274] Tang, G. and A. Nehorai. 2010. "Performance Analysis for Sparse Support Recovery". *IEEE Trans. Inform. Theory*. 56(3): 1383–1399.

[275] Taubock, G. and F. Hlawatsch. 2008. "A Compressed Sensing Technique for OFDM Channel Estimation in Mobile Environments: Exploiting Channel Sparsity for Reducing Pilots". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Las Vegas, NV.

[276] Tian, C. and J. Chen. 2009. "Remote Vector Gaussian Source Coding With Decoder Side Information Under Mutual Information and Distortion Constraints". *IEEE Trans. Inform. Theory*. 55(10): 4676–4680.

[277] Tian, Z., G. Leus, and V. Lottici. 2009. "Detection of Sparse Signals under Finite-Alphabet Constraints". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Taipei, Taiwan. 2349–2352.

[278] Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society*. 58(1): 267–288.

[279] Tipping, M. E. 2001. "Sparse Bayesian Learning and the Relevance Vector Machine". *Journal of Machine Learning Research*. 1(Sept.): 211–244.

[280] Tseng, P. 2001. "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization". *J. Opt. Theory Appl.* 109(3): 475–494.

[281] Turek, J. S., I. Yavneh, and M. Elad. 2011. "On MMSE and MAP Denoising Under Sparse Representation Modeling Over a Unitary Dictionary". *IEEE Trans. Signal Processing*. 59(8): 3526–3535.

[282] Valenzuela, R. A. 1994. "Ray Tracing Prediction of Indoor Radio Propagation". In: *Proc. IEEE 5th Int. Symp. on Personal Indoor and Mobile Radio Comm.* The Hague, Netherlands. 140–144.

[283] Van Huffel, S. and J. Vandewalle. 1991. *The Total Least Squares Problem: Computational Aspects and Analysis*. Vol. 9. *Frontier in Applied Mathematics*. Philadelphia: SIAM.

[284] Vaswani, N. 2008. "Kalman Filtered Compressed Sensing". In: *Proc. IEEE Int. Conf. Image Proc.* San Diego, CA. 893–896.

[285] Vaswani, N. 2010. "LS-CS-Residual (LS-CS): Compressive Sensing on Least Squares Residual". *IEEE Trans. Signal Processing*. 58(8): 4108–4120.

[286] Vaswani, N. and W. Lu. 2010. "Modified-CS: Modifying Compressive Sensing for Problems With Partially Known Support". *IEEE Trans. Signal Processing*. 58(9): 4595–4607.

[287] Vaswani, N. and W. Lu. 2014. "Recursive Reconstruction of Sparse Signal Sequences". In: *Compressed Sensing & Sparse Filtering*. Ed. by A. Y. Carmi, L. Mihaylova, and S. J. Godsill. *Signals and Communication Technology*. Springer Berlin Heidelberg. Chap. 11. 357–380.

[288] Verdu, S. 1998. *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press.

[289] Vetterli, M., P. Marziliano, and T. Blu. 2002. "Sampling Signals with Finite Rate of Innovation". *IEEE Trans. Signal Processing*. 50(6): 1417–1428.

[290] Wainwright, M. J. 2009. "Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ℓ_1-Constrained Quadratic Programming (Lasso)". *IEEE Trans. Inform. Theory*. 55(5): 2183–2202.

[291] Wang, J., S. Tang, B. Yin, and X.-Y. Li. 2012. "Data Gathering in Wireless Sensor Networks through Intelligent Compressive Sensing". In: *Proc. IEEE Int. Conf. on Comp. Commun.* 603–611.

[292] Wang, W., M. Garofalakis, and K. Ramchandran. 2007. "Distributed Sparse Random Projections for Refinable Approximation". In: *Proc. IEEE Int. Symp. on Inform. Proc. in Sensor Networks*. 331–339.

[293] Weidmann, C. and M. Vetterli. 2012. "Rate Distortion Behavior of Sparse Sources". *IEEE Trans. Inform. Theory*. 58(8): 4969–4992.

[294] Wernersson, N., J. Karlsson, and M. Skoglund. 2009. "Distributed Quantization over Noisy Channels". *IEEE Trans. Commun.* 57(6): 1693–1700.

[295] White, G. P., Y. V. Zakharov, and J. Liu. 2008. "Low-Complexity RLS Algorithms using Dichotomous Coordinate Descent Iterations". *IEEE Trans. Signal Processing*. 56(July): 3150–3161.

[296] Wijewardhana, U. L. and M. Codreanu. 2017. "A Bayesian Approach for Online Recovery of Streaming Signals From Compressive Measurements". *IEEE Trans. Signal Processing*. 65(1): 184–199.

[297] Wilson, J. and N. Patwari. 2010. "Radio Tomographic Imaging with Wireless Networks". 9(5): 621–632.

[298] Wilson, J. and N. Patwari. 2011. "See-Through Walls: Motion Tracking Using Variance-Based Radio Tomography Networks". 10(5): 612–621.

[299] Wimalajeewa, T., Y. C. Eldar, and P. K. Varshney. 2015. "Subspace Recovery From Structured Union of Subspaces". *IEEE Trans. Inform. Theory*. 61(4): 2101–2114.

[300] Wimalajeewa, T. and P. K. Varshney. 2014. "OMP Based Joint Sparsity Pattern Recovery Under Communication Constraints". *IEEE Trans. Signal Processing*. 62(19): 5059–5072.

[301] Wipf, D. P. and B. D. Rao. 2004. "Sparse Bayesian Learning for Basis Selection". *IEEE Trans. Signal Processing*. 52(8): 2153–2164.

[302] Wipf, D. and S. Nagarajan. 2010. "Iterative Reweighted l_1 and l_2 Methods for Finding Sparse Solutions". *IEEE J. Select. Topics Signal Processing*. 4(2): 317–329.

[303] Witsenhausen, H. 1980. "Indirect Rate Distortion Problems". *IEEE Trans. Inform. Theory*. 26(5): 518–521.

[304] Wolf, J. and J. Ziv. 1970. "Transmission of Noisy Information to a Noisy Receiver with Minimum Distortion". *IEEE Trans. Inform. Theory*. 16(4): 406–411.

[305] Wölfle, G., R. Wahl, P. Wertz, P. Wildbolz, and F. Landstorfer. 2005. "Dominant Path Prediction Model for Indoor Scenarios". In: *Proc. German Microwave Conf.* 176–179.

[306] Wood, R. 1969. "On Optimum Quantization". *IEEE Trans. Inform. Theory*. 15(2): 248–252.

[307] Wright, J., A. Ganesh, K. Min, and Y. Ma. 2013. "Compressive Principal Component Pursuit". *Information and Inference*. 2(1): 32–68.

[308] Wu, T. T. and K. Lange. 2008. "Coordinate Descent Algorithms for Lasso Penalized Regression". *Annals of Applied Statistics*. 2(Mar.): 224–244.

[309] Xiang, L., J. Luo, and C. Rosenberg. 2013. "Compressed Data Aggregation: Energy-Efficient and High-Fidelity Data Collection". *IEEE/ACM Trans. Networking*. 21(6): 1722–1735.

[310] Xiao, J.-J., A. Ribeiro, Z.-Q. Luo, and G. Giannakis. 2006. "Distributed Compression-Estimation Using Wireless Sensor Networks". *IEEE Signal Processing Mag.* 23(4): 27–41.

[311] Xiong, Z., A. Liveris, and S. Cheng. 2004. "Distributed Source Coding for Sensor Networks". *IEEE Signal Processing Mag.* 21(5): 80–94.

[312] Xu, W. and B. Hassibi. 2012. "Fundamental Thresholds in Compressed Sensing: A High-Dimensional Geometry Approach". In: *Compressive Sensing: Theory and Applications*. Ed. by Y. C. Eldar and G. Kutyniok. Cambridge University Press. Chap. 7. 305–347.

[313] Yang, G., V. Y. F. Tan, C. K. Ho, S. H. Ting, and Y. L. Guan. 2013. "Wireless Compressive Sensing for Energy Harvesting Sensor Nodes". *IEEE Trans. Signal Processing*. 61(18): 4491–4505.

[314] Yuan, M. and Y. Lin. 2006a. "Model Selection and Estimation in Regression with Grouped Variables". *J. Royal Stat. Soc – Series B*. 68(1): 49–67.

[315] Yuan, M. and Y. Lin. 2006b. "Model Selection and Estimation in Regression with Grouped Variables". *Journal Royal Stat. Society: Series B*. 68: 49–67.

[316] Zanella, A., N. Bui, A. Castellani, L. Vangelista, and M. Zorzi. 2014. "Internet of Things for Smart Cities". *IEEE Internet of Things J.* 1(1): 22–32.

[317] Zeger, K. and A. Gersho. 1988. "Vector Quantizer Design for Memoryless Noisy Channels". In: *Proc. IEEE Int. Conf. Commun.* Philadelphia, PA, USA. 1593–1597.

[318] Zhan, J. and N. Vaswani. 2015. "Time Invariant Error Bounds for Modified-CS based Sparse Signal Sequence Recovery". *IEEE Trans. Inform. Theory*. 61(3): 1389–1409.

[319] Zhao, Q. and M. Effros. 2003a. "Lossless and Near-Lossless Source Coding for Multiple Access Networks". *IEEE Trans. Inform. Theory*. 49(1): 112–128.

[320] Zhao, Q. and M. Effros. 2003b. "Low Complexity Code Design for Lossless and Near-Lossless Side Information Source Codes". In: *Proc. IEEE Data Compr. Conf.* 3–12.

[321] Zheng, L., A. Maleki, H. Weng, X. Wang, and T. Long. 2017. "Does ℓ_p-Minimization Outperform ℓ_1-Minimization?" *IEEE Trans. Inform. Theory*. 63(11): 6896–6935.

[322] Zhu, H. and G. B. Giannakis. 2011. "Exploiting Sparse User Activity in Multiuser Detection". *IEEE Trans. Commun.* 59(2): 454–465.

[323] Zhu, H., G. B. Giannakis, and G. Leus. 2011a. "Weighted and Structured Sparse Total Least-Squares for Perturbed Compressive Sampling". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Prague, Czech Republic,

[324] Zhu, H., G. Leus, and G. B. Giannakis. 2011b. "Sparsity-Cognizant Total Least-Squares for Perturbed Compressive Sampling". *IEEE Trans. Signal Processing*. 59(5): 2002–2016.

[325] Zhu, J., D. Baron, and F. Krzakala. 2017. "Performance Limits for Noisy Multimeasurement Vector Problems". *IEEE Trans. Signal Processing*. 65(9): 2444–2454.

[326] Ziniel, J. and P. Schniter. 2010. "Tracking and Smoothing of Time-Varying Sparse Signals via Approximate Belief Propagation". In: *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* Pacific Grove, CA. 808–812.

[327] Ziniel, J. and P. Schniter. 2013. "Dynamic Compressive Sensing of Time-Varying Signals Via Approximate Message Passing". *IEEE Trans. Signal Processing*. 61(21): 5270–5284.

[328] Ziv, J. 1985. "On Universal Quantization". *IEEE Trans. Inform. Theory*. 31(3): 344–347.

[329] Ziv, J. and A. Lempel. 1978. "Compression of Individual Sequences via Variable-Rate Coding". *IEEE Trans. Inform. Theory*. 24(5): 530–536.

[330] Zordan, D., G. Quer, M. Zorzi, and M. Rossi. 2011. "Modeling and Generation of Space-Time Correlated Signals for Sensor Network Fields". In: *Proc. IEEE Global Telecommun. Conf.* Houston, Texas, USA. 1–6.

[331] Zou, H. 2006. "The Adaptive Lasso and its Oracle Properties". *J. of the American Stat. Assoc.* Dec.: 1418–1429.

[332] Zou, H. and R. Li. 2008. "One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models". *Annals of Statistics*. 36(4): 1509–1533.

[333] Zymnis, A., S. Boyd, and E. J. Candés. 2010. "Compressed Sensing With Quantized Measurements". *IEEE Signal Processing Lett.* 17(2): 149–152.