

Hamza Djelouat

OPTIMIZING MASSIVE
RANDOM ACCESS:
LEVERAGING
CORRELATION MODELS
AND SPARSE RECOVERY
ALGORITHMS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING



ACTA UNIVERSITATIS OULUENSIS
C Technica 936

HAMZA DJELOUAT

**OPTIMIZING MASSIVE RANDOM
ACCESS: LEVERAGING
CORRELATION MODELS AND
SPARSE RECOVERY ALGORITHMS**

Academic dissertation to be presented with the assent of the Doctoral Programme Committee of Information Technology and Electrical Engineering of the University of Oulu for public defence in the Oulun Puhelin auditorium (L5), Linnanmaa, on 17 April 2024, at 12 noon

UNIVERSITY OF OULU, OULU 2024

Copyright © 2024
Acta Univ. Oul. C 936, 2024

Supervised by
Professor Markku Juntti
Doctor Markus Leinonen

Reviewed by
Professor Krishna Narayanan
Doctor Maxime Guillaud

Opponent
Docent Mikko Vehkaperä

ISBN 978-952-62-4053-4 (Paperback)
ISBN 978-952-62-4054-1 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

PUNAMUSTA
TAMPERE 2024

Djelouat, Hamza, Optimizing Massive Random Access: Leveraging Correlation Models and Sparse Recovery Algorithms

University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering

Acta Univ. Oul. C 936, 2024

University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Massive machine-type communications (mMTC) serve as the backbone for future internet of things (IoT) applications. Characterized by sporadic traffic from a massive number of connected devices, mMTC presents unique challenges for traditional channel access protocols. To address this, grant-free random access has been widely adopted as a key enabler of mMTC. The major advantage of grant-free random access is its ability to eliminate the four-step handshaking protocol used in previous communications systems. However, it introduces pilot collisions that renders device activity detection and channel estimation difficult. The central challenge in implementing grant-free access therefore lies in the joint activity detection and channel estimation (JADCE) problem. Although extensive research exists on solving the JADCE as a sparse recovery, most studies employ overly simplistic assumptions regarding both the structure of the propagation channel and the traffic patterns within mMTC networks.

This thesis addresses the JADCE problem by focusing on two distinct categories of mMTC networks. The first category involves mMTC networks that operate under spatially correlated channels. The aim here is to go beyond the often-used simplistic models in the literature by employing more realistic and practical channel models that take into account the correlated nature of multi-input multi-output (MIMO) channels. The second category considers mMTC networks where device activities are correlated, capturing a more nuanced realization of real-world device interactions within the large-scale IoT network.

To address these issues, this thesis aims to propose, develop and optimize various sparse recovery algorithms for the JADCE solution. These solutions are developed within both deterministic and Bayesian frameworks to account for different levels of available prior information, both in terms of channel statistics and device activity patterns. Furthermore, the thesis goes beyond its algorithmic contributions to provide an in-depth theoretical analysis of the device activity problem.

By leveraging the correlation structures in practical mMTC, this thesis develops multiple solutions that achieve equivalent performance metrics to current state-of-the-art solutions but with reduced signaling overhead and computationally efficient implementation.

Keywords: activity detection, channel estimation, correlated activity pattern, massive machine-type communications, sparse recovery algorithm, spatially correlated channel, sporadic activity

Djelouat, Hamza, Massiivisen satunnaisverkkoliittynän optimointi: korrelaatiomallien ja harvojen palautusalgoritmien hyödyntäminen

Oulun yliopiston tutkijakoulu; Oulun yliopisto, tieto- ja sähkötekniikan tiedekunta

Acta Univ. Oul. C 936, 2024

Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Massiivinen konetyyppinen tietoliikenne (mMTC) toimii tulevaisuuden esineiden internetin (IoT) sovellusten selkärangana. Massiiviselle konetyyppiselle viestinnälle on ominaista satunnainen liikenne valtavasta määrästä yhdistettyjä laitteita. Se tuo uusia haasteita perinteisille kanavayhteysprotokollille. Tämän ratkaisemiseksi skeduloimaton satunnaiskäyttö on mMTC:n keskeinen mahdollistaja. Skeduloimattoman satunnaiskäytön etu on sen kyky poistaa aiemmissa viestintäjärjestelmissä käytetty nelivaiheinen kättelyprotokolla. Se tuo kuitenkin mukanaan pakettitörmäyksiä, jotka vaikeuttavat laitetoiminnan havaitsemista ja kanavaestimointia. Näin ollen keskeinen haaste skeduloimattoman käytön toteuttamisessa on yhteisen toiminnan havaitsemisen ja kanavaestimoinnin (joint activity detection and channel estimation, JADCE) ongelma. Vaikka ongelman ratkaisemista harvana signaalin rekonstruktiona on tutkittu laajasti, useimmissa tutkimuksissa käytetään liian yksinkertaisia oletuksia sekä radiokanavasta että mMTC-verkkojen liikennemalleista.

Tässä väitöskirjassa JADCE-ongelmaa käsitellään keskittymällä kahteen erityiseen mMTC-verkkojen luokkaan. Ensimmäinen luokka koostuu mMTC-verkoista, jotka toimivat tilatasossa korreloivissa kanavissa. Niiden osalta syvennetään kirjallisuudessa usein käytettyjä yksinkertaistettuja malleja hyödyntämällä realistisempia ja käytännöllisempiä kanavamalleja, joissa otetaan huomioon MIMO-kanavien (Multiple-Input Multiple-Output) korreloiva luonne. Toisessa luokassa tarkastellaan mMTC-verkkoja, joissa laitetoiminnot korreloivat keskenään. Siinä otetaan tarkemmin ja monipuolisemmin huomioon todelliset laitevuorovaikutukset suuren mittakaavan IoT-verkossa.

Yllä kuvattujen tutkimuskysymysten selvittämiseksi tämän väitöskirjan tavoitteena on suunnitella, kehittää ja optimoida erilaisia harvaa rekonstruktioita käyttäviä algoritmeja JADCE-ratkaisua varten. Nämä ratkaisut on kehitetty sekä deterministisessä että bayesilaisessa viitekehyksessä, jotta voidaan ottaa huomioon saatavilla olevan tiedon eri tasot sekä kanavastatistiikkojen että laitetoiminnan mallien osalta. Algoritmeja koskevien havaintojen lisäksi väitöskirja tarjoaa syvällisen teoreettisen analyysin laiteaktiivisuuden havaitsemisesta.

Hyödyntämällä korrelaatorakenteita käytännön massiivisessa konetyyppisessä viestinnässä tämä väitöskirja kehittää useita ratkaisuja, joilla saavutetaan vastaava suorituskyky kuin nykyisillä huipputaso ratkaisuilla mutta joilla on pienempi signaloiminnan yleisrasite ja laskennallisesti tehokas toteutus.

Asiasanat: harva rekonstruktio, kanavaestimointi, korreloiva laiteaktiivisuus, laiteaktiivisuuden havaitseminen, massiivinen konetyyppinen viestintä, satunnainen toiminta, tilatasossa korreloiva kanava

To my family

Preface

The research for this thesis was carried out at the Centre for Wireless Communications (CWC) at the University of Oulu in Finland between March 2019 and October 2023.

Praise be to “Allah” for for granting me the faith, the opportunity, and the strength to complete this thesis.

I wish to express my deepest gratitude to my principal supervisor, Prof. Markku Juntti, for providing me with the opportunity to pursue a doctoral degree and for his unwavering confidence in my capabilities throughout my doctoral training. His consistent support and profound technical knowledge have significantly broadened my research perspectives and inspired me to pursue challenging goals.

This thesis is dedicated to the memory of my late co-supervisor, Dr. Markus Leinonen, to whom I owe immense gratitude for his continuous support, valuable guidance, and encouragement during my postgraduate journey. His passing leaves a significant void; however, his academic legacy endures through the work he inspired and the profound impact he had on our field.

I would like to express my deep appreciation to all my co-authors and collaborators for their invaluable contributions. Special thanks are owed to Prof. Mikko J. Silanpää for the fruitful discussions and outstanding collaboration during the final year of my studies. My appreciation goes to the thesis pre-examiners, Dr. Maxime Guillaud from INRIA, France, and Prof. Krishna Narayanan from Texas A&M University, USA. I would also like to extend my gratitude to Dr. Mikko Vehkaperä from Nokia for acting as the prestigious opponents in my doctoral thesis defense. Furthermore, I am deeply grateful to Dr. Italo Atzeni and Dr. Nurul Huda Mahmood for serving as members of my follow-up group and for their continual support. I am also grateful to our administrative staff for assisting me in all sorts of work duties.

I am grateful to my colleagues at the CWC for creating such a supportive and enjoyable research environment over the years. Special thanks are extended to Prof. Mehdi Bennis, Marata, Ehsan, Mohammed, Abolfazel, Saeed, Bikshpatai, Abdelmomen, and Abanoub. Additionally, my sincere appreciation goes to my Algerian friends in Oulu for their delightful company, and to my friends in Qatar– Atman, Rushdy, Hachemi and Badro– for maintaining our friendship across the continents.

This thesis represents the culmination of a long journey that began long before my arrival at the CWC. I am particularly thankful for the friendships I forged at INELEC with Amine, Cherif, Issam, Mohammed, Saleh, Selman, and Tarek. Furthermore, I am deeply grateful to the numerous individuals who have profoundly impacted my

academic career, paving the way for this endeavor. I am eternally grateful to my former advisors, Dr. Khalida Ghanem and Prof. Abbes Amira, whose guidance and mentorship have been invaluable.

Most importantly, I would not have been able to embark on this journey without the support and the encouragement of my family. I am grateful beyond words to my parents for their endless and unconditional support throughout my life, and to my brothers and sisters for always keeping me in their hearts and minds. Last but not least, I wish to express my heartfelt gratitude to my wife, Karima, for her unwavering love and support throughout all the ups and downs of this journey, and to my daughter, Selma, whose presence has been the delight of our lives and a constant source of joy and inspiration. To all my family: I owe all my success to you, and to you, this work is dedicated.

Acknowledgements

The work in this thesis has been supported by the Academy of Finland Flagship program (grant 318927), the ROHM project (grant 319485) and the 6GWiCE project (grant 24304415). Additionally, I am deeply honored to have received several personal research grants from the following foundations (in alphabetical order): The Nokia Foundation, Riitta and Jorma J. Takanen Foundation, and Tauno Tönning Foundation.

List of abbreviations

Acronyms

3GPP	<i>Third-generation partnership project</i>
5G	<i>Fifth generation</i>
6G	<i>Sixth generation</i>
ADMM	<i>Alternating direction method of multipliers</i>
AMP	<i>Approximate message passing</i>
ASD	<i>Angular standard deviation</i>
BS	<i>Base station</i>
BP	<i>Basis pursuit</i>
CDI	<i>Channel distribution information</i>
CDF	<i>Cumulative distribution function</i>
CLM	<i>Central limit theorem</i>
CoSaMP	<i>Compressive sampling matching pursuit</i>
CS	<i>Compressive sensing</i>
CSI	<i>Channel state information</i>
EM	<i>Expectation maximization</i>
eMBB	<i>Enhanced mobile broadband</i>
EP	<i>Expectation propagation</i>
HBOMP	<i>Hierarchical block Orthogonal matching pursuit</i>
HTC	<i>Human-type communications</i>
i.i.d	<i>Independent and identically distributed</i>
ISTA	<i>Iterative soft thresholding algorithm</i>
JADCE	<i>Joint device detection and channel estimation</i>
KL	<i>Kullback-Leibler</i>
LASSO	<i>Least absolute shrinkage and selection operator</i>
LBF	<i>Loopy belief propagation</i>
LS	<i>Least square</i>
MIMO	<i>Multiple input multiple output</i>
MM	<i>Majorization-minimization</i>
MMSE	<i>Minimum mean square error</i>
mmWave	<i>Millimeter-wave</i>
MMV	<i>Multiple measurement vector</i>
MTC	<i>Machine-type communications</i>
MTD	<i>Machine-type device</i>

NMSE	<i>Normalized mean square error</i>
NOMA	<i>Non-orthogonal multiple access</i>
OAMP	<i>Orthogonal AMP</i>
OFDM	<i>Orthogonal frequency division multiple access</i>
PDF	<i>Probability density function</i>
PFA	<i>Probability of false alarm</i>
PMD	<i>Probability of miss detection</i>
SBL	<i>Sparse Bayesian learning</i>
RIS	<i>Reconfigurable intelligent surface</i>
SMV	<i>Single measurement vector</i>
SSR	<i>Sparse signal recovery</i>
SRR	<i>Support recovery rate</i>
QoS	<i>Quality of service</i>

Mathematical symbols and operators:

$\{0, 1\}^n$	<i>the set of binary column vector of length n</i>
\otimes	<i>the Kronecker product</i>
\mathbf{A}	<i>a matrix containing column vectors as $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$</i>
\mathbf{a}_i	<i>the ith column of the matrix \mathbf{A}</i>
\mathbf{A}^T	<i>the transpose of \mathbf{A}</i>
\mathbf{A}^H	<i>the conjugate transpose of \mathbf{A}</i>
\mathbf{A}	<i>the conjugate of \mathbf{A}</i>
\mathbf{A}^{-1}	<i>the inverse of \mathbf{A}</i>
$\mathbf{0}$	<i>a vector with all entries 0</i>
$\mathbf{1}$	<i>a vector with all entries 1</i>
$\ \cdot\ _0$	<i>the operator that counts the number of non-zero element of a vector</i>
$\ \cdot\ _1$	<i>the ℓ_1-norm</i>
$\ \cdot\ _2$	<i>the ℓ_2-norm</i>
$\ \cdot\ _F$	<i>the Frobenius norm</i>
$ \cdot $	<i>absolute value</i>
$\det(\mathbf{A})$	<i>determinant of matrix \mathbf{A}</i>
$\text{vec}(\mathbf{A})$	<i>an operator to stack the matrix $\text{vec}(\mathbf{A}) = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$ into a column vector $\mathbf{a} = [\mathbf{a}_1^T \dots, \mathbf{a}_N^T]^T$</i>
\mathbf{I}_M	<i>the $M \times M$ identity matrix</i>
$ \mathcal{A} $	<i>the cardinality of the set \mathcal{A}</i>
$\text{diag}(\cdot)$	<i>a (block) diagonal matrix</i>
$\mathbb{E}(\cdot)$	<i>the expectation of the argument</i>
$\text{Var}(\cdot)$	<i>the variance of the argument</i>
$\max\{a, b\}$	<i>the maximum value of a and b</i>
\mathbb{C}^M	<i>the set of column vectors of length M</i>
$\mathbb{C}^{M \times N}$	<i>the set of $M \times N$ complex matrices</i>
\mathbb{R}_+	<i>the set of non-negative real numbers</i>
$\text{Tr}(\mathbf{A})$	<i>the trace of matrix \mathbf{A}</i>
$\langle \mathbf{A}, \mathbf{B} \rangle$	<i>the inner product of two matrices \mathbf{A} and \mathbf{B} defined as $\text{Tr}(\mathbf{A}^H \mathbf{B})$</i>
$\mathcal{N}_{\mathbb{C}}(\mathbf{a}; \mathbf{0}, \mathbf{R})$	<i>denotes a circular Gaussian distribution with zero mean and correlation matrix</i>
$\mathcal{B}(a; p_a)$	<i>denote a the Bernoulli distribution with $\text{Pr}(a = 1) = p_a$</i>
χ_2^2	<i>Chi-squared distribution with two degrees of freedom</i>
$\bar{\gamma}$	<i>lower incomplete gamma function.</i>
$\Gamma(\cdot)$	<i>the gamma function</i>
$\text{Pr}(\xi)$	<i>probability of event ξ</i>

$p(\mathbf{x})$	<i>probability density function of random variable \mathbf{x}</i>
$p(\mathbf{x} \mathbf{y})$	<i>conditional probability density function of random variable \mathbf{x} given \mathbf{y}</i>
$p(\mathbf{x}; \boldsymbol{\mu})$	<i>probability density function of random variable \mathbf{x} parameterized by the deterministic variable $\boldsymbol{\mu}$</i>
$\text{KL}(\cdot \cdot)$	<i>KL divergence distance</i>
$\delta(\cdot)$	<i>Dirac function</i>

Roman letters:

a	<i>the ULA steering vector</i>
\mathbf{B}_l	<i>scale matrix in the (inverse) Wishart distribution</i>
C	<i>a local matrix involved in the soft-thresholding operator</i>
c	<i>cluster activity indicator vector</i>
D	<i>a local matrix involved at the \mathbf{Z} update in cov-IRW-ADMM</i>
e	<i>noise term in the equivalent linear model of the AMP algorithm</i>
$f(\cdot)$	<i>the true joint posterior distribution in the EP framework</i>
$f_s(\cdot)$	<i>approximate function for the step function</i>
\mathbf{f}_{c_i}	<i>Taylor approximation of the non-convex term of the ith sub-problem in cov-ADMM</i>
\mathbf{F}_q	<i>Taylor approximation of a quadratic term</i>
f_w	<i>normalization term for the inverse Wishart distribution</i>
g	<i>vector containing the reweighting parameters for the $\ell_{2,1}$-norm</i>
\bar{g}	<i>a function involved in the CDF of a complex quadratic function</i>
\mathbf{h}_i	<i>the channel response of the ith device</i>
$\hat{\mathbf{h}}_i$	<i>estimated channel response of the ith device in the training phase</i>
$\mathbb{I}(\cdot)$	<i>indicator function</i>
i	<i>device index</i>
$J(\cdot)$	<i>the cost function associated with the argument</i>
$J_c(\cdot)$	<i>cluster-sparsity-promoting prior</i>
$J_s(\cdot)$	<i>sparsity-promoting prior</i>
k	<i>the ADMM iteration index</i>
k_{\max}	<i>maximum number of iterations in ADMM</i>
k_c	<i>iteration index in the outer ADMM loop in corr-MAP-ADMM</i>
$k_{c\max}$	<i>maximum number of iterations in the ADMM outer loop in corr-MAP-ADMM</i>
K_c	<i>number of iterations in the outer-loop in the corr-MAP-ADMM</i>
k_u	<i>iteration index in the inner ADMM loop in corr-MAP-ADMM</i>

$k_{u_{\max}}$	<i>maximum number of iterations in the ADMM inner loop in corr-MAP-ADMM</i>
K	<i>number of active devices at each coherence interval</i>
\hat{K}	<i>number of estimated active devices</i>
l	<i>MM iteration index</i>
L	<i>number of devices in each cluster</i>
l_{\max}	<i>maximum number of MM iterations</i>
M	<i>number of BS antennas</i>
\mathbf{m}_i	<i>the posterior mean of the ith element in the approximated posterior distribution in the EP framework</i>
$\mathbf{m}_{1,i}$	<i>the mean of the ith element from the first approximate site in the EP framework</i>
$\bar{\mathbf{m}}_1$	<i>the vector form of the posterior mean of the first approximate site in the EP framework</i>
$\mathbf{m}_{2,i}$	<i>the mean of the ith element from the second approximate site in the EP framework</i>
N	<i>total number of devices</i>
N_c	<i>total number of clusters</i>
P_a	<i>probability of device activity</i>
P_c	<i>probability of cluster activity</i>
p^{FA}	<i>probability of false alarm</i>
p^{MD}	<i>probability of miss detection</i>
P_i	<i>number of signal physical paths in the channel response</i>
$Q(\cdot)$	<i>the global approximation of the joint posterior distribution in the EP framework</i>
$Q^{\setminus 2,l}(\cdot)$	<i>the cavity distribution of the global approximation without the contribution of the lth term in the second approximate site in the EP framework</i>
\mathbf{R}_i	<i>the true channel covariance matrix of the ith device</i>
$\hat{\mathbf{R}}_i$	<i>the estimated channel covariance matrix of the ith device</i>
$\tilde{\mathbf{R}}_i$	<i>the scaled version of the estimated channel covariance matrix of the ith device</i>
$\bar{\mathbf{R}}_i$	<i>positive definite matrix to model the spatial correlation between the entries of \mathbf{x}_i</i>
t	<i>AMP iteration index</i>
T	<i>number of channel responses needed to estimate the second-order statistics at the BS</i>
T_c	<i>coherence interval</i>

\mathbf{u}	<i>vector of auxiliary optimization variables for the log-sum penalty</i>
\mathbf{V}	<i>splitting variable for the ADMM such that</i>
ν_w	<i>parameter to control the degrees of freedom of the inverse Wishart distribution</i>
$\hat{\nu}$	<i>a random variable following a second-order Chi-square distribution</i>
\mathbf{W}	<i>white Gaussian noise at the BS</i>
\mathbf{w}	<i>white Gaussian noise in vector form</i>
\mathbf{x}_i	<i>effective channel of the ith device</i>
\mathbf{X}	<i>effective channel matrix</i>
$\mathbf{X}^{(k)}$	<i>estimated effective channel matrix at the kth iteration</i>
$\bar{\mathbf{x}}$	<i>effective channel in vector form</i>
\mathbf{Y}	<i>received pilot signal matrix</i>
\mathbf{y}	<i>received pilot signal in vector form</i>
\mathbf{Z}	<i>splitting variable for the ADMM</i>
$\tilde{\mathbf{Z}}$	<i>residual matrix for AMP</i>

Greek letters:

α_i	<i>regularization parameter involved in the soft thresholding operator</i>
β_1	<i>sparsity regularization parameter</i>
β_2	<i>positive parameter on the covariance matching term</i>
β_3	<i>positive parameter on the inverse Wishart prior</i>
γ	<i>activity indicator vector</i>
$\tilde{\gamma}$	<i>activity hyper-parameter vector</i>
$\hat{\gamma}$	<i>estimated activity indicator vector</i>
Δ_r	<i>normalized spacing between the adjacent BS antennas</i>
ε	<i>positive parameter that denotes a tolerable error in the LASSO and BP formulations</i>
ε_0	<i>positive stability parameter for the log-sum penalty.</i>
ε_{stp}	<i>a stopping criterion for any particular algorithm</i>
ε_{thr}	<i>a small positive pre-defined threshold for any particular algorithm to detect if the device is active or inactive</i>
ζ_B	<i>damping parameter for generating the prior guess on the average covariance matrix $\mathbf{R}_{\mathcal{G}_i}$</i>
ζ_{EP}	<i>small positive regularization parameter that ensures all covariance matrices within the EP framework as positive definite</i>
$\zeta_{i,p}^{\text{ASD}}$	<i>deviation from the incident angle</i>
$\eta(\cdot, \cdot)$	<i>denoiser function of the Bayesian AMP algorithm</i>

θ	<i>a generic random variable</i>
κ	<i>small positive parameter to adjust the steepness of $f_s(\cdot)$</i>
Λ	<i>denotes the set of dual variable matrices $\{\Lambda_1, \dots, \Lambda_N\}$ in the cov-ADMM algorithm</i>
Λ_z	<i>denotes the set of dual variable matrix in cov-IRW-ADMM, MAP-ADMM and corr-MAP-ADMM algorithms</i>
Λ_v	<i>denotes the set of dual variable matrix in cov-IRW-ADMM, MAP-ADMM and corr-MAP-ADMM algorithms</i>
ρ	<i>ADMM penalty parameter</i>
$\rho_{i,p}^{\text{path}}$	<i>path gain of the pth path from the ith device</i>
ρ_i^{UL}	<i>transmit power assigned to the i device</i>
σ^2	<i>the noise variance at a BS</i>
σ_ψ	<i>Angular standard deviation</i>
Σ_i	<i>the posterior covariance matrix of the ith element in the approximated posterior distribution in the EP framework</i>
$\Sigma_{1,i}$	<i>the covariance matrix of the ith element from the second approximate site in the EP framework</i>
$\bar{\Sigma}_{i,1}$	<i>a block diagonal matrix such that $\bar{\Sigma}_1 = \text{diag}(\Sigma_{1,1}, \dots, \Sigma_{1,N})$</i>
$\Sigma_{2,i}$	<i>the covariance matrix of the ith element from the second approximate site in the EP framework</i>
Σ_θ	<i>AMP's State Evolution matrix</i>
τ	<i>regularization term to stabilize the Taylor approximation in cov-ADMM</i>
τ_c	<i>number of symbols transmitted at each coherence time</i>
τ_s	<i>number of coherence intervals where the second-order statistics are considered to be fixed</i>
τ_p	<i>pilot sequence length</i>
ϕ_i	<i>the pilot sequence of the ith device</i>
Φ	<i>the pilot sequence matrix</i>
Ξ	<i>set of hyper-parameters</i>
$\Psi_{i,p}$	<i>the AoA between the BS and the ith device from the pth path</i>
$\bar{\Psi}_i$	<i>incident angle between the ith device and the BS</i>
Ψ_l	<i>random positive-definite Hermitian matrix to model the error in the prior knowledge of the average covariance matrix $\mathbf{R}_{\mathcal{C}_i}$</i>

Calligraphy letter notations:

\mathcal{A} *the set active devices*

$\hat{\mathcal{A}}$	<i>the set of the estimated active devices obtained by any particular algorithm</i>
\mathcal{B}	<i>denotes the Bernoulli Distribution</i>
\mathcal{C}_l	<i>denotes the set of the devices belonging to the lth cluster.</i>
\mathcal{C}_l	<i>denotes the set of the devices belonging to the lth cluster.</i>
\mathcal{IW}	<i>denotes the inverse Wishart distribution</i>
\mathcal{L}	<i>augmented Lagrangian function</i>
\mathcal{J}	<i>set of estimated active clusters</i>
\mathcal{O}	<i>big \mathcal{O} notation</i>

Contents

Abstract	
Tiivistelmä	
Preface	9
Acknowledgements	11
List of abbreviations	13
Contents	21
1 Introduction	25
1.1 Motivation and significance	25
1.2 Grant-free access in mMTC	27
1.3 Device activity detection and channel estimation	30
1.4 Literature review	32
1.4.1 Compressive sensing overview	32
1.4.2 CS-based JADCE	38
1.5 Thesis objectives	40
1.6 Thesis outline and contributions	42
1.7 Author's contributions	44
2 JADCE under spatially correlated channels: A deterministic approach	45
2.1 System model and problem formulation	45
2.1.1 mMTC uplink system model	45
2.1.2 Pilot assignment	47
2.1.3 Data transmission	47
2.2 JADCE without side information	49
2.2.1 JADCE via group LASSO	49
2.2.2 JADCE via reweighted $\ell_{2,1}$ -norm minimization	50
2.3 JADCE with channel covariance information	51
2.4 Generalized solution via ADMM	53
2.4.1 ADMM via Taylor approximation	53
2.4.2 Exact solution	57
2.5 A note on the availability of second-order statistics	59
2.6 Numerical results	60
2.6.1 Simulation setup	60
2.6.2 Performance without side information	62
2.6.3 Performance with side information	64
2.7 Conclusion	67

3	JADCE under spatially correlated channels: A Bayesian approach	71
3.1	System model	72
3.2	Problem formulation	72
3.3	MAP-ADMM solution	75
3.4	Simulation results	80
3.4.1	The impact of exploiting channel statistics	81
3.4.2	Effect of the number of BS antennas	84
3.4.3	Impact of imperfect knowledge of the channel covariance matrix	87
3.5	Conclusion	87
4	Hierarchical mMTC activity detection and channel estimation with unknown spatial covariance	91
4.1	JADCE with correlated activity and partially known CDI	92
4.1.1	System model	92
4.1.2	Bayesian inference setup	95
4.1.3	Hierarchical spike-and-slab prior	95
4.2	JADCE via EM-EP	96
4.2.1	Main idea of expectation propagation	98
4.2.2	E-Step: posterior approximation via EP	98
4.2.3	M-Step: hyper-parameter update	101
4.2.4	Algorithm implementation	103
4.3	Alternative solution via ADMM	103
4.3.1	JADCE as MAP estimation	104
4.3.2	Cluster-sparsity promoting prior via log-sum	105
4.3.3	Proposed ADMM solution	106
4.4	Simulation result	110
4.4.1	Simulation setup	110
4.4.2	Correlated activity pattern	110
4.4.3	Robustness to model mismatch	114
4.4.4	Convergence behaviour	117
4.5	Conclusion	117
5	Activity detection performance via AMP	119
5.1	System model and problem formulation	119
5.2	AMP for JADCE with spatially correlated channels	120
5.3	Activity detection performance	122
5.3.1	Decision threshold	122
5.3.2	Probabilities of miss detection and false alarm	123

5.4	Simulation results.....	125
5.5	Conclusion.....	125
6	Conclusion and future work	129
6.1	Conclusions.....	129
6.2	Future work.....	131
	References	133
	Appendices	141

1 Introduction

1.1 Motivation and significance

The rapid progress of 5th Generation (5G) wireless networks has paved the way for delivering ubiquitous and reliable connectivity to a massive number of Internet of Things (IoT) devices. As a result, there has been an exponential explosive growth in the number of connected IoT devices, catalyzing the proliferation of widespread IoT applications across various domains, including autonomous mobility [1], healthcare [2], smart factories [3], home automation [4], and more. By the end of 2022, the global count of interconnected IoT devices had reached a staggering 13.2 billion, with projections from the Ericsson Mobility Report [5] forecasting an increase to 34.7 billion by 2028. This exponential growth underscores the critical role of IoT in reshaping our world and underscores the importance of efficient and effective IoT communications systems.

To fully harness the advantages of massive IoT networks, the key lies in deploying reliable wireless access technology. Conventional solutions have often relied on cost-effective technologies such as Bluetooth [6, 7] and WiFi [8]. However, these technologies have limitations when deployed on a large scale, as they are primarily designed to accommodate a limited number of devices within a short-range wireless network. This necessitates the design of seamless channel access schemes for the new era of massive IoT networks [9].

To ensure dependable connectivity for a vast number of IoT devices, with densities reaching up to 1 million devices per square kilometer, the third generation partnership project (3GPP) has designated massive machine-type communications (mMTC) as a primary use case within the framework of 5G wireless networks [10]. In addition, two distinct cellular technologies have been proposed to facilitate the deployment of mMTC: narrow-band IoT (NB-IoT), tailored for fixed and low-rate scenarios, and LTE-machine (LTE-M), designed for mobile and high-rate applications [11, 12]. Therefore, the existing architecture and technology of cellular networks serve as a robust foundation for the practical realization of massive IoT [13].

In addition to the mMTC, 5G encompasses two other essential use cases: enhanced mobile broadband (eMBB) and ultra-reliable low-latency communications (URLLC) [14, 15]. In particular, eMBB is primarily designed for human-type communications (HTC) or human-to-human communications. Its goal is to provide higher data throughput to a relatively small number of users, enabling access to multimedia content, services, and data such as 3D video streaming and augmented reality experiences. On the other

hand, URLLC places stringent requirements on capabilities such as throughput, latency, and availability. It aims to deliver guaranteed quality of service (QoS) for critical missions such as tele-surgery and intelligent transportation in which ultra-reliable and low-latency communications is vital.

In contrast with both eMBB and URLLC, the practical mMTC networks are envisioned to embody the following key characteristics [9, 10, 16]

- **Sporadic transmission:** Unlike traditional communications systems, which often involve continuous data streams, mMTC is geared towards facilitating sporadic data transmissions. IoT devices, many of which have single antennas due to size and power limitations, transmit data only when specific events occur or at periodic intervals.
- **Small-size packet communications:** In many IoT applications, the data packets transmitted are quite small. For example, a temperature sensor might send only a few bytes of data at a time.
- **Predominance of uplink traffic:** Given the nature of IoT deployments, the data traffic is usually from the devices towards a centralized multiple-antenna base station (BS). This configuration enables efficient data aggregation and processing at the BS, which often possesses greater computational power and energy resources than individual IoT devices.
- **Moderate constraints on data reliability and latency:** While some applications may require high reliability and low latency, mMTC generally operates under moderate constraints for these metrics. This means the system is designed to tolerate a certain level of data loss or delays, which makes it adaptable to a wide range of IoT applications—from industrial sensors to smart agriculture—that may not have stringent real-time requirements.

The inherent characteristics of mMTC, such as sporadic transmission, small data sizes, and tolerance of moderate delays, requiring a rethinking QoS metrics and traffic management strategies. The efficiency of the allocation of resources to accommodate this sporadic and often asynchronous communications behaviour is a non-trivial task. For instance, how efficiently allocating resources to accommodate this sporadic and often asynchronous communications behavior is a non-trivial task. Thus, novel resource allocation algorithms that strike a balance between conserving energy, maximizing network capacity, and ensuring timely delivery of data are required. Secondly, rethinking the channel access protocols may be necessary to effectively accommodate the massive concurrency and sporadic nature of mMTC traffic.

In conclusion, while mMTC networks hold great promise for revolutionizing various industries through the connectivity of a massive number of devices, they introduce a

new set of challenges. To harness their full potential, significant upgrades in resource allocation strategies, channel access protocols, and traffic management techniques are essential. Addressing these challenges is imperative for the successful deployment and operation of mMTC networks in the 5G and beyond wireless ecosystem.

1.2 Grant-free access in mMTC

Conventional cellular network solutions have predominantly relied on orthogonal multiple access (OMA) and grant-based access protocols [9, 17]. These techniques have traditionally been developed with a focus on serving HTC requirements, prioritizing high data rates for a limited number of users. Nevertheless, since high data rates are not a major concern in mMTC, conventional channel access protocols prove to be ill-suited for mMTC. In conventional channel access protocols, each device is allocated a dedicated transmission resource. However, given that IoT devices are not continuously active, this approach often leads to inefficient resource allocation. In such cases, numerous resource blocks may remain unused, resulting in sub-optimal network performance [17].

Optimizing channel access protocols to meet the distinctive demands of mMTC, including high network density, energy efficiency, and minimal signaling overhead, presents a substantial challenge. In 5G networks, any device seeking to transmit data to the BS must initially establish a connection using a designated random access protocol. Consequently, the development of efficient random channel access protocols has been recognized as the primary performance bottleneck in mMTC [18].

In practice, the procedure of a typical grant-based random access protocol, in LTE/LTE-advanced (LTE-A), for example, includes four transmission steps [19], as depicted in Fig. 1: 1) Each active device randomly selects a pilot sequence from a predefined set of orthogonal sequences and informs the BS that it has become active; 2) the BS responds to each device with a grant to transmit its data; 3) each active device sends a connection request containing the selected pilot sequence to request resources for data transmission; 4) if a pilot sequence is selected by only one active device, the BS authorizes the corresponding request and sends a contention-resolution message to inform the active device about the allocated resources. Otherwise, the access request is not granted.

The main advantage of the grant-based random access protocol is the simple processing at the BS. However, in the context of massive access, it has several shortcomings. First, due to the limited coherence time and sequence length, the number of orthogonal sequences is finite. Subsequently, there is a high probability that two devices will select the same pilot sequence, causing access failures and delays. More importantly,

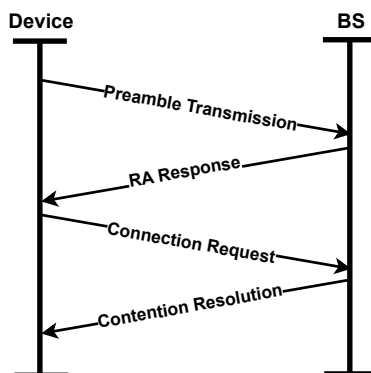


Fig. 1. A typical grant-based random access protocol in LTE/LTE-A.

the access delay inevitably increases as the number of devices increases. Second, the first and third phases include only pilot sequence and no information data. Thus, the signalling overhead would be prohibitively large, given that the actual information data in practice are very short. Third, the high signalling overhead will also increase the power consumption of battery-operated IoT devices, hence reducing their lifespan. Finally, allocating entire resource blocks to each device is not feasible, given the large number of devices. Moreover, it is inefficient, as IoT devices are not always active, leaving several resource blocks unused and rendering performance sub-optimal.

Given the complexities and unique requirements of mMTC networks, two critical questions arise that require further exploration:

1. What are the most effective methods for efficiently allocating the available channel resources among a diverse and dynamically changing set of connected or active IoT devices?
2. How can channel access protocols be optimized to not only meet the requirements of mMTC but also significantly improve energy efficiency and reduce signaling overhead?

Resource allocation has been addressed using non-orthogonal multiple access (NOMA) protocols, where different connected devices share the same radio resource block [17]. This is particularly useful in massive IoT networks characterized by sporadic traffic and a demand for energy-efficient protocols. In such settings, inactive devices are likely to opt for an idle mode, reducing the likelihood of simultaneous channel resource use. The fundamental premise of NOMA is to support non-orthogonal resource allocation among all the connected devices at the cost of complicated receiver

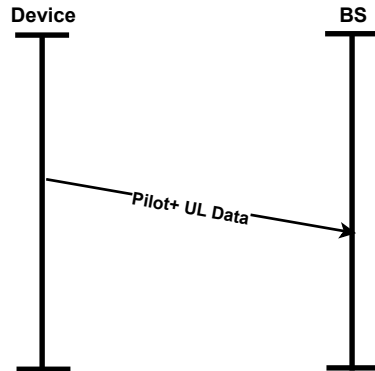


Fig. 2. Grant-free access mechanism as described in [22].

design. Several NOMA solutions have been investigated in the literature, which can be categorized into either power-domain NOMA or code-domain NOMA [20].

While NOMA would solve the resource allocation issue, it fails to efficiently address the issue of massive channel access. Recently, a significant amount of research focused on enhancing random access protocols to address the challenges associated with massive random access. Among the newly proposed protocols, grant-free random access protocols [18, 21, 22, 23] have received increasing attention. For example, in a practical grant-free access protocol, the active devices access the channel and transmit their uplink (UL) data in a single shot without undergoing four-steps channel access protocols [22], as shown in Fig. 2. The inherent advantage of combining pilot sequence and information data within a single step in grant-free access is the reduced signaling overhead and the improved energy efficiency of the IoT devices.

The paramount challenge in grant-free access protocol is to reliably identify the set of active devices that have transmitted their pilot sequences to the BS. Moreover, the BS may need to estimate the active devices channels' information to preform coherent data detection. Indeed, the problem of joint activity detection and channel estimation (JADCE) is the result of the deployment of non-orthogonal pilot sequences and the massive number of devices, meaning the received pilot signal at the BS suffers from high co-channel interference [9]. It is therefore necessary to utilize sophisticated signal processing techniques to design and optimize computationally-efficient JADCE solution.

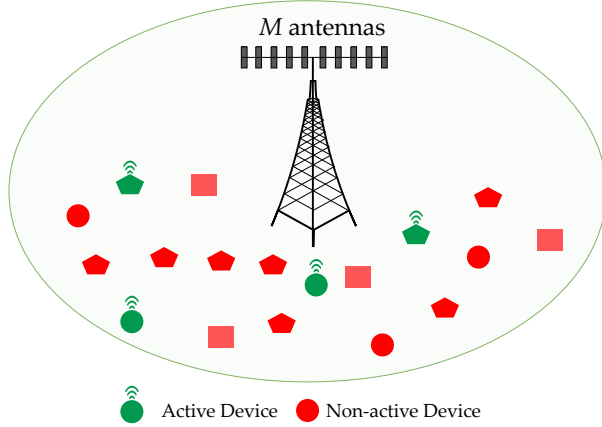


Fig. 3. Illustration of a standard uplink mMTC scenario with a single BS and sporadically active devices.

1.3 Device activity detection and channel estimation

This section introduces a general mathematical model for uplink transmission in grant-free access. This model will serve as the foundational framework for the problem explored throughout the thesis.

This thesis considers the conventional massive random access problem in mMTC, as depicted in Fig. 3. More precisely, we consider a single cell mMTC network consisting of a single BS equipped with M antennas serving a set of $\mathbf{N} = \{1, \dots, N\}$ devices in a sporadic uplink only communication.

Activity model

During each coherence interval, there is generally at most $K \ll N$ devices that are in an active state, i.e., they need to transmit their UL data to the BS. The active devices are considered to belong to set $\mathcal{A} \subseteq \mathcal{N}$, $|\mathcal{A}| = K$. The device activity state is modeled by the *activity state binary vector* $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_N]$, such that

$$\gamma_i = \begin{cases} 1, & i \in \mathcal{A} \\ 0, & \text{otherwise.} \end{cases}$$

Since only K devices are active at any particular coherence interval, the vector $\boldsymbol{\gamma}$ exhibits a *sparse* structure.

Exploiting the sparsity within the activity of devices is a pivotal step in designing efficient and reliable solutions for the JADCE problem. Therefore, it is of utmost importance to leverage any structure within the underlying sparse activity patterns. In practical scenarios, two broad categories of activity patterns are commonly observed: 1) *Independent activity pattern*: This is the most widely considered in mainstream discussions of the massive random access problem. It models scenarios such as an IoT network where each device is assigned an independent task, resulting in device activities that are unrelated and can be considered mutually independent. 2) *Correlated activity pattern*: This arises naturally in large-scale and dense IoT applications following an event-triggered traffic model. It is particularly relevant in practical MTC networks where devices are organized into clusters for specific monitoring tasks. For instance, consider a sensor-equipped manufacturing facility monitoring automated machinery. When a significant event occurs in any machine, the corresponding sensors often transmit uplink data to the BS simultaneously, resulting in a correlated activity pattern.

Pilot sequences assignment

In grant-free access, each device receives a pre-assigned unique pilot sequence, denoted here as ϕ_i , that serves as an identifier. Given that the number of mMTC devices is expected to be extremely large, employing orthogonal pilot sequences such as Zadoff-Chu (ZC) sequences may be impractical for massive access. Specifically, the constraints of the limited coherence interval, the large number of connected IoT devices, and the finite set of possible (sub)-orthogonal pilot sequence combinations make it unfeasible to allocate a unique orthogonal sequence to each device. To address this issue, the pilot sequences are drawn from a large pool of non-orthogonal sequences to form the pilot sequence code book, denoted here as $\Phi = [\phi_1, \dots, \phi_N]$.

Uplink transmission

In grant-free access, each transmission interval is divided into two stages. In the first stage, the active devices transmit their identifier to the BS. The remaining transmission interval is used to transmit the UL data¹. Subsequently, based on its received signal, the BS aims to detect the active devices and estimate their instantaneous channel state information by detecting which pilot sequences have been transmitted.

¹We assume that all the devices and the BS are perfectly synchronized.

The received signal during the pilot transmission phase, denoted as \mathbf{Y} , is given by

$$\mathbf{Y} = \sum_{i=1}^N \gamma_i \phi_i \mathbf{h}_i^T + \mathbf{W} = \Phi \mathbf{X}^T + \mathbf{W}, \quad (1)$$

where \mathbf{h}_i denotes *the channel response* between the i th device and the BS, $\mathbf{x}_i = \gamma_i \mathbf{h}_i$ denotes *the effective channel vector*, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, and \mathbf{W} represents the additive noise at the BS.

While the use of non-orthogonal signature sequences as device identifiers removes the need for contention resolution, it makes the task of joint device activity detection and channel estimation at the BS more challenging. For instance, if the sequences are generated from an orthogonal pilot code book, the JADCE can be simply solved by applying a matched filter on the received signal. However, due to the non-orthogonality of the pilot sequence, a simple solution like a matched filter is no longer applicable to solve the JADCE problem. Therefore, advanced signal processing techniques are therefore required.

The sparse device activity pattern induced by the sporadic transmissions in mMTC motivates the formulation of JADCE as a compressive sensing (CS) problem [24, 25, 26]. Traditionally, CS is characterized by an under-determined system of linear equations aimed at recovering a high-dimensional *sparse signal* vector from a low-dimensional single measurement vector (SMV). Furthermore, as the BS senses the same sparse activity from different antennas, the JADCE problem extends to the multiple measurement vector (MMV) CS framework. One such extension, known as the MMV problem, generalizes sparse signal recovery or the SMV problem to cases where a group of measurement vectors has been obtained from a set of signal vectors that are assumed to be jointly sparse, sharing a common support, i.e., the positions of the nonzero rows of the unknown sparse matrix.

1.4 Literature review

The JADCE problem has gained significant attention, leading to a substantial body of research that addresses it as a sparse recovery problem. To set the stage, we first present a concise literature review of the main techniques deployed to solve such problem.

1.4.1 Compressive sensing overview

Compressive sensing (CS) is a signal processing technique with applications in various fields such as communications, imaging, and data science [27]. The fundamental idea

behind CS is the efficient acquisition and reconstruction of a signal that is sparse or compressible in some basis.

In classical signal processing, the Shannon-Nyquist sampling theorem dictates the sampling rate necessary to capture all the information contained in a signal. According to the theorem, the signal must be sampled at a rate that is at least at twice its highest frequency to guarantee its perfect reconstruction. However, this classical notion is challenged by the CS theory, which states that, under certain conditions, a sparse signal can be sampled at much lower rates than dictated by the Nyquist-Shannon theorem and still be perfectly reconstructed [27].

Mathematically speaking, the CS problem aims to recover a high-dimensional sparse vector \mathbf{x} from its low-dimensional measurement vector \mathbf{y} . The transformation of \mathbf{x} into \mathbf{y} is a mapping through a matrix ϕ , known as the sensing matrix, as follows

$$\mathbf{y} = \phi \mathbf{x} + \mathbf{w}. \quad (2)$$

The canonical CS recovery formulation, in the noiseless case, is given as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \Phi \mathbf{x} = \mathbf{y}, \quad (3)$$

where $\|\cdot\|_0$ denotes the ℓ_0 -norm defined as the number of nonzero elements of the argument. However, since the observation is practically always noisy in real-world applications, one can relax the equality constraint to rewrite the noiseless CS problem in its noisy version as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon_{\ell_1}, \quad (4)$$

where ε_{ℓ_1} is a positive small parameter that denotes a tolerable error due to the additive noise \mathbf{w} .

Solving the constrained optimization problems outlined in (3) and (4) is typically an NP-hard task. Exact solutions are therefore unattainable using traditional computational techniques, especially in the context of large-scale optimization. To address this issue, the literature presents various algorithms designed to relax the optimization problems (3) and (4). Broadly speaking, the most prominent algorithms for sparse recovery problems can be categorized into the following classes: 1) convex optimization approaches [28]; 2) iterative methods such as approximate message passing (AMP) [29]; 3) sparse Bayesian learning (SBL) [30]; and 4) greedy algorithms [31]. The current literature suggests that greedy pursuit techniques are generally outperformed by mixed-norm minimization approaches. The latter are in turn surpassed by AMP and SBL methods [29]. In the following subsection, we will briefly introduce each of the main classes of sparse recovery algorithms.

Greedy algorithm

One way to avoid solving the non-convex optimization problem (3) is to utilize greedy algorithms. Instead of solving the non-convex problem explicitly, greedy algorithms employ an iterative approach that seeks to make locally optimal decisions in each iteration, leading to sub-optimal solutions.

Since the sparse signal \mathbf{x} contains only a few nonzero elements, in the absence of noise, the measurement matrix \mathbf{y} is in a space supported by a sub-matrix of Φ determined by the support of the vector \mathbf{x} . Greedy algorithms estimate the *unknown* support of \mathbf{x} by evaluating the correlation between a residual from the measurement matrix \mathbf{y} and the sensing matrix Φ . The most widely used greedy algorithm is the orthogonal matching pursuit (OMP) [32], which solves the sparse signal recovery problem in the SMV setup by detecting a single nonzero element at each iteration.

Additionally, several variants of the OMP have been proposed, including the compressive sampling matching pursuit (CoSaMP) [33] and subspace pursuit (SP) [34]. Both of these algorithms aim to accelerate algorithm convergence by detecting more than one nonzero element per iteration. Furthermore, an extension tailored to the MMV setup has been introduced in [31] to harness the joint sparsity pattern in the unknown sparse signal, capitalizing on both the intra-signal and inter-signal sparsity structures. Furthermore, the group orthogonal matching pursuit (GOMP) [35] builds on the OMP by integrating any previous knowledge of the signal's block-sparsity structure to improve support detection accuracy.

While greedy algorithms are generally simple to implement, their performance often falls short compared to state-of-the-art algorithms in noisy scenarios. This is because they lack a feedback mechanism to rectify errors made in previous iterations.

Convex optimization

Another strategy for finding a feasible solution to the CS problems (3) and (4) is to relax the combinatorial ℓ_0 -norm penalty using the convex ℓ_1 -norm instead. Indeed, due to the convex nature of ℓ_1 -norm and its tendency to promote sparse solutions, ℓ_1 -norm minimization has been a focus in the development of sparse recovery methods.

Chen *et al.* proposed the basis pursuit (BP) algorithm to solve the optimization problem (3) by formulating the CS problem as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \Phi\mathbf{x} = \mathbf{y}. \quad (5)$$

Furthermore, the BP has been modified to accommodate the noisy scenario through the basis pursuit denoising (BPDN) algorithm [36], which reformulates the CS problem as

$$\min_{\mathbf{x}} \|\Phi\mathbf{x} - \mathbf{y}\|_2 + \lambda \|\mathbf{x}\|_1, \quad (6)$$

for some positive parameter λ that controls the sparsity of the solution. Indeed, the BPDN formulation bears resemblance to Tibshirani's seminal work [37], in which he introduced the least absolute shrinkage and selection operator (LASSO) that solves

$$\min_{\mathbf{x}} \|\Phi\mathbf{x} - \mathbf{y}\|_2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \varepsilon_{\ell_1}. \quad (7)$$

Given that BP, BPDN, and LASSO are convex problems, they are readily solvable using standard optimization tools.

While many studies have explored ℓ_1 -norm minimization, it was the pioneering studies in [38, 39] that truly ignited a proliferation in related works and applications leveraging CS recovery algorithms. More precisely, these studies provided rigorous theoretical conditions, focusing on the structure of the sensing matrix and the number of measurements within the observation signal, to achieve perfect sparse recovery from a limited set of linear, non-adaptive measurements. For example, Donoho showed in [38] that in the noiseless case, the non-convex problem (3) could be equivalently solved via the BP algorithm. In addition, Candés *et al.* [39] showed that for a *sufficiently* sparse signal, an ℓ_1 -norm penalty could lead to an optimal solution for the CS problem even in the noisy case.

Alternative to the ℓ_0 -norm are the iterative hard-thresholding methods [40], the non-convex ℓ_p pseudo-norm ($0 < p < 1$) minimization [41], and the iteratively reweighted ℓ_1 -norm [42]. For the MMV case, a mixed norm approach has been proposed, such as the group LASSO [43], the $\ell_{2,1}$ -norm [44], the atomic norm [45], and iterative reweighted ℓ_2 [46].

Sparse Bayesian learning

While optimization approaches offer significant improvements over greedy algorithms, they predominantly rely on deterministic regularization. This focus on deterministic methods neglects any prior knowledge of the data, potentially leading to sub-optimal performance in sparse recovery. To address this limitation, numerous studies have reformulated the CS problem within a Bayesian framework [47]. The Bayesian approach brings several unique advantages to traditional formulations, including the ability to make probabilistic predictions, the automatic incorporation and estimation of model parameters, and the provision of theoretical means to assess the quality of the recovered signal.

In a typical sparse recovery problem from the Bayesian inference perspective, the parameters to be estimated—specifically, the components of the sparse signal \mathbf{x} —are treated as random variables (vectors) and assigned with a probability distribution function (PDF) to reflect any available prior knowledge of these parameters. One of the most prominent approaches in Bayesian compressed sensing (BCS) is the SBL algorithm [48, 49]. SBL adapts the relevance vector machine (RVM) concept proposed in [49] to the CS problem. A key feature of the RVM mechanism is its ability to suppress elements that do not contribute significantly to the reconstructed signal, thereby yielding a sparse solution.

In the SBL framework, the sparse signal \mathbf{x} is treated as a random variable and modeled in accordance with a sparsity-promoting distribution parameterized by a set of hyper-parameters Ξ . Two popular signal priors that have already been explored in the CS Bayesian literature are the Bernoulli-Gaussian prior (also known as the spike-and-slab-prior) and the Gaussian prior. Subsequently, instead of finding a point estimate for \mathbf{x} , the SBL framework aims to infer the posterior distribution $p(\mathbf{x}|\Xi^*, \mathbf{y})$ where Ξ^* maximizes the evidence function $p(\mathbf{y}|\Xi^*)$.

One main advantage of SBL compared to other conventional CS methods is that it does not require prior knowledge of signal sparsity or noise levels, as both are inferred automatically using the RVM mechanism through the inference process [48]. Note that many CS recovery schemes can be viewed from the Bayesian perspective. For example, the LASSO problem can be interpreted as maximum *a posteriori* probability (MAP) estimation under a Bayesian model with a sparseness-inducing prior drawn from a Laplacian distribution [48]. For the MMV case, both MSBL [30] and T-SBL [50] have been proposed to exploit the joint sparsity and the temporal correlation of the ensemble of the sparse signals respectively.

A known limitation of SBL is the prohibitively high computational costs of the inference algorithm due to the requirement to evaluate a dense covariance matrix at each iteration. SBL (and its variants) may therefore not always be suitable when applied to large-scale problems such as activity detection and channel estimation in mMTC.

Approximate message passing

The AMP algorithm introduced by Donoho *et al.* in [51] has emerged as a significant and efficient algorithmic framework to solve the CS problems. AMP is an iterative algorithm that operates as follows

$$\mathbf{z}^{(t)} = \mathbf{y} - \Phi \mathbf{x}^{(t)} + \mathbf{u}^{(t)}, \quad (8)$$

$$\mathbf{x}^{(t+1)} = \eta^{(t)}(\Phi^H \mathbf{z}^{(t)} + \mathbf{x}^{(t)}), \quad (9)$$

where (t) denotes the iteration index, $\eta^{(t)}(\cdot)$ is a component-wise separable *denoising function*, and $\mathbf{u}^{(t)}$ is a correction term that is unique to the AMP, and the so-called Onsager term. While the AMP shares similarities with the iterative soft thresholding algorithm (ISTA) (where $\mathbf{u}^{(t)} = \mathbf{0}$), it distinguishes itself through the incorporation of the Onsager term. This inclusion notably accelerates the convergence rate of AMP, also enhancing its robustness [52].

One of the key features of AMP lies in its flexibility to design the denoising function $\eta(\cdot)$ in a way that promotes sparsity. For example, AMP can be considered a relaxed version of the BP algorithm when no prior knowledge about the distribution of the sparse signal \mathbf{x} is available. In this scenario, the denoising function $\eta(\cdot)$, adheres to a min-max framework via the soft thresholding operator [53]. Conversely, Donoho et al. [54] introduced a Bayesian variant of AMP, which is applicable when the distribution of the sparse signal is known a priori. Here, the denoising function $\eta(\cdot)$ is designed based on the minimum mean-squared error (MMSE) criterion. Further extending this work, Vila [55] proposed an innovative AMP-based algorithm aimed at learning both the signal distribution and the noise variance while concurrently recovering the sparse signal. Remarkably, the quality of reconstruction achieved through this method is comparable to an AMP algorithm developed within a comprehensive Bayesian framework, where both the signal distribution and noise levels are known.

The cornerstone of the AMP algorithm success and popularity is its state evolution (SE) analysis. This analytical tool allows the prediction of the mean squared error (MSE) per iteration under large i.i.d. sub-Gaussian sensing matrices [52] and finite size i.i.d. sub-Gaussian sensing matrices [56]. Nonetheless, AMP has been noted to suffer from divergence for non i.i.d sub-Gaussian sensing matrices [57]. To address this issue, several variants of the AMP have been proposed in the literature. For instance, the orthogonal AMP (OAMP) has been proposed in [58] for general unitary-invariant matrices. However, due to the matrix inversion required at each iteration, OAMP suffers from high computational complexity. Rangan *et al.* proposed in [59] the vector approximate message passing (VAMP) algorithm, which provides a robust performance for a much broader class of large random matrices. Furthermore, they provide a rigorous theoretical analysis for the SE under right-orthogonal invariant random matrices.

Covariance approach

A new class of sparse recovery algorithm, namely the covariance-based approach has been introduced in [60, 61]. Indeed, the covariance approach was originally proposed to

tackle the activity detection problem in massive access. Haghghatshoar *et al.* [60] leveraged the fact that the BS is equipped with a large number of antennas in practice, allowing an accurate covariance matrix estimate of the received signal at the BS. The covariance approach formulates the detection problem as a maximum likelihood (ML) problem, treating the signal coefficients as deterministic unknowns.

One of the key advantages of the covariance-based approach is its ability to identify the support of a sparse signal, using short pilot sequences. More precisely, a scaling law has been established, stating that the support of the signal can be reliably recovered when the cardinality of the signal support is in the quadratic order of the pilot sequences length under a non-negative least squares (NNLS) formulation [61]. A similar conclusion was reached using a maximum likelihood approach and Fisher information analysis, as detailed in [62]. On the other hand, unlike traditional CS sparse recovery algorithms, the covariance-based approach focuses primarily on detecting the support of the sparse signal, without estimating its coefficients.

1.4.2 CS-based JADCE

Numerous solutions for JADCE that employ greedy algorithms have been reported in the literature [63, 64, 65, 66, 67]. For instance, one of the earliest solutions to the JADCE problem in an SMV setup, i.e, when the BS is equipped with just one antenna, have been proposed in [63], using a hierarchical block orthogonal matching pursuit (HBOMP). Additionally, Wang *et al.* [68] introduced a CS-based device detection scheme that leverages the structured sparsity across consecutive coherence intervals to improve both active device and data detection. Meanwhile, Du *et al.* [65] have proposed a block sparsity adaptive SP, which assimilates activity patterns from prior time slots to boost the accuracy of device activity detection in the current coherence interval.

From a convex optimization perspective, the JADCE has been formulated as an $\ell_{2,1}$ -norm minimization problem and solved via various methods such as the alternating direction method of multipliers (ADMM) [69], augmented Lagrangian-based alternating direction inexact Newton (ALADIN) [70], and block coordinate descent (BCD) [69, 71, 72]. Li *et al.* addressed JADCE in mMTC with asynchronous transmission using both LASSO and sparsity-constrained methods [73]. Yuan *et al.* [74] addressed the JADCE problem in a distributed mMTC system with mixed-analog-to-digital converters under two different user traffic classes. Zhu *et al.* [75] addressed JADCE in millimeter-wave (mmWave) channels where they leveraged the angular domain sparsity inherent in these channels to formulate the JADCE problem as an atomic-norm minimization problem and designed a solution based on ADMM. The authors in [76] provided a theoretical

study of the performance of JADCE under a group LASSO formulation. More precisely, they conducted a phase transition analysis to determine the minimum length of the pilot sequence required to achieve a high probability of correct support recovery.

The covariance-based approach has also been applied to the problem of activity detection when channel coefficients are known [61, 77, 62, 78]. For example, the work in [61] formulated the device activity detection in a single-cell network as an ML estimation problem and proposed a BCD-based approach to obtain a stationary point for the ML estimation. To enhance the convergence rate of BCD, Dong *et al.* [78] proposed a multi-arm bandit approach that aimed to select the coordinates that would result in a more aggressive increase in the ML objective function. The same ML estimation problem has been solved using gradient descent [77]. On the other hand, Chen *et al.* [62] utilized the same formulation presented in [61] to solve joint activity and data detection problem. An extension to multi-cell network has been proposed in [79, 80, 81]. While the work in [79, 80] ignored inter-cell interference when designing activity detection solutions, the work in [70] proposed a cooperative approach, where the BS from different cells jointly estimated both active devices and inter-cell interference.

From the Bayesian perspective, the authors exploited the path loss information and the joint sparsity structure in in [82], proposing a Bayesian CS-based algorithm to efficiently detect device activity in an uplink cloud radio access network. The authors in [83] deployed pattern-coupled sparse Bayesian learning [84] to exploit the temporal correlation of the device activity state between consecutive coherence intervals. Furthermore, the appealing properties of the Bayesian AMP have garnered significant attention as a solution for the JADCE problem, as evidenced in numerous studies such as [9, 85, 86, 87, 88, 89]. For instance, Chen *et al.* addressed the user activity detection problem in grant-free mMTC using AMP and derived analytical performance of the proposed AMP algorithm in both SMV and MMV setups [9]. Liu *et al.* [85, 86] extended the analysis of [9], conducting an asymptotic performance analysis for activity detection, channel estimation, and achievable rate under the assumption of known LSFC. Ke *et al.* [87] considered an enhanced mobile broadband network, investigating the performance of the generalized AMP algorithm that exploits the structured sparsity in the multiple-input multiple-output (MIMO) channels. Senel and Larsson [88] designed a non-coherent detection scheme for very short packet transmission by jointly detecting the active users and transmitted information bits. An extension to cell-free mMTC networks proposed in [90] utilized the prior knowledge of the devices position to assign adequate prior distribution and solved the JADCE problem via expectation maximization-AMP (EM-AMP). The authors in [89] addressed the JADCE problem in a mMTC network assisted with a reconfigurable intelligent surface (RIS). Furthermore, they proposed an

AMP-based solution that exploited both the sparsity in the device activity pattern and the sparsity of the RIS-to-BS channel in the angular domain.

Recently, conventional sparse recovery algorithms have been increasingly adapted to develop machine-learning solutions that address the JADCE problem [91, 92, 93, 94, 95, 96, 97]. For instance, [92, 91] introduced data-driven auto-encoder architectures capable of jointly optimizing the measurement matrix and the support recovery algorithm. On the other hand, Cui *et al.* focused on harnessing the inherent sparsity patterns in mMTC networks and proposed a model-driven framework for JADCE in [98]. This framework includes two distinct decoders, one based on MAP problem and another that utilizes a group LASSO penalty, thereby enabling the joint design of pilot sequences and active device detection. Additionally, model-driven JADCE solutions that employ known pilot sequence code books have been proposed [94, 95, 97]. These solutions feature various decoder architectures designed based on AMP [94], SBL [97], and LASSO [95]. It is noteworthy that while machine-learning-based approaches often excel in terms of accuracy and computational efficiency, they typically require extensive training on large data sets and do not provide theoretical performance guarantees.

1.5 Thesis objectives

It is evident from the literature that CS-based JADCE solutions are a very promising candidate for enabling grant-free access in mMTC. Given the plethora of available work for solving the JADCE problem, it is therefore natural to investigate what, if any, improvements can be made. After a careful analysis of the mainstream literature, two notable *gaps* become apparent:

1. **Uncorrelated fading channels:** A prevalent simplifying assumption in many of these studies is that the wireless channels between the devices and the BS are *spatially uncorrelated*. This assumption often facilitates tractable performance analysis and low-complexity algorithmic implementations. Nevertheless, this assumption does not always hold in practice, as the MIMO channels are usually spatially correlated [99, 100].
2. **Independent activity pattern:** Most existing works on JADCE address mMTC networks with independent activity patterns among devices, typically modeling scenarios where each device monitors an independent process and is randomly activated. This does not fully capture the intricacies of more realistic IoT network deployments. In practice, IoT devices are often organized into clusters, each monitoring the same phenomena, which results in *correlated* activity patterns rather than independent ones.

In this thesis, we aim to propose, develop, and optimize several solutions to address the JADCE problem in mMTC. These solutions are developed to investigate, address, and exploit the spatial correlation of the MIMO channels between the connected devices and the BS. Furthermore, the proposed solutions cater for various assumptions on the device activity pattern and the availability of channel distributions information (CDI). To achieve these objectives, we employ a range of well-established mathematical methodologies, including:

- Convex optimization and decomposition techniques: These tools will be utilized to formulate and solve optimization problems, breaking them down into more manageable sub-problems when necessary.
- CS theory: We will apply principles from CS theory to address problems related to sparse signal representation and reconstruction, which are often encountered in the JADCE problem.
- Bayesian inference: Employing Bayesian methods will allow us to incorporate prior knowledge into our algorithms, thus improving their robustness and accuracy.

The efficacy of the newly developed algorithms will be rigorously assessed through extensive Monte-Carlo computer simulations, implemented in Matlab, to provide a comprehensive numerical evaluation of their performance. By taking a multifaceted approach that combines these mathematical tools and techniques, this thesis aims to deliver innovative and practical solutions that advance the state-of-the-art in JADCE for mMTC. The main objectives of this thesis are the following:

1. **JADCE under spatially correlated channels:** Initially, we aim to tackle the principal shortcoming present in the existing solutions, which is the lack of accounting for spatial correlations in channels between the BS and the devices. Addressing the JADCE problem while considering the spatial correlation structure of channels is imperative. This importance stems from two factors: first, practical MIMO channels inherently exhibit such spatial structures, and second, the performance of previously developed JADCE solutions may vary depending on the specific correlation structures encountered in real-world scenarios [98]. Indeed, to realistically evaluate the performance of the JADCE problem, it is essential to employ channel models that accurately reflect the core features of massive MIMO networks.
2. **Efficient JADCE solutions for diverse scenarios:** Our second objective is to provide a range of versatile and robust JADCE solutions tailored to various system model assumptions and prior information scenarios, regarding both the device activity pattern and the availability of exact or limited CDI at the BS. For each of these scenarios, our approach includes a thorough investigation and comparison of the

proposed algorithms against state-of-the-art counterparts designed for similar setups and similar levels of prior side information.

3. **Theoretical analysis:** Beyond algorithmic contributions, we aim to provide theoretical insights by analyzing the performance of AMP-based JADCE in the asymptotic regime. Specifically, we use the state evolution of the AMP algorithm to characterize user activity detection performance, deriving closed-form expressions for missed detection and false alarm error probabilities.

1.6 Thesis outline and contributions

In this section, we outline the structure of the thesis and provide an overview of the main contributions of each chapter.

Chapter 2

The main contributions of Chapter 2 are based on the journal publication [101] and the conference paper [102]. The contribution on this chapter are twofold. First, we provide a detailed mathematical overview of the JADCE problem in spatially correlated channels in grant-free mMTC. Second, it addresses the JADCE problem under two different assumptions of the availability of the second-order statistics at the BS. We therefore made the following contribution: 1) For unknown CDI, we investigate the use of iterative reweighted $\ell_{2,1}$ -minimization (IRW- $\ell_{2,1}$). The core idea of the reweighted $\ell_{2,1}$ -norm is to suppress the magnitude dependency associated with the use of the ℓ_1 -norm so that the reweighted ℓ_1 -minimization becomes closer to the sparsity optimal ℓ_0 -minimization. 2) When the second order statistics are available to the BS, we formulate a novel deterministic JADCE problem based on an iterative reweighted $\ell_{2,1}$ -norm optimization problem augmented with a covariance matching penalty. Furthermore, we propose two computationally efficient methods based on the ADMM framework to solve the JADCE problem. The proposed approaches are shown, via computer simulations, to significantly enhance the device activity detection accuracy and therefore the channel estimation performance. Moreover, the proposed approaches achieve the same performance as baseline *deterministic* MMV sparse recovery algorithms, yet with a smaller signaling overhead.

Chapter 3

The content in Chapter 3 is primarily based on the journal paper [103]. This chapter maintains the same system model as Chapter 2 and addresses the JADCE problem under the assumption that the CDI are available at the BS. In addition, Chapter 3 adopts a Bayesian perspective to address the JADCE problem. In this context, we introduce a sparsity-inducing prior that characterizes the sporadic activity of the devices, transforming the JADCE problem into a MAP estimation framework. This formulation allows us to incorporate prior knowledge of the CDI through the Mahalanobis distance measure. Moreover, through the application of an appropriate problem decomposition technique, we transform the non-convex Bayesian MAP estimation into a multi-block convex optimization problem. As a result, we derive a computationally efficient solution based on ADMM that exhibits a closed-form update rule for the involved sub-problem.

Chapter 4

Chapter 4, based on conference paper [104], introduces two key departures from the preceding chapters. First, it addresses the JADCE in spatially correlated MIMO channels *without prior knowledge of the exact CDI*. Second, the chapter shifts the focus to mMTC networks with correlated device activities. This scenario gives rise to a specific hierarchical sparse activity model, encompassing both *cluster-level* and *intra-cluster* structures.

To effectively address these distinctions, we propose two different solutions from a Bayesian perspective. Specifically, we rely on the selection of appropriate prior functions to model the underlying aspects of this particular system. Initially, to estimate the unknown CDI, we employ a prior function following a Wishart distribution [105]. Furthermore, leveraging the concept of strong sparsity inducing prior, we introduce a novel hierarchical *spike-and-slab prior* to account for both the cluster-level and intra-cluster activity patterns to the model. Furthermore, the proposed solution builds on the integration of an expectation propagation (EP) algorithm [106] within an EM framework [105]. This proposed solution effectively navigates the complexities of approximating an intractable joint posterior distribution. Furthermore, it offers estimations pertaining to active devices, their associated channels, and their CDI.

In addition to the EP-based solution, we propose an alternative solution to the JADCE problem by relaxing the spike-and-slab prior with a log-sum prior [107] that provides a more flexible approach to encode the hierarchical activity pattern. Subsequently, we formulate the JADCE as a MAP problem and derive a solution based on the ADMM

framework iteratively solve an approximated version of the MAP problem via a sequence of closed-form updates.

Chapter 5

Chapter 5 is based on conference paper [108]. While Chapters 2 through 4 employ algorithmic approaches to formulate and optimize solutions for the JADCE problem, Chapter 5 offers a detailed theoretical analysis of activity detection accuracy using the AMP algorithm. Chapter 6 presents a detailed performance analysis for JADCE in spatially correlated fading channels. In particular, Chapter 6 conducts a detailed theoretical performance analysis of MMV-AMP-based JADCE in the asymptotic regime. More precisely, by using the state evolution analysis of the AMP algorithm, Chapter 5 derives closed-form expressions for both miss detection and false alarm probabilities.

Chapter 6

Chapter 6 serves as a culmination of the extensive research and findings presented throughout the thesis. In this chapter, we aim to provide a detailed summary of the key contributions, insights, and implications arising from our investigation into the JADCE problem. Additionally, we outline potential directions for future work and research in this domain.

1.7 Author's contributions

This thesis is structured as a monograph and based on the following original publications, including two already-published journal articles [101, 103], one submitted journal article [109], and four conference papers [102, 108, 104, 110]. The author of this thesis had the main responsibilities in conceiving the original concepts behind the developed methods, formulating the mathematical problems, deriving their analytical solutions, compiling the MATLAB codes to run empirical performance analysis simulations, and writing the original international peer-reviewed publications. The collaborating co-authors provided invaluable guidance for shaping research directions, contributed ideas pertinent to the developed methods and algorithms, and offered constructive feedback during the manuscript preparation phase to enhance the overall quality.

2 JADCE under spatially correlated channels: A deterministic approach

This chapter addresses the JADCE problem in single-cell mMTC networks under spatially correlated fading MIMO channels. In particular, the JADCE problem in spatially correlated MIMO channels aims to provide a realistic assessment of the performance of the proposed JADCE solutions. Addressing the JADCE under such a channel model is significant for two primary reasons:

- The MIMO channels in practical scenarios are inherently spatially correlated [111].
- Spatial correlation can enhance multi-user MIMO system performance if devices possess sufficiently diverse spatial correlation matrices [111, 112]. Furthermore, given that channel spatial correlation varies more slowly than channel realizations, the channel covariance matrices for all devices can be accurately estimated [113]. This intrinsic information within spatial configuration matrices can therefore be leveraged to boost JADCE performance.

This chapter addresses the JADCE problem through a deterministic framework under two distinct assumptions of the availability of prior knowledge of the channel's second-order statistics at the BS. When the channel statistics are unavailable, we propose an iterative reweighted $\ell_{2,1}$ -norm optimization problem that depends solely on the sparsity of the channel matrix and is robust to different channel distributions. Conversely, when the covariance matrices are known to the BS, we introduce a novel JADCE formulation that incorporates this side information to enhance the JADCE performance. For both cases, we derive two ADMM-based solutions that feature computationally efficient closed-form expressions, which can be calculated via simple analytical formulas.

2.1 System model and problem formulation

2.1.1 mMTC uplink system model

Consider a single cell uplink communication scheme, shown in Fig. 4. (a), consisting of a single BS equipped with a uniform linear array (ULA) containing M antennas serving a set of N uniformly distributed single-antenna devices $\mathcal{N} = \{1, \dots, N\}$.

We consider a block fading channel response over each coherence period T_c . Furthermore, to model the propagation channels between the devices and the BS, we

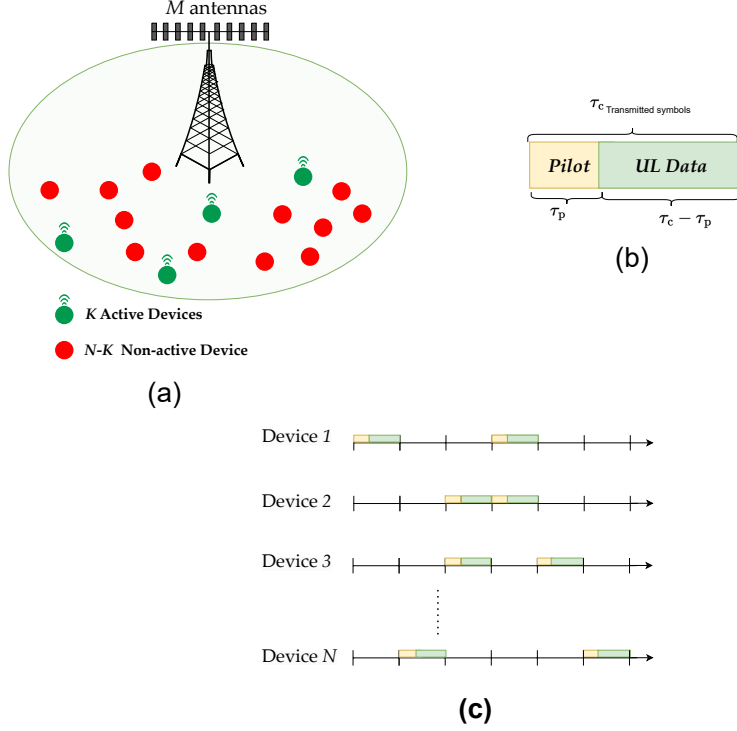


Fig. 4. Depiction of a standard mMTC network: (a) an uplink mMTC system featuring K active devices and $N - K$ inactive devices; (b) grant-free access packet; (c) sporadic device activity.

employ a *one-ring* channel model [100, Sect. 2.6]. In such channel model, the BS is assumed to be positioned at an elevated location compared to the devices, and it thus has no scatterers in its near-field.

The channel response vector for each device \mathbf{h}_i , $i \in \mathcal{N}$ is the superposition of P_i physical signal paths, each reaching the BS as a plane wave, and modelled as

$$\mathbf{h}_i = \frac{1}{\sqrt{P_i}} \sum_{p=1}^{P_i} \rho_{i,p}^{\text{path}} \mathbf{a}(\psi_{i,p}), \quad (10)$$

where $\psi_{i,p}$ and $\rho_{i,p}^{\text{path}} \in \mathbb{C}$ denote the angle of arrival (AoA) and the gain of the p th path respectively, and $\mathbf{a}(\psi_{i,p}) \in \mathbb{C}^M$ represents the steering vector of the ULA, given as

$$\mathbf{a}(\psi_{i,p}) = [1, e^{-j2\pi\Delta_r \cos(\psi_{i,p})}, \dots, e^{-j2\pi(M-1)\Delta_r \cos(\psi_{i,p})}]^T, \quad (11)$$

where Δ_r denotes the normalized spacing between the adjacent BS antennas. We consider that $\psi_{i,p} = \bar{\psi}_i + \zeta_{i,p}^{\text{ASD}}$, where $\bar{\psi}_i \in [-\pi/2, \pi/2]$ represents the (deterministic) incident

angle between the i th device and the BS, and $\zeta_{i,p}^{\text{ASD}}$ denotes a (random) deviation from the incident angle with angular standard deviation (ASD) σ_ψ .

The channel realizations $\mathbf{h}_i, i \in \mathcal{N}$, are assumed to be independent between different coherence intervals T_c . Additionally, in this work, we consider devices with low mobility, which is justified in the context of mMTC [114]. Therefore, we adopt the common assumption that the channels are wide-sense stationary [113]. Subsequently, the set of channels covariance matrices $\{\mathbf{R}_i\}_{i=1}^N$ vary in a slower timescale compared to the channel realizations [99]. Accordingly, $\{\mathbf{R}_i\}_{i=1}^N$ are assumed to remain fixed for τ_s coherence intervals, where τ_s can be in the order of thousands [111]. Nevertheless, in practice, the BS does not know the exact channel covariance matrices $\{\mathbf{R}_i\}_{i=1}^N$. Rather, during a training phase, the BS acquires an estimate of the second-order statistics of the channel, $\{\hat{\mathbf{R}}_i\}_{i=1}^N$ [99]. Thus, for each device $i \in \mathcal{N}$, the BS obtains T estimates of channel response \mathbf{h}_i from different coherence intervals, denoted as $\hat{\mathbf{h}}_i^1, \dots, \hat{\mathbf{h}}_i^T$. We define the estimated channel covariance for each device i as $\hat{\mathbf{R}}_i = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{h}}_i^t \hat{\mathbf{h}}_i^{tH}$. We note that this assumption can be challenging in mMTC; where some devices are inactive for a longer period. We will therefore elaborate further on this issue in Section 2.5.

2.1.2 Pilot assignment

In grant-free mMTC, the BS assigns a unique unit-norm pilot sequence $\phi_i \in \mathbb{C}^{\tau_p}$ to each device $i \in \mathcal{N}$. Due to the large number of devices, it is impractical to assign orthogonal pilot sequences to all connected devices. This would require a pilot length comparable to the number of devices in the network. Instead, the BS uses a pilot codebook containing non-orthogonal sequences. These sequences can be generated from an i.i.d. Gaussian or Bernoulli distribution. In this context, we consider pilot sequences generated from a complex symmetric Bernoulli distribution. This approach minimizes the probability of pilot collision, i.e., the likelihood of assigning the same pilot to two distinct devices, making it negligible [88].

2.1.3 Data transmission

Owing to the sporadic activity patterns inherent to mMTC networks, only a subset of devices—specifically $K \ll N$ —are active during each coherence interval T_c . The remaining $N - K$ devices are inactive, as illustrated in Fig. 4(b). Furthermore, τ_c symbols are transmitted at each coherence over two phases, as shown in Fig. 4(c)². During the first phase, each active device sends a pilot sequence of length τ_p to the BS.

²We make the assumption that all devices and the Base Station (BS) are synchronized.

Subsequently, in the second phase, the active devices transmit their UL data to the BS using the remaining $\tau_c - \tau_p$ symbols. Within each coherence interval T_c , the BS employs the pilot sequences received during the first phase to identify which devices are active. It then estimates their respective channel states to accurately decode the information data transmitted in the second phase.

Furthermore, to counteract the discrepancies in the channel gain amongst the devices, a power control policy is implemented. Each device $i \in \mathcal{N}$ transmits with a power ρ_i^{UL} that is inversely proportional to the average channel gain [88, 111]. This strategy serves to equalize the effective channel conditions across all devices, thereby enhancing the overall system performance [88].

The received pilot signal at the BS, denoted as $\mathbf{Y} \in \mathbb{C}^{\tau_p \times M}$, is given by

$$\mathbf{Y} = \sum_{i=1}^N \gamma_i \sqrt{\rho_i^{\text{UL}}} \phi_i \mathbf{h}_i^T + \mathbf{W}, \quad (12)$$

where $\mathbf{W} \in \mathbb{C}^{\tau_p \times M}$ is an additive white Gaussian noise with i.i.d. elements, each following $\mathcal{CN}(0, \sigma^2)$, and $\gamma_i \in \mathbb{B}$ is the i th element of the binary device activity indicator vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_N]^T$, defined as

$$\gamma_i = \begin{cases} 1, & i \in \mathcal{A} \\ 0, & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{N}, \quad (13)$$

where $\mathcal{A} \subseteq \{1, \dots, N\}$, $|\mathcal{A}| = K$, is the set of active devices. We further assume that no prior knowledge of the device activity level $\frac{K}{N}$ is available at the BS.

Let us define the effective channel of the i th device as $\mathbf{x}_i = \gamma_i \sqrt{\rho_i^{\text{UL}}} \mathbf{h}_i$, and subsequently, the *effective channel matrix* as $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{C}^{M \times N}$. The pilot sequence matrix is defined as $\Phi = [\phi_1, \dots, \phi_N] \in \mathbb{C}^{\tau_p \times N}$. Accordingly, we can rewrite the received pilot signal in (12) as

$$\mathbf{Y} = \Phi \mathbf{X}^T + \mathbf{W}. \quad (14)$$

The columns of the effective channel matrix \mathbf{X} corresponding to the inactive devices are zero. Hence, \mathbf{X}^T has a *row-sparse* structure. Therefore, joint device identification and channel estimation reduces to the detection of the support of \mathbf{X}^T (i.e., the nonzero columns of \mathbf{X}), as well as estimating their coefficients. Thus, JADCE can be modeled as a sparse MMV reconstruction problem.

2.2 JADCE without side information

The canonical form of JADCE in the MMV setup is given as

$$\min_{\mathbf{X}} \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_F^2 + \beta_1 \|\mathbf{X}^T\|_{2,0}, \quad (15)$$

where $\|\mathbf{X}^T\|_{2,0} = \sum_{i=1}^N 1(\|\mathbf{x}_i\|_2 \neq 0)$ represents the sparsity-promoting penalty, and β_1 controls the trade-off between the emphasis on the measurement matching term and the sparsity-promoting penalty. However, as discussed in the first chapter, solving the optimal JADCE problem using ℓ_0 -“norm” minimization is an intractable combinatorial NP-hard problem. Various solutions have therefore been proposed to address this challenge. The existing solutions can be categorized based on the type of prior information they require about the unknown sparse signal. For example, AMP-based and SBL-based solutions require prior information about the CDI. On the other hand, mixed-norm minimization and most greedy algorithms operate solely on the assumption that the signal is sparse. Although AMP-based and SBL-based sparse recovery algorithms have been shown to offer superior performance in the literature [29], they come with the trade-off of higher computational costs.

In this chapter, we aim to address the JADCE problem by utilizing deterministic sparse recovery frameworks. The proposed frameworks aim to provide computationally efficient solutions, as well as being adaptable to different assumptions regarding the availability of channel statistics at the BS.

2.2.1 JADCE via group LASSO

Without a priori knowledge of the effective channel statistics, it is logical to replace the canonical ℓ_0 -norm with the tractable and convex ℓ_1 -norm penalty. This is a sensible approach, given that ℓ_1 -norm minimization has the best theoretically established recovery threshold among polynomial-complexity decoding algorithms for sparse signals [115].

To this end, and to account the row-sparsity of \mathbf{X}^T , JADCE can be formulated as a group LASSO as follows

$$\min_{\mathbf{X}} \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_F^2 + \beta_1 \|\mathbf{X}^T\|_{2,1}, \quad (16)$$

where $\|\mathbf{X}^T\|_{2,1} = \sum_{i=1}^N \|\mathbf{x}_i\|_2$, and parameter β_1 introduces a trade-off between the measurement fidelity $\|\Phi \mathbf{X}^T - \mathbf{Y}\|_F^2$ and the sparsity level in \mathbf{X} .

2.2.2 JADCE via reweighted $\ell_{2,1}$ -norm minimization

Unlike the democratic ℓ_0 -norm, which penalizes all nonzero coefficients equally, the ℓ_1 -norm is biased toward coefficients with larger magnitudes. In other words, coefficients with larger magnitudes are penalized more heavily than those with smaller ones. As a result, although the group LASSO formulation in equation (16) promotes sparsity and effectively identifies the locations of large nonzero elements in \mathbf{X} , it may fail to detect smaller nonzero elements, leading to inaccurate estimates for these nonzero components.

To achieve improved sparsity recovery, we invoke the *log-sum* penalty [107] to formulate a JADCE problem, that is both efficient and tractable as

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{u}} \quad & \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_F^2 + \beta_1 \sum_{i=1}^N \log(u_i + \varepsilon_0) \\ \text{s.t.} \quad & \|\mathbf{x}_i\|_2 \leq u_i, \forall i \in \mathcal{N}, \end{aligned} \quad (17)$$

where $\mathbf{u} = [u_1, u_2, \dots, u_N]^T$ is a vector of auxiliary optimization variables, and ε_0 is a small positive stability parameter. The log-sum penalty resembles most closely the ℓ_0 -norm penalty when $\varepsilon_0 \rightarrow 0$. However, a practical, and numerically robust choice is to set ε_0 to be slightly less than the expected norm of the nonzero rows in \mathbf{X}^T [107].

As the objective function in (17) is a sum of convex and concave functions, it is not convex in general. We therefore rely on the majorization-minimization (MM) framework [116], and we approximate the concave penalty, i.e., the log-sum penalty, by its first-order Taylor expansion to transform (17) to a convex problem. To this end, we proceed as follows. Given the starting points $(\mathbf{X}^{(l)}, \mathbf{u}^{(l)})$, where (l) denotes the MM iteration index, the MM algorithm iterates the solution by approximating (17) as follows

$$\begin{aligned} (\mathbf{X}^{(l+1)}, \mathbf{u}^{(l+1)}) = \quad & \min_{\mathbf{X}, \mathbf{u}} \frac{1}{2} \|\Phi \mathbf{X} - \mathbf{Y}\|_F^2 + \beta_1 \sum_{i=1}^N \frac{u_i}{(u_i^{(l)} + \varepsilon_0)} \\ \text{s.t.} \quad & \|\mathbf{x}_i\|_2 \leq u_i, \forall i \in \mathcal{N}. \end{aligned} \quad (18)$$

Removing the auxiliary variable \mathbf{u} , the JADCE problem in (18) can be equivalently written as the following *iterative reweighted (IRW)- $\ell_{2,1}$* norm minimization problem

$$\mathbf{X}^{(l+1)} = \min_{\mathbf{X}} \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_F^2 + \sum_{i=1}^N \beta_1 g_i^{(l)} \|\mathbf{x}_i\|_2, \quad (19)$$

where the weights $g_i^{(l)}$, $i \in \mathcal{N}$, are defined as

$$g_i^{(l)} = (\varepsilon_0 + \|\mathbf{x}_i^{(l)}\|_2)^{-1}, \forall i \in \mathcal{N}. \quad (20)$$

By applying the MM approximation, the optimization problem (19) is *convex*. Hence, it can be solved optimally using standard convex optimization techniques. However, as the mMTC networks are large-scale systems by nature, the standard techniques can suffer from high computational complexity. As a remedy, we utilize ADMM [117] to solve (19) iteratively in a computationally efficient manner at each MM iteration (l), as we will show in section 2.4.

2.3 JADCE with channel covariance information

While the group Lasso and the iterative group Lasso formulations in equation (16) and equation (19), respectively, leverage the sparse nature of the effective channel matrix \mathbf{X} , they neglect the information embedded in the estimated channel covariance matrices $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. We therefore propose a novel JADCE problem formulation that exploits both the row sparsity and the covariance information available at the BS.

The foundational ideas behind our proposed formulation are driven by the following considerations: 1) Given that the channel covariance matrices vary very slowly, the sample covariance matrix $\mathbf{x}_i\mathbf{x}_i^H$ for the effective channel computed over a single coherence time for each active device $i \in \mathcal{A}$ should carry similar information to the estimated covariance matrix $\hat{\mathbf{R}}_i$ that has been computed by the BS in the training phase. 2) In practical one-ring models, the covariance matrix is typically low-rank, with many eigenvalues being negligibly small [100, Sect. 2.6]. The results reported in [100] indicate that the effective rank of the covariance matrix is approximately $M/3$ for an ASD of 10° and decreases for a smaller ASD. Consequently, in scenarios where the number of BS antennas M is not large, and the ASD is small, it is feasible to impose a regularization penalty on the mismatch between the one-rank single covariance matrix and the low-rank covariance matrix.

Based on the above argument, we augment the log-sum-based formulation in (17) with a regularization term that penalizes the deviation of the sample covariance matrix $\mathbf{x}_i\mathbf{x}_i^H$ from the estimated scaled covariance matrix $\tilde{\mathbf{R}}_i = \rho_i^{\text{UL}}\hat{\mathbf{R}}_i$, $i \in \mathcal{A}$. Subsequently, the modified JADCE problem, which capitalizes on prior knowledge of the second-order channel statistics, can be formulated as follows

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{u}} \quad & \frac{1}{2} \|\Phi\mathbf{X}^T - \mathbf{Y}\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N \log(u_i + \epsilon_0) + \beta_2 \sum_{i=1}^N \mathbb{I}(u_i) \|\mathbf{x}_i\mathbf{x}_i^H - \tilde{\mathbf{R}}_i\|_{\text{F}}^2 \\ \text{s.t.} \quad & \|\mathbf{x}_i\|_2 \leq u_i, \quad \forall i \in \mathcal{N}, \end{aligned} \quad (21)$$

where the parameter β_2 controls the penalty on the covariance deviation term, and $\mathbb{I}(u_i)$ is an indicator function given by

$$\mathbb{I}(u_i) = \begin{cases} 1, & u_i > 0, \\ 0, & u_i = 0, \end{cases} \quad \forall i \in \mathcal{N}. \quad (22)$$

Note that $\mathbb{I}(u_i)$ ensures that only the estimated active devices are penalized with the covariance regularization term. The indicator function is hard to handle due to its *combinatorial* nature. Therefore, we relax (22) with a smooth approximation by utilizing the function $f_s(\cdot)$, which approximates the sign step function for positive values v , and defined as

$$f_s(v; \kappa) = \frac{\log(1 + \kappa v)}{\log(1 + \kappa)}, \quad (23)$$

where κ is a positive parameter to adjust the steepness of the function for small input values [118]. The choice of the function f_s to approximate the indicator function is a viable. First, it is generally magnitude-independent, so it penalizes the estimated active devices almost equally. 2) It allows a closed-form solution to a sub-problem involved in the optimization procedure, as shown later.

The optimization problem (21) is approximated as

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{u}} \quad & \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N \log(u_i + \epsilon_0) \\ & + \beta_2 \sum_{i=1}^N f_s(u_i; \kappa) \|\mathbf{x}_i \mathbf{x}_i^H - \tilde{\mathbf{R}}_i\|_{\text{F}}^2 \quad \text{s.t.} \quad \|\mathbf{x}_i\|_2 \leq u_i, \quad \forall i \in \mathcal{N}. \end{aligned} \quad (24)$$

Note that since both the log-sum penalty and $f_s(u_i; \kappa)$ are concave functions, we again rely on the MM approach, and we approximate the problem in (24) by its first-order Taylor expansion at $\mathbf{u}^{(l)}$. Subsequently, with the use of some simple manipulations as shown in the previous section, we solve (24) as the following iterative reweighted problem given at the l th MM iteration as

$$\begin{aligned} \mathbf{X}^{(l+1)} = \quad & \min_{\mathbf{X}} \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 \\ & + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i\|_2 \|\mathbf{x}_i \mathbf{x}_i^H - \tilde{\mathbf{R}}_i\|_{\text{F}}^2 \end{aligned} \quad (25)$$

with $q_i^{(l)} = \frac{\kappa}{\log(1 + \kappa)} \frac{1}{1 + \kappa \|\mathbf{x}_i^{(l)}\|_2}$, $\forall i \in \mathcal{N}$. The problem (25) is still a *non-convex* problem. In the next section, we present the proposed solution to this problem.

2.4 Generalized solution via ADMM

It is noteworthy that both the group LASSO approach (16) and the iterative reweighted $\ell_{2,1}$ -norm minimization problem (19) are specific instances of the general optimization problem described by equation (25). To elaborate, equation (19) is equivalent to (25) when $\beta_2 = 0$. Similarly, equation (16) matches (25) when $\beta_2 = 0$ and $\mathbf{g}^{(l)}$ is a unit vector. Within this context, we propose a *generalized* algorithm solution leveraging the ADMM to efficiently solve the optimization problems (25), (19), and (16). ADMM is a rather simple yet efficient framework that is well suited for high-dimensional optimization problems. For instance, Deng *et al.* [119] reported that ADMM outperformed coordinate gradient descent for sparse recovery from single measurement vector in terms of both signal recovery quality and convergence rate.

ADMM has gained wide acceptance for solving computationally intensive problems related to sparse signal recovery [120]. Beyond applications in signal reconstruction, ADMM has found utility in the specific context of device activity detection in mMTC. For instance, Cirik *et al.* [121] introduced an ADMM-based solution for multi-user support and signal detection in an SMV model, incorporating prior information about the recovered signal from previous transmission instances. Additionally, Ying *et al.* [98] proposed an approximation step in ADMM, similar to [120], and solved the ensuing sub-problems through a model-driven deep learning decoder.

In this section, we introduce two ADMM-based solutions for the JADCE problem. The first approach is in line with existing works [98, 120], and it provide an approximate solution for the JADCE problem. Indeed, this solution utilizes the Taylor approximation to linearize any involved quadratic and non-convex functions in the objective function, hence facilitating the derivation of closed-form update rules. Conversely, the second approach to solving the JADCE problem *exactly* employs a strategic splitting technique to convert the non-convex optimization problem into a set of convex sub-problems. Each of these sub-problems has the advantage of admitting a closed-form solution.

2.4.1 ADMM via Taylor approximation

By introducing a splitting variable $\mathbf{Z} \in \mathbb{C}^{M \times N}$, i.e., a copy of optimization variable \mathbf{X} , we can decompose the objective function in (25) into two separate components. The first term is a quadratic function on the measurement fidelity term over \mathbf{Z} , and the second term is a reweighted $\ell_{2,1}$ -norm penalty applied to \mathbf{X} . Subsequently, we rewrite (25)

equivalently as

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z}_1, \dots, \mathbf{Z}_N} \quad & \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\mathbb{F}}^2 + \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i\|_2 \|\mathbf{Z}_i - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2, \\ \text{s.t.} \quad & \mathbf{Z}_i = \mathbf{x}_i \mathbf{x}_i^H, i \in \mathcal{N}. \end{aligned} \quad (26)$$

To proceed with the ADMM solution, we first write the augmented Lagrangian of (26) as

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \Lambda) = & \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 + \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\mathbb{F}}^2 + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i\|_2 \|\mathbf{Z}_i - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2 \\ & + \sum_{i=1}^N \langle \Lambda_i, \mathbf{x}_i \mathbf{x}_i^H - \mathbf{Z}_i \rangle + \frac{\rho}{2} \sum_{i=1}^N \|\mathbf{x}_i \mathbf{x}_i^H - \mathbf{Z}_i\|_{\mathbb{F}}^2, \end{aligned} \quad (27)$$

where $\Lambda = \{\Lambda_i \in \mathbb{C}^{M \times M}\}_{i=1}^N$ denotes the set containing the ADMM dual variable matrices, and ρ is a positive parameter for adjusting the convergence of the ADMM. Combining the last two terms in (27) and with some manipulations, we can express (27) in the following form, known as the scaled augmented Lagrangian:

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \Lambda) = & \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 + \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\mathbb{F}}^2 - \frac{\rho}{2} \|\frac{\Lambda_i}{\rho}\|_{\mathbb{F}}^2 \\ & + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i\|_2 \|\mathbf{Z}_i - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2 + \frac{\rho}{2} \|\mathbf{x}_i \mathbf{x}_i^H - \mathbf{Z}_i + \frac{1}{\rho} \Lambda_i\|_{\mathbb{F}}^2. \end{aligned} \quad (28)$$

The basic idea of ADMM is to iterate the optimization through three steps sequentially. First, $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \Lambda)$ is minimized over the set of primal variables \mathbf{Z} for fixed \mathbf{X} and Λ . Second, $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \Lambda)$ is minimized over primal variable \mathbf{X} for fixed \mathbf{Z} and Λ . Finally, we update the set of Lagrange multipliers Λ using a gradient ascent step.

Let (k) denotes the ADMM iteration index. The ADMM algorithm iterates as follows [117, Eqs. (3.2)–(3.4)]:

$$\begin{aligned} \mathbf{Z}_i^{(k+1)} := \min_{\mathbf{Z}_i} \left\{ \beta_2 q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \|\mathbf{Z}_i - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2 \right. \\ \left. + \frac{\rho}{2} \|\mathbf{x}_i^{(k)} \mathbf{x}_i^{(k)H} - \mathbf{Z}_i + \frac{1}{\rho} \Lambda_i^{(k)}\|_{\mathbb{F}}^2 \right\}, \forall i \in \mathcal{N}, \end{aligned} \quad (29)$$

$$\begin{aligned} \mathbf{X}^{(k+1)} := \min_{\mathbf{X}} \left\{ \sum_{i=1}^N \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \frac{1}{2} \|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\mathbb{F}}^2 \right. \\ \left. + \frac{\rho}{2} \sum_{i=1}^N \|\mathbf{x}_i \mathbf{x}_i^H - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho} \Lambda_i^{(k)}\|_{\mathbb{F}}^2 \right\}, \end{aligned} \quad (30)$$

where $\alpha_i^{(k)} = (\beta_1 g_i^{(l)} + \beta_2 \|\mathbf{Z}_i^{(k+1)} - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2)$, $\forall i \in \mathcal{N}$, and

$$\Lambda_i^{(k+1)} := \Lambda_i^{(k)} + \rho (\mathbf{x}_i^{(k+1)} \mathbf{x}_i^{(k+1)H} - \mathbf{Z}_i^{(k+1)}), \forall i \in \mathcal{N}. \quad (31)$$

We present the derivations of the ADMM steps (29) and (30) in detail below.

Z-update: Finding the optimal solution for (29) decouples into solving N convex sub-problems in parallel. Thus, the unique optimal solution is obtained by setting the derivative of (29) with respect to \mathbf{Z}_i , $i \in \mathcal{N}$, to zero, resulting in

$$\mathbf{Z}_i^{(k+1)} = \frac{2\beta_2 q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \tilde{\mathbf{R}}_i + \Lambda_i^{(k)} + \rho \mathbf{x}_i^{(k)} \mathbf{x}_i^{(k)H}}{2\beta_2 q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 + \rho}, \forall i \in \mathcal{N}. \quad (32)$$

X-update: First, it is important to recognize that the objective function in (30), which includes a fourth-order polynomial term, is non-convex. While it is feasible to tackle equation (30) using traditional iterative methods like steepest descent, this would necessitate an additional inner loop of iterations within each primary ADMM iteration (denoted by (k)). This approach becomes computationally burdensome for large-scale problems. To circumvent this issue, we resort to Taylor expansion to approximate the solution to (30).

The use of Taylor expansion serves three critical purposes. First, it linearizes both the quadratic and non-convex terms in the equation. Second, it enables the decomposition of the overarching optimization problem into N individual sub-problems. Finally, each of these N sub-problems is structured to have a closed-form solution, making the computational process more efficient.

The quadratic term in (30) is approximated by the first-order Taylor expansion at $\mathbf{X}^{(k)}$ as

$$\|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\mathbb{F}}^2 \approx \|\Phi \mathbf{X}^{(k)} - \mathbf{Y}\|_{\mathbb{F}}^2 + \langle \mathbf{F}_q^{(k)} - \mathbf{X}^{(k)}, \mathbf{X} - \mathbf{X}^{(k)} \rangle, \quad (33)$$

where $\mathbf{F}_q^{(k)} = 2(\mathbf{X}^{(k)} \Phi^* - \mathbf{Y}) \Phi^T$ is the gradient of $\|\Phi \mathbf{X}^T - \mathbf{Y}\|_{\mathbb{F}}^2$ with respect to \mathbf{X} at point $\mathbf{X}^{(k)}$. Similarly, the non-convex term in (30) is approximated by the first-order Taylor expansion as

$$\|\mathbf{x}_i \mathbf{x}_i^H - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho} \Lambda_i\|_{\mathbb{F}}^2 \approx \|\mathbf{x}_i^{(k)} \mathbf{x}_i^{(k)H} - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho} \Lambda_i\|_{\mathbb{F}}^2 + \langle \mathbf{f}_i^{(k)}, \mathbf{x}_i - \mathbf{x}_i^{(k)} \rangle, \forall i \in \mathcal{N}, \quad (34)$$

where $\mathbf{f}_i^{(k)} = 4(\mathbf{x}_i^{(k)} \mathbf{x}_i^{(k)H} - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho} \Lambda_i) \mathbf{x}_i$.

Due to the linearization, the Taylor approximations in (33) and (34) are accurate only for \mathbf{X} near $\mathbf{X}^{(k)}$. We therefore add a quadratic proximal regularization term $\frac{1}{\tau} \|\mathbf{X} - \mathbf{X}^{(k)}\|_{\mathbb{F}}^2$ with parameter $\tau > 0$ to the objective function of (30) [122]. Subsequently, we reformulate (30) as follows

$$\begin{aligned} \mathbf{X}^{(k+1)} &\approx \min_{\mathbf{X}} \sum_i^N \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \frac{1}{2} \langle 2(\mathbf{X}^{(k)} \Phi^* - \mathbf{Y}) \Phi^T, \mathbf{X} - \mathbf{X}^{(k)} \rangle \\ &+ \sum_{i=1}^N \frac{\rho}{2} \langle \mathbf{f}_{c_i}, \mathbf{x}_i - \mathbf{x}_i^{(k)} \rangle + \frac{1}{\tau} \|\mathbf{X} - \mathbf{X}^{(k)}\|_{\mathbb{F}}^2. \end{aligned} \quad (35)$$

The optimization in (35) can be decomposed into N separable sub-problems as

$$\mathbf{x}_i^{(k+1)} \approx \min_{\mathbf{x}_i} \left\{ \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \frac{1}{2} \langle \mathbf{f}_{q_i} + \rho \mathbf{f}_{c_i}, \mathbf{x}_i - \mathbf{x}_i^{(k)} \rangle + \frac{1}{\tau} \|\mathbf{x}_i - \mathbf{x}_i^{(k)}\|_2^2 \right\}, \forall i \in \mathcal{N}, \quad (36)$$

where $\mathbf{f}_{q_i} \in \mathbb{C}^M$ is the i th row of \mathbf{F}_{q_i} in (33). By adding an appropriate constant term and applying some manipulations, we can reformulate the objective function in (36) as

$$\frac{\tau \alpha_i^{(k)}}{2} \|\mathbf{x}_i\|_2 + \frac{1}{2} \|\mathbf{x}_i - \mathbf{c}_i^{(k)}\|_2^2, \quad \forall i \in \mathcal{N}, \quad (37)$$

where $\mathbf{c}_i^{(k)} = \mathbf{x}_i^{(k)} - \frac{\tau}{2} (\mathbf{f}_{q_i} + \rho \mathbf{f}_{c_i})$. The expression in (37) admits a closed-form solution to (36), given by the soft thresholding operator [53] as

$$\mathbf{x}_i^{(k+1)} = \frac{\max \left\{ 0, \|\mathbf{c}_i^{(k)}\|_2 - \frac{\tau \alpha_i^{(k)}}{2} \right\}}{\|\mathbf{c}_i^{(k)}\|_2} \mathbf{c}_i^{(k)}, \quad \forall i \in \mathcal{N}. \quad (38)$$

Λ -update: Finally, we update the set of Lagrange multipliers in (31) to obtain $\Lambda_i^{(k+1)}$ for $i = 1, \dots, N$.

Algorithm implementation

The details for the proposed algorithm in this section, termed cov-ADMM, are summarized in Algorithm 1. cov-ADMM is run until the \mathbf{X} -update is converged, measured as $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_{\mathbb{F}}^2 < \varepsilon_{\text{stp}}$, with a predefined tolerance parameter $\varepsilon_{\text{stp}} > 0$, or until a maximum number of iterations k_{max} is reached, where k_{max} denotes the maximum number of ADMM iterations.

Algorithm 1: cov-ADMM

Input: 1) Pilot matrix Φ , 2) parameters $\beta_1, \beta_2, \rho, \tau, \varepsilon_0, \varepsilon_{\text{stp}}, k_{\text{max}}$

Output: $\hat{\mathbf{X}}$

Initialization: $\mathbf{X}^{(0)}, \mathbf{Z}^{(0)}, \Lambda^{(0)}, k = 1$

- 1 Receive \mathbf{Y} at the BS
 - 2 **while** $k < k_{\text{max}}$ or $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_{\mathbb{F}}^2 < \varepsilon_{\text{stp}}$ **do**
 - 3 Update $\mathbf{Z}^{(k+1)}$ using (32)
 - 4 Update $\mathbf{X}^{(k+1)}$ using (38)
 - 5 $\Lambda^{(k+1)} = \Lambda^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)})$
 - 6 $k \leftarrow k + 1$
-

2.4.2 Exact solution

In contrast with the approximate solution presented above, this section derives an *exact solution* to the JADCE formulation proposed in (25) by adopting a different variable splitting strategy. More precisely, the proposed splitting technique decomposes the objective function in (25) into three separable convex functions that can be solved efficiently via simple analytical formulas. In particular, we derive a set of update rules to solve (25) iteratively in a sequential fashion over multiple convex sub-problems, where each sub-problem admits a *closed-form solution*, as we will show next.

First, we introduce the *two splitting variables* $\mathbf{Z} \in \mathbb{C}^{M \times N}$ and $\mathbf{V} \in \mathbb{C}^{M \times N}$. Hence, we rewrite the objective function in (25) as

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\Phi \mathbf{Z}^T - \mathbf{Y}\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i\|_2 \|\mathbf{z}_i \mathbf{v}_i^{\text{H}} - \tilde{\mathbf{R}}_i\|_{\text{F}}^2 \\ \text{s.t.} \quad \mathbf{x}_i = \mathbf{z}_i, \quad \mathbf{x}_i = \mathbf{v}_i, \quad \forall i \in \mathcal{N}. \end{aligned} \quad (39)$$

The optimization problem in (39) is *block multi-convex*, i.e., the problem is convex in one set of variables, while all the other variables are fixed. Since ADMM implicitly exploits the block multi-convexity nature of (39), utilizing ADMM to solve (39) is a reasonable choice. Accordingly, the scaled augmented Lagrangian associated with (39) is given by

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{V}, \mathbf{X}} \frac{1}{2} \|\Phi \mathbf{Z}^T - \mathbf{Y}\|_{\text{F}}^2 + \sum_{i=1}^N \beta_1 g_i^{(l)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i\|_2 \|\mathbf{z}_i \mathbf{v}_i^{\text{H}} - \tilde{\mathbf{R}}_i\|_{\text{F}}^2 + \\ + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z} + \frac{\Lambda_z}{\rho}\|_{\text{F}}^2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{V} + \frac{\Lambda_v}{\rho}\|_{\text{F}}^2 - \frac{\|\Lambda_v\|_{\text{F}}^2}{2\rho} - \frac{\|\Lambda_z\|_{\text{F}}^2}{2\rho}, \end{aligned} \quad (40)$$

where $\Lambda_z = [\lambda_{z_1}, \dots, \lambda_{z_N}]$ and $\Lambda_v = [\lambda_{v_1}, \dots, \lambda_{v_N}]$ are the ADMM dual variable matrices.

The ADMM solution to the optimization problem (39) is achieved by sequentially minimizing $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{V}, \Lambda_z, \Lambda_v)$ over the primal variables $(\mathbf{Z}, \mathbf{V}, \mathbf{X})$, followed by dual variable (Λ_z, Λ_v) updates as follows.

First, the \mathbf{Z} -subproblem, i.e., minimizing (40) with respect to \mathbf{Z} , is given by

$$\begin{aligned} \mathbf{Z}^{(k+1)} := \min_{\mathbf{Z}} \frac{1}{2} \|\Phi \mathbf{Z}^T - \mathbf{Y}\|_{\text{F}}^2 + \frac{\rho}{2} \|\mathbf{X}^{(k)} - \mathbf{Z} + \frac{\Lambda_z^{(k)}}{\rho}\|_{\text{F}}^2 \\ + \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \|\mathbf{z}_i \mathbf{v}_i^{(k)\text{H}} - \tilde{\mathbf{R}}_i\|_{\text{F}}^2. \end{aligned} \quad (41)$$

The objective function in (41) is convex, and the solution is obtained by setting the gradient with respect to \mathbf{Z} to zero, resulting in

$$\mathbf{Z}^{(k+1)} = (\mathbf{Y}^T \Phi^* + \mathbf{d}^{(k)}) (\Phi^T \Phi^* + (2\beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \|\mathbf{v}_i^{(k)}\|_2^2 + \rho) \mathbf{I}_M)^{-1}, \quad (42)$$

where $\mathbf{d}_i^{(k)} = 2\beta_2 q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \tilde{\mathbf{R}}_i \mathbf{v}_i^{(k)} + \rho \mathbf{x}_i^{(k)} + \lambda_{z_i}^{(k)}$ is the i th column of matrix $\mathbf{D}^{(k)}$.

Second, the \mathbf{V} -update solves the minimization problem given by

$$\mathbf{V}^{(k+1)} := \min_{\mathbf{V}} \beta_2 \sum_{i=1}^N q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \|\mathbf{z}_i^{(k+1)} \mathbf{v}_i^H - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2 + \frac{\rho}{2} \|\mathbf{X}^{(k)} - \mathbf{V} + \frac{\Lambda_{\mathbf{V}}^{(k)}}{\rho}\|_{\mathbb{F}}^2. \quad (43)$$

The optimization problem in (43) can be decoupled into N convex sub-problems, each a unique solution given by:

$$\mathbf{v}_i^{(k+1)} = \frac{2\beta_2 q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \tilde{\mathbf{R}}_i \mathbf{z}_i^{(k+1)} + \rho \mathbf{x}_i^{(k)} + \lambda_{v_i}^{(k)}}{2\beta_2 q_i^{(l)} \|\mathbf{x}_i^{(k)}\|_2 \|\mathbf{z}_i^{(k+1)}\|_2^2 + \rho}, \quad \forall i \in \mathcal{N}. \quad (44)$$

Next, with some manipulations, the \mathbf{X} -update solves the following convex optimization problem

$$\mathbf{X}^{(k+1)} := \min_{\mathbf{X}} \sum_{i=1}^N \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \rho \|\mathbf{X} - \mathbf{C}^{(k)}\|_{\mathbb{F}}^2, \quad (45)$$

where $\mathbf{C}^{(k)} = \frac{1}{2} (\mathbf{Z}^{(k+1)} + \mathbf{V}^{(k+1)} - \frac{\Lambda_{\mathbf{Z}}^{(k)} + \Lambda_{\mathbf{V}}^{(k)}}{\rho})$, and $\alpha_i^{(k)} = \beta_1 g_i^{(l)} + \beta_2 q_i^{(l)} \|\mathbf{z}_i^{(k+1)} \mathbf{v}_i^{(k+1)H} - \tilde{\mathbf{R}}_i\|_{\mathbb{F}}^2$. The optimal solution to (45) has a closed-form expression given by

$$\mathbf{x}_i^{(k+1)} = \frac{\max\{0, \|\mathbf{c}_i^{(k)}\|_2 - \frac{\alpha_i^{(k)}}{2\rho}\}}{\|\mathbf{c}_i^{(k)}\|_2} \mathbf{c}_i^{(k)}, \quad \forall i \in \mathcal{N}. \quad (46)$$

Finally, the dual variables are updated as follows

$$\Lambda_{\mathbf{Z}}^{(k+1)} := \Lambda_{\mathbf{Z}}^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)}). \quad (47)$$

$$\Lambda_{\mathbf{V}}^{(k+1)} := \Lambda_{\mathbf{V}}^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{V}^{(k+1)}). \quad (48)$$

Algorithm implementation

We term the algorithm presented in this section as cov-IRW-ADMM, and its details are summarized in Algorithm 2. As one stopping criterion, Algorithm 1 is run until $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_{\mathbb{F}}^2 < \epsilon_{\text{stp}}$, or until a maximum number of iterations $l_{\text{max}} k_{\text{max}}$ is reached, where l_{max} and k_{max} denotes the maximum number of MM iterations and ADMM iterations respectively. Furthermore, note that:

- If we set $\beta_2 = 0$, Algorithm 2 solves the IRW- $\ell_{2,1}$ norm minimization in (19), and we refer to the algorithm as IRW-ADMM.
- If we set $\beta_2 = 0$, and the weight vector is fixed to $\mathbf{g}^{(l)} = \mathbf{1}$, $l = 1, 2, \dots$, Algorithm 2 provides the ADMM solution for the group LASSO problem (16), which we term ADMM henceforth.

Algorithm 2: cov-IRW-ADMM

Input: 1) Pilot matrix Φ , 2) parameters $\beta_1, \beta_2, \rho, \epsilon_0, \epsilon_{\text{stp}}, l_{\text{max}}, k_{\text{max}}$

Output: $\hat{\mathbf{X}}$

Initialization: 1) $\mathbf{X}^{(0)}, \mathbf{V}^{(0)}, \mathbf{Z}^{(0)}, \Lambda_z^{(0)}, \Lambda_v^{(0)}, k = 1, l = 1$, and 2)

$$(\Phi^T \Phi^* + \rho \mathbf{I}_N)^{-1}$$

1 Receive \mathbf{Y} at the BS, and compute and store $\mathbf{Y}^T \Phi^*$

2 **while** $l < l_{\text{max}}$ **do**

3 **while** $k < k_{\text{max}}$ or $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_{\text{F}}^2 < \epsilon_{\text{stp}}$ **do**

4 Update $\mathbf{Z}^{(k+1)}$ using (41)

5 Update $\mathbf{v}^{(k+1)}$ using (43)

6 Update $\mathbf{X}^{(k+1)}$ using (46)

7 $\Lambda_z^{(k+1)} := \Lambda_z^{(k)} + \rho(\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)})$

8 $\Lambda_v^{(k+1)} := \Lambda_v^{(k)} + \rho(\mathbf{X}^{(k+1)} - \mathbf{V}^{(k+1)})$

9 $k \leftarrow k + 1$

10 $\mathbf{X}^{(l)} \leftarrow \mathbf{X}^{(k+1)}$

11 $g_i^{(l)} = (\epsilon_0 + \|\mathbf{x}_i^{(l)}\|_2)^{-1}, i \in \mathcal{N}$

12 $q_i^{(l)} = \frac{\kappa}{\log(1 + \kappa)} \frac{1}{1 + \kappa \|\mathbf{x}_i^{(l)}\|_2}, \forall i \in \mathcal{N}$

13 $l \leftarrow l + 1$

2.5 A note on the availability of second-order statistics

The proposed algorithm operates on the assumption that the BS knows the second-order statistics of the individual channels, i.e., $\{\mathbf{R}\}_{i=1}^N$. We note that the assumption that the BS knows $\{\mathbf{R}_i\}_{i=1}^N$ is widely accepted in the massive MIMO literature [100, 113]. Furthermore, a similar assumption on the availability of the second-order statistics has been adopted in [123] that addresses JADCE in mMTC with sporadic user activity.

The acquisition of the second-order statistics at the BS may be challenging, especially for devices that are inactive for a long period. Therefore, a possible solution to circumvent such an issue can be realized by deploying a training phase to estimate the second-order statistics. The training phase can be implemented over separate channel resource blocks that are solely dedicated to estimating the second-order statistics. In practice, the BS would consume a set of available channel resources to obtain an estimate of all the channel covariance matrices, denoted as $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. In particular, at different time intervals, specific of devices transmit pre-assigned orthogonal training pilots to the BS over T coherence intervals, and subsequently, the BS employs conventional MIMO

channel estimation techniques to obtain T estimates of channel responses \mathbf{h}_i , denoted as $\hat{\mathbf{h}}_i^1, \dots, \hat{\mathbf{h}}_i^T$. Subsequently, the BS computes the estimated channel covariance matrix³ for the i th UE as $\hat{\mathbf{R}}_i = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{h}}_i^t \hat{\mathbf{h}}_i^{tH}$.

The frequency of updating the second-order statistics at the BS depends on the mobility and the activity level of the devices, as well as the changes in the multi-path environment. Therefore, the estimated $\hat{\mathbf{R}}_i$ can be used over several coherence intervals because: 1) the channel covariance matrices vary on a slower timescale compared to the channel coherence time, and 2) the devices have very low mobility in many practical mMTC systems. Consequently, learning the second-order statistics does not consume a disproportionate amount of resources. As we will show in the next chapter, numerically, the BS does not require a large number of training samples T to estimate $\{\mathbf{R}_i\}_{i=1}^N$. In fact, the BS needs roughly $T = 2M$ samples to provide near-optimal results in terms of the mean square error for channel estimation, and we note that similar conclusion has been reported in [100, Sect. 3.3.3].

2.6 Numerical results

In this section, we provide numerical results to evaluate the performance of the proposed method. Specifically, we focus on three key metrics: device activity detection accuracy; channel estimation quality; and convergence rate. These metrics are compared with various MMV sparse recovery algorithms.

2.6.1 Simulation setup

System setup: The BS employs a ULA to serve a total of $N = 200$ devices that are randomly distributed in a cell with a radius of 100 m. We set the channel ASD $\sigma_\psi = 5^\circ$, and we fix the number of propagation channel paths between the i th device and the BS to $P_i = 200^4$, $\forall i \in \mathcal{N}$. At each coherence interval T_c , $K = 10$ devices are active. Each device $i \in \mathcal{N}$ is assigned with a unique normalized quadratic phase shift keying (QPSK) sequence ϕ_i . The QPSK pilot symbols are drawn from an i.i.d. complex Bernoulli distribution. In addition, the BS uses $T = 100$ channel realizations in the training phase to estimate $\{\hat{\mathbf{R}}_i\}_{i=1}^N$.

³For a recent review of the different channel estimation and channel covariance estimation techniques in massive MIMO networks, we refer the reader to [100, Sect. 3.2-3.3].

⁴Note that knowledge of the number of paths is apriori unknown to the BS, and it does not affect the performance of the proposed algorithms.

Baselines: The performance of the proposed method is compared to the following algorithms that solve the sparse MMV problem in (16):

- F-ADM algorithm [120], which utilizes the same Taylor approximation approach as in cov-ADMM, while ignoring the prior second-order statistics.
- SPARROW, which reformulates (16) as a semi-definite programming problem [28, Eq. (22)], and we solve it using the CVX toolbox [124].
- SOMP [31], which represents the extension of the classical OMP to account for the row-sparsity of the matrix \mathbf{X}^T .

In addition, we use the oracle least square (LS) that is provided “oracle” knowledge on the true set of active UEs, i.e., set \mathcal{A} . The oracle LS estimator provides a good benchmark for channel estimation when no prior knowledge of the signal statistics is available at the BS.

Metrics: The channel estimation accuracy of the proposed JADCE solutions is quantified in terms of NMSE, which is defined as

$$\text{NMSE} = \frac{\mathbb{E} [\|\mathbf{X}_{\mathcal{A}} - \hat{\mathbf{X}}_{\mathcal{A}}\|_{\mathbb{F}}^2]}{\mathbb{E} [\|\mathbf{X}_{\mathcal{A}}\|_{\mathbb{F}}^2]}, \quad (49)$$

where $\mathbf{X}_{\mathcal{A}}$ and $\hat{\mathbf{X}}_{\mathcal{A}}$ denote the original and estimated channel matrix respectively, restricted to the true active support \mathcal{A} . Note that the expectation in (49) is computed via Monte-Carlo averaging over the randomness of effective channel matrix \mathbf{X} , the pilot sequence matrix Φ , and noise \mathbf{W} ; thus, the NMSE is presented as the *normalized average square error* (NASE)⁵.

The activity detection accuracy is evaluated in terms of the support recovery rate (SRR), defined as

$$\text{SRR} = \frac{|\mathcal{A} \cap \hat{\mathcal{A}}|}{|\mathcal{A} - \hat{\mathcal{A}}| + K}, \quad (50)$$

where $\hat{\mathcal{A}} = \{i \mid \|\mathbf{x}_i\|_2 > \epsilon_{\text{thr}}, \forall i \in \mathcal{N}\}$ denotes the detected support for a pre-defined threshold parameter ϵ_{thr} . Thus, $|\mathcal{A} \cap \hat{\mathcal{A}}|$ denotes the number of correctly identified active devices, whereas $|\mathcal{A} - \hat{\mathcal{A}}|$ accounts both for the number of misdetected active devices and falsely identified inactive devices. The SRR rate approaches 1 when $\hat{\mathcal{A}}$ is close to the true \mathcal{A} .

⁵To maintain clarity and consistency throughout the manuscript, the terms NMSE and NASE are used interchangeably.

Parameter tuning: The recovery performance of the algorithms depends on their hyper-parameters. For all algorithms, the sparsity of the solution is controlled by β_1 which is usually selected via a trial and error approach. In our simulations, we set β_1 based on the noise variance σ^2 as $\beta_1 = \sqrt{\frac{\sigma^2}{2}}$ because it provided the most robust convergence for the SNR range $[0 - 20]$ dB. The proposed approach also depends on β_2 , which controls the emphasis on the covariance deviation term. We used empirical tuning and found a robust choice for β_2 to 25 % of the average norm of the effective channels.

2.6.2 Performance without side information

First, we examine the scenario when no prior knowledge of the channel statistics is available to the BS. To this end, we compare the performance of the proposed ADMM and IRW-ADMM algorithms (Algorithm 1) with F-ADM, SPARROW, SOMP, and oracle LS.

Fig. 5 illustrates the obtained SRR against SNR for the different sparse recovery algorithms. The obtained results indicate that our proposed IRW-ADMM outperforms other methods by achieving the highest device activity detection accuracy. Specifically, with a pilot sequence length of $\tau_p = 20$, IRW-ADMM using pilot sequence length $\tau_p = 20$ attains an SRR of = 0.95 when $\text{SNR} > 8$ dB. Remarkably, even when the pilot sequence length is reduced by 25 % to $\tau_p = 15$, IRW-ADMM continues to surpass other MMV sparse recovery algorithms by a significant margin. Furthermore, Fig. 5 demonstrates that even the proposed ADMM, which solves the problem exactly, delivers superior performance compared to other algorithms solving the conventional group LASSO problem, including SOMP and SPARROW. Nonetheless, it achieves the same performance as F-ADM. It is noteworthy that ADMM uses fewer regularization parameters than F-ADM, potentially making it more resilient to parameter tuning.

Fig. 6 depicts the channel estimation performance for the different recovery algorithms in terms of NASE against SNR, including the comparison to the genie-aided LS benchmark. It can be readily seen that for $\tau_p = 20$, and the performance of the proposed IRW-ADMM matches the performance of the genie-aided LS. Notably, even with a reduced pilot sequence length of $\tau_p = 15$, IRW-ADMM still outperforms ADMM, F-ADM, SOMP, and SPARROW for $\text{SNR} > 8$ dB. Similarly to the SRR performance, both ADMM and F-ADMM provide the same NASE, while outperforming SPARROW and SOMP.

Figure 7 illustrates the typical convergence behavior of ADMM, IRW-ADMM, and F-ADM at an SNR of $\text{SNR} = 16$ dB. The figure indicates the number of iterations required for each algorithm to reach optimal performance. The results show that

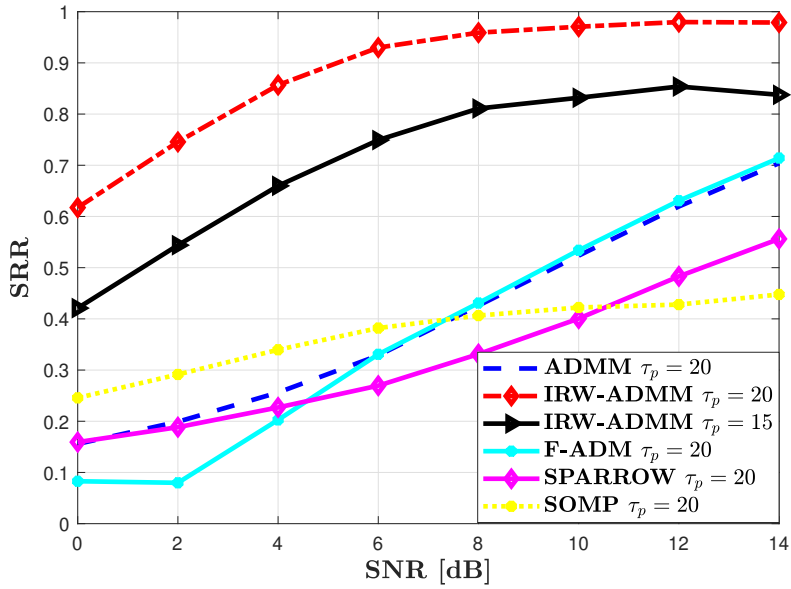


Fig. 5. Activity detection performance of JADCE with no side information in terms of SRR for $N = 200$, $M = 20$, and $K = 10$.

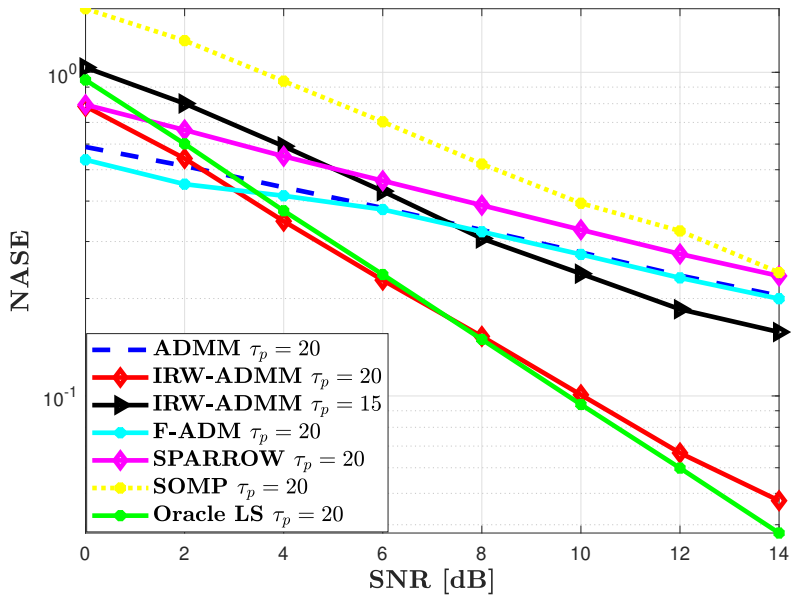


Fig. 6. Channel estimation performance of JADCE with no side information in terms of NASE for $N = 200$, $M = 20$, and $K = 10$.

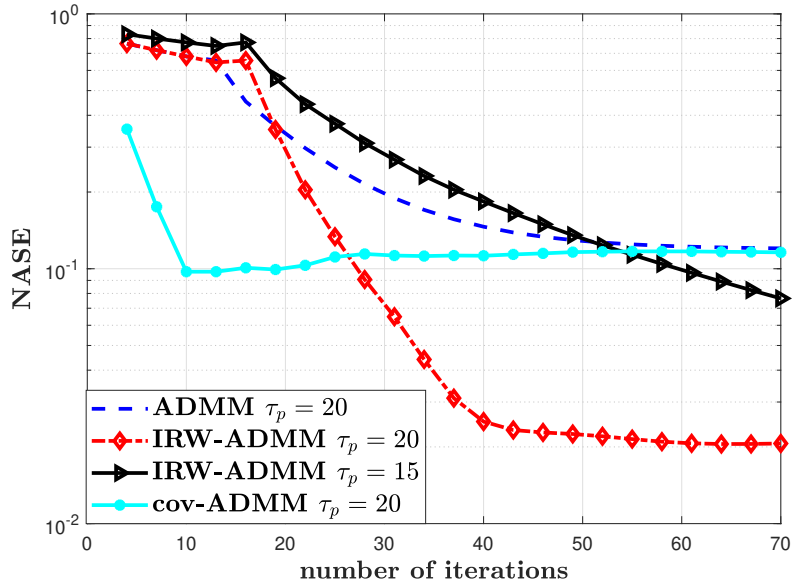


Fig. 7. Convergence behaviour in terms of NASE versus the number of iterations at SNR = 16 dB for $N = 200$, $M = 20$, and $K = 10$.

when using a pilot sequence length of $\tau_p = 20$, IRW-ADMM takes approximately 450 iterations to converge. The rate of convergence decreases when the pilot sequence length is reduced to $\tau_p = 15$. Interestingly, unlike in the cases of SRR and NASE performance, where ADMM and F-ADM exhibited similar efficacy, F-ADM converges more rapidly, taking only about 15 iterations. This is in contrast to the proposed ADMM, which requires approximately 50 iterations to reach convergence.

2.6.3 Performance with side information

In this section, we quantify the impact of exploiting the available prior knowledge of the second-order statistics of the channel in designing the JADCE solution.

Fig. 8 depicts the activity detection performance in terms of SRR. The results show that for pilot sequence length $\tau_p = 20$, the proposed methods (cov-ADMM and IRW-ADMM) indisputably provide the best activity detection quality, where they can both identify the set of true active devices \mathcal{A} perfectly at SNR > 8 dB. While the performance of cov-ADMM and IRW-ADMM would naturally degrade when reducing the pilot length by a factor of 25 % (i.e., $\tau_p = 15$), they still slightly outperform IRW-ADMM that utilizes $\tau_p = 15$. Hence, the significant reduction in signaling overhead clearly

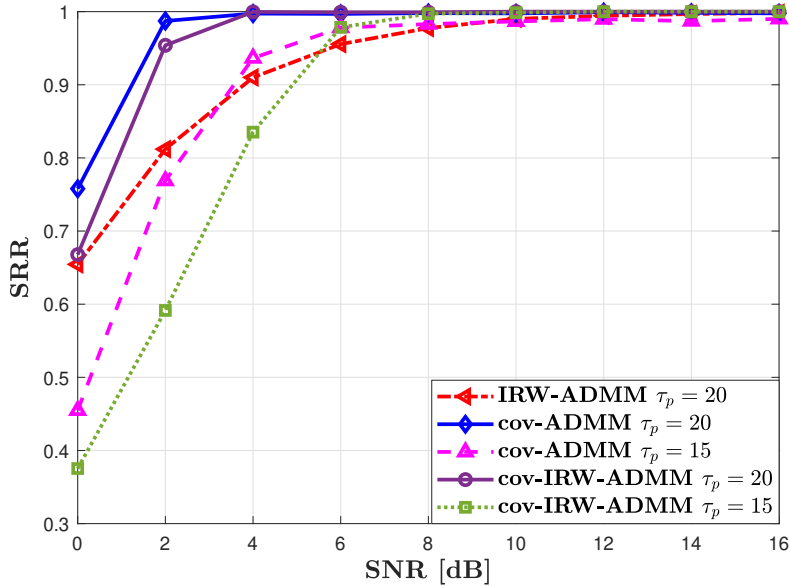


Fig. 8. Impact of prior knowledge on the activity detection performance in terms of NASE, for $N = 200$, $M = 20$, and $K = 10$.

outweighs the minor performance degradation. The obtained results imply that utilizing the prior channel covariance information is highly beneficial in the identification of the true set of active devices.

We now focus on the performance of channel estimation. Fig. 9 compares the NASE performance for the different algorithms. It can be seen that for the pilot sequence length of $\tau_p = 20$, the proposed cov-IRW and cov-IRW-ADMM provide the best performance, and they practically the same performance. Additionally, Fig. 9 illustrates one advantageous feature of utilizing prior channel statistics information: For $\text{SNR} < 16$ dB, both cov-ADMM and cov-IRW-ADMM with $\tau_p = 15$ still outperform the oracle LS, that uses a pilot length of $\tau_p = 20$ with a gain of 6 – 8 dB. Furthermore, the gains from employing our iterative reweighted approach become particularly pronounced for shorter pilot sequence lengths. This is evidenced by the fact that cov-IRW-ADMM offers a 2 dB gain in performance compared to its cov-ADMM counterpart.

The results in Figs. 8 and Figs. 9 show that the performance of the proposed approach depends heavily on the length of the pilot sequence. Although device activity detection achieves near-optimal performance at $\tau_p = 1.5K$, channel estimation accuracy continues to improve significantly with the increase of τ_p . Nevertheless, increasing τ_p will come at the cost of a large signaling overhead. Therefore, a trade-off has to be made between

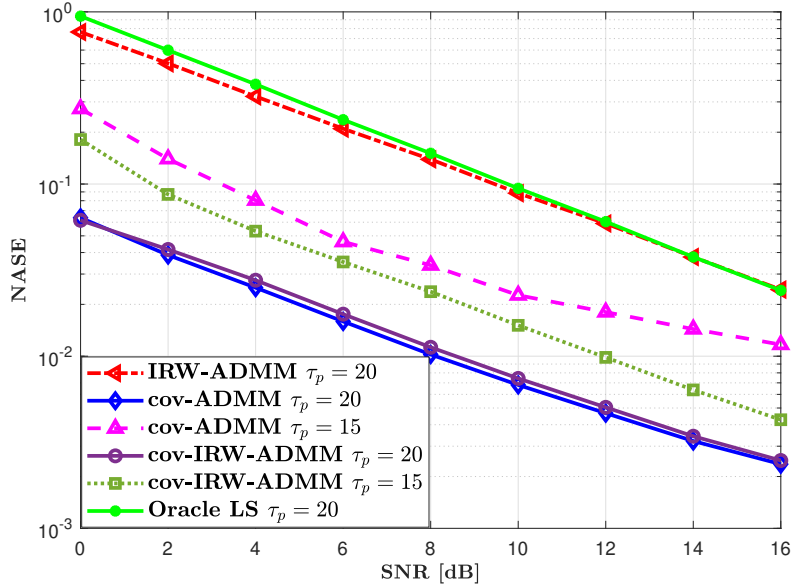


Fig. 9. Impact of prior knowledge on the channel estimation performance in terms of NASE, for $N = 200$, $M = 20$, and $K = 10$.

recovery performance and signaling overhead in the design of the optimal pilot sequence length.

Fig. 10 displays the typical convergence behavior of the proposed algorithms SNR = 20 dB. The results reveal that for $\tau_p = 20$, cov-ADMM exhibits the fastest convergence rate, as it requires about 20 iterations for convergence, whereas cov-IRW-ADMM converges in 35 iterations. Again, the performance gap between cov-IRW-ADMM and cov-ADMM becomes more pronounced when the pilot sequence length is shortened to $\tau_p = 15$. Specifically, reducing the pilot length leads to a slower convergence rate for cov-ADMM, which then requires twice the number of iterations to converge. Interestingly, cov-IRW-ADMM does not experience a similar decline in the convergence rate when the pilot sequence length is reduced.

Fig. 11 shows the active device detection performance as a function of SNR for different numbers of BS antennas M using cov-ADMM. A key observation from the results is the diminishing returns experienced with increasing M . Specifically, the performance gains are more pronounced when M is increased from $M = \frac{K}{2}$ to $M = K$, as compared to the increase from $M = K$ to $M = 3K$. This result can be interpreted as the relative improvement in device activity detection gradually declines as the number of antennas M continues to rise. It suggests an important trade-off in system design: there

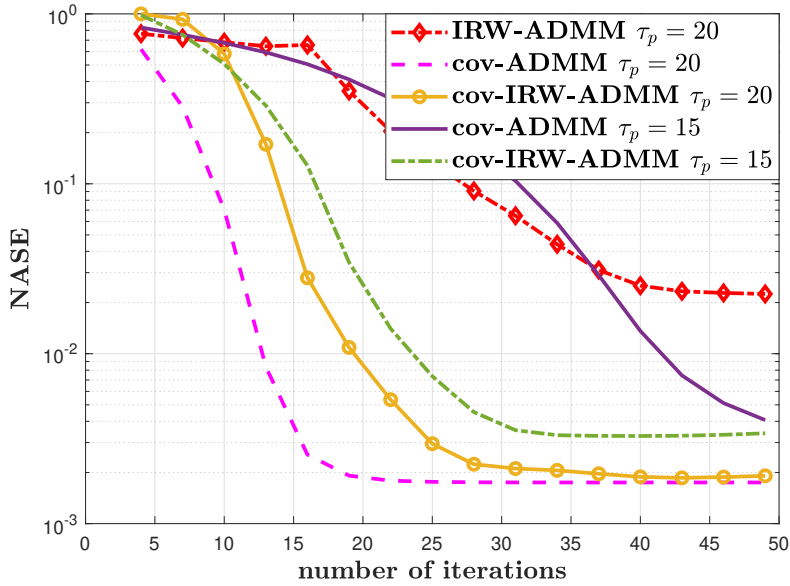


Fig. 10. Impact of prior knowledge on the convergence rate in terms of NASE vs. the number of iterations at SNR = 20 dB.

appears to be an optimal number of antennas beyond which additional increases do not contribute as substantially to performance improvement.

This trend is particularly insightful when considering massive MIMO configurations, where $M > K$. In such regimes, our results indicate that focusing on increasing the pilot sequence length could be a more effective strategy for boosting performance than merely increasing the number of antennas M .

2.7 Conclusion

This chapter addressed the JADCE problem in single-cell mMTC under spatially correlated fading MIMO channels. This channel model is crucial for realistically evaluating the solutions designed for grant-free access in mMTC. The goal of this chapter was to develop a sparse recovery problem that takes into account any side information available at the BS to offer acceptable performance while minimizing signaling overhead. To this end, we addressed JADCE in two distinct scenarios: one without prior information on second-order channel statistics and another that leverages this prior knowledge for enhanced performance. This chapter introduced novel JADCE formulations that effectively capitalize on prior knowledge about second-order channel

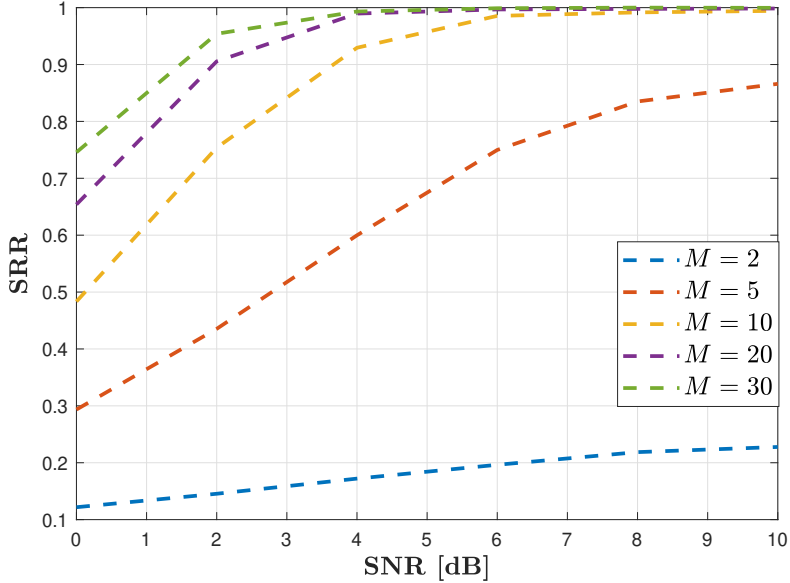


Fig. 11. Device activity detection performance vs. the number of BS antennas. (Under CC BY 4.0 license from Paper [101] ©2021 Authors)

statistics. Furthermore, two ADMM-based solutions have been derived based on deploying different optimization problem decomposition techniques.

The numerically obtained results highlight the following key findings:

- When channel statistics are unknown, our results underscore the substantial benefits of formulating JADCE as an iterative reweighted $\ell_{2,1}$ -norm minimization problem compared to classical deterministic sparse recovery methods like ℓ_1 -norm minimization and greedy algorithms.
- If prior knowledge of the second-order statistics of the channel is available, our proposed formulations show great potential in reducing signaling overhead. Specifically, the proposed cov-IRW-ADMM and cov-ADMM methods yield significant gains, particularly with short pilot sequences. Given that the length of the pilot sequence is directly proportional to the signaling overhead, reducing this length while maintaining good performance is a valuable gain.
- The performance of JADCE dramatically improves when transitioning from the conventional MIMO regime to the massive MIMO regime.

A limitation of the research presented in this chapter lies in the specific applicability of our proposed approach. Specifically, the solutions are tailored for spatially correlated channels that adhere to a one-ring model with a narrow ASD. This raises

an important question: Can we extend our methodology to accommodate a broader spectrum of spatially correlated channel models?

Recognizing this limitation, the next chapter aims to address this gap by presenting a more generalized approach. Specifically, we introduce a novel methodology rooted in Bayesian principles to tackle this challenge. By adopting a Bayesian perspective, we strive to create a more robust and adaptable framework capable of handling a variety of spatially correlated channel models, thereby broadening the scope and solutions provided in this work.

3 JADCE under spatially correlated channels: A Bayesian approach

In contrast with the previous chapter, which employed a deterministic approach for solving the JADCE problem, this chapter takes a fundamentally different approach by adopting a Bayesian perspective. Specifically, we focus on scenarios in which the CDI are known to the BS, thus allowing the treatment of JADCE from a Bayesian perspective. While deterministic approaches generally yield fixed-point solutions, offering consistency and ease of computation, they may not fully capitalize on the prior knowledge concerning the statistical properties inherent to the spatially correlated MIMO channels. The Bayesian framework, on the other hand, allows for a probabilistic formulation of the problem, incorporating prior knowledge and uncertainties directly into the model. This results in more robust and potentially more accurate solutions, especially when dealing with broader channel models.

Concurrent with this work, there has been a growing interest in recent years in tackling the JADCE problem under spatially correlated channels using Bayesian methods. For instance, Cheng *et al.* exploited in [123] both the spatial and temporal correlation of the propagation channels and proposed the orthogonal AMP algorithm to solve the JADCE problem. Rajoriya *et al.* [125] proposed a Bayesian solution that coupled AMP and SBL to provide an algorithm that enjoyed the low complexity of AMP and the good performance of SBL. Bai *et al.* [126] proposed a distributed AMP algorithm in cell-free MTC networks that aimed to reduce the complexity of AMP by distributing the computation load over several access points.

By deploying a Bayesian framework, we can take into account the prior information about the CDI and the devices sparse activity, thus creating an adaptive and resilient estimation process. This framework allows better exploitation of the spatial correlations in MIMO channels, thereby potentially yielding more accurate device detection accuracy and improved channel estimation. To this end, this chapter expands the scope of the JADCE problem by introducing a Bayesian approach tailored for spatially correlated MIMO channels, aiming to provide a more comprehensive and robust solution compared to deterministic methods.

3.1 System model

This chapter continues the discussion of the system model initially described in Section 2.1. We further extend our analysis by introducing several generalizations to the existing channel model. First, we align our study with the prevailing assumptions frequently encountered in the MIMO literature. Specifically, we adopt the *local scattering model* as the basis for describing the propagation channel between each device and the Base Station (BS). This choice is particularly well-suited for characterizing multi-antenna channels, as it encapsulates key features commonly observed in conventional MIMO systems. Within the local scattering model, it is assumed that each device is surrounded by a large number of scatterers, which significantly influence the propagation characteristics. Following this, we leverage the central limit theorem to derive the statistical properties of the channel vector \mathbf{h}_i , $i \in \mathcal{N}$, can be modelled as a *complex Gaussian random vector* with zero mean and covariance matrix $\mathbf{R}_i = \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^H] \in \mathbb{C}^{M \times M}$ [100, Sect. 2.6], i.e.,

$$\mathbf{h}_i \sim \mathcal{CN}(0, \mathbf{R}_i), \quad P_i \rightarrow \infty, \quad \forall i \in \mathcal{N}, \quad (51)$$

where P_i is the number of propagation paths.

We consider devices with low mobility, which is justified in the context of mMTC [114]. Hence, we adopt the common assumption that the channels are wide-sense stationary [113]. Thus, the channel covariance matrices $\{\mathbf{R}_i\}_{i=1}^N$ vary in a slower timescale compared to the channel realizations [99]. Accordingly, $\{\mathbf{R}_i\}_{i=1}^N$ are assumed to remain fixed for τ_s coherence intervals. Furthermore, we make the assumption that an estimated set of covariance matrices $\{\hat{\mathbf{R}}_i\}_i^N$ is known to the BS.

Recall the received signal model $\mathbf{Y} \in \mathbb{C}^{\tau_p \times M}$, is given by

$$\mathbf{Y} = \sum_{i=1}^N \gamma_i \sqrt{\rho_i^{\text{UL}}} \phi_i \mathbf{h}_i^T + \mathbf{W} = \Phi \mathbf{X}^T + \mathbf{W},$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{C}^{M \times N}$, such that $\mathbf{x}_i = \gamma_i \sqrt{\rho_i^{\text{UL}}} \mathbf{h}_i$, $\gamma_i \in \{0, 1\}$, $i \in \mathcal{N}$, and $\mathbf{W} \in \mathbb{C}^{\tau_p \times M}$ is an additive white Gaussian noise with i.i.d. elements, each following $\mathcal{CN}(0, \sigma^2)$.

3.2 Problem formulation

The JADCE formulation presented in Chapter 2 can be viewed as a joint support and signal recovery problem with a deterministic sparsity regularization. Its robustness stems from its invariance with channel statistics, making it widely applicable across

various channel distributions. However, this formulation does not take into account any available side information concerning the CDI. Alternatively, if the CDI is available, the JADCE problem can be formulated in a Bayesian framework to account for the fact that each unknown channel to be estimated is a realization of a random variable (vector) with the known distribution. Such a Bayesian sparse signal recovery framework has great potential to provide certain advantages over deterministic formulations, as discussed in [127].

A Bayesian approach to JADCE is enabled by leveraging two key factors: 1) the fact that the propagation channels \mathbf{h}_i , $i \in \mathcal{N}$, are modeled by Gaussian distributions as shown in (51), and 2) the relatively slowly changing covariance matrices $\{\mathbf{R}_i\}_{i=1}^N$ which can be estimated with high accuracy. In the rest of the paper, we consider the common assumption that $\{\mathbf{R}_i\}_{i=1}^N$ are known to the BS [113].

Next, we utilize the prior information on the CDI and derive a Bayesian formulation for the JADCE problem. JADCE performs two tasks in a joint fashion: 1) the identification of the support of the device activity indicator vector γ , and 2) the estimation of the effective channel matrix \mathbf{X} , relying on the current estimate of γ . In the Bayesian modeling, we treat the two variables to be estimated, γ and \mathbf{X} , as unknown quantities with such prior distributions that best model our knowledge of their true distributions, that is: 1) the sparse distribution of the device activity indicator vector γ , and 2) the effective channel \mathbf{x}_i , $\forall i \in \mathcal{N}$, which is a random vector consisting of a multiplication of γ_i and the complex Gaussian random vector \mathbf{h}_i (i.e., $\mathbf{x}_i = \sqrt{\rho_i^{\text{UL}}}\gamma_i\mathbf{h}_i$).

We derive joint MAP estimates $\{\hat{\mathbf{X}}, \hat{\gamma}\}$ by making an explicit use of the prior knowledge of the fact that the propagation channels between the devices and the BS follow complex Gaussian distributions given in (51), under the assumption that the BS knows the estimates of the second-order statistics of the channels, i.e., the matrices $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. To this end, the joint MAP estimates $\{\hat{\mathbf{X}}, \hat{\gamma}\}$ with respect to the posterior density given the measurement matrix \mathbf{Y} are given by

$$\begin{aligned}
\{\hat{\mathbf{X}}, \hat{\gamma}\} &= \max_{\mathbf{X}, \gamma} p(\mathbf{X}, \gamma | \mathbf{Y}) \\
&= \max_{\mathbf{X}, \gamma} \frac{p(\gamma)p(\mathbf{X}|\gamma)p(\mathbf{Y}|\mathbf{X}, \gamma)}{p(\mathbf{Y})} \\
&\stackrel{(a)}{=} \max_{\mathbf{X}, \gamma} p(\gamma)p(\mathbf{X}|\gamma)p(\mathbf{Y}|\mathbf{X}) \tag{52} \\
&= \min_{\mathbf{X}, \gamma} -\log p(\mathbf{Y}|\mathbf{X}) - \log p(\mathbf{X}|\gamma) - \log p(\gamma) \\
&\stackrel{(b)}{=} \min_{\mathbf{X}, \gamma} \frac{1}{\sigma^2} \|\mathbf{Y} - \Phi\mathbf{X}\|_{\text{F}}^2 - \log p(\mathbf{X}|\gamma) - \log p(\gamma),
\end{aligned}$$

where (a) follows from the Markov chain $\gamma \rightarrow \mathbf{X} \rightarrow \mathbf{Y}$ and because $p(\mathbf{Y})$ does not affect the maximization and (b) follows from the additive Gaussian noise model in (12). The

term $p(\mathbf{X}|\gamma)$ denotes the conditional probability density function (PDF) of the effective channel \mathbf{X} given the device activity state vector γ , whereas the term $p(\gamma)$ represents the prior belief concerning the distribution of the device activity.

Next, we elaborate in detail the choice of the prior $p(\gamma)$ and the definition of the conditional PDF $p(\mathbf{X}|\gamma)$. Then, having fixed these quantities, we derive an ADMM algorithm to find an approximate solution to the MAP estimation in (52).

Sparsity-inducing prior $p(\gamma)$

Through the model assumption of the sporadic device activity, the device activity indicator vector γ exhibits a sparse structure ($\gamma_i = 0, \forall i \notin \mathcal{A}$). Therefore, in the context of sparse recovery, we impose a *sparsity prior* $p(\gamma)$ on γ . For instance, given a continuous-magnitude random vector $\theta \in \mathbb{C}^N$, a sparsity-inducing prior can be given by $p(\theta) \propto \exp(-\sum_{i=1}^N |\theta_i|^p)$, where $p \in [0, 1]$ [50].

Note that setting $p = 1$ results in the ℓ_1 -norm penalty, corresponding to the *Laplace* density function. On the other hand, setting $p = 0$ renders the optimal sparsity-inducing penalty corresponding to the ℓ_0 -norm. Since γ is a vector of binary elements, setting $p = 0$ is equivalent to $p = 1$, as it imposes the same sparsity prior $p(\gamma)$. Subsequently, we select the prior $p(\gamma)$ as the ℓ_0 -norm penalty as

$$p(\gamma) \propto \exp\left(-\sum_{i=1}^N \mathbb{I}(\gamma_i)\right). \quad (53)$$

Conditional probability $p(\mathbf{X}|\gamma)$

Since the device activity is controlled by γ , the conditional probability $p(\mathbf{X}|\gamma)$ is defined as follows. First, we note that the activity patterns of the different devices are mutually independent. Hence, the conditional PDF factorizes as $p(\mathbf{X}|\gamma) = \prod_{i=1}^N p(\mathbf{x}_i|\gamma_i)$. In addition, for each device $i \in \mathcal{N}$, we distinguish the two possible cases for $p(\mathbf{x}_i|\gamma_i)$ as follows:

1. Conditioned on $\gamma_i = 1$, the i th device is active, and \mathbf{x}_i follows a Gaussian distribution, i.e., $p(\mathbf{x}_i|\gamma_i = 1) = p_{\mathbf{x}_i}$, where $p_{\mathbf{x}_i} \sim \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \tilde{\mathbf{R}}_i)$, and $\tilde{\mathbf{R}}_i$ denotes the scaled covariance matrix defined as $\tilde{\mathbf{R}}_i = \rho_i^{\text{UL}} \hat{\mathbf{R}}_i$.
2. Conditioned on $\gamma_i = 0$, the i th device is inactive, and \mathbf{x}_i is a deterministic all-zero vector $\mathbf{x}_i = \mathbf{0}$ with probability 1, i.e., $Pr(\mathbf{x}_i = \mathbf{0}|\gamma_i = 0) = 1$.

Therefore, $p(\mathbf{x}_i|\gamma_i)$ can be modeled as

$$p(\mathbf{x}_i|\gamma_i) = \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \tilde{\mathbf{R}}_i)^{\mathbb{I}(\gamma_i)}. \quad (54)$$

Subsequently, and based on the assumption that the device activity is independent across the devices, $p(\mathbf{X}|\gamma)$ is given by

$$p(\mathbf{X}|\gamma) = \prod_{i=1}^N p(\mathbf{x}_i|\gamma_i) = \prod_{i \in \mathcal{A}} p_{\mathbf{x}_i}. \quad (55)$$

By applying the log transformation to $p(\gamma)$ in (53) and to $p(\mathbf{X}|\gamma)$ in (55), and by dropping the constant terms that do not depend on γ and \mathbf{X} , the joint MAP estimation problem (52) can be equivalently written as

$$\{\hat{\mathbf{X}}, \hat{\gamma}\} = \min_{\mathbf{X}, \gamma} \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}^T\|_F^2 + \beta_1 \sum_{i=1}^N 1(\gamma_i) + \beta_2 \sum_{i=1}^N \mathbf{x}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{x}_i, \quad (56)$$

where regularization weights β_1 and β_2 balance the emphasis on the priors, both in relation to each other and to the measurement fidelity term. The third term in (56) applies a *quadratic Mahalanobis distance* measure⁶, $\mathbf{x}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{x}_i$, $i \in \mathcal{N}$, for active devices to incorporate the knowledge of the spatial correlation matrices of the devices into the optimization process.

3.3 MAP-ADMM solution

The non-convex problem (56) is a mixed-integer programming problem due to the involvement of the binary optimization variables γ , and is thus difficult to solve for large N . In this section, we develop a computationally efficient ADMM algorithm, which is numerically illustrated to achieve great performance in Section 3.4.

We start by noting that the recovery of effective channel $\hat{\mathbf{X}}$ implicitly renders the vector γ , i.e., finding the index set $\{i \mid \gamma_i \neq 0, i \in \mathcal{N}\}$ is equivalent to finding the index set $\{i \mid \|\mathbf{x}_i\|_2 > 0, i \in \mathcal{N}\}$. Therefore, we solve a relaxed version of the MAP estimation (56) by approximating the penalty term that depends on γ by penalty term that depends on $\|\mathbf{x}_i\|_2$, $\forall i \in \mathcal{N}$.

Note that the second term $\sum_{i=1}^N 1(\gamma_i)$ in (56) is equivalent to an $\|\mathbf{X}\|_{2,0}$ penalty in the sense that it enforces the row sparsity of the matrix \mathbf{X}^T . Therefore, $\sum_{i=1}^N 1(\gamma_i)$ can be relaxed by the log-sum penalty $\sum_{i=1}^N \log(\|\mathbf{x}_i\|_2 + \varepsilon_0)$.

Subsequently, we can eliminate γ and approximate (56) as

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{u}} \quad & \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}^T\|_F^2 + \beta_1 \sum_{i=1}^N \log(u_i + \varepsilon_0) + \beta_2 \sum_{i=1}^N \mathbf{x}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{x}_i \\ \text{s.t.} \quad & \|\mathbf{x}_i\|_2 \leq u_i, \forall i \in \mathcal{N}. \end{aligned} \quad (57)$$

⁶The Mahalanobis distance between a vector θ and the Gaussian distribution with mean μ and covariance matrix \mathbf{R} is defined as $\sqrt{(\theta - \mu)^H \mathbf{R}^{-1} (\theta - \mu)}$. It measures the distance between the vector θ and the mean of the distribution (μ), measured along the principal component axes determined by the covariance matrix \mathbf{R} .

Again, we utilize MM and linearize the concave penalty term by its first-order Taylor expansion at point $\mathbf{u}^{(l)}$. Thus, an approximate solution to (57) is found by iteratively solving the problem

$$\hat{\mathbf{X}}^{(l+1)} = \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}^T\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{i=1}^N \mathbf{x}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{x}_i, \quad (58)$$

where the weight vector $\mathbf{g}^{(l)} = [g_1^{(l)}, g_2^{(l)}, \dots, g_N^{(l)}]^T$ is given according to (20). The optimization problem (58) can be seen as an iterative reweighted $\ell_{2,1}$ -norm minimization augmented with an additional penalty function that incorporates the spatial correlation knowledge to the optimization process by applying a Mahalanobis distance penalty to the active devices.

The objective function in (58) is a sum of convex functions, hence, the optimization problem (58) is convex. Therefore, aiming to provide a computationally efficient solution, we develop an ADMM framework that solves (58) through a set of sequential update rules, each computed in closed-form. In particular, to decompose (58) into a set of separate functions, we introduce two splitting variables $\mathbf{Z}, \mathbf{V} \in \mathbb{C}^{M \times N}$ and rewrite the optimization problem as

$$\begin{aligned} (\hat{\mathbf{X}}^{(l+1)}, \hat{\mathbf{Z}}^{(l+1)}, \hat{\mathbf{V}}^{(l+1)}) = & \min_{\mathbf{X}, \mathbf{Z}, \mathbf{V}} \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{Z}^T\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 \\ & + \beta_2 \sum_{i=1}^N \mathbf{v}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{v}_i \\ \text{s.t.} \quad & \mathbf{x}_i = \mathbf{z}_i, \quad \mathbf{x}_i = \mathbf{v}_i, \quad \forall i \in \mathcal{N}. \end{aligned} \quad (59)$$

The optimization problem (59) is *block multi-convex*, i.e., the problem is convex in one set of variables, while all the other variables are fixed. Since ADMM implicitly exploits the block multi-convexity nature of (59), utilizing ADMM to solve (59) is a reasonable choice. Accordingly, the augmented Lagrangian associated with (59) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{V}, \Lambda_z, \Lambda_v) = & \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{Z}^T\|_{\text{F}}^2 + \beta_1 \sum_{i=1}^N g_i^{(l)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{i=1}^N \mathbf{v}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{v}_i \\ & + \frac{\rho}{2} \|\mathbf{X} - \mathbf{V} + \frac{1}{\rho} \Lambda_v\|_{\text{F}}^2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z} + \frac{1}{\rho} \Lambda_z\|_{\text{F}}^2 - \frac{\|\Lambda_z\|_{\text{F}}^2}{2\rho} - \frac{\|\Lambda_v\|_{\text{F}}^2}{2\rho}, \end{aligned} \quad (60)$$

where $\Lambda_z = [\lambda_{z1}, \dots, \lambda_{zN}] \in \mathbb{C}^{M \times N}$, and $\Lambda_v = [\lambda_{v1}, \dots, \lambda_{vN}] \in \mathbb{C}^{M \times N}$ are the matrices of the ADMM dual variables.

The ADMM solution to the optimization problem (58) at the (l) th MM iteration is achieved by sequentially minimizing $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{V}, \Lambda_z, \Lambda_v)$ over the primal variables

$(\mathbf{Z}, \mathbf{V}, \mathbf{X})$, followed by dual variable (Λ_z, Λ_v) updates as follows:

$$\mathbf{Z}^{(k+1)} := \min_{\mathbf{Z}} \mathcal{L}(\mathbf{X}^{(k)}, \mathbf{Z}, \mathbf{V}^{(k)}, \Lambda_z^{(k)}, \Lambda_v^{(k)}) \quad (61)$$

$$:= \min_{\mathbf{Z}} \frac{1}{2} \|\Phi \mathbf{Z}^T - \mathbf{Y}\|_{\mathbb{F}}^2 + \frac{\rho}{2} \|\mathbf{X}^{(k)} - \mathbf{Z} + \frac{1}{\rho} \Lambda_z^{(k)}\|_{\mathbb{F}}^2$$

$$\mathbf{V}^{(k+1)} := \min_{\mathbf{V}} \mathcal{L}(\mathbf{X}^{(k)}, \mathbf{Z}^{(k+1)}, \mathbf{V}, \Lambda_z^{(k)}, \Lambda_v^{(k)}) \quad (62)$$

$$:= \min_{\mathbf{V}} \beta_2 \sum_{i=1}^N \mathbf{v}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{v}_i + \frac{\rho}{2} \|\mathbf{X}^{(k)} - \mathbf{V} + \frac{\Lambda_v^{(k)}}{\rho}\|_{\mathbb{F}}^2$$

$$\mathbf{X}^{(k+1)} := \min_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{Z}^{(k+1)}, \mathbf{V}^{(k+1)}, \Lambda_z^{(k)}, \Lambda_v^{(k)}) \quad (63)$$

$$:= \min_{\mathbf{X}} \sum_{i=1}^N \beta_1 g_i^{(l)} \|\mathbf{x}_i\|_2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z}^{(k+1)} + \frac{1}{\rho} \Lambda_z^{(k)}\|_{\mathbb{F}}^2 \\ + \frac{\rho}{2} \|\mathbf{X} - \mathbf{V}^{(k+1)} + \frac{1}{\rho} \Lambda_v^{(k)}\|_{\mathbb{F}}^2$$

$$\Lambda_z^{(k+1)} := \Lambda_z^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)}), \quad (64)$$

$$\Lambda_v^{(k+1)} := \Lambda_v^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{V}^{(k+1)}). \quad (65)$$

We present the derivations of the ADMM steps (61), (62), and (63) in detail below.

Z-update ADMM updates the primal variable \mathbf{Z} by solving the convex optimization problem (61). Thus, $\mathbf{Z}^{(k+1)}$ is obtained by setting the gradient of the objective function in (61) with respect to \mathbf{Z} to zero, resulting in

$$\mathbf{Z}^{(k+1)} = (\rho \mathbf{X}^{(k)} + \Lambda_z^{(k)} + \mathbf{Y}^T \Phi^*) (\Phi^T \Phi^* + \rho \mathbf{I}_N)^{-1}. \quad (66)$$

Note that the matrix inversion $(\Phi^* \Phi^T + \rho \mathbf{I}_N)^{-1}$ and the product $\mathbf{Y}^T \Phi^*$ need to be computed only once, so they can be stored, reducing the overall algorithm complexity.

V-update We can easily show that the \mathbf{V} -update in (62) can be decoupled into N convex sub-problems, given by

$$\mathbf{v}_i^{(k+1)} = \min_{\mathbf{v}_i} \beta_2 \mathbf{v}_i^H \tilde{\mathbf{R}}_i^{-1} \mathbf{v}_i + \frac{\rho}{2} \|\mathbf{x}_i^{(k)} - \mathbf{v}_i + \frac{\lambda_{\mathbf{v}_i}^{(k)}}{\rho}\|_2^2, \quad \forall i \in \mathcal{N}. \quad (67)$$

The solution for (67) is obtained by setting the derivative of the objective function with respect to \mathbf{v}_i to zero, resulting in

$$\mathbf{v}_i^{(k+1)} = \frac{1}{\beta_2} \tilde{\mathbf{R}}_i \left(\frac{\rho}{\beta_2} \tilde{\mathbf{R}}_i + \mathbf{I}_M \right)^{-1} (\rho \mathbf{x}_i^{(k)} + \lambda_{\mathbf{v}_i}^{(k)}), \quad \forall i \in \mathcal{N}. \quad (68)$$

X-update We can rewrite (63) as

$$\begin{aligned}
\mathbf{X}^{(k+1)} &= \min_{\mathbf{X}} \sum_{i=1}^N \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \frac{\rho}{2} \text{Tr} \left(2\mathbf{X}\mathbf{X}^H + \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^H + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^H \right. \\
&\quad \left. - \mathbf{X}^H(\tilde{\mathbf{Z}} + \tilde{\mathbf{V}}) - (\tilde{\mathbf{Z}} + \tilde{\mathbf{V}})^H\mathbf{X} \right) \\
&= \min_{\mathbf{X}} \sum_{i=1}^N \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \rho \text{Tr} \left(\mathbf{X}\mathbf{X}^H + \frac{1}{2}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^H + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^H) \right. \\
&\quad \left. - \frac{1}{2}\mathbf{X}^H(\tilde{\mathbf{Z}} + \tilde{\mathbf{V}}) - \frac{1}{2}(\tilde{\mathbf{Z}} + \tilde{\mathbf{V}})^H\mathbf{X} \right). \tag{69}
\end{aligned}$$

By subtracting the constant term $\frac{\rho}{4}\|\tilde{\mathbf{V}} - \tilde{\mathbf{Z}}\|_F^2$ from (69) and denoting $\mathbf{C}^{(k)} = \frac{1}{2}(\tilde{\mathbf{V}} + \tilde{\mathbf{Z}})$, the **X**-update solves the following convex optimization problem:

$$\mathbf{X}^{(k+1)} := \min_{\mathbf{X}} \sum_{i=1}^N \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \rho \|\mathbf{X} - \mathbf{C}^{(k)}\|_F^2, \tag{70}$$

where $\alpha_i^{(k)} = \beta_1 g_i^{(l)}$ and $\{\mathbf{C}^{(k)} = \frac{1}{2}(\mathbf{Z}^{(k+1)} + \mathbf{V}^{(k+1)} - \frac{\Lambda_z^{(k)} + \Lambda_v^{(k)}}{\rho})\}$. The problem (70) decouples into N sub-problems, each admitting the closed-form solution

$$\mathbf{x}_i^{(k+1)} = \frac{\max\{0, \|\mathbf{c}_i^{(k)}\|_2 - \frac{\alpha_i^{(k)}}{2\rho}\}}{\|\mathbf{c}_i^{(k)}\|_2} \mathbf{c}_i^{(k)}, \quad \forall i \in \mathcal{N}. \tag{71}$$

Algorithm implementation The details of the proposed MAP-based JADCE, termed MAP-ADMM, are summarized in Algorithm 3. We note that the **Z**-update (61) and the **V**-update (62) are independent of each other. Hence, they can be performed fully in parallel. MAP-ADMM is run until $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_F^2 < \varepsilon_{\text{stp}}$, or until a maximum number of iterations $l_{\text{max}} k_{\text{max}}$ is reached.

Algorithm computational complexity In a typical mMTC scenario, where the number of connected devices is very large, the complexity of the recovery algorithms is an important issue to address. Indeed, for the implementation of the proposed algorithms, the computational complexity determines the hardware processing cost. Next, we analyze the complexity of the proposed JADCE algorithms in terms of the number of required complex multiplications per iteration using the big $\mathcal{O}(\cdot)$ notation. The complexity analysis is summarized in Table 1, which also shows the exact number of matrix multiplications.

At the **Z**-update step of IRW-ADMM and MAP-ADMM, for fixed ρ , the quantity $(\Phi^T \Phi^* + \rho \mathbf{I}_N)^{-1}$ is computed only once at an algorithm initialization. Similarly,

Algorithm 3: MAP-ADMM

Input: 1) Pilot matrix Φ , covariance matrices $\{\tilde{\mathbf{R}}_i\}_{i=1}^N$, $(\Phi^T \Phi^* + \rho \mathbf{I}_N)^{-1}$,
 $\frac{1}{\beta_2} \tilde{\mathbf{R}}_i (\frac{\rho}{\beta_2} \tilde{\mathbf{R}}_i + \mathbf{I}_M)^{-1}$, $\forall i \in \mathcal{N}$
2) parameters $\beta_1, \beta_2, \rho, \epsilon_0, \epsilon_{\text{stp}}, \kappa, l_{\text{max}}, k_{\text{max}}$,
Output: $\hat{\mathbf{X}}$
Initialization: $\mathbf{X}^{(0)}, \mathbf{V}^{(0)}, \mathbf{Z}^{(0)}, \Lambda_{\mathbf{v}}^{(0)}, \Lambda_{\mathbf{z}}^{(0)}, k = 1, l = 1$.

- 1 Receive \mathbf{Y} at the BS, and compute and store **while** $l < l_{\text{max}}$ **do**
- 2 **while** $k < k_{\text{max}}$ or $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\| < \epsilon$ **do**
- 3 Update $\mathbf{Z}^{(k+1)}$ using (66)
- 4 Update $\mathbf{V}^{(k+1)}$ using (68)
- 5 Update $\mathbf{X}^{(k+1)}$ using (71)
- 6 $\Lambda_{\mathbf{z}}^{(k+1)} = \Lambda_{\mathbf{z}}^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)})$
- 7 $\Lambda_{\mathbf{v}}^{(k+1)} = \Lambda_{\mathbf{v}}^{(k)} + \rho (\mathbf{X}^{(k+1)} - \mathbf{V}^{(k+1)})$
- 8 $k \leftarrow k + 1$
- 9 $\mathbf{X}^{(l)} \leftarrow \mathbf{X}^{(k+1)}$
- 10 Update $g_i^{(l)}$ using (20)
- 11 $l \leftarrow l + 1$

the term $\mathbf{Y}^T \Phi^*$ is computed only once upon receiving the pilot signal \mathbf{Y} . Therefore, computing $\mathbf{Z}^{(k+1)}$ requires $(M+1)N^2$ complex multiplications. For the \mathbf{V} -update of MAP-ADMM, the terms $\frac{1}{\beta_2} \tilde{\mathbf{R}}_i (\frac{\rho}{\beta_2} \tilde{\mathbf{R}}_i + \mathbf{I}_M)^{-1}$, $\forall i \in \mathcal{N}$, in (68) need to be computed only once and can subsequently be used for several coherence intervals. Hence, MAP-ADMM requires $N(M^2 + 2M)$ complex multiplications to compute $\mathbf{V}^{(k+1)}$. The soft-threshold operators in (71) for the \mathbf{X} -update require $2MN$ complex multiplications. Finally, the weight vector $\mathbf{g}^{(l)}$ is computed only at the outer MM iteration level (l), and it requires MN complex multiplications. Therefore, the overall complexity for IRW-ADMM and MAP-ADMM is $\mathcal{O}(MN^2)$ and $\mathcal{O}(MN^2 + NM^2)$ respectively.

Table 1 also compares the complexity MAP-ADMM to the four baseline algorithms that we consider in the numerical experiments: the fast alternating direction method (F-ADM) [120], IRW-ADMM presented in Chapter 2, SOMP [31], and T-SBL [50]. F-ADM has computational complexity of $\mathcal{O}(M\tau_p N)$. The greedy SOMP is reported

in [87] to exhibit computational complexity of $\mathcal{O}(M\tau_p N)$. T-SBL has computational complexity⁷ of $\mathcal{O}(N^2 M^3 \tau_p)$.

In summary, incorporating the channel spatial correlation information results in *increased* computational complexity. For instance, MAP-ADMM has higher computational complexity per iteration compared to IRW-ADMM due to the incorporation of the spatial structure information in the **V**-update step. In addition, as the proposed IRW-ADMM and MAP-ADMM aim to providing an exact solution to the JADCE problem, they are more computationally complex than F-ADM. Nevertheless, the additional cost of the proposed algorithms is compensated for by the convergence to a more accurate solution, as we will show in the next section.

Table 1. Computational complexity for different recovery algorithms, where (k) is the iteration index, and \hat{K} is the estimated number of nonzero element at any particular iteration (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

Algorithm	complex multiplications per iteration	$\mathcal{O}(\cdot)$
IRW-ADMM	$(M+1)N^2 + M(N + 2\hat{K} + N)$	$\mathcal{O}(MN^2)$
MAP-ADMM	$(M+1)N^2 + M(N + 2\hat{K} + N) + NM^2$	$\mathcal{O}(MN^2 + NM^2)$
F-ADM [120]	$4\tau_p MN + 5MN$	$\mathcal{O}(\tau_p MN)$
SOMP [31]	$(2\tau_p + 1)MN + \tau_p M^2 + (M+1)\tau_p(k)^2 + (k)^3$	$\mathcal{O}(\tau_p MN)$
T-SBL [50]	$2M^3 N^2 \tau_p + M^2 \tau_p^2 + NM^2 \tau_p$	$\mathcal{O}(N^2 M^3 \tau_p)$

3.4 Simulation results

In this section, we evaluate the performance of the proposed algorithm via numerical simulations in terms of device activity detection accuracy, channel estimation quality, and convergence rate, and compare them to existing MMV reconstruction algorithms.

Simulation setup We consider a single-cell of a radius of 50 m, where the BS is surrounded by $N = 200$ uniformly distributed devices, of which $K = 10$ devices are active at each coherence interval T_c . The propagation channel between the i th device and the BS in (51) consists of $P_i = 200$ paths with an angular spread deviation of $\sigma_\psi = 10^\circ$. Each device $i = 1, \dots, N$ is assigned with a unique normalized quadratic phase shift keying (QPSK) sequence ϕ_i , where the QPSK pilot symbols are drawn from an i.i.d. complex Bernoulli distribution.

⁷The authors in [50] also devised a low-complexity version of T-SBL relying on approximate updates which was shown to work well in the high SNR regime. However, since we are interested in a broader SNR range, this implementation is not readily applicable to our JADCE problem.

Baselines We compare the performance of the proposed algorithms against the following algorithms that solve any MMV sparse recovery problem: 1) IRW-ADMM; 2) T-SBL[50] where both the second-order statistics and the noise variance are known at the BS, and the sparse recovery is performed using the update rules given by [50, Eqs. (6), (7), (12)] (i.e., “**B**-update” in [50, Eq. (13)] is not performed, because we provide the covariance matrices $\{\hat{\mathbf{R}}_i\}_{i=1}^N$). In addition, we use the oracle joint minimum mean square error (MMSE) estimator, shown in Appendix 1, where the MMSE estimator is provided “oracle” knowledge of the true set of active devices. The oracle joint MMSE estimator provides a lower bound on channel estimation performance when both the CDI and the noise variances are available at the BS.

Parameter tuning The sparse recovery algorithms require their regularization parameters to be fine-tuned to yield their best estimates. While the regularization parameters depend on the different system parameters such as N , M , K , τ_p , and σ^2 , in practice, they are selected empirically by cross-validation. For a fair comparison, all the parameters have been empirically tuned in advance and then fixed such that they provide the best overall performance in terms of NASE for the SNR range $[0 - 16]$ dB. For instance, β_1 depends highly on the ratio $\frac{K}{N}$. However, since the K is not known to the BS in general, we tuned β_1 based on the noise variance σ^2 as $\beta_1 = \sqrt{\frac{\sigma^2}{2}}$, since it provided the most robust convergence. Furthermore, we set β_2 and log-sum stability parameter ε_0 to $\beta_2 = 1\%$ and $\varepsilon_0 = 0.07 - 0.12\%$ of the average norm of the effective channels. Moreover, since ADMM typically converges in few tens of iterations, a maximum number of iterations of $l_{\max} = 12$, $k_{\max} = 5$, and stopping criterion $\varepsilon = 10^{-3}$ were found to be sufficient for the ADMM-based algorithms to converge to their best performance. All the optimization variables for MAP-ADMM (\mathbf{X} , \mathbf{V} , \mathbf{Z} , Λ_z , and Λ_v) and for IRW-ADMM (\mathbf{X} , \mathbf{Z} , and Λ) are initialized as zero matrices. The results are obtained by averaging over 10^3 random channel realizations.

3.4.1 *The impact of exploiting channel statistics*

Since we have shown the superiority of IRW-ADMM over conventional sparse recovery algorithms where no knowledge of the CDI is used, we next investigate the effect of incorporating the CDI on the JADCE performance.

First, we quantify the activity detection accuracy performance of the proposed MAP-ADMM and compare it to IRW-ADMM and T-SBL. Fig. 12 presents the SRR against SNR for the proposed algorithms for the different pilot length values. The results clearly show that MAP-ADMM provides superior performance compared to

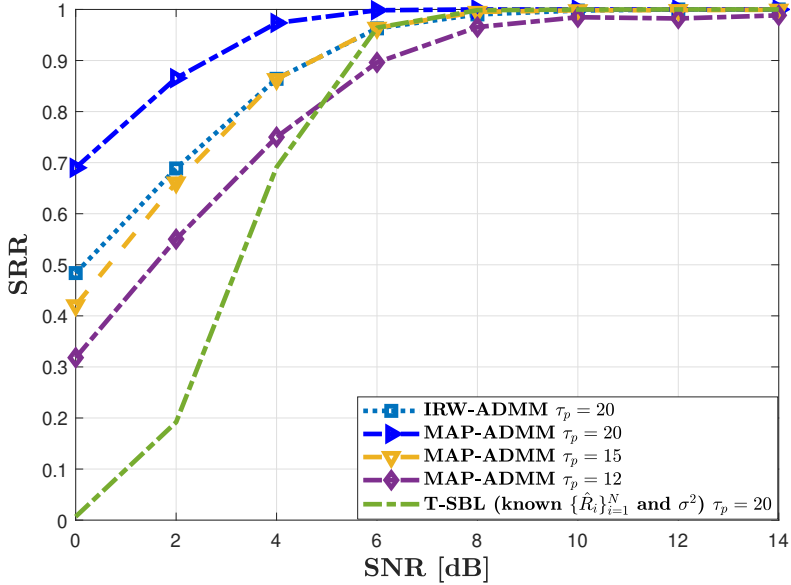


Fig. 12. Activity detection performance of the proposed MAP-ADMM in terms of SRR and NASE against SNR for $N = 200$, $M = 20$, and $K = 10$. (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

IRW-ADMM. For instance, MAP-ADMM identifies the set of true active devices \mathcal{A} perfectly for $\text{SNR} > 8$ dB, using a pilot length $\tau_p = 20$. Furthermore, reducing the pilot length by a factor of 25% (i.e., $\tau_p = 15$) affects the performance of MAP-ADMM only moderately, and optimal performance is achieved for $\text{SNR} > 10$ dB. More interestingly, the results indicate that even with 40% reduction in the pilot sequence length (i.e., $\tau_p = 12$), MAP-ADMM provides 95% SRR rate for $\text{SNR} > 10$ dB. Finally, the results show that while T-SBL suffers from an inferior performance in the low SNR regime, it provides optimal activity detection performance when $\text{SNR} > 8$ dB.

Fig. 13 illustrates the channel estimation performance in terms of NASE for MAP-ADMM against SNR for different pilot lengths and compares it to IRW-ADM, T-SBL, and the oracle MMSE benchmark. The proposed MAP-ADMM indisputably provides superior performance and significant improvement over IRW-ADMM. For example, given the same pilot sequence length of $\tau_p = 20$, MAP-ADMM achieves the same performance as IRW-ADMM while using up to 6 dB lower SNR. Furthermore, Fig. 13 reveals one advantageous feature of utilizing available CDI: even with 25% reduction in the pilot length, i.e., $\tau_p = 15$, MAP-ADMM still provides a 2 dB gain compared to IRW-ADMM. Comparing the performances between the MAP-ADMM and T-SBL algorithms, we distinguish two cases: 1) For $\text{SNR} < 8$ dB, T-SBL does not provide a

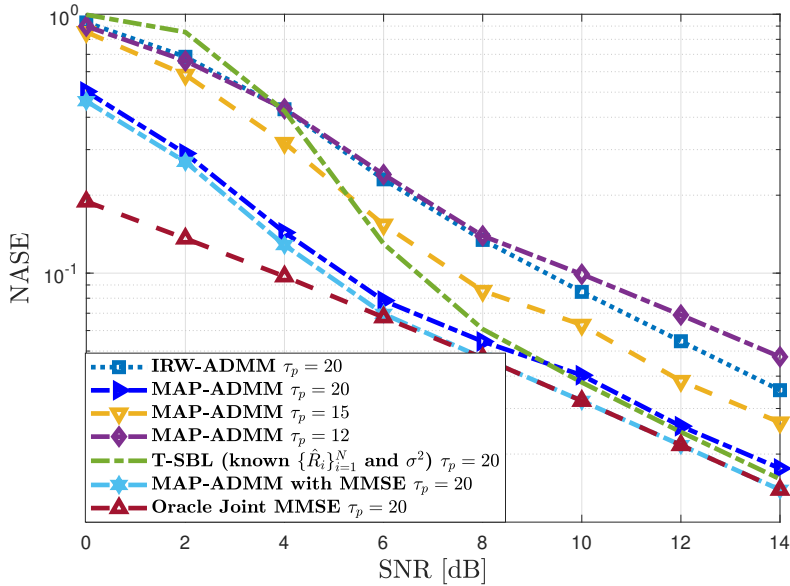


Fig. 13. Channel estimation performance of the proposed MAP-ADMM in terms of SRR and NASE against SNR for $N = 200$, $M = 20$, and $K = 10$. (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

reliable performance and MAP-ADMM outperforms T-SBL by a large margin, or so; 2) In the high SNR regimes, i.e., $\text{SNR} \geq 8$ dB, T-SBL slightly outperforms MAP-ADMM. These results can be explained by the fact that, in contrast to MAP-ADMM, T-SBL knows and uses the exact noise variance σ^2 . However, when the BS has the exact knowledge of the noise variance σ^2 , as well as the CDI, the slight gap in the NASE performance between T-SBL and MAP-ADMM can be compensated for by utilizing a joint MMSE estimator on the received signal associated with the estimated active device set $\hat{\mathcal{A}}$ obtained by MAP-ADMM. Fig. 4(b) shows that using the joint MMSE estimator on the estimated active devices provides even the same performance as the oracle joint MMSE estimator starting from $\text{SNR} > 8$ dB, which consolidate the results from with Fig. 12 where perfect recovery is attained at $\text{SNR} > 8$ dB. The results shown in Fig. 12 and Fig. 13 clearly highlight the advantages of exploiting the prior information about the channel to improve the JADCE performance in terms of activity detection accuracy and channel estimation quality.

Fig. 14 presents a typical convergence behavior of NASE versus the number of ADMM iterations for MAP-ADMM using different pilot sequence lengths at $\text{SNR} = 16$ dB. The results reveal that MAP-ADMM, using $\tau_p = 20$ requires about 40 iterations to converge, which is similar to IRW-ADMM performance. On the other

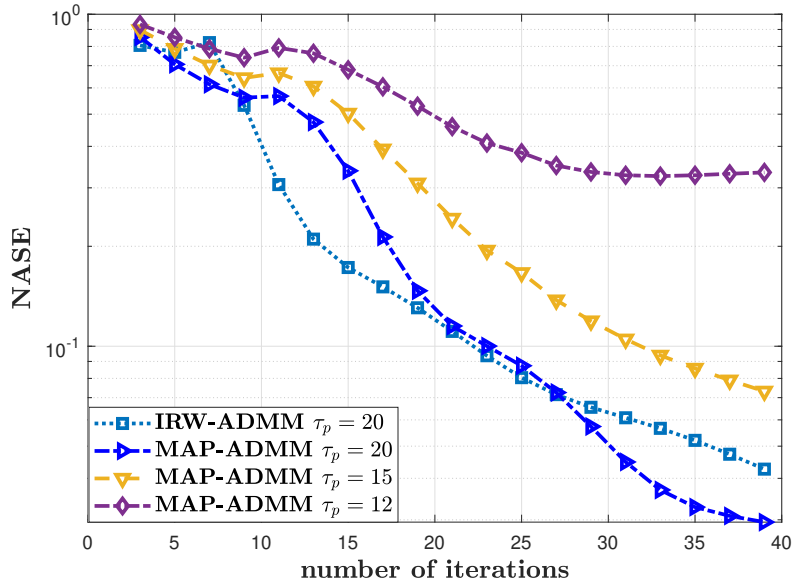


Fig. 14. cov-ADMM performance in terms of NASE versus number of ADMM iterations (k). (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

hand, the results show that reducing the pilot length also affects the convergence rate of MAP-ADMM, as it takes more iterations to converge.

Fig. 15 plots the SRR performance versus the average number of ADMM iterations using different pilot lengths. MAP-ADMM using $\tau_p = 20$ achieves the perfect activity detection in 20 iterations, whereas it takes up to 40 iterations to achieve the same performance for $\tau_p = 15$. This result is interesting because MAP-ADMM needs not to run until convergence in the NASE domain, where it takes up to 40 iterations. Rather, MAP-ADMM can be run for a few iterations until it detects perfectly the set of active devices (20 iterations on average), as shown in Fig. 15. Afterward, the joint MMSE estimator (136) can be applied to the estimated set of active devices to provide the optimal channel estimation quality for the effective channel matrix, as shown in Fig. 13.

3.4.2 Effect of the number of BS antennas

Next, we focus on quantifying the effect of the number of the BS antennas on the JADCE performance. We fix both the pilot sequence length $\tau_p = 20$, and SNR = 16 dB.

Fig. 16 illustrates the SRR of MAP-ADMM versus the number of BS antennas M . It is clear that increasing the number of BS antennas significantly improves the active

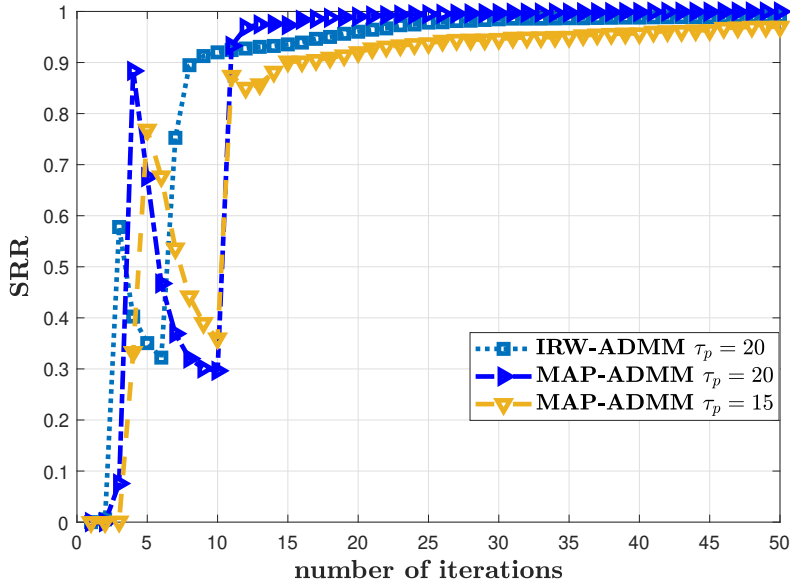


Fig. 15. The performance of the cov-ADMM in terms of SRR and versus the number of ADMM iterations. (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

device detection accuracy. Moreover, for the low SNR regime, i.e., $\text{SNR} < 8$ dB, the results show the significance of increasing the number of antennas to be greater than the number of active devices, i.e., $M > K$. However, the SRR performance starts to saturate gradually when the number of BS antennas M is increased. In fact, increasing the number of BS antennas from $M = 8$ to $M = 16$ provides more gains than increasing from $M = 24$ to $M = 32$; *this means that the gain in SRR gradually decreases as M increases.*

Fig. 17 depicts the channel estimation performance as a function of the number of BS antennas M at $\text{SNR} = 12$ dB. First, as expected, increasing M improves NASE for all the algorithms. Moreover, by increasing M , the channel estimation quality obtained by MAP-ADMM significantly improves and it approaches the lower bound offered by the oracle joint MMSE. More interestingly, in contrast with activity detection accuracy where the performance saturates when $M > 2K$, channel estimation quality improves considerably with the increase of M . Fig. 17 illustrates the effects of operating in a massive MIMO regime, i.e., $M > K$: while device activity detection accuracy saturates around $M > 2K$ as depicted in Fig. 16, channel estimation quality consistently improves when moving to the large numbers of BS antennas M .

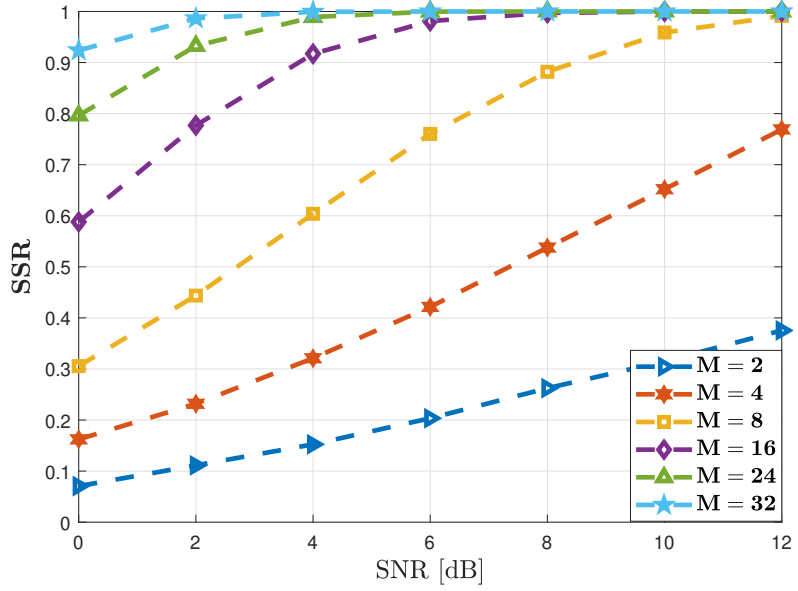


Fig. 16. cov-ADMM performance in terms of SRR versus the number of BS antennas M . (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

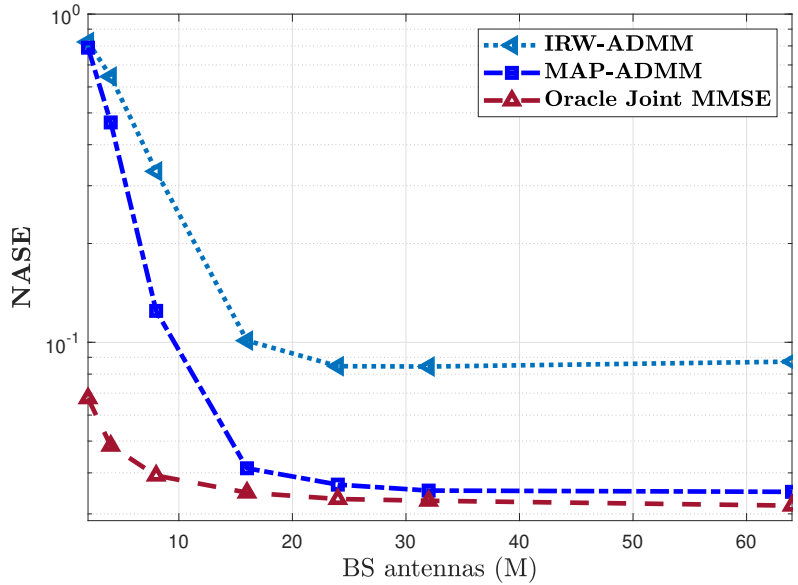


Fig. 17. cov-ADMM performance in terms of NASE versus the number of BS antennas M . (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

3.4.3 Impact of imperfect knowledge of the channel covariance matrix

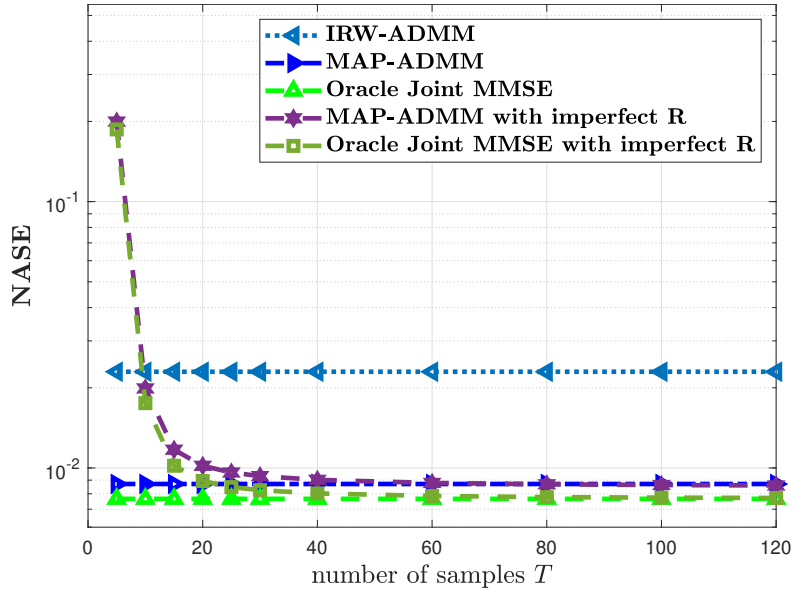
This section explores the impact of the training phase on the estimation of the channel's second-order statistics, denoted as $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. Specifically, we examine the impact of varying the number of training samples T on the NASE performance of the MAP-ADMM and the oracle joint MMSE estimator. It is important to note that once the set of covariance matrices is generated with a given number of samples T , it is directly inputted into MAP-ADMM. As such, the BS does not need to update these matrices during each MAP-ADMM iteration.

Fig. 18 depicts the NASE versus the number of samples T used to generate $\{\hat{\mathbf{R}}_i\}_{i=1}^N$ for $M = 20$ and $M = 40$ at SNR = 16 dB. The regularization parameters for MAP-ADMM and IRW-ADMM are fixed to the ones providing the best results when perfect knowledge of $\{\hat{\mathbf{R}}_i\}_{i=1}^N$ is available. First, Fig. 18 indicates that using a low number of training samples T is detrimental to the performance of MAP-ADMM and the joint MMSE estimator, as they require at least $T > \frac{M}{2}$ training samples to achieve the same performance as IRW-ADMM. Second, as expected, increasing the number of samples T improves the channel estimation quality for both MAP-ADMM and the joint MMSE estimator as their NASE asymptotically approaches the lower bounds achieved by their counterparts that rely on perfect knowledge of $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. More interestingly, the results show that MAP-ADMM and joint MMSE require around $T = 2M$ samples to achieve the same NASE performance to their optimal lower bound. This result indicates that MAP-ADMM is not highly sensitive to imperfect channel statistics. Finally, we note that a similar conclusion on the required number of samples T to achieve near-optimal performance for the MMSE estimator is reported in [100, Sect. 3.3.3].

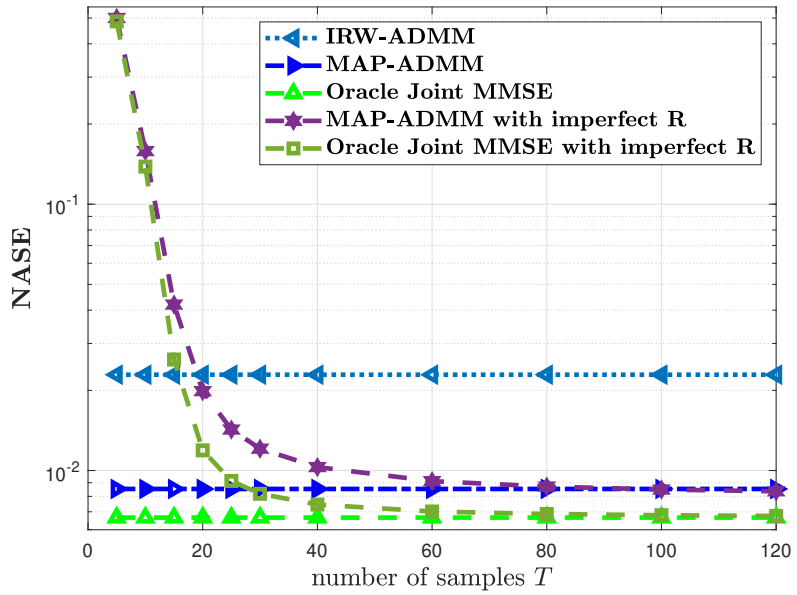
3.5 Conclusion

The chapter addressed the JADCE problem in grant-free access in mMTC under spatially correlated fading channels. We approached JADCE from a Bayesian perspective and proposed a novel JADCE formulation based on MAP estimation that exploit available CDI at the BS. Furthermore, we derived ADMM-based algorithms that featured computationally efficient closed-form solutions that could be computed via simple analytical formulas.

The obtained numerical results highlight the following key findings: 1) While incorporating the spatial correlation of the channels increases the computational complexity of the recovery algorithms, it results in significant gains, even with a smaller signalling overhead. 2) The performance of JADCE improve dramatically when



(a) $M = 20$.



(b) $M = 40$.

Fig. 18. Channel estimation performance versus the number of samples T , $N = 200$, $K = 10$, and $\tau_p = 20$, SNR = 16 dB. (Under CC BY 4.0 license from Paper [103] ©2021 Authors)

moving from the conventional MIMO regime to the massive MIMO regime. 3) The training phase for estimating the second-order statistics of the channel does not require a substantial amount of resources. Furthermore, MAP-ADMM is robust against imperfect channel statistics knowledge, which is conducive for practical use cases.

MAP-ADMM relies on the knowledge of the CDI at the BS, which may be challenging to acquire in practice. To this end, in the next chapter, we derive two Bayesian inference frameworks that aim to design a sparse recovery algorithm that estimates the second-order statistics of the channels within the recovery process.

4 Hierarchical mMTC activity detection and channel estimation with unknown spatial covariance

This chapter addresses the JADCE problem in mMTC under the practical spatially correlated channels model with *unknown covariance matrices*. Furthermore, we consider an mMTC network with *correlated activity* patterns following an event-triggered traffic mode. While the prior works reported in [123, 101, 102, 103, 125, 126, 108, 128] addressed the JADCE problem under the same spatially correlated MIMO channels, they make the assumption that the CDI for all the devices are *fully known to the BS at any transmission instance*. However, this assumption can be challenging to fulfill in realistic scenarios, as the BS may be unable to track the CDI of the devices with long inactive status without spending additional communication resources .

Furthermore, existing research on the JADCE problem has predominantly focused on the generic random device activity pattern. These models typically assume that the devices are independently active, capturing scenarios where each device is monitoring an independent process and activates in a random manner. For example, one might envision an mMTC network where various devices are employed to monitor disparate environmental variables such as temperature, humidity, and air quality, where each device acts independently based on its specific monitoring requirements.

mMTC networks with independent activity are well-suited for capturing the randomness and diversity inherent in large-scale mMTC networks, and much of the existing literature has developed methods and algorithms that operate effectively under such assumptions. However, this model fails to capture the complexities of real-world deployment scenarios, where devices are often organized into clusters. In these cluster-based setups, all devices within a particular cluster are tasked with monitoring the same phenomenon, thus breaking the assumption of independent activity. This leads to what is known as “event-triggered traffic” models. In these models, the activities of devices within a cluster are highly correlated, often triggering the transmission of simultaneous updates to the BS.

Consider an industrial IoT network deployed in a manufacturing plant. In such a setting, multiple sensors may be distributed across the plant’s different zones, each zone acting as a cluster. If one sensor in a zone detects an anomaly such as a sudden temperature spike or equipment failure, it is highly likely that other sensors in the same zone will detect the same event. As a result, the BS would receive simultaneous updates

from multiple sensors within that cluster, requiring more sophisticated solutions that effectively incorporate this prior information to improve the detection of the active sensors, as well as estimating their channel coefficient for possible data detection.

Cluster-based, event-triggered traffic patterns introduce new challenges in the design and optimization of JADCE algorithms. They necessitate the development of new frameworks that can adapt to the correlation in device activities, offering both efficient and accurate detection and estimation performance under these more realistic conditions.

The event-triggered traffic model results in a *hierarchical* sparse activation pattern constituting in both *cluster-level* sparsity and *intra-cluster* sparsity. The cluster-level sparsity arises because the event epicenters are concentrated around a small subset of clusters (referred to as active clusters), causing only the devices from those active clusters to be prompt for activity. On the other hand, each event would trigger only a subset of devices to be active in practice, resulting in an intra-cluster sparsity model⁸. To the best of our knowledge, only three works have investigated the JADCE problem with a correlated devices activity pattern. Liu *et al.* [129] addressed only the activity detection problem and proposed a solution based on preamble selection under different assumptions of the availability of the prior knowledge of the activity correlation distribution. Becirovic *et al.* [130] proposed two sparsity-promoting priors based on the $\ell_{2,1}$ -norm and total variation and proposed a non-negative least squares algorithm to solve a relaxed version of an ℓ_0 -norm minimization. Furthermore, the authors in [131] modeled the correlated activity via a Beta distribution and derived a GAMP-based solution.

4.1 JADCE with correlated activity and partially known CDI

4.1.1 System model

We consider a single-cell uplink network consisting of a set $\mathcal{N} = \{1, 2, \dots, N\}$ devices served by a single BS equipped with a uniform linear array containing M antennas, as depicted in Fig. 19. The devices are geographically distributed over N_c clusters, where each device belongs to a unique cluster. We denote the index set of the l th cluster as $\mathcal{C}_l \subseteq \mathcal{N}$, $l = 1, \dots, N_c$, where each cluster contains L devices⁹, such that $N = LN_c$.

with the majority of the literature on grant-free access, which considers an independent devices activation pattern, in this paper, we consider an mMTC network operating

⁸This can occur if the devices in the same cluster do not perform the same task, or when some of the devices are not triggered by the event when they are far from the event epicentre.

⁹For the simplicity of presentation, yet without loss of generality, we assume that all clusters contain the same number of devices.

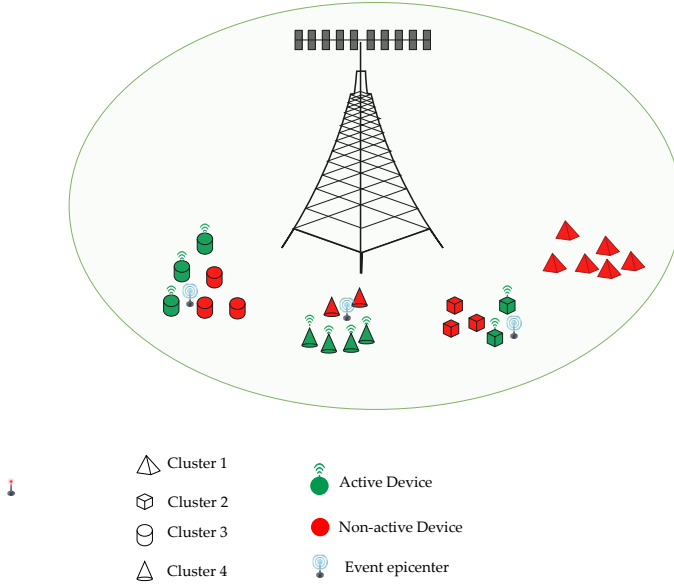


Fig. 19. Illustration of an mMTC network with an activity pattern centered around few number of clusters.

in an event-triggered traffic model inducing a *correlated activation pattern* between the devices. In particular, the event-triggered traffic model arises in the practical mMTC network, where the devices are grouped into clusters, such that each cluster is associated with a monitoring task. Subsequently, we make herein the following *distinctions* based on technical observations on the devices activation pattern under the event-triggered traffic model:

- The devices activity is triggered by events concentrated around a small subset of *active* clusters, thus giving rise to a *cluster-level* sparsity structure. We therefore define the cluster-level vector $\mathbf{c} = [c_1, \dots, c_L]^T$, where $c_l = 1$ if the l th cluster is active, and $c_l = 0$, otherwise.
- Within each active cluster, a subset containing at most $L_c \leq L$ devices will be active, thus inducing a correlation between the devices activity within the same cluster, in the form of *intra-cluster* sparsity. Thus, to model the intra-cluster sparsity, we introduce $\gamma_i \in \{0, 1\}$, $i \in \mathcal{N}$, where $\gamma_i = 1$ if the i th device is active and $\gamma_i = 0$, otherwise. Note that intra-cluster sparsity imposes a special structure on $\gamma = \underbrace{[\gamma_1, \dots, \gamma_L]}_{\text{Cluster 1}}, \underbrace{[\gamma_{L+1}, \dots, \gamma_{2L}]}_{\text{Cluster 2}}, \dots, \underbrace{[\gamma_{N-L+1}, \dots, \gamma_N]}_{\text{Cluster } c}$, where the elements belonging to a given cluster ($\gamma_i, i \in \mathcal{C}_l$) are

assumed to be correlated. The correlation in the devices activity will be discussed in the following sections.

We consider that the channel response \mathbf{h}_i between the i th device and the BS follows a local scattering model [100]. Thus, $\mathbf{h}_i \in \mathbb{C}^M$, $\forall i \in \mathcal{N}$, is modeled as a zero-mean complex Gaussian random variable, i.e., $\mathbf{h}_i \sim \mathcal{CN}(0, \mathbf{R}_i)$, such that $\mathbf{R}_i \in \mathbb{C}^{M \times M}$ represents the channel covariance matrix computed as $\mathbf{R}_i = \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^H]$. Additionally, we adopt the common assumption that the channels are wide-sense stationary. Thus, the changes in the covariance matrices $\mathbf{R} = \{\mathbf{R}_i\}_{i=1}^N$ occur less frequently compared to the variations in the channel realizations [99].

At any coherence interval T_c , each active device transmits a total of τ_c symbols to the BS over two phases. In the first phase, each active device transmits a τ_p -length pilot sequence to the BS, whereas in the second phase, each active device transmits its information data to the BS using the remaining $\tau_c - \tau_p$ symbols. Subsequently, to decode the information data transmitted from the active devices, the BS utilizes the received signal from the pilot transmission phase to perform the JADCE task. To this end, the pilot transmission phase is realized by assigning to each device i , $\forall i \in \mathcal{N}$, a unique unit-norm pilot sequence $\phi_i \in \mathbb{C}^{\tau_p}$, and a transmit power ρ_i^{UL} inversely proportional to its average channel gain to reduce the disparity in the channel's gain amongst the devices [132, 111].

Recall that the received signal during the pilot transmission phase $\mathbf{Y} \in \mathbb{C}^{\tau_p \times M}$ is given by

$$\mathbf{Y} = \sum_{i=1}^N \gamma_i \sqrt{\rho_i^{\text{UL}}} \phi_i \mathbf{h}_i^T + \mathbf{W} = \Phi \mathbf{X}^T + \mathbf{W},$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ represents the effective channel matrix with $\mathbf{x}_i = \gamma_i \sqrt{\rho_i^{\text{UL}}} \mathbf{h}_i$, $\Phi = [\phi_1, \dots, \phi_N] \in \mathbb{C}^{\tau_p \times N}$ is the pilot sequence matrix, and $\mathbf{W} \in \mathbb{C}^{\tau_p \times M}$ is an additive white Gaussian noise with i.i.d. elements, where each element is drawn from $\mathcal{CN}(0, \sigma^2)$.

The joint detection of the active devices and estimating their channel reduces to estimating the (unknown) *row-sparse* effective channel matrix \mathbf{X}^T based on the received pilot signal in (1). JADCE can be formulated as a sparse recovery problem from an MMV setup. Prior works in the literature showed that sparse recovery algorithms derived from a Bayesian perspective usually yield the best performance [29].

4.1.2 Bayesian inference setup

In the Bayesian framework, the unknown unknown variable to be estimated, i.e., \mathbf{X} , is modeled using a PDF which incorporates and encodes both the prior knowledge of \mathbf{X} and the prior knowledge of its hidden hyper-parameters. Subsequently, utilizing prior functions that promote sparsity while incorporating the hierarchical sparse structures of \mathbf{X}^T is the key to achieving accurate solutions for the JADCE problem. In particular, sparsity-promoting priors can be categorized as either weak or strong priors. For instance, the Laplace distribution is a weak sparsity prior, as it promotes sparsity, but it does not strictly assign a zero probability to the case where the vector $\mathbf{x}_i = 0$. On the other hand, strong sparsity priors such as the spike-and-slab distributions (with the delta peak at zero) are more stringent, as they assign a strictly positive probability to the case of $\mathbf{x}_i = 0$ [133]. Therefore, for the JADCE problem in this paper, the priors should be designed to encode (i) the spatial correlation of each \mathbf{x}_i , $\forall i$, (ii) the cluster-level sparsity between clusters, and (iii) the intra-cluster sparsity structure. We will next discuss, how to design the priors to encode (i), (ii), and (iii).

4.1.3 Hierarchical spike-and-slab prior

In the following, we leverage the concept of strong sparsity priors, and we propose a *hierarchical spike-and-slab sparsity-promoting prior* to model the cluster-level sparsity, the intra-cluster sparsity, and the spatially correlated channels. First, we introduce the following parameters:

Cluster-level activity prior: To impose the cluster-level activity, we model each c_l as a Bernoulli random variable with $\Pr(c_l = 1) = p_c$ and $\Pr(c_l = 0) = 1 - p_c$. Furthermore, to account for the independence amongst the clusters activity, we express the probability mass function of \mathbf{c} as

$$p(\mathbf{c}) = \prod_{l=1}^{N_c} \mathcal{B}(c_l; p_c), \quad (72)$$

where \mathcal{B} denotes the Bernoulli distribution.

Intra-cluster sparsity prior: We define the hyper-parameters $\tilde{\gamma}_i \in \mathbb{R}^+$, $\forall i$, that impose a row-sparsity structure on \mathbf{X} . To this end, we design $p(\tilde{\gamma})$ such that it promotes sparse solution. For example, $p(\tilde{\gamma})$ can be drawn from the Laplacian distribution as

$$p(\tilde{\gamma}) \propto \exp\left(-\sum_{i=1}^N \tilde{\gamma}_i\right). \quad (73)$$

Channels spatial correlation prior: We introduce the set of covariance matrices, namely, $\{\bar{\mathbf{R}}_i\}_{i=1}^N$, such that each $\bar{\mathbf{R}}_i$ is a positive definite matrix that captures the spatial correlation between the entries of i th row in \mathbf{X}^T . A common and physically grounded prior for an $\bar{\mathbf{R}}_i$ of the Gaussian random variable \mathbf{x}_i is given by the inverse Wishart distribution [105], defined as

$$p(\bar{\mathbf{R}}_i) \sim \mathcal{IW}(\bar{\mathbf{R}}_i; \mathbf{B}_i, \nu_w) = f_w(\mathbf{B}_i, \nu_w) \det(\bar{\mathbf{R}}_i)^{-d} \exp[-\text{Tr}(\mathbf{B}_i \bar{\mathbf{R}}_i^{-1})], \quad (74)$$

where $f_w(\mathbf{B}_i, \nu_w)$ is a normalization constant, $d = \nu_w - M + 1 > 0$, ν_w controls the degrees of freedom of the distribution, such that $\nu_w - M + 1 > 0$, and $\mathbf{B}_i \in \mathbb{C}^{M \times M}$ is a symmetric positive-definite matrix that represents the prior information for the covariance matrix $\bar{\mathbf{R}}_i$ [105].

By utilizing the definitions in (72), (73), and (74), we model the effective channel \mathbf{x}_i , $\forall i \in \mathcal{N}$, using the spike-and-slab prior as

$$p(\mathbf{x}_i | c_i, \tilde{\gamma}_i, \bar{\mathbf{R}}_i) = (1 - c_i) \delta(\mathbf{x}_i) + c_i \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \tilde{\gamma}_i \bar{\mathbf{R}}_i). \quad (75)$$

The main idea in (75) can be summarized as follows:

- If $c_i = 0$, the vector \mathbf{x}_i would have only the spike component, i. e., delta function, from (75), thus leading to an estimation of a zero-vector ($\mathbf{x}_i = \mathbf{0}$).
- If $c_i = 1$, \mathbf{x}_i would only have the slab component from (75) in the form be a Gaussian random vector with covariance matrix $\tilde{\gamma}_i \bar{\mathbf{R}}_i$. Therefore, if $\tilde{\gamma}_i \approx 0$, the covariance matrix $\tilde{\gamma}_i \bar{\mathbf{R}}_i$ of the slab component in (75) would be very small, allowing us to safely estimate that $\mathbf{x}_i \approx \mathbf{0}$, whereas if $\tilde{\gamma}_i > 0$, \mathbf{x}_i would be a nonzero Gaussian random vector.

Finally, to encode (ii)–(iii) into the Bayesian formulation, we propose the following *Hierarchical spike and slab prior* on $p(\mathbf{X} | \mathbf{c}, \mathbf{R})$ as

$$p(\mathbf{X} | \mathbf{c}, \mathbf{R}) = \prod_{l=1}^{N_c} p(\mathbf{X}_{\mathcal{C}_l} | c_l) = \prod_{l=1}^{N_c} \left[(1 - c_l) \delta(\mathbf{X}_{\mathcal{C}_l}) + c_l \prod_{i \in \mathcal{C}_l} \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}_i) \right], \quad (76)$$

where $\delta(\mathbf{X}_{\mathcal{C}_l}) = \prod_{i \in \mathcal{C}_l} \delta(\mathbf{x}_i)$, and, for simplicity, we denoted $\mathbf{R}_i = \tilde{\gamma}_i \bar{\mathbf{R}}_i$.

from the exponential family by solving a series of local variational problems in an iterative fashion.

4.2 JADCE via EM-EP

A common approach in Bayesian inference when the hyper-parameter set $\Xi = \{\tilde{\gamma}_i, \bar{\mathbf{R}}_i\}_{i=1}^N$ is not known is to maximize the likelihood function $p(\mathbf{Y} | \Xi)$. However, in many cases,

the likelihood $p(\mathbf{Y}|\Xi)$ is a non-convex function of Ξ and its global maximum cannot be found in a closed form. Thus, Ξ can be obtained through type-II maximum likelihood estimation by finding a local maximum using the expectation-maximization (EM) framework. The classical EM iteratively alternates between two steps, namely, the E-step and the M-step. In the E-step, the current values of the hyper-parameters are used to evaluate the posterior distribution of interest. Subsequently, the hyper-parameters are estimated again using the current statistics of the posterior distribution in the M-step [105].

In the JADCE context, for a given Ξ , the joint posterior distribution $p(\mathbf{X}, \mathbf{c}|\mathbf{Y})$ is expressed as a product for three factors, namely, $f_1(\mathbf{X})$, $f_2(\mathbf{X}, \mathbf{c})$ and $f_3(\mathbf{c})$ as

$$\begin{aligned}
f(\mathbf{X}, \mathbf{c}) &= p(\mathbf{X}, \mathbf{c}|\mathbf{Y}) = \frac{1}{p(\mathbf{Y})} p(\mathbf{Y}|\mathbf{X}) p(\mathbf{X}|\mathbf{c}) p(\mathbf{c}) \\
&= \frac{1}{p(\mathbf{Y})} \underbrace{\mathcal{L} \mathcal{N}(\Phi \mathbf{X}^T, \sigma^2 \mathbf{I}_m)}_{f_1(\mathbf{X})} \underbrace{\prod_{l=1}^{N_c} [(1 - c_l) \delta(\mathbf{X}_{\mathcal{C}_l}) + c_l \prod_{i \in \mathcal{C}_l} \mathcal{L} \mathcal{N}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}_i)]}_{f_2(\mathbf{X}, \mathbf{c})} \\
&\quad \underbrace{\prod_{l=1}^{N_c} \mathcal{B}(\mathbf{c}_l; p_c)}_{f_3(\mathbf{c})}.
\end{aligned} \tag{77}$$

The disadvantage of the spike-and-slab prior is that it renders the evaluation of the posterior distribution in (77) a computationally demanding Task. In particular, $p(\mathbf{X}, \mathbf{c}|\mathbf{Y})$ cannot be computed exactly when N is large, and, thus, it has to be estimated numerically. To this end, we resort to the expectation propagation (EP) algorithm to find a tractable approximation for the true posterior distribution $p(\mathbf{X}, \mathbf{c}|\mathbf{Y})$ in (77).

Next, we will derive a novel JADCE solution based on coupling the EP algorithm within the EM framework. More precisely, at any EM iteration (k): (i) given the set of hyper-parameter $\Xi^{(k-1)}$, the EP framework is utilized to approximate the intractable posterior distribution $p(\mathbf{X}, \mathbf{c}|\mathbf{Y})$ and subsequently, compute the posterior mean, denoted as $\mathbf{m}_i^{(k-1)}$, and covariance matrix, denoted as $\Sigma_i^{(k-1)}$, of the effective channel \mathbf{x}_i , $\forall i \in \mathcal{N}$. (ii) In the M-step, by utilizing the sufficient statistics of \mathbf{X} obtained from the previous E-step, we compute the new values of the hyper-parameters $\Xi^{(k)}$ by minimizing an upper-bound on the negative log-likelihood $p(\mathbf{Y}|\Xi)$. Next, we will present the details of the proposed algorithm.

4.2.1 Main idea of expectation propagation

The EP algorithm is a deterministic framework for approximating probability distributions. In particular, in the EP framework, the true distribution is approximated by another distribution that is drawn from the exponential family by solving a series of local variational problems in an iterative fashion.

The EP algorithm aims to approximate iteratively the true posterior distribution $f(\mathbf{X}, \mathbf{c})$ by a simpler distribution $Q(\mathbf{X}, \mathbf{c})$ that belongs to an exponential family. More precisely, EP aims to approximate the factors $f_1(\cdot)$, $f_2(\cdot)$, $f_3(\cdot)$ in (77) by $q_1(\cdot)$, $q_2(\cdot)$, $q_3(\cdot)$, respectively, such that

$$f(\mathbf{X}, \mathbf{c}) \approx Q(\mathbf{X}, \mathbf{c}) = \frac{1}{K^{\text{EP}}} q_1(\mathbf{X}) q_2(\mathbf{X}, \mathbf{c}) q_3(\mathbf{c}). \quad (78)$$

Each factor $q_k(\cdot)$, $k = 1, 2, 3$, of the joint approximation $Q(\mathbf{X}, \mathbf{c})$ is obtained by minimizing iteratively the Kullback-Leibler divergence [105] as

$$\tilde{q}_k^* = \min_{q_k} \text{KL} \left(f_k(\cdot) Q^{\setminus k}(\cdot) \parallel q_k(\cdot) Q^{\setminus k}(\cdot) \right), k = 1, 2, 3, \quad (79)$$

where $Q^{\setminus k}(\cdot) = \frac{Q(\cdot)}{q_k(\cdot)}$ is termed the cavity distribution. The optimization problem (79) is convex, with a unique global optimum solution that is obtained by matching the expected values of the sufficient statistics of $q_k(\cdot) Q^{\setminus k}(\cdot)$ to those of $f_k(\cdot) Q^{\setminus k}(\cdot)$ [105].

In the following section, we show in detail how to derive the approximation factor $Q(\mathbf{X}, \mathbf{c})$ through the EP framework.

4.2.2 E-Step: posterior approximation via EP

The choice of the approximating factors in EP is not stringent, rather, it is flexible. We therefore design $q_1(\cdot)$, $q_2(\cdot)$, $q_3(\cdot)$ such that they: 1) offer tractability and closed-form updates; 2) capture the important features of the true posterior distribution, such as cluster-level and intra-cluster sparsity. To this end, we design the approximate factors as follows

$$q_1(\mathbf{X}) = \prod_{i=1}^N q_1(\mathbf{x}_i) = \prod_{i=1}^N \mathcal{C} \mathcal{N}(\mathbf{x}_i; \mathbf{m}_{1,i}, \Sigma_{1,i}), \quad (80)$$

$$q_2(\mathbf{X}, \mathbf{c}) = \prod_{l=1}^{N_c} q_2(\mathbf{X}_{\mathcal{C}_l}, c_l) = \prod_{l=1}^{N_c} c_l \prod_{i \in \mathcal{C}_l} \mathcal{C} \mathcal{N}(\mathbf{x}_i; \mathbf{m}_{2,i}, \Sigma_{2,i}), \quad (81)$$

$$q(\mathbf{c}) = \prod_{l=1}^{N_c} \mathcal{B}(c_l; p_c). \quad (82)$$

Thus, we write the global approximation $Q(\mathbf{X}, \gamma, \mathbf{c})$ as

$$\begin{aligned}
Q(\mathbf{X}, \gamma, \mathbf{c}) &\propto q_1(\mathbf{X})q_2(\mathbf{X}, \mathbf{c})q_3(\mathbf{c}) \\
&\propto \prod_{i=1}^N \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{m}_{1,i}, \Sigma_{1,i}) \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{m}_{2,i}, \Sigma_{2,i}) \prod_{l=1}^{N_c} c_l \mathcal{B}(c_l; p_c) \\
&\propto \prod_{i=1}^N \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{m}_i, \Sigma_i) \prod_{l=1}^{N_c} c_l \mathcal{B}(c_l; p_c),
\end{aligned} \tag{83}$$

with

$$\begin{aligned}
\Sigma_i &= (\Sigma_{1,i}^{-1} + \Sigma_{2,i}^{-1})^{-1}, \quad i = 1, \dots, N, \\
\mathbf{m}_i &= \Sigma_i (\Sigma_{1,i}^{-1} \mathbf{m}_{1,i} + \Sigma_{2,i}^{-1} \mathbf{m}_{2,i}), \quad i = 1, \dots, N,
\end{aligned} \tag{84}$$

where $\mathbf{m} = [\mathbf{m}_1, \dots, \mathbf{m}_N]$ and $\Sigma = \{\Sigma_i\}_{i=1}^N$ are obtained by applying the product of two Gaussian terms rule, as shown in Appendix 2.1. We note that since $q_3(\mathbf{c})$ is the same as $f_3(\mathbf{x})$, it can be obtained directly, and we need to estimate only $q_1(\mathbf{X})$ and $q_2(\mathbf{X}, \mathbf{c})$, as we show next.

Estimation of $q_1(\mathbf{X})$

We now describe how to compute the $\{\mathbf{m}_{1,i}, \Sigma_{1,i}\}_{i=1}^N$ of the first approximate term $q_1(\mathbf{X})$. Note that since both the $f_1(\mathbf{X})$ and $q_1(\mathbf{X})$ have a Gaussian form, $f_1(\mathbf{X})$ can be approximated exactly by $q_1(\mathbf{X})$, independently of the values of the other approximate factors $q_2(\cdot)$ and $q_3(\cdot)$. Subsequently, we only have to set $q_1(\cdot) = f_1(\cdot)$ at the start of the EP algorithm, and it can be kept constant afterwards.

Now, let us rewrite the vector form of the received signal as

$$\mathbf{y} = \Theta \bar{\mathbf{x}} + \mathbf{w}, \tag{85}$$

where $\bar{\mathbf{x}} = \text{vec}(\mathbf{X})$, $\mathbf{w} = \text{vec}(\mathbf{W}^T) \in \mathbb{C}^{\tau_p M}$, $\mathbf{y} = \text{vec}(\mathbf{Y}^T) \in \mathbb{C}^{\tau_p M}$, and $\Theta = \Phi \otimes \mathbf{I}_M \in \mathbb{C}^{M \tau_p \times NM}$, where the operation $\text{vec}(\cdot)$ stacks the columns of the matrix vertically. Subsequently, the vector-form of the likelihood function $f_1(\cdot)$ is given by

$$f_1(\bar{\mathbf{x}}) = p(\mathbf{y}|\bar{\mathbf{x}}) \sim \mathcal{C}\mathcal{N}(\mathbf{y}; \Theta \bar{\mathbf{x}}, \sigma^2 \mathbf{I}_{M \tau_p}). \tag{86}$$

Similarly, we can write the vector form of $q_1(\mathbf{X})$ as

$$q_1(\bar{\mathbf{x}}) \sim \mathcal{C}\mathcal{N}(\bar{\mathbf{x}}; \bar{\mathbf{m}}_1, \bar{\Sigma}_1), \tag{87}$$

where $\bar{\Sigma}_1 = \text{diag}(\Sigma_{1,1}, \dots, \Sigma_{1,N})$ is a block diagonal matrix and $\bar{\mathbf{m}}_1 = [\mathbf{m}_{1,1}^T, \dots, \mathbf{m}_{1,N}^T]^T$. Note that $f_1(\bar{\mathbf{x}})$ is a distribution of \mathbf{y} conditioned on $\bar{\mathbf{x}}$, whereas $q_1(\bar{\mathbf{x}})$ is a function of $\bar{\mathbf{x}}$ that depends on \mathbf{y} , $\bar{\mathbf{m}}$, and $\bar{\Sigma}$. Thus, by writing the full Gaussian distributions in

(86) and (87), and rearranging a few terms, the first and second moments of $q_1(\bar{\mathbf{x}})$ are parameterized as follows

$$\bar{\Sigma}_1^{-1} = \frac{1}{\sigma^2} \Theta^H \Theta, \quad \bar{\Sigma}_1^{-1} \bar{\mathbf{m}}_1 = \frac{1}{\sigma^2} \Theta^H \mathbf{y}. \quad (88)$$

Estimation of $q_2(\cdot)$

Note that $q_2(\mathbf{X}, \mathbf{c})$ in (81) factorizes into N_c independent mixed Gaussian-Bernoulli distributions $q_2(\mathbf{X}_{\mathcal{C}_l}, c_l)$, $l = 1, \dots, N_c$, allowing for the parallel updates of $q_2(\mathbf{X}, \mathbf{c})$. In the following, we present in detail how to update each term $q_2(\mathbf{X}_{\mathcal{C}_l}, c_l)$.

Update $Q^{\setminus 2,l}(\cdot)$ First, we compute the marginal cavity distribution $Q^{\setminus 2,l}(\mathbf{X}_{\mathcal{C}_l}, c_l)$ by removing the contribution of $q_2(\mathbf{X}_{\mathcal{C}_l}, \mathbf{c}_l)$ from the the global approximation $Q(\mathbf{X}, \mathbf{c})$ in (83) as

$$Q^{\setminus 2,l}(\mathbf{X}_{\mathcal{C}_l}, c_l) = \frac{Q(\mathbf{X}, \mathbf{c})}{q_2(\mathbf{X}_{\mathcal{C}_l}, c_l)} \propto \prod_{i \in \mathcal{C}_l} \mathcal{C} \mathcal{N}(\mathbf{x}_i; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l}) \mathcal{B}(c_l; p_c), \quad (89)$$

where $\hat{\Sigma}_i^{\setminus 2,l}$ and $\hat{\mathbf{m}}_i^{\setminus 2,l}$ are obtained by utilizing the rule of the fraction of two Gaussian terms, as shown in Appendix 2.1, and they are given as

$$\begin{aligned} \hat{\Sigma}_i^{\setminus 2,l} &= (\Sigma_i^{-1} - \Sigma_{2,i}^{-1})^{-1}, \quad \forall i \in \mathcal{C}_l, \\ \hat{\mathbf{m}}_i^{\setminus 2,l} &= \hat{\Sigma}_i^{\setminus 2,l} (\Sigma_i^{-1} \mathbf{m}_i - \Sigma_{2,i}^{-1} \mathbf{m}_{2,i}), \quad \forall i \in \mathcal{C}_l. \end{aligned} \quad (90)$$

Update Q^{new} Next, we update $Q^{\text{new}(\cdot)} = q_2(\cdot) Q^{\setminus 2,l}(\cdot)$ by minimizing a KL divergence distance given as $\text{KL} \left(\frac{1}{G_{l,0}} f_2(\cdot) Q^{\setminus 2,l}(\cdot) \parallel q_2(\cdot) Q^{\setminus 2,l}(\cdot) \right)$, where $G_{l,0}$ is the normalizing constant needed to ensure that the $f_2(\cdot) Q^{\setminus 2,l}(\cdot)$ integrates to unity, given as (Please refer to Appendix 2.2)

$$G_{l,0} = \sum_{\mathbf{c}} \int_{\mathbf{X}} f_2(\mathbf{X}_{\mathcal{C}_l}, c_l) Q^{\setminus 2,l}(\mathbf{X}_{\mathcal{C}_l}, \mathbf{c}) d\mathbf{X} d\mathbf{c} = b_l + a_l, \quad (91)$$

where $b_l = (1 - p_c) \prod_{i \in \mathcal{C}_l} \mathcal{C} \mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l})$ and $a_l = p_c \prod_{i \in \mathcal{C}_l} \mathcal{C} \mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l} + \mathbf{R}_i)$.

Since all the terms in $Q(\cdot)$ and $f_2(\cdot)$ are drawn from a Gaussian distribution, the solution to the previous KL-divergence is given by matching the moments of $Q^{\text{new}(\cdot)}$ and $f_{2,l}(\cdot) Q^{\setminus 2,l}(\cdot)$. To this end, we compute the sufficient statistics of $f_{2,l}(\cdot) Q^{\setminus 2,l}(\cdot)$ with respect to both c_l and \mathbf{x}_i , $i \in \mathcal{C}_l$. Subsequently, the sufficient central moments of $Q^{\text{new}(\cdot)}$

are given as

$$\begin{aligned}
\mathbb{E}_{f_{2,l}Q^{2,l}}[c_l] &= \frac{a_l}{a_l + b_l} \\
\mathbb{E}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i] &= \frac{1}{G_{l,0}} (a_l \mathbf{R}_i (\mathbf{R}_i + \hat{\Sigma}_i^{2,l})^{-1} \hat{\mathbf{m}}_i^{2,l}) \\
\text{Var}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i] &= \frac{a_l}{G_{l,0}} (\hat{\Sigma}_i^{2,l^{-1}} + \mathbf{R}_i^{-1})^{-1} + \left(\frac{1}{G_{l,0} a_l} - 1 \right) \mathbb{E}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i] \mathbb{E}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i]^H.
\end{aligned} \tag{92}$$

The details of the derivations to obtain (92) are presented in Appendix 2.2.

Update $q_2(\mathbf{X}_{\mathcal{C}_l}, c_l)$ Finally, the updated $q_2(\cdot)$ is computed as

$$q_2(\mathbf{X}_{\mathcal{C}_l}, c_l) = \frac{Q^*(\mathbf{X}, \mathbf{c})}{Q^{2,l}(\mathbf{X}_{\mathcal{C}_l}, c_l)} \propto \prod_{i \in \mathcal{C}_l} \mathcal{C} \cdot \mathcal{N}(\mathbf{x}_i; \mathbf{m}_{2,i}, \Sigma_{2,i}) \mathcal{B}(c_l; \frac{a_l}{a_l + b_l}), \tag{93}$$

where the mean and the covariance matrices of the updated $q_2(\mathbf{X}_{\mathcal{C}_l}, c_l)$ are computed as

$$\Sigma_{2,i} = (\text{Var}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i]^{-1} - \hat{\Sigma}_{2,i}^{2,l^{-1}})^{-1}, i \in \mathcal{C}_l, \tag{94}$$

$$\mathbf{m}_{2,i} = \Sigma_{2,i}^{-1} (\text{Var}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i]^{-1} \mathbb{E}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i] - \hat{\Sigma}_{2,i}^{2,l^{-1}} \hat{\mathbf{m}}_i^{2,l})^{-1}, i \in \mathcal{C}_l. \tag{95}$$

4.2.3 M-Step: hyper-parameter update

Once $q_2(\mathbf{X}, \mathbf{c})$ is updated using the previous $\Xi^{(k-1)}$, the new posterior mean $\mathbf{m}^{(k)}$ and posterior covariance $\Sigma^{(k)}$ of $Q(\mathbf{X}, \mathbf{c})$ are updated using (84). Subsequently, the M-step at the k th EM iteration is carried out as follows

$$\begin{aligned}
\Xi^{(k)} &= \arg \max_{\Xi} \mathbb{E}_{p(\mathbf{X}, \mathbf{c} | \mathbf{Y}, \Xi^{(k-1)})} \log p(\mathbf{Y}, \mathbf{X}, \mathbf{c}, \Xi) \\
&\stackrel{(a)}{\approx} \arg \max_{\Xi} \mathbb{E}_{Q(\mathbf{X}, \mathbf{c} | \mathbf{Y}, \Xi^{(k-1)})} \log p(\mathbf{Y}, \mathbf{X}, \mathbf{c}, \Xi) \\
&\stackrel{(b)}{=} \arg \max_{\Xi} \mathbb{E}_{Q(\mathbf{X}, \mathbf{c} | \mathbf{Y}, \Xi^{(k-1)})} \left[\log p(\mathbf{X}, \mathbf{c} | \bar{\gamma}, \bar{\mathbf{R}}) + \log p(\bar{\gamma}) + \log p(\bar{\mathbf{R}}) \right] \\
&\stackrel{(c)}{\propto} \arg \max_{\Xi} \sum_{i=1}^N -M \log(\bar{\gamma}_i) - \log |\bar{\mathbf{R}}_i| - \text{Tr} \left[\bar{\mathbf{R}}_i^{-1} \frac{\mathbb{E}_{\mathbf{X}, \mathbf{c} | \mathbf{Y}, \Xi}[\mathbf{x}_i \mathbf{x}_i^H]}{\bar{\gamma}_i} \right] + \log p(\bar{\gamma}) + \log p(\bar{\mathbf{R}}) \\
&\stackrel{(d)}{\propto} \arg \max_{\Xi} \sum_{i=1}^N -M \log(\bar{\gamma}_i) - \log |\bar{\mathbf{R}}_i| + \log p(\bar{\gamma}) + \log p(\bar{\mathbf{R}}) \\
&\quad - \text{Tr} \left[\bar{\gamma}_i^{-1} \bar{\mathbf{R}}_i^{-1} (\mathbf{m}_i^{(k)} \mathbf{m}_i^{(k)H} + \Sigma_i^{(k)}) \right],
\end{aligned} \tag{96}$$

where (a) is obtained by approximating the expectation over the intractable $p(\mathbf{X}, \mathbf{c}|\mathbf{Y}, \Xi^{(k-1)})$ by its EP approximation $Q(\mathbf{X}, \mathbf{c})$, (b) from the joint distribution chain rule, and $\tilde{\gamma}$ and $\{\bar{\mathbf{R}}_i\}_{i=1}^N$ are independent by design, and by dropping the terms that do not depend on Ξ from the objective function, (c) by using the conditional PDF of \mathbf{X} as defined in (76), and (d) by using the moments of the global EP approximation (83).

$\tilde{\gamma}$ -update

Note that the final expression in (96) shows that the prior $p(\tilde{\gamma})$ plays a role in the M-step. However, it has been shown that even a non-uniform prior will lead to a sparse vector $\tilde{\gamma}$, thanks to the RVM mechanism [49]. Thus we will make the same simplification and drop $p(\tilde{\gamma})$ from (96). Subsequently, (96) decouples into N subproblem across $\tilde{\gamma}$ as

$$\tilde{\gamma}_i^{(k)} = \max_{\tilde{\gamma}_i \geq 0} -M \log(\tilde{\gamma}_i) - \text{Tr} \left[\tilde{\gamma}_i^{-1} \bar{\mathbf{R}}_i^{-1} (\mathbf{m}_i^{(k)} \mathbf{m}_i^{(k)\text{H}} + \Sigma_i^{(k)}) \right], \quad \forall i. \quad (97)$$

Thus, the optimal solution is obtained by setting the gradients of the objective function in (97) with respect to $\tilde{\gamma}_i$ to zero, resulting in the following update rule:

$$\tilde{\gamma}_i^{(k)} = \frac{1}{M} \text{Tr} \left[\bar{\mathbf{R}}_i^{-1} (\mathbf{m}_i^{(k)} \mathbf{m}_i^{(k)\text{H}} + \Sigma_i^{(k)}) \right], \quad \forall i. \quad (98)$$

Covariance matrix $\{\bar{\mathbf{R}}_i\}$ -update

Taking inspiration from the argument presented in [50], we recognize that attempting to estimate all N covariance matrices $\bar{\mathbf{R}}_i$ using the data available at the BS would result in overfitting. Therefore, instead of estimating N covariance matrices $\bar{\mathbf{R}}_i$, we estimate only N_c covariance matrices $\bar{\mathbf{R}}_{\mathcal{C}_l}$, $l = 1, \dots, N_c$. In particular, we make the assumption that all effective channels within a single cluster share the same spatial correlation structure¹⁰, i.e., $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \tilde{\gamma}_i \bar{\mathbf{R}}_{\mathcal{C}_l})$, $\forall i \in \mathcal{C}_l, \forall l = 1, \dots, N_c$. Consequently, the optimization problem (96) with respect to $\bar{\mathbf{R}}_{\mathcal{C}_l}$ decoupled into L subproblems as

$$\begin{aligned} \bar{\mathbf{R}}_{\mathcal{C}_l}^{(k)} &= \max_{\bar{\mathbf{R}}_{\mathcal{C}_l} \succ \mathbf{0}} -(L + Ld) \log \det(\bar{\mathbf{R}}_{\mathcal{C}_l}) \\ &+ \text{Tr} \left[\bar{\mathbf{R}}_{\mathcal{C}_l}^{-1} \left(\sum_{i \in \mathcal{C}_l} \frac{\mathbf{m}_i^{(k)} \mathbf{m}_i^{(k)\text{H}} + \Sigma_i^{(k)}}{\tilde{\gamma}_i} + L\mathbf{B}_{\mathcal{C}_l} \right) \right], \quad l = 1, \dots, N_c. \end{aligned} \quad (99)$$

¹⁰It is noteworthy that this assumption is not only technical but also highly justified in the context of MIMO channels, where closely located devices are expected to have approximately the same covariance matrices [134]

Therefore, by applying the first-order optimality condition, $\bar{\mathbf{R}}_{\mathcal{C}_l}$ is given as

$$\bar{\mathbf{R}}_{\mathcal{C}_l}^{(k)} = \frac{1}{L + Ld} \left(\sum_{i \in \mathcal{C}_l} \frac{\mathbf{m}_i^{(k)} \mathbf{m}_i^{(k)H} + \Sigma_i^{(k)}}{\tilde{\gamma}_i} + L\mathbf{B}_{\mathcal{C}_l} \right), l = 1, \dots, N_c. \quad (100)$$

Subsequently, we set $\bar{\mathbf{R}}_i^{(k)} = \bar{\mathbf{R}}_{\mathcal{C}_l}^{(k)}, \forall i \in \mathcal{C}_l$.

4.2.4 Algorithm implementation

The details of the proposed algorithm, termed EM-EP, are summarized in Algorithm 4. EM-EP is run until $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_{\text{F}}^2 < \epsilon_{\text{stp}}$, or until a maximum number of iterations k_{max} is reached. Next, we outline two practical implementation considerations of the EM-EP algorithm.

Non-positive covariance matrix

While $Q(\cdot)$ and $Q^{\text{new}}(\cdot)$ must have to be proper distributions in the EP framework, the approximated terms $q_1(\cdot)$ and $q_2(\cdot)$, on the other hand, are not constrained to be proper distributions. In practice, some factors $q_2(\cdot)$ may become improper, resulting in non-positive definite covariance matrices. In this work, for any non-positive definite matrix $\Sigma_{2,i}, \forall i$, we add a small regularization parameter $\zeta_{\text{EP}} > 0$ to its diagonal elements to ensure that the covariance matrix is positive definite. For more alternative solutions, we refer the reader to [135].

Pruning

In practice, to overcome the computational complexity of EM-EP, at each iteration, we reduce the search space by ignoring any non-active devices when updating the approximate factors. To this end, we use the posterior mean of the activity cluster indicator $\mathbb{E}_{f_{2,l}Q^{2,l}}[c_l]$ as a measure to prune non-active clusters. More precisely, for any cluster l with $\mathbb{E}_{f_{2,l}Q^{2,l}}[c_l] < \epsilon_{\text{thr}}$, all the effective channels are set to zero, i.e., $\mathbf{X}_{\mathcal{C}_l} = \mathbf{0}$, and they are pruned from the model.

4.3 Alternative solution via ADMM

Owing to the proposed *Hierarchical spike-and-slab* prior and the efficient joint posterior approximation obtained using the EP framework, the proposed EM-EP algorithm provides superior performance compared to the state-of-the-art algorithm, as we will

Algorithm 4: EM-EP

Input: received signal \mathbf{Y} , pilot sequence matrix Φ , noise variance σ^2 , p_c , ε_{stp} ,

$\{\mathbf{B}_l\}_{l=1}^{N_c}, k_{\text{max}}$.

Initialization: $\mathbf{m}_2^{(0)}, \Sigma_2^{(0)}, k = 1$

- 1 Compute the $\hat{\Sigma}_1^{-1} = \frac{1}{\sigma^2} \Theta^H \Theta$ $\mathbf{V}_1^{-1} \hat{\mathbf{m}}_1 = \frac{1}{\sigma^2} \Theta^H \mathbf{y}$
- 2 **while** $k < k_{\text{max}}$ or $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\| < \varepsilon_{\text{stp}}$ **do**
- 3 Compute the covariance $\hat{\Sigma}_i^{2,l}$ and the mean $\hat{\mathbf{m}}_i^{2,l}$ of the cavity distribution \mathbf{m}_2 using (90)
- 4 Compute the normalization term $G_{l,0}$ using (91)
- 5 Compute the covariance $\text{Var}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i]$ and the mean $\mathbb{E}_{f_{2,l}Q^{2,l}}[\mathbf{x}_i]$ using (92)
- 6 Compute the new covariance $\hat{\Sigma}_{2,i}$ and the mean $\hat{\mathbf{m}}_{2,i}$ of $q_2(\mathbf{X}_{\mathcal{C}_l}, c_l)$ using (94) and (95) respectively
- 7 update the new posterior mean \mathbf{m}_i and covariance Σ_i using (84)
- 8 Update the set of hyper-parameters Ξ using (98) and (100)
- 9 $k \leftarrow k + 1$

Output: $\mathbf{X} = \mathbf{m}$

show in the simulation analysis. However, the EM-EP algorithm, like most EP-based methods, comes with the burden of high computational complexity. Therefore, in this section, we provide an alternative solution to the JADCE problem by relaxing the *strong* spike-and-slab prior with a weak prior based on the *log-sum penalty* [107], which still captures the essence of the *hierarchical structure* of the device's activity pattern. Furthermore, we reformulate JADCE as a MAP estimation that is solved via ADMM to offer a computationally-efficient algorithm with closed-form updates that can be computed via simple analytical expressions. Next, we present the proposed solution in detail.

4.3.1 JADCE as MAP estimation

In this section, we formulate JADCE as a MAP estimation problem to: 1) identify the active clusters, and their corresponding active devices; 2) estimate the effective channel matrix \mathbf{X} ; and 3) estimate the unknown covariance matrices $\{\mathbf{R}_i\}_{i=1}^N$. Subsequently, given the received signal \mathbf{Y} , the MAP estimation problem is expressed as

$$\begin{aligned}
\{\hat{\mathbf{X}}, \hat{\boldsymbol{\gamma}}, \hat{\mathbf{R}}\} &= \max_{\mathbf{X}, \boldsymbol{\gamma}, \mathbf{R}} p(\mathbf{X}, \boldsymbol{\gamma}, \mathbf{R} | \mathbf{Y}) \\
&= \max_{\mathbf{X}, \boldsymbol{\gamma}, \mathbf{R}} \frac{p(\boldsymbol{\gamma}, \mathbf{R}) p(\mathbf{X} | \boldsymbol{\gamma}, \mathbf{R}) p(\mathbf{Y} | \mathbf{X}, \boldsymbol{\gamma}, \mathbf{R})}{p(\mathbf{Y})} \\
&\stackrel{(a)}{=} \max_{\mathbf{X}, \boldsymbol{\gamma}, \mathbf{R}} p(\boldsymbol{\gamma}, \mathbf{R}) p(\mathbf{X} | \boldsymbol{\gamma}, \mathbf{R}) p(\mathbf{Y} | \mathbf{X}) \\
&= \min_{\mathbf{X}, \boldsymbol{\gamma}, \mathbf{R}} -\log p(\mathbf{Y} | \mathbf{X}) - \log p(\mathbf{X} | \boldsymbol{\gamma}, \mathbf{R}) - \log p(\boldsymbol{\gamma}, \mathbf{R}) \\
&\stackrel{(b)}{=} \min_{\mathbf{X}, \boldsymbol{\gamma}, \mathbf{R}} \frac{1}{\sigma^2} \|\mathbf{Y} - \Phi \mathbf{X}^T\|_F^2 - \log p(\mathbf{X} | \boldsymbol{\gamma}, \mathbf{R}) - \log p(\mathbf{R}) - \log p(\boldsymbol{\gamma}),
\end{aligned} \tag{101}$$

where (a) follows from the Markov chain $\{\boldsymbol{\gamma}, \mathbf{R}\} \rightarrow \mathbf{X} \rightarrow \mathbf{Y}$, and since the maximization is independent from $p(\mathbf{Y})$, and (b) follows from the likelihood function of the received signal model (1). Note that an alternative expression to the conditional PDF of \mathbf{X} in (76) is given as

$$p(\mathbf{X} | \boldsymbol{\gamma}, \mathbf{R}) = \prod_{i=1}^N p(\mathbf{x}_i | \boldsymbol{\gamma}_i, \mathbf{R}_i) = \prod_{i=1}^N \mathcal{C} \mathcal{N}(\mathbf{x}_i; \mathbf{0}, \rho_i^{\text{UL}} \mathbf{R}_i)^{\mathbb{I}(\boldsymbol{\gamma}_i=0)}, \tag{102}$$

where $\mathbb{I}(a)$ is an indicator function that takes the value 1 if $a \neq 0$, and 0 otherwise. Note that $p(\mathbf{x}_i | \boldsymbol{\gamma}_i, \mathbf{R}_i)$ in (102) implies that if $\boldsymbol{\gamma}_i = 0$, then \mathbf{x}_i equals 0 with probability one, and in the case of $\boldsymbol{\gamma}_i = 1$, the effective channel \mathbf{x}_i follows a complex Gaussian distribution with a zero mean and a covariance matrix $\rho_i^{\text{UL}} \mathbf{R}_i$.

4.3.2 Cluster-sparsity promoting prior via log-sum

Recall the definition of the effective channel $\mathbf{x}_i = \sqrt{\rho_i^{\text{UL}}} \boldsymbol{\gamma}_i \mathbf{h}_i$, where the true indicator variable $\boldsymbol{\gamma}_i \in \{0, 1\}$ controls the sparsity of i th device. Thus, assigning a sparsity-promoting prior to $p(\boldsymbol{\gamma})$ is the key to obtaining a sparse solution to (101). A conventional choice for a *tractable* sparsity-promoting prior $p(\boldsymbol{\gamma})$ is the *log-sum* penalty prior $\sum_{i=1}^N \log(\boldsymbol{\gamma}_i + \epsilon_0)$, as it resembles most closely the canonical ℓ_0 -norm, when $\epsilon_0 \rightarrow 0$. Subsequently, we define the following sparsity prior $p(\boldsymbol{\gamma})$

$$p(\boldsymbol{\gamma}) \propto J_s(\boldsymbol{\gamma}) = \sum_{i=1}^N \log(\boldsymbol{\gamma}_i + \epsilon_0). \tag{103}$$

Although the prior (103) is an appropriate choice as it 1) promotes sparsity, and 2) is separable across the devices, it ignores the hierarchical structure of the devices' activity pattern. Therefore, to account for the cluster-level sparsity, we propose the *cluster-sparsity-promoting prior* $J_c(\cdot)$ that correlates the device activity indicator variables

belonging to the same cluster, i.e., $\gamma_i, i \in \mathcal{C}_l$, with the same log-sum penalty. More precisely, we propose the following prior function:

$$p(\boldsymbol{\gamma}) \propto J_c(\boldsymbol{\gamma}) = \sum_{l=1}^{N_c} \log \left(\sum_{i \in \mathcal{C}_l} \gamma_i + \varepsilon_0 \right). \quad (104)$$

Note that $J_c(\boldsymbol{\gamma})$ promotes quite stringently solutions that have clustered sparsity, as it has the tendency to enforce all devices within each cluster to be detected as active even if only one device is active, thereby being susceptible to high false alarm error rate. Thus, $J_c(\boldsymbol{\gamma})$ would face robustness issues in instances where the devices' activity pattern does not exhibit a clustered structure.

4.3.3 Proposed ADMM solution

Before we derive the ADMM-based solution for (101), we make two technical choices: 1) the binary nature of γ renders the objective function in (101) intractable for large N . To overcome this challenge, we note that finding the index set $\{i \mid \gamma_i \neq 0, i \in \mathcal{N}\}$ is equivalent to finding the index set $\{i \mid \|\mathbf{x}_i\|_2 > 0, i \in \mathcal{N}\}$. We can thus eliminate the variable γ from the MAP problem by approximating each γ_i by $\|\mathbf{x}_i\|_2$ and by relaxing $p(\boldsymbol{\gamma})$ by an equivalent prior function $p(\mathbf{X})$ which depends on $\|\mathbf{x}_i\|_2, \forall i \in \mathcal{N}$. 2) Similarly to Section 4.2.3, we resort to estimating a shared covariance matrix for all the devices within the same cluster.

By using the aforementioned arguments and introducing the regularization weights β_1, β_2 , and β_3 that control the emphasis on the priors with respect to the measurement fidelity term, the MAP estimation problem (101) can be equivalently rewritten as

$$\begin{aligned} \{\hat{\mathbf{X}}, \hat{\mathbf{R}}\} = \min_{\mathbf{X}, \mathbf{R}} \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}^T\|_{\text{F}}^2 - \beta_1 \log p(\mathbf{X}) + \beta_2 \sum_{l=1}^{N_c} \sum_{i \in \mathcal{C}_l} \frac{\mathbf{x}_i^H \mathbf{R}_{\mathcal{C}_l}^{-1} \mathbf{x}_i}{\rho_i^{\text{UL}}} \\ - \beta_2 \sum_{l=1}^{N_c} \log \det(\mathbf{R}_{\mathcal{C}_l}) \sum_{i \in \mathcal{C}_l} \rho_i^{\text{UL}M} \|\mathbf{x}_i\|_2 - \beta_3 L \sum_{l=1}^{N_c} (d \log \det(\mathbf{R}_{\mathcal{C}_l}) + \text{tr}(\mathbf{B}_l \mathbf{R}_{\mathcal{C}_l}^{-1})). \end{aligned} \quad (105)$$

Now, we propose an iterative solution based on a hierarchical algorithm with two loops, an *outer loop* and an *inner loop*. The central idea is to alternate $p(\mathbf{X})$ over $J_c(\cdot)$ (104) and $J_s(\cdot)$ (103). More precisely, in the outer loop, the algorithm enforces the detection of active clusters via the cluster-level sparsity-promoting function J_c . Subsequently, the algorithm runs an inner loop over the just-estimated active clusters to detect the individual active devices belonging to them by using the sparsity-promoting prior J_s . The algorithm details are presented next.

Outer loop

As we aim to detect the set of the active clusters first, we enforce $p(\mathbf{X})$ to promote the cluster-level sparsity by $-\log p(\mathbf{X}) = \sum_{l=1}^{N_c} \log(\sum_{i \in \mathcal{C}_l} \|\mathbf{x}_i\|_2 + \varepsilon_0)$. Since $-\log p(\mathbf{X})$ is concave, we apply the MM approximation [116] to linearize the cluster-level sparsity as

$$-\log p(\mathbf{X}) \approx \sum_{i=1}^N q_i^{(k_c)} \|\mathbf{x}_i\|_2, \quad (106)$$

where $q_i^{(k_c)} = (\sum_{i \in \mathcal{C}_l} \|\mathbf{x}_i^{(k_c)}\| + \varepsilon_0)^{-1}$, and k_c is the MM iteration index for the outer loop. Thus, the relaxed version of the problem (105) can be solved iteratively as

$$\begin{aligned} \{\hat{\mathbf{X}}^{(k_c+1)}, \mathbf{R}^{(k_c+1)}\} = \min_{\mathbf{X}, \mathbf{R}} & \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}^T\|_F^2 + \beta_1 \sum_{i=1}^N q_i^{(k_c)} \|\mathbf{x}_i\|_2 \\ & + \beta_2 \sum_{l=1}^{N_c} \sum_{i \in \mathcal{C}_l} \mathbf{x}_i^H \mathbf{R}_{\mathcal{C}_l}^{-1} \mathbf{x}_i + \sum_{l=1}^{N_c} \mu^{(k_c)} \log \det(\mathbf{R}_{\mathcal{C}_l}) + \beta_3 L \sum_{l=1}^{N_c} \text{tr}(\mathbf{B}_l \mathbf{R}_{\mathcal{C}_l}^{-1}), \end{aligned} \quad (107)$$

where $\mu^{(k_c)} = (\beta_2 \sum_{i \in \mathcal{C}_l} \rho_i^{\text{ULM}} q_i^{(k_c)} \|\mathbf{x}_i^{(k_c)}\| + \beta_3 L d)$.

We develop a computationally efficient ADMM solution for (107) through a set of sequential update rules, each computed in a closed-form. To this end, we introduce two splitting variables $\mathbf{Z}, \mathbf{V} \in \mathbb{C}^{M \times N}$ and the Lagrange dual variable Λ_v, Λ_z and define the set of variables to be estimated as $\Omega = \{\mathbf{X}, \mathbf{R}, \mathbf{Z}, \mathbf{V}, \Lambda_z, \Lambda_v\}$. Subsequently, we write the augmented Lagrangian as

$$\begin{aligned} \mathcal{L}(\Omega) = & \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{Z}^T\|_F^2 + \beta_1 \sum_{i=1}^N q_i^{(k_c)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{i=1}^N \mathbf{v}_i^H \mathbf{R}_{\mathcal{C}_l}^{-1} \mathbf{v}_i + \frac{\rho}{2} \|\mathbf{X} - \mathbf{V}\|_F^2 \\ & + \sum_{l=1}^{N_c} (\mu^{(k_c)} \log \det(\mathbf{R}_{\mathcal{C}_l}) + \beta_3 L \text{tr}(\mathbf{B}_l \mathbf{R}_{\mathcal{C}_l}^{-1})) + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z}\|_F^2 - \frac{\|\Lambda_z\|_F^2}{2\rho} - \frac{\|\Lambda_v\|_F^2}{2\rho}. \end{aligned} \quad (108)$$

The ADMM solves the optimization problem (107) by minimizing the augmented Lagrangian $\mathcal{L}(\Omega)$ in (108) over the primal variables $(\mathbf{Z}, \mathbf{V}, \mathbf{X}, \Sigma)$, followed by updating

the dual variables (Λ_z, Λ_v) [117]. The primal variable update is given as

$$\begin{cases} \mathbf{Z}^{(k_c+1)} = \min_{\mathbf{Z}} \frac{1}{2} \|\Phi \mathbf{Z}^T - \mathbf{Y}\|_{\mathbb{F}}^2 + \frac{\rho}{2} \|\mathbf{X}^{(k_c)} - \mathbf{Z} + \frac{1}{\rho} \Lambda_z^{(k_c)}\|_{\mathbb{F}}^2 \\ \mathbf{V}^{(k_c+1)} = \min_{\mathbf{V}} \beta_2 \sum_{i=1}^N \mathbf{v}_i^H \mathbf{R}_{\mathcal{C}_l}^{(k_c)} \mathbf{v}_i + \frac{\rho}{2} \|\mathbf{X}^{(k_c)} - \mathbf{V} + \frac{\Lambda_v^{(k_c)}}{\rho}\|_{\mathbb{F}}^2, \\ \mathbf{X}^{(k_c+1)} = \min_{\mathbf{X}} \sum_{i=1}^N \alpha_i^{(k_c)} \|\mathbf{x}_i\| + \frac{\rho}{2} \|\mathbf{X} - \mathbf{C}^{(k_c+1)}\|_{\mathbb{F}}^2, \\ \mathbf{R}_{\mathcal{C}_l}^{(k_c+1)} = \min_{\mathbf{R}_{\mathcal{C}_l}} \beta_2 \sum_{i \in \mathcal{C}_l} \mathbf{v}_i^{(k_c+1)H} \mathbf{R}_{\mathcal{C}_l}^{-1} \mathbf{v}_i^{(k_c+1)} + \mu^{(k_c+1)} \log \det(\mathbf{R}_{\mathcal{C}_l}) \\ \quad + \beta_3 L \text{tr}(\mathbf{B}_l \mathbf{R}_{\mathcal{C}_l}^{-1}), \forall l, \end{cases} \quad (109)$$

where $\alpha_i^{(k_c)} = q_i^{(k_c)} (\beta_1 - \beta_2 \log \det(\mathbf{R}_{\mathcal{C}_l}^{(k_c)}))$, and

$$\mathbf{C}^{(k_c)} = \frac{1}{2} (\mathbf{Z}^{(k_c+1)} + \mathbf{V}^{(k_c+1)} - \frac{\Lambda_z^{(k_c)} + \Lambda_v^{(k_c)}}{\rho}).$$

Inner loop

After running the outer loop for some pre-defined K_c iterations, we detect the set of the estimated active clusters $\hat{\mathcal{A}} = \bigcup_{j \in \mathcal{J}} \mathcal{C}_j$, where $l \in \mathcal{J}$ if there exists $i \in \mathcal{C}_l$ such that $\|\mathbf{x}_i\|_2 > \varepsilon_{\text{thr}}$.

In the inner loop, the proposed algorithm aims to detect the active devices belonging to $\hat{\mathcal{A}}$ by using the separable sparsity-promoting prior $-\log p(\mathbf{X}) \propto \sum_{i \in \hat{\mathcal{A}}} \log(\|\mathbf{x}_i\|_2 + \varepsilon_0)$. Furthermore, we apply the MM approximation to linearize the concave function as

$$-\log p(\mathbf{X}) \approx \sum_{i \in \hat{\mathcal{A}}} g_i^{(k_u)} \|\mathbf{x}_i\|_2, \quad (110)$$

where k_u is the inner loop iteration index, and $g_i^{(k_u)} = (\|\mathbf{x}_i^{(k_u)}\|_2 + \varepsilon_0)^{-1}$. Subsequently, the optimization problem for the inner loop is given by

$$\begin{aligned} \{\hat{\mathbf{X}}_{\hat{\mathcal{A}}}^{(k_u+1)}, \mathbf{R}_{\hat{\mathcal{A}}}^{(k_u+1)}\} = \min_{\mathbf{X}_{\hat{\mathcal{A}}}, \mathbf{R}} \frac{1}{2} \|\mathbf{Y} - \Phi_{\hat{\mathcal{A}}} \mathbf{X}_{\hat{\mathcal{A}}}^T\|_{\mathbb{F}}^2 + \beta_1 \sum_{i \in \hat{\mathcal{A}}} g_i^{(k_u)} \|\mathbf{x}_i\|_2 + \beta_2 \sum_{l \in \mathcal{J}} \sum_{i \in \mathcal{C}_l} \mathbf{x}_i^H \mathbf{R}_{\mathcal{C}_l}^{-1} \mathbf{x}_i \\ + \beta_2 \sum_{l \in \mathcal{J}} \log \det(\mathbf{R}_{\mathcal{C}_l}) \left(\sum_{i \in \mathcal{C}_l} g_i^{(k_u)} \|\mathbf{x}_i\|_2 + \beta_3 L d \right) + \beta_3 L \sum_{l \in \mathcal{J}} \text{tr}(\mathbf{B}_l \mathbf{R}_{\mathcal{C}_l}^{-1}), \end{aligned} \quad (111)$$

where $\Phi_{\hat{\mathcal{A}}}$ and $\mathbf{X}_{\hat{\mathcal{A}}}$ denote the matrices Φ and \mathbf{X} respectively, restricted to the set $\hat{\mathcal{A}}$.

Algorithm 5: corr-MAP-ADMM

Input: $\Phi, \{\mathbf{B}_l\}_{l=1}^{N_c}, \beta_1, \beta_2, \beta_3, \rho, \varepsilon_0, \varepsilon_{\text{stp}}, k_{u_{\text{max}}}, k_{c_{\text{max}}}, K_c$.

Initialization: $\mathbf{X}^{(0)}, \mathbf{V}^{(0)}, \mathbf{Z}^{(0)}, \Lambda_{\mathbf{v}}^{(0)}, \Lambda_{\mathbf{z}}^{(0)}, k_u = 1, k_c = 1$.

- 1 BS receives \mathbf{Y} , compute and store $(\Phi^T \Phi^* + \rho \mathbf{I}_N)^{-1}$
- 2 **while** $k_c < k_{c_{\text{max}}}$ or $\|\mathbf{X}^{(k_c)} - \mathbf{X}^{(k_c-1)}\| < \varepsilon_{\text{stp}}$ **do**
- 3 $\mathbf{Z}^{(k_c+1)} = (\rho \mathbf{X}^{(k_c)} + \Lambda_{\mathbf{z}}^{(k_c)} + \mathbf{Y}^T \Phi^*) (\Phi^T \Phi^* + \rho \mathbf{I}_N)$
- 4 $\mathbf{v}_i^{(k_c+1)} = (\beta_2 \Sigma_i^{(k_c)} + \rho \mathbf{I}_M)^{-1} (\rho \mathbf{x}_i^{(k_c)} + \lambda_{v_i}^{(k_c)}), i = 1, \dots, N$
- 5 $\mathbf{x}_i^{(k_c+1)} = \frac{\max\{0, \|\mathbf{c}_i^{(k_c)}\|_2 - \frac{\alpha_i^{(k_c)}}{2\rho}\} \mathbf{c}_i^{(k_c)}}{\|\mathbf{c}_i^{(k_c)}\|_2}, i = 1, \dots, N$
- 6 $\mathbf{R}_{\mathcal{C}_l}^{(k_c)+1} = \frac{1}{\mu^{(k_c)}} (\beta_2 \sum_{i \in \mathcal{C}_l} \mathbf{v}_i \mathbf{v}_i^H + \beta_3 \mathbf{L} \mathbf{B}_l), l = 1, \dots, N_c$
- 7 $\Lambda_{\mathbf{z}}^{(k_c+1)} = \Lambda_{\mathbf{z}}^{(k_c)} + \rho (\mathbf{X}^{(k_c+1)} - \mathbf{Z}^{(k_c+1)})$
- 8 $\Lambda_{\mathbf{v}}^{(k_c+1)} = \Lambda_{\mathbf{v}}^{(k_c)} + \rho (\mathbf{X}^{(k_c+1)} - \mathbf{V}^{(k_c+1)})$
- 9 **if** $(k_c \bmod K_c) = 0$ **then**
- 10 $\mathcal{S} = \bigcup_{j \in \mathcal{J}} \mathcal{C}_j, \{l \in \mathcal{J} : \exists i \in \mathcal{C}_l, \|\mathbf{x}_i\|_2 > p_c\}$
- 11 **while** $k_u < k_{u_{\text{max}}}$ **do**
- 12 Solve (111) using the similar update rules as steps 3–6, but using
- 13 $\mathbf{g}_i^{(k_u)} = (\|\mathbf{x}_i^{(k_u)} + \varepsilon_0\|)^{-1}$
- 14 $\Lambda_{\mathbf{z}_{\mathcal{J}}}^{(k_c+1)} = \Lambda_{\mathbf{z}_{\mathcal{J}}}^{(k_c)} + \rho (\mathbf{X}_{\mathcal{J}}^{(k_c+1)} - \mathbf{Z}_{\mathcal{J}}^{(k_c+1)})$
- 15 $\Lambda_{\mathbf{v}_{\mathcal{J}}}^{(k_c+1)} = \Lambda_{\mathbf{v}_{\mathcal{J}}}^{(k_c)} + \rho (\mathbf{X}_{\mathcal{J}}^{(k_c+1)} - \mathbf{V}_{\mathcal{J}}^{(k_c+1)})$
- 16 $k_u \leftarrow k_u + 1$
- 17 $\mathbf{X}_{\mathcal{S}}^{(k_c)} = \mathbf{X}_{\mathcal{S}}^{(k_u)}, \mathbf{Z}_{\mathcal{S}}^{(k_c)} = \mathbf{Z}_{\mathcal{S}}^{(k_u)}, \mathbf{V}_{\mathcal{S}}^{(k_c)} = \mathbf{V}_{\mathcal{S}}^{(k_u)}, \Lambda_{\mathbf{z}_{\mathcal{S}}}^{(k_c)} = \Lambda_{\mathbf{z}_{\mathcal{S}}}^{(k_u)}, \Lambda_{\mathbf{v}_{\mathcal{S}}}^{(k_c)} = \Lambda_{\mathbf{v}_{\mathcal{S}}}^{(k_u)}$
- 18 $k_u = 1,$
- 19 $k_c \leftarrow k_c + 1$

ADMM update rules

The details of the proposed algorithm in Section 4.3, termed corr-MAP-ADMM, are summarized in Algorithm 5. We note that owing to the *proposed splitting techniques* in (108), all the sub-problems in (109) are convex, and therefore can be solved analytically via *closed-form formulas*. Further, the optimization over \mathbf{X} , \mathbf{V} , and Σ is separable over the devices and the clusters, allowing *parallel updates*. Thus, the exact solution to (109) is given in steps 3-6 in Algorithm 2, and we follow the same analogy to solve (111) in the inner loop.

4.4 Simulation result

This section quantifies the performance and robustness of the proposed algorithms and compares them to existing sparse recovery algorithms in terms of active devices identification accuracy, channel estimation quality, and convergence behavior.

4.4.1 Simulation setup

We consider a network with a single BS serving $N = 200$ devices distributed equally over $N_c = 20$ clusters, with a total of active $K = 16$ at each coherence interval T_c . Each device is assigned a unique unit-norm pilot sequence drawn from an i.i.d. complex Bernoulli distribution. We set each $\mathbf{B}_l, \forall l$, as $\mathbf{B}_l = \zeta_B \Psi_l + (1 - \zeta_B) \frac{1}{L} \sum_{i \in \mathcal{C}_l} \mathbf{R}_i$, where Ψ_l is a random positive-definite Hermitian matrix to model the error in the prior knowledge of the covariance matrix $\mathbf{R}_{\mathcal{C}_l}$, whereas the parameter ζ_B controls the level of average mismatch between \mathbf{B}_l and $\mathbf{R}_i, \forall i \in \mathcal{C}_l$, and it is set as $\zeta_B = 0.1$.

The performance is evaluated in terms of NMSE, SRR, as well as the convergence rate. We compare the performances of the proposed algorithms against three algorithms that solve MMV sparse recovery problem, namely, 1) IRW- $\ell_{2,1}$, 2) MAP-ADMM, and 3) T-SBL [50]. Note that for MAP-ADMM and T-SBL, the second-order statistics are known at the BS. Furthermore, for an optimal MMSE benchmark, we consider the oracle minimum mean square error (MMSE) estimator.

The regularization parameters of the sparse recovery algorithms needs to be fine-tuned. For a fair comparison, all the parameters have been empirically tuned via cross-validation in advance and then fixed such that they provide their best performance.

4.4.2 Correlated activity pattern

First, we investigate the performance of the proposed EM-EP and corr-MAP-ADMM and compare it to the baseline algorithms. To this end, the 16 active devices are distributed over 2 active clusters, each with 8 active devices.

Effect of pilot length

First, we quantify the performance of the proposed algorithms in comparison to MAP-ADMM, T-SBL, and IRW- $\ell_{2,1}$ for different pilot sequence lengths with fixed SNR = 16 dB.

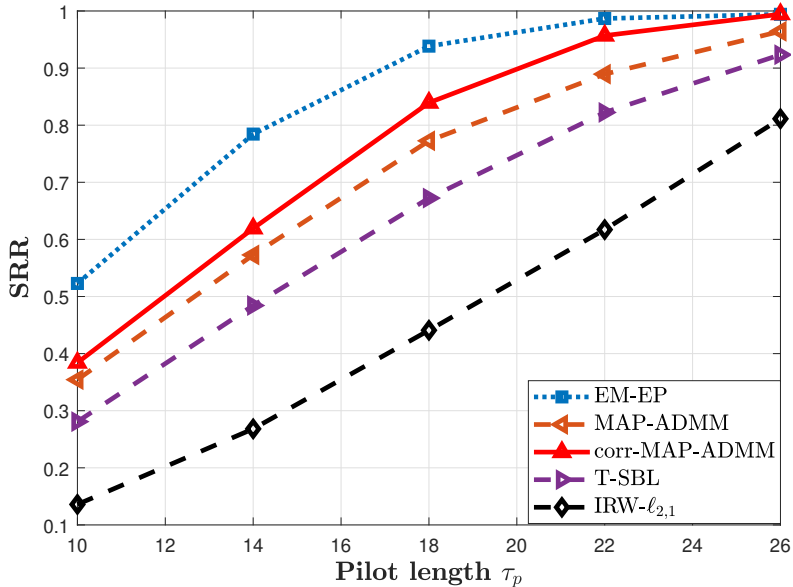


Fig. 20. Activity detection performance in correlated activity scenario in terms of SRR against pilot sequence length τ_p for SNR = 16 dB, and $M = 10$ dB.

Fig. 20 depicts the activity detection performance in terms of SRR. The results show clearly that the proposed EM-EP and corr-MAP-ADMM algorithms provide the best activity detection accuracy, as they achieve the highest SSR compared to all other algorithms. For instance, EM-EP achieves the same SRR as MAP-ADMM and T-SBL using 25% shorter pilot sequence length. Furthermore, the results show that the EM-EP provides a comparable gain over corr-MAP-ADMM in the low pilot length regime.

Fig. 21 depicts the channel estimation performance for the different recovery algorithms in terms of NMSE. Again, the obtained results highlight the remarkable gains in channel estimation accuracy obtained by using EM-EP, which significantly outperforms all other algorithms. Indeed, even with a 30% reduction in the pilot length used, EM-EP achieves the same performance as corr-MAP-ADMM, MAP-ADMM, and T-SBL. Moreover, the proposed corr-MAP-ADMM moderately outperforms T-SBL and provides almost similar performance to MAP-ADMM. These results can be explained by the fact that EM-EP, using the hierarchical spike-and-slab prior, efficiently captures the intrinsic structure of the sparsity model to improve the solution in terms of recovery rate and channel estimation.

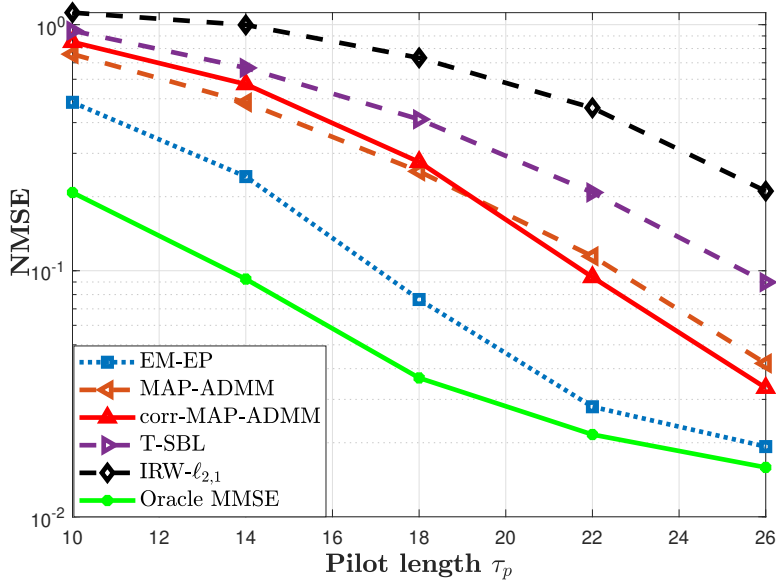


Fig. 21. Channel estimation performance in correlated activity scenario in terms of NMSE against pilot sequence length τ_p for SNR = 16 dB, and $M = 10$ dB.

Effect of SNR

In this section, we focus on analyzing the effects of SNR on the performance of various algorithms in activity detection and channel estimation. We specifically set the pilot sequence length to $\tau_p = 24$ for this part of the analysis.

Fig. 22 provides insights into the SRR achieved by different algorithms under various SNR conditions. Notably, the proposed algorithms, EM-EP and corr-MAP-ADMM, outperform the other algorithms in terms of activity detection accuracy. These algorithms exhibit a substantial SRR improvement, reaching an SRR = 0.9 when operated at an SNR of 8 dB. In contrast, the alternative algorithms T-SBL and MAP-ADMM necessitate a considerably higher SNR level—specifically, an SNR of 16 dB—to attain a comparable level of activity detection performance. This demonstrates the efficiency of our proposed algorithms in achieving accurate activity detection, even in lower SNR environments.

Further evaluation of channel estimation quality is shown in Fig. 23, which quantifies the performance in terms of the NMSE. Interestingly, the corr-MAP-ADMM algorithm shows performance metrics that are closely aligned with those of MAP-ADMM and T-SBL, implying that these algorithms are roughly on a par in this aspect of evaluation. However, it is crucial to point out the distinct advantage of the proposed EM-EP

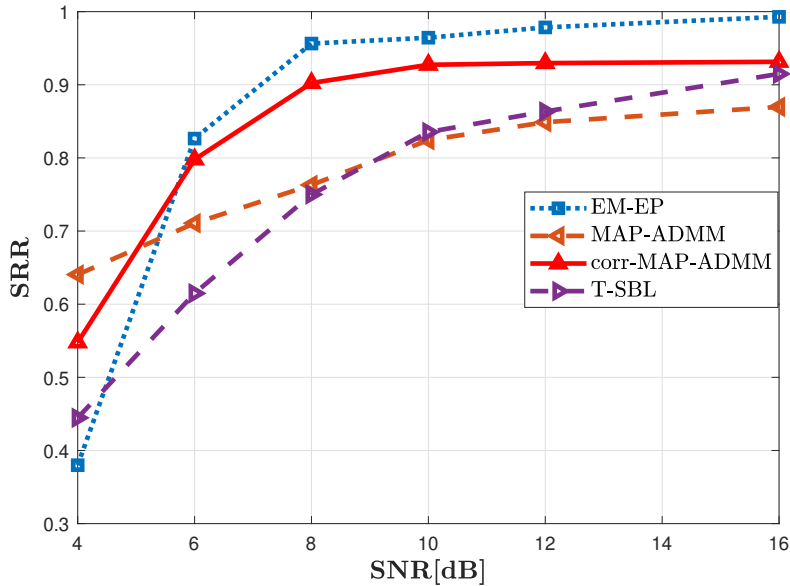


Fig. 22. Effect of the SNR on the activity detection accuracy in mMTC with correlated activity for $\tau_p = 24$, and $M = 10$ dB..

algorithm in terms of channel estimation quality. Specifically, EM-EP offers a noticeable gain of around 8 dB in comparison to both MAP-ADMM and T-SBL.

These observations collectively indicate that the proposed algorithms are particularly effective in scenarios with high levels of noise. Their superior performance becomes increasingly pronounced when compared to state-of-the-art algorithms under these challenging conditions.

Effect of the number of BS antennas

Fig. 24 illustrates the effect of the number of BS antennas (M) on the performance of the proposed algorithms in terms of NMSE. As expected, all the algorithms experience an improvement in the channel estimation quality as M increases. Notably, the EM-EP algorithm substantially outperforms the other sparse recovery algorithms, with remarkable gains in the lower range for BS antennas, i.e., $M < 12$. In addition, corr-MAP-ADMM provides the second best performance for $M < 12$. However, it is important to note that as M increases, the performance gap between all algorithms gradually decreases. This can be attributed to the fact that having more measurements

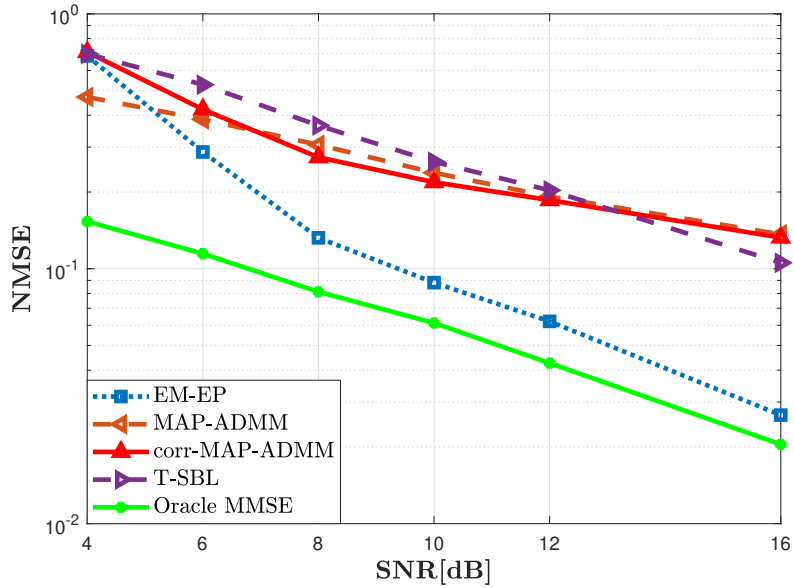


Fig. 23. Effect of the SNR on the channel estimation quality in mMTC with correlated activity for $\tau_p = 24$, and $M = 10$ dB.

from additional antennas would yield more gains compared to incorporating any side information, such as a correlated activity structure.

In summary, the results presented in this section clearly demonstrate the substantial gains achieved by incorporating the structured activity pattern when designing solutions for the JADCE problem. Indeed, even with no prior knowledge of the CDI, by carefully exploiting the correlated activity pattern, the proposed algorithms outperform state-of-the-art recovery algorithms that utilize prior knowledge of the full CDI.

4.4.3 Robustness to model mismatch

In this section, we investigate the performance of the proposed algorithms under a model mismatch, which refers to the discrepancies between the actual scenario and the model imposed by the algorithm. In our specific context, this could describe, for example, transmission instance where the activity is independent of all the devices in the network. A sparsity mismatch therefore occurs if the devices' activity is independent rather than exhibiting a correlated pattern. Next, we show that the proposed algorithms exhibit robustness in such scenarios and maintain good performance.

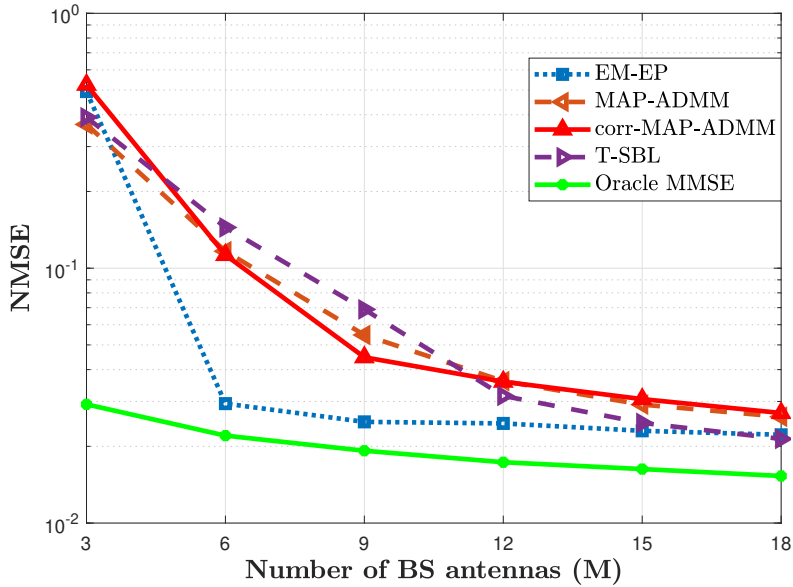


Fig. 24. JADCE performance in terms of NASE versus the number of BS antennas M for $\tau_p = 20$, and SNR = 15 dB.

We now investigate the robustness of the proposed algorithms in scenarios of sparsity model mismatch where discrepancies exist between the actual scenario and the imposed algorithm model. To this end, in this section, the assumption of the hierarchical activity is not met, and *the activity is independent of the devices*.

Fig. 25 illustrates the SRR against the pilot length for the different algorithms in *an independent activity pattern*. The results show that EM-EP and corr-MAP-ADMM clearly outperform IRW- $\ell_{2,1}$ and provide a slight improvement over T-SBL and match MAP-ADMM performance for $\tau_p > 20$.

Fig. 26 demonstrates that the proposed algorithms outperform IR- $\ell_{2,1}$, with EM-EP delivering inferior NMSE performance compared to MAP-ADMM but outperforming T-SBL. On the other hand, both corr-MAP-ADMM and T-SBL provide roughly the same NMSE performance.

The results shown in this section demonstrate that the proposed algorithms yield comparable results to state-of-the-art algorithms, even in *an activity pattern model mismatch and no prior knowledge of the CDI*. More precisely, even though the proposed *strong* hierarchical spike-and-slab (75) and the *weak* log-sum-based cluster-sparsity (104) priors were designed to favor clustered sparse solution, they are still flexible and *robust* to changes in the underlying sparsity structure.

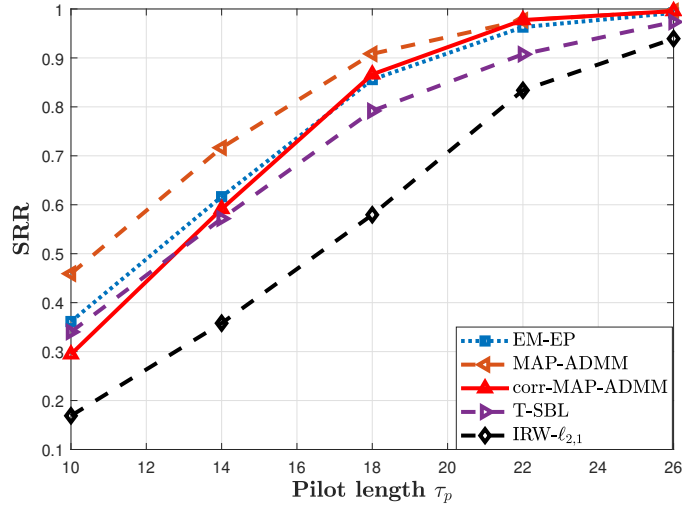


Fig. 25. Performance of the proposed algorithms in random activity pattern in terms of SRR against pilot sequence length.

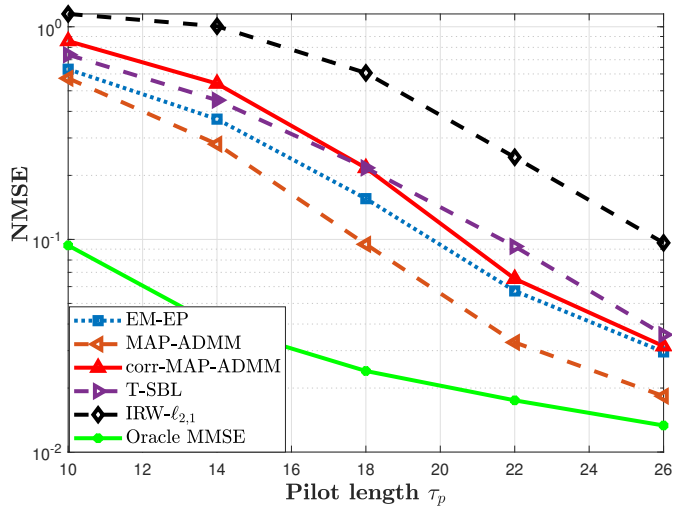


Fig. 26. Performance of the proposed algorithms in random activity pattern in terms of NMSE against pilot sequence length.

4.4.4 Convergence behaviour

The computational complexity per iteration of EM-EP is mainly dominated by the Σ update in (84) and requires $\mathcal{O}(M^3N^3)$ complex multiplications. In contrast, corr-MAP-ADMM's primary computational complexity arises from the \mathbf{V} -update and the \mathbf{R} -update, resulting in computational complexity in the order of $\mathcal{O}(NM^2 + N_cM^3)$. For reference, MAP-ADMM and T-SBL require $\mathcal{O}(MN^2 + NM^2)$ and $\mathcal{O}(N^2M^3\tau_p)$ complex multiplications respectively. Nonetheless, despite being more computationally demanding than to corr-MAP-ADMM, EM-EP exhibits a faster convergence rate, requiring only about 10 iterations to converge, in contrast to corr-MAP-ADMM which takes up to 60 iterations, as shown in Fig. 27.

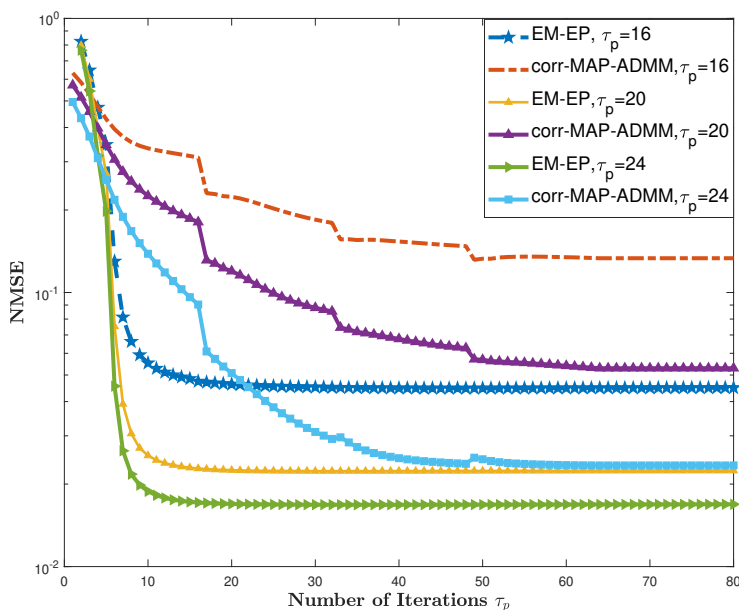


Fig. 27. NMSE versus number of iterations for the proposed algorithms.

4.5 Conclusion

This chapter proposed two Bayesian inference approaches to solve the JADCE problem with hierarchical devices activity pattern in spatially correlated MIMO channels with a limited knowledge of the CDI. We have developed two algorithms from a Bayesian

perspective by utilizing appropriate structured sparsity-promoting priors to capture both cluster-level and intra-cluster sparsity.

First, we proposed a *hierarchical spike-and-slab* prior that incorporated both the cluster-level and intra-cluster sparsity patterns. Subsequently, we derived an EP-based solution that iteratively approximates the intractable joint posterior distribution and provides an estimate on the active devices, their channels, and their CDI.

Second, we proposed an alternative solution to the JADCE problem by relaxing the spike-and-slab prior with a cluster-level log-sum sparsity-inducing prior that was able encode the hierarchical activity pattern. Subsequently, we formulated the JADCE as a MAP problem and derived a solution based on the ADMM framework to solve iteratively an approximated version of the MAP problem via a sequence of closed-form updates.

We numerically showed that the proposed algorithms provided a huge gain compared to state-of the art algorithms, as well as exhibiting a strong robustness against activity pattern model mismatch.

5 Activity detection performance via AMP

This chapter aims to provide a detailed theoretical analysis of the device activity detection problem in the uplink single-cell mMTC network in spatially correlated channels with known covariance matrices.

The existing theoretical work addressing the activity detection problem has primarily focused on the AMP algorithm. This focus arises from the AMP's state evolution analysis, which provides means of tracking several performance metrics at each iteration. For example, Chen *et al.* addressed the device activity detection problem in grant-free mMTC using AMP, deriving analytical performance metrics for the proposed AMP algorithm in both SMV and MMV setups [9]. Liu *et al.* extended this analysis and conducted an asymptotic performance analysis for activity detection, channel estimation, and achievable rate [85, 86]. However, these works have considered spatially-uncorrelated MIMO channels. However, it is of utmost importance to consider the channel's spatial correlation when evaluating the performance of a JADCE solution from a sparse recovery perspective [98].

Therefore, to bridge this gap in the literature, this chapter first extends the uncorrelated vector AMP [85] to its correlated vector AMP. Furthermore, it provides a detailed theoretical analysis of the activity detection performance of the AMP-based JADCE algorithm. This includes deriving closed-form expressions for the probabilities of both miss detection and false alarm.

5.1 System model and problem formulation

We consider a single-cell uplink mMTC network consisting of a set $\mathcal{N} = \{1, \dots, N\}$ of uniformly distributed single-antenna devices communicating with a BS equipped with a ULA containing M antennas. We consider a block fading channel response over each coherence period T_c . Let $\mathbf{h}_i \in \mathbb{C}^M$ denotes the channel response from the i th device to the BS.

Recall the local scattering model discussed in Chapter 3. The propagation channel $\mathbf{h}_i, i \in \mathcal{N}$ can be equivalently written as

$$\mathbf{h}_i = \mathbf{R}_i^{\frac{1}{2}} \bar{\mathbf{h}}_i, \quad \forall i \in \mathcal{N}, \quad (112)$$

where $\mathbf{R}_i = \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^H] \in \mathbb{C}^{M \times M}$ is the channel covariance matrix, and $\bar{\mathbf{h}}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Furthermore, $\{\mathbf{R}_i\}_{i=1}^N$ are assumed to be known to the BS [123].

Due to the sporadic traffic in mMTC, the activity state of each device is controlled by $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_N]^T \in \{0, 1\}^N \mathbb{B}$, where each γ_i can be statistically modeled as

$$\begin{cases} \Pr(\gamma_i = 1) = p_a \\ \Pr(\gamma_i = 0) = 1 - p_a, \end{cases} \quad \forall i \in \mathcal{N}. \quad (113)$$

Subsequently, we define the effective channel of the i th device as $\mathbf{x}_i = \gamma_i \mathbf{h}_i$ ¹¹. Thus, the effective channel \mathbf{x}_i can be modeled as a mixed Gaussian-Bernoulli distribution, given as

$$p(\mathbf{x}_i | p_a, \mathbf{R}_i) = (1 - p_a) \delta(\mathbf{x}_i) + p_a \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}_i). \quad (114)$$

According to (114):

- The effective channel for the i th device would be a zero-vector ($\mathbf{x}_i = \mathbf{0}$), with probability $(1 - p_a)$.
- The effective channel for the i th follows a complex Gaussian distribution, i.e., $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i)$ with probability p_a .

During the pilot transmission phase, each active device transmits its pilot sequence $\phi_i \in \mathbb{C}^{\tau_p}$ to the BS. We consider that the pilot sequences are drawn from a complex symmetric Bernoulli distribution. This choice is motivated by the fact that: 1) pilot sequences generated from a complex symmetric Bernoulli distribution are practical, as they can be deployed using QPSK modulation; 2) matrices drawn from a Bernoulli distribution are well-suited for AMP-based support and signal recovery [88, 136], as we will discuss later. The transmitted pilot are used to perform activity detection and channel estimation.

The received signal associated with the transmitted pilots at the BS, denoted by $\mathbf{Y} \in \mathbb{C}^{\tau_p \times M}$, is given by

$$\begin{aligned} \mathbf{Y} &= \sum_{i=1}^N \gamma_i \phi_i \mathbf{h}_i^T + \mathbf{W} \\ &= \Phi \mathbf{X}^T + \mathbf{W}. \end{aligned} \quad (115)$$

5.2 AMP for JADCE with spatially correlated channels

AMP is an iterative sparse recovery algorithm that was originally proposed in [51] for the general sparse recovery problem in an SMV setup and extended to an MMV setup in [29]. Subsequently, AMP has been deployed to solve the JADCE problem in [9, 85, 86, 88, 87]. In this paper, we adopt a Bayesian AMP originally proposed in [137]

¹¹Without any loss of generality, we drop herein the power control term for simplicity of presentation

for solving an MMV sparse recovery problem. This section provides a description of the design of AMP for solving JADCE in spatially correlated channels.

The AMP algorithm for sparse signal recovery from an MMV setup can be expressed by the following iterations [137]:

$$\hat{\mathbf{x}}_i^{(t+1)} = \eta(\tilde{\mathbf{Z}}^{(t)\text{T}} \Phi_i^* + \hat{\mathbf{x}}_i^{(t)}; \Sigma_\theta^{(t)}), \quad (116)$$

$$\tilde{\mathbf{Z}}^{(t+1)} = \mathbf{Y} - \Phi \mathbf{X}^{(t+1)\text{T}} + \frac{N}{\tau_p} \sum_{i=1}^N \frac{\eta'(\tilde{\mathbf{Z}}^{(t)\text{T}} \Phi_i^* + \mathbf{x}_i^{(t)})}{N}, \quad (117)$$

where $t = 1, 2, \dots$ is the iteration index, $\hat{\mathbf{X}}^t = [\hat{\mathbf{x}}_1^{(t)}, \dots, \hat{\mathbf{x}}_N^{(t)}]$ is the estimate of \mathbf{X} at iteration t , $\tilde{\mathbf{Z}}^{(t)} \in \mathbb{C}^{\tau_p \times M}$ is the residual matrix initialized with $\tilde{\mathbf{Z}}^{(1)} = \mathbf{Y}$, and $\Sigma_\theta^{(t)} \in \mathbb{C}^{M \times M}$ denotes a covariance matrix that can be tracked using the state evolution, as we discuss later. Function $\eta(\cdot)$ represents the denoising function that operates on each row of \mathbf{X}^T individually, and $\eta'(\cdot)$ is the first-order derivative of $\eta(\cdot)$. The third term in (117) is called the Onsager term, and it is the key component in determining the performance of AMP [51].

For the matrix Φ drawn from a Bernoulli distribution and under the assumption that $N, \tau_p \rightarrow \infty$ with a fixed ratio $\frac{\tau_p}{N}$, the term $\theta_i^{(t)} = \tilde{\mathbf{Z}}^{(t)\text{T}} \Phi_i^* + \mathbf{x}_i^{(t)}$, $\forall i \in \mathcal{N}$, is statistically equivalent to the sum of the true effective channel \mathbf{x}_i and a colored noise term $\mathbf{e}^{(t)} \sim \mathcal{CN}(\mathbf{0}, \Sigma_\theta^{(t)})$ as follows

$$\theta_i^{(t)} = \mathbf{x}_i + \mathbf{e}^{(t)}, \quad \forall i \in \mathcal{N}. \quad (118)$$

Given the linear signal model (118) and by exploiting the fact that the CDI is known to the BS, a minimum mean square error (MMSE) based denoiser function $\eta(\theta_i^{(t)}; \Sigma_\theta^{(t)})$ is calculated as

$$\begin{aligned} \eta(\theta_i^{(t)}; \Sigma_\theta^{(t)}) &= \mathbb{E}[\mathbf{x}_i | \theta_i^{(t)}] \\ &= \boldsymbol{\psi}(\theta_i^{(t)}; \Sigma_\theta^{(t)}) \mathbf{R}_i (\mathbf{R}_i + \Sigma_\theta^{(t)})^{-1} \theta_i^{(t)}, \forall i \in \mathcal{N}, \end{aligned} \quad (119)$$

such that

$$\boldsymbol{\psi}(\theta_i^{(t)}; \Sigma_\theta^{(t)}) = \left(1 + \frac{1 - p_a}{p_a} \exp(\mu_i^{(t)} - \theta_i^{(t)\text{H}} \Omega_i^{(t)} \theta_i^{(t)}) \right)^{-1}, \quad (120)$$

where $\Omega_i^{(t)} = \Sigma_\theta^{(t)-1} - (\mathbf{R}_i + \Sigma_\theta^{(t)})^{-1}$, and $\mu_i^{(t)} = \log \det(\mathbf{R}_i + \mathbf{I}_M)$. For more details on the derivation of MMSE denoiser function (119), please refer to Appendix 3.

The covariance matrix of the noise term $\Sigma_\theta^{(t)}$ can be tracked in the asymptotic regime via the state evolution. More precisely, the matrix $\Sigma_\theta^{(t)}$ is updated at each iteration t

using the following update rules [137]

$$\begin{aligned}\Sigma_{\theta}^{(1)} &= \sigma^2 \mathbf{I}_M + \mathbb{E}[\mathbf{X}\mathbf{X}^H] \\ \Sigma_{\theta}^{(t+1)} &= \sigma^2 \mathbf{I}_M + \frac{N}{\tau_p} \frac{1}{N} \sum_{i=1}^N (\psi_i^{(t)} - \psi_i^{(t)^2}) \mathbf{v}_i^{(t)} \mathbf{v}_i^{(t)H} \\ &\quad + \psi_i^{(t)} \Sigma_{\theta}^{(t)} \mathbf{R}_i (\mathbf{R}_i + \Sigma_{\theta}^{(t)})^{-1},\end{aligned}\tag{121}$$

where we $\psi_i^{(t)} = \psi(\theta_i^{(t)}; \Sigma_{\theta}^{(t)})$, and $\mathbf{v}_i^{(t)} = \mathbf{R}_i (\mathbf{R}_i + \Sigma_{\theta}^{(t)})^{-1} \theta_i^{(t)}$.

5.3 Activity detection performance

In this section, we derive closed-form expressions for the probabilities of miss detection and false alarm achieved by AMP in spatially correlated fading channels. The derivation hinges mainly on the equivalent system model (118) and the state evolution matrix (121). While the Gaussian system model in (118) holds in the asymptotic regime, it can provide a useful insight into the performance of the AMP for a practical mMTC setup where the number of connected devices N is large, yet finite.

5.3.1 Decision threshold

Here, we discuss the decision rule for the device activity detector on the AMP output. Let us examine the denoising function $\eta(\cdot)$ given in (119). Note that for any finite p_a , $\psi(\theta_i^{(t)}; \Sigma_{\theta}^{(t)}) \in [0, 1]$. Thus, a close examination shows that the denoising function $\eta(\cdot, \cdot)$ consists of an activity indicator estimate $\psi(\theta_i^{(t)}; \Sigma_{\theta}^{(t)}) \in [0, 1]$ and a conventional MMSE estimate term $\mathbf{R}_i (\mathbf{R}_i + \Sigma_{\theta}^{(t)})^{-1} \theta_i^{(t)}$, $\forall i \in \mathcal{N}$. Therefore, in order to set the decision rule for the device activity detector, one can use $\psi(\cdot; \cdot)$ to determine the activity status of each device. More precisely, if $\psi_i^{(t)} \rightarrow 1$, the i th device is declared active, and if $\psi_i^{(t)} \rightarrow 0$, the i th device is declared not active. While the activity detection performance can be characterized at each iteration (t), it is typically more interesting to discuss performance upon AMP convergence. We therefore omit the iteration index (t) in the following derivations for the sake of brevity.

In a practical scenario, one would use a vector of pre-defined threshold values $\zeta = [\zeta_1, \dots, \zeta_N]^T$, such that $\zeta_i \in [0, 1]$, $\forall i \in \mathcal{N}$. The activity detector will declare the i th device to be active if $\psi_i \geq \zeta_i$, and inactive otherwise. The values of ζ_i , $\forall i \in \mathcal{N}$, can be selected based on the cost of miss detection and the cost of a false alarm for each device. This paper proposes the following decision rule on the device's activity:

$$\hat{\gamma}_i = \begin{cases} 1, & \theta_i^H \Omega_i \theta_i \geq v_i = \mu_i - \log\left(\frac{p_a(1-\zeta_i)}{\zeta_i(1-p_a)}\right) \\ 0, & \theta_i^H \Omega_i \theta_i < v_i = \mu_i - \log\left(\frac{p_a(1-\zeta_i)}{\zeta_i(1-p_a)}\right) \end{cases}, \forall i \in \mathcal{N}.\tag{122}$$

The decision rule in (122) provides a threshold-based decision rule for AMP-based recovery at each iteration (t) by comparing $\psi_i^{(t)}$ to the predefined threshold value ζ_i .

5.3.2 Probabilities of miss detection and false alarm

The activity detection performance is quantified using two types of error probabilities: First, the probability of false alarm, which represents the probability of declaring an inactive device to be active; Second, the probability of miss detection, which represents the probability of declaring an active device as inactive.

The equivalent signal model (118) suggests that the term $\theta_i, \forall i \in \mathcal{N}$, follows a zero-mean complex Gaussian distribution, i.e., $\theta_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{S}_i)$. The covariance matrix \mathbf{S}_i depends on the value of the true γ_i , and it is given as

$$\mathbf{S}_i = \begin{cases} \Sigma_\theta + \mathbf{R}_i, & \gamma_i = 1 \\ \Sigma_\theta, & \gamma_i = 0 \end{cases}, \forall i \in \mathcal{N}, \quad (123)$$

$$\begin{aligned} p(\theta_i^t | \gamma_i \neq 0) &\sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i + \Sigma_\theta^t) \\ p(\theta_i^t | \gamma_i = 0) &\sim \mathcal{CN}(\mathbf{0}, \Sigma_\theta^t). \end{aligned} \quad (124)$$

To characterize the activity detection performance, we refer to (122) and define two complex quadratic Gaussian random variables, $Q_{1,i}$ and $Q_{0,i}, \forall i \in \mathcal{N}$, as

$$\begin{aligned} Q_{1,i} &= (\theta_i^H \Omega_i \theta_i | \gamma_i = 1) \\ Q_{0,i} &= (\theta_i^H \Omega_i \theta_i | \gamma_i = 0) \end{aligned}, \forall i \in \mathcal{N}. \quad (125)$$

Next, by using the decision rule in (122) and the Gaussian quadratic random variables in (125), the probabilities of a miss detection and a false alarm for each device i are defined respectively, as

$$P_i^{\text{MD}} = \Pr(\hat{\gamma}_i = 0 | \gamma_i = 1) = \Pr(Q_{1,i} \leq v_i), \forall i \in \mathcal{N}. \quad (126)$$

$$P_i^{\text{FA}} = \Pr(\hat{\gamma}_i = 1 | \gamma_i = 0) = \Pr(Q_{0,i} > v_i), \forall i \in \mathcal{N}. \quad (127)$$

To evaluate (126) and (127), we need to evaluate how to obtain the cumulative distribution function (CDF) $\Pr(Q_{1,i}^{(t)} \leq v_i)$.

Now, let us consider a general complex Gaussian quadratic form $Q = \theta^H \Omega \theta$ for a random variable $\theta \sim \mathcal{CN}(\mathbf{0}, \mathbf{S}) \in \mathbb{C}^M$. By using some algebraic transformations, Q can be expressed as a linear combination of chi-squared random variables, as we show next. First, we write θ as

$$Q = \bar{\mathbf{z}}^H \mathbf{S}^{\frac{1}{2}} \Omega \mathbf{S}^{\frac{1}{2}} \bar{\mathbf{z}}, \quad (128)$$

where $\boldsymbol{\theta} = \mathbf{S}^{\frac{1}{2}} \bar{\mathbf{z}}$ for $\bar{\mathbf{z}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Let \mathbf{U} and $\Lambda_{\mathbf{q}} = \text{diag}(\lambda_{q_1}, \dots, \lambda_{q_M})$ denote the eigenvectors and the eigenvalues respectively, associated with $\mathbf{S}^{\frac{1}{2}H} \boldsymbol{\Omega} \mathbf{S}^{\frac{1}{2}}$. Thus, we can further express Q as

$$Q = \bar{\mathbf{z}}^H \mathbf{U} \Lambda \mathbf{U}^H \bar{\mathbf{z}} = \bar{\mathbf{v}}^H \Lambda \bar{\mathbf{v}} = \sum_{m=1}^M \lambda_{q_m} |\bar{v}_m|^2, \quad (129)$$

where $\bar{\mathbf{v}} = [\bar{v}_1, \dots, \bar{v}_M]^T = \mathbf{U}^H \bar{\mathbf{z}}$. Note that since \mathbf{U} is a unitary matrix, $\bar{\mathbf{v}}$ follows the same distribution as $\bar{\mathbf{z}}$, i.e., $\bar{\mathbf{v}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. We can rewrite $\bar{v}_m = \frac{1}{\sqrt{2}}(a_m + jb_m)$, where $a_m, b_m \sim \mathcal{N}(0, 1)$, for $m = 1, \dots, M$. Therefore, Q can be finally expressed by a linear combination of zero-mean independent squared Gaussian random variables as

$$Q = \sum_{m=1}^M \frac{1}{2} \lambda_{q_m} (a_m^2 + b_m^2) = \sum_{m=1}^M \frac{1}{2} \lambda_{q_m} \hat{v}_m, \quad (130)$$

where $\hat{v}_m = a_m^2 + b_m^2$, $m = 1, \dots, M$. Note that Q can be viewed as a linear combination of M independent Chi-squared random variables $\hat{v}_m \sim \chi_2^2(1)$, with two degrees of freedom. Subsequently, a closed-form expression for the cumulative distribution function (CDF) of Q in (130) is given as [138, Sect. 4.3]

$$F_Q(v) = \Pr(Q \leq v) = \sum_{m=1}^M \bar{g}(\lambda_{q_m}; m) \frac{\bar{\gamma}(1, \frac{2v}{\lambda_{q_m}})}{\Gamma(1)}, \quad (131)$$

where

$$\bar{g}(\lambda_{q_m}; m) = \prod_{j=1}^M \left(\frac{-1}{\lambda_{q_j}} \right) \lambda_{q_m} \lim_{k \rightarrow \frac{1}{\lambda_{q_m}}} \left(\left(k - \frac{1}{\lambda_{q_m}} \right) \prod_{j=1}^M \left(k - \frac{1}{\lambda_{q_j}} \right)^{-1} \right), \quad (132)$$

with $\Gamma(\cdot)$ denoting the gamma function, and $\bar{\gamma}(\cdot; \cdot)$ denoting the lower incomplete gamma function.

Having provided a closed-form expression for the CDF of a general complex quadratic form Q given in (128), we now present Proposition 1, which characterizes the activity detection performance based on the state evolution matrix $\Sigma_{\boldsymbol{\theta}}$.

Proposition 1. *Consider device activity detection by the Bayesian AMP for mMTC in spatially correlated channels with finite $\frac{N}{\tau_p}$ and largely enough N and τ_p so that the equivalent signal model (118) holds. Using the CDF expression (131), the probabilities of miss detection and a false alarm per device are given respectively, as*

$$P_i^{\text{MD}} = \sum_{m=1}^M \bar{g}(\hat{\lambda}_i; m) \frac{\bar{\gamma}(1, 2v_i \hat{\lambda}_{i,m})}{\Gamma(1)} \quad (133)$$

$$P_i^{\text{FA}} = 1 - \sum_{m=1}^M \bar{g}(\bar{\lambda}_i; m) \frac{\gamma(1, 2\nu_i/\bar{\lambda}_{i,m})}{\Gamma(1)}, \quad (134)$$

where $\hat{\lambda}_i = [\hat{\lambda}_{i,1}, \dots, \hat{\lambda}_{i,M}]^T$ and $\bar{\lambda}_i = [\bar{\lambda}_{i,1}, \dots, \bar{\lambda}_{i,M}]^T$ are the eigenvalues associated with $(\mathbf{R}_i + \Sigma_\theta)^{\frac{1}{2}} \Omega_i (\mathbf{R}_i + \Sigma_\theta)^{\frac{1}{2}}$ and $\Sigma_\theta^{\frac{1}{2}} \Omega_i \Sigma_\theta^{\frac{1}{2}}$ respectively.

5.4 Simulation results

We consider a single cell with a radius of 100 m with a BS equipped with M antennas and surrounded by $N = 1000$ uniformly distributed single-antenna devices. We consider an activity level $p_a = 0.05$ at each T_c . The channel responses \mathbf{h}_i , for $i \in \mathcal{N}$, are generated using a local scattering model for the channel covariance matrices [100, Sect. 2.2]. To mitigate the channel gain differences between the devices, we assume that each device has a unit average channel gain, i.e., $\frac{1}{M} \text{Tr}(\mathbf{R}_i) = 1, \forall i \in \mathcal{N}$. Furthermore, each device is assigned a normalized QPSK sequence ϕ_i drawn from an i.i.d. complex Bernoulli distribution.

The activity detection performance is quantified in terms of the average miss detection $P^{\text{MD}} = \frac{1}{N} \sum_{i=1}^N P_i^{\text{MD}}$ and false alarm $P^{\text{FA}} = \frac{1}{N} \sum_{i=1}^N P_i^{\text{FA}}$ probabilities.

Fig. 28 and Fig. 29 compare the the simulated and predicted P^{MD} and P^{FA} . The simulated performance is obtained by running the AMP algorithm and deploying the decision rule (122) to detect the active devices, whereas the predicted performance is obtained from the analytical expressions (133) and (134).

Fig. 28 shows the activity detection performance versus the pilot sequence length τ_p with $M = 32$, and a signal-to-noise ratio (SNR) = 10 dB. The obtained results suggest that increasing τ_p results in continual improvement in the activity detection performance. More interestingly, the results show clearly that the probabilities of miss detection and false alarm derived in (133) and (134) match the simulation results obtained by AMP.

Fig. 29 shows the same activity detection performance but with respect to the number of BS antennas M . As expected, the activity detection performance improves significantly with increasing M . Furthermore, similar to the results in Fig. 28, the simulated results provide the same performance as the theoretical ones.

5.5 Conclusion

This chapter presented an AMP-based solution for the JADCE problem in mMTC in spatially correlated fading channels. By utilizing the properties of the state evolution of the AMP algorithm, we derived a closed-form expression for both the miss detection

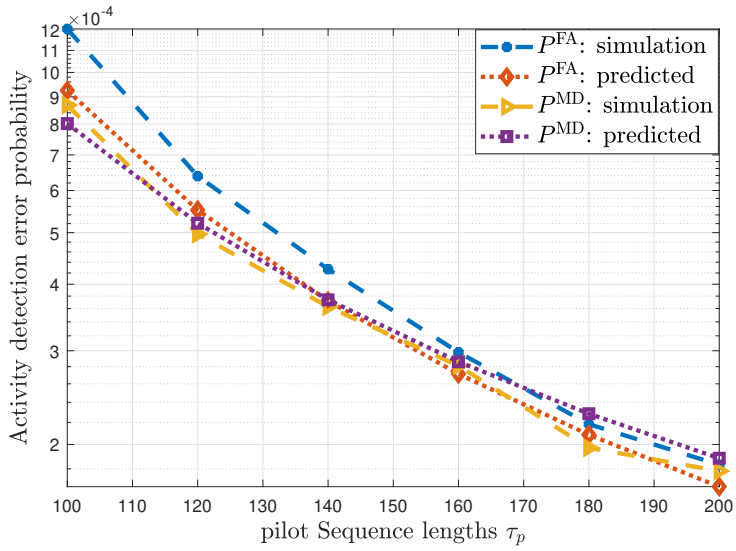


Fig. 28. Probabilities of miss detection and false alarm obtained by the AMP algorithm (simulation) and the derived theoretical results (predicted) versus: the pilot length τ_p . (Reprinted with permission from [108] ©2021, IEEE)

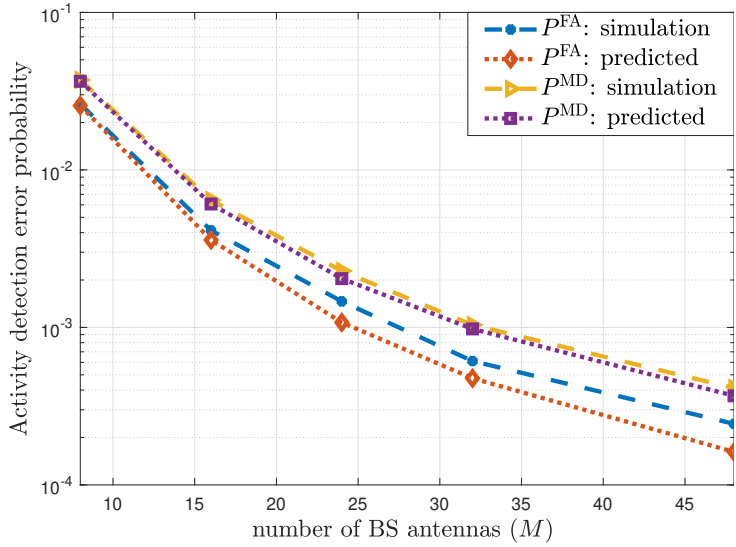


Fig. 29. Probabilities of miss detection and false alarm obtained by the AMP algorithm (simulation) and the derived theoretical results (predicted) versus the number of BS antennas M . (Reprinted with permission from [108] ©2021, IEEE)

and false alarm probabilities. The simulation experiments showed that the theoretical analysis provided in this paper matched the numerical results.

6 Conclusion and future work

6.1 Conclusions

This thesis considered the massive random access problem in mMTC networks, characterized by massive device connectivity, uplink-dominant traffic, sporadic device activity, and relaxed data rate requirements. These characteristics necessitate innovative solutions for efficient channel access. Grant-free access protocols have been highlighted as a pivotal factor in reducing signaling overhead, allowing devices to transmit data as required without conventional channel access schemes. On the other hand, grant-free access poses the challenge of detecting which device is active and estimating their channel coefficient. To enable reliable active device identification, each device is assigned a unique pilot sequence drawn from a large pool of non-orthogonal sequences, which serves as its unique identifier. In the random access phase, all the active devices transmit their pilot sequences simultaneously to the BS. The latter aims to identify the set of the active devices by detecting which sequences are transmitted. In addition, these pilot sequences are used to deploy pilot-based channel estimation.

This thesis focused on the problem of jointly identifying which devices are active and estimating their channel coefficients, a problem referred to as JADCE throughout the thesis. A distinction has been made from the rich body of work in the literature in this thesis by addressing the following crucial research areas: 1) The consideration of spatially correlated channel setups, which have a significant impact on the communication environment; and 2) the incorporation of correlated activity patterns among the devices, which further provide a more realistic assessment of practical mMTC networks. By examining these two fundamental aspects, we aimed to provide a comprehensive understanding of the challenges posed by massive random access and devise effective solutions for them.

The first part of this thesis focused on the JADCE problem in spatially correlated channels, considering two scenarios of the availability of prior information about the channel statistics at the BS. In the absence of side information, we employed deterministic mixed-norm minimization, specifically, the iterative reweighted $\ell_{2,1}$ -norm optimization. This approach excels in robustness and outperforms mainstream sparse recovery algorithms with similar prior knowledge. When channel statistics are available at the BS, the problem has been approached from a deterministic perspective, and we propose with a novel covariance matching penalty combined with $\ell_{2,1}$ -norm optimization. Additionally, from a Bayesian standpoint, we leveraged the available CDI to derive

a joint estimation and detection problem based on the MAP estimator. Furthermore, for each problem formulation, we proposed optimization decomposition techniques and the ADMM framework, enabling computationally efficient iterative solutions with closed-form update rules calculated via simple analytical formulas.

In the second part of the thesis, we employed a full Bayesian approach to tackle the JADCE problem under two critical constraints: 1) a correlated (structured) activity pattern, and 2) unknown second-order statistics of the channel. Overcoming these constraints requires well-defined and thoughtfully designed priors to achieve highly accurate solutions. To address these challenges, we introduced structured sparsity-promoting priors based on the spike-and-slab family. These priors effectively encode the correlated structure of device activity. Additionally, we incorporated a Wishart prior to model the unknown second-order statistics of the channel. Furthermore, we developed a comprehensive Bayesian framework by combining the EP algorithm with the EM framework. This framework aims to accomplish two primary objectives: detecting active devices and jointly estimating the channel coefficients and second-order channel statistics.

The third part of this thesis provided a theoretical evaluation of the activity detection problem, leveraging the AMP algorithm to analyze activity detection accuracy. More specifically, the thesis derived closed-form expressions for the probabilities of miss detection and false alarm. While maintaining brevity, the provided analysis ensured a thorough comprehension of the AMP algorithm's operation and its significance in the context of activity detection.

The key results obtained throughout the thesis demonstrate the following:

- Incorporating spatial correlation into the design of the JADCE solution holds paramount importance within mMTC networks. Acknowledging and accommodating this correlation in the design of JADCE solutions is not only pivotal but also highly advantageous, as it leads to enhanced accuracy in activity detection and elevates the quality of channel estimation.
- Having prior information about spatial correlation can lead to substantial improvements across various performance metrics. Thus, if the BS possesses prior knowledge about the spatial correlation, it can leverage this information to improve and optimize the JADCE solutions. This optimization translates into superior system performance, increased reliability, and enhanced resource utilization.
- A comparison between Bayesian and deterministic approaches, particularly in scenarios lacking prior spatial correlation information, reveals that the Bayesian approach significantly outperforms the deterministic approach across various performance metrics. However, it is important to note that these gains come with the caveat of

increased computational complexity. While Bayesian approaches excel in accuracy, they require more computational resources. This trade-off between performance and complexity must be carefully considered when selecting an approach for JADCE in such scenarios.

6.2 Future work

While the JADCE problem has garnered significant attention in the literature, there remain more challenging issues that warrant further investigation. Drawing on the results presented in this thesis, as well as insights from the existing literature, several possible research directions emerge, as outlined below.

First, extending the discussion to cell-free MIMO networks, which rely on distributed access points instead of traditional cell-based architectures, would be highly beneficial. Indeed, for cell-free MIMO network, the JADCE solution can be effectively designed to leverage both the spatial diversity and cooperative processing. For example, the EP-EM solution can be made more efficient by distributing the computational load among the different access points in the network.

Second, more realistic models of mMTC traffic should be incorporated into the design of the JADCE solution. For instance, comprehensive characterization of the mMTC traffic pattern would include a statistical modeling of the traffic generation of the mMTC devices to account for packet size, packet inter-arrival time, and the distribution of packet arrivals. Incorporating these realistic models into the design of the JADCE solution serves two crucial purposes: 1) It allows for the design of solutions that are better suited to the dynamic and sporadic nature of mMTC traffic, ultimately leading to more effective and reliable communications systems; 2) It ensures that performance assessment is closely aligned with the behavior of practical mMTC networks.

Finally, one assumption that has been made throughout the thesis is the synchronized transmission from the devices. However, achieving such synchronization is not always guaranteed without deploying additional synchronization signaling overhead. Therefore, the JADCE problem in realistic scenarios must consider the possible asynchronous transmission from the devices. When addressing the detection of device activity in situations where received signature sequences from different devices lack synchronization, a common approach is to detect both the transmitted sequences and their respective delays. This is done by expanding the sequence pool to include all possible time-shifted versions. However, it is important to note that this approach comes with the drawback of potential performance degradation due to increased complexity in delay estimation. Additionally, computational complexity escalates significantly due to the larger sequence pool.

References

- [1] N. H. Motlagh, M. Bagaa, and T. Taleb, "UAV-based IoT platform: A crowd surveillance use case," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 128–134, 2017.
- [2] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiquzzaman, "Privacyprotector: Privacy-protected patient data collection in IoT-based healthcare systems," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 163–168, 2018.
- [3] J. Cheng, W. Chen, F. Tao, and C.-L. Lin, "Industrial IoT in 5G environment towards smart manufacturing," *Journal of Industrial Information Integration*, vol. 10, pp. 10–19, 2018.
- [4] W. Ejaz, M. Naeem, A. Shahid, A. Anpalagan, and M. Jo, "Efficient energy management for the internet of things in smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 84–91, 2017.
- [5] S. Ericsson, "Ericsson mobility report: June 2023," in *Ericsson Mobility Report*. Ericsson, 2023. [Online]. Available: <https://www.ericsson.com/en/reports-and-papers/mobility-report/reports>
- [6] P. Spachos and K. Plataniotis, "BLE beacons in the smart city: Applications, challenges, and research opportunities," *IEEE Internet Things. Mag.*, vol. 3, no. 1, pp. 14–18, 2020.
- [7] K.-H. Chang, "Bluetooth: a viable solution for IoT? [industry perspectives]," *IEEE Wirel. Commun.*, vol. 21, no. 6, pp. 6–7, 2014.
- [8] A. Sun, W. Gong, R. Shea, and J. Liu, "A castle of glass: Leaky IoT appliances in modern smart homes," *IEEE Wirel. Commun.*, vol. 25, no. 6, pp. 32–37, 2018.
- [9] Z. Chen, F. Sahrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Processing*, vol. 66, no. 7, pp. 1890–1904, 2018.
- [10] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, 2016.
- [11] A. Rico-Alvarino, M. Vajapeyam, H. Xu, X. Wang, Y. Blankenship, J. Bergman, T. Tirronen, and E. Yavuz, "An overview of 3GPP enhancements on machine to machine communications," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 14–21, 2016.
- [12] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Select. Areas Commun.*, vol. 39, no. 3, pp. 615–637, 2020.
- [13] A. Hoglund, X. Lin, O. Liberg, A. Behravan, E. A. Yavuz, M. Van Der Zee, Y. Sui, T. Tirronen, A. Ratilainen, and D. Eriksson, "Overview of 3GPP release 14 enhanced NB-IoT," *IEEE Network*, vol. 31, no. 6, pp. 16–22, 2017.
- [14] J. Navarro-Ortiz, P. Romero-Díaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 2, pp. 905–929, 2020.
- [15] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, 2019.
- [16] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the standardization of grant-free operation and the associated NOMA strategies in 3GPP," *IEEE Commun. Stand. Mag.*, vol. 3, no. 4, pp. 60–66, 2019.
- [17] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-orthogonal multiple access: Common myths and critical questions," *IEEE Wirel. Commun.*, vol. 26, no. 5, pp. 174–180, 2019.

- [18] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys & Tutorials*, 2020.
- [19] H. Seo, J.-P. Hong, and W. Choi, "Low latency random access for sporadic MTC devices in internet of things," *IEEE Internet of Things J.*, vol. 6, no. 3, pp. 5108–5118, 2019.
- [20] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [21] Z. Zhang, X. Wang, Y. Zhang, and Y. Chen, "Grant-free rateless multiple access: A novel massive access scheme for internet of things," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2019–2022, 2016.
- [22] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. De Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Processing Mag.*, vol. 35, no. 5, pp. 88–99, 2018.
- [23] J. Ding, M. Nemati, S. R. Pookhrel, O.-S. Park, J. Choi, and F. Adachi, "Enabling grant-free URLLC: An overview of principle and enhancements by massive MIMO," *IEEE Internet of Things J.*, vol. 9, no. 1, pp. 384–400, 2022.
- [24] E. J. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [25] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [26] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.
- [27] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing." 2012.
- [28] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $l_{2,1}$ mixed-norm minimization problem," *IEEE Trans. Signal Processing*, vol. 66, no. 6, pp. 1483–1497, 2018.
- [29] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Processing*, vol. 61, no. 2, pp. 340–354, 2012.
- [30] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.
- [31] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part I: Greedy pursuit," *Signal processing*, vol. 86, no. 3, pp. 572–588, 2006.
- [32] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [33] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
- [34] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [35] G. Swirszcz, N. Abe, and A. C. Lozano, "Grouped orthogonal matching pursuit for variable selection and prediction," in *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., vol. 22. Curran Associates, Inc., 2009.

- [36] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [37] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Royal Stat. Soc – Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [38] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci.*, vol. 102, no. 27, pp. 9446–9451, 2005.
- [39] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [40] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
- [41] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Lett.*, vol. 14, no. 10, pp. 707–710, 2007.
- [42] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. IEEE, 2008, pp. 3869–3872.
- [43] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Stat. Soc – Series B*, vol. 68, no. 1, pp. 49–67, 2006.
- [44] M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 2009.
- [45] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2015.
- [46] Z. Zhang and B. D. Rao, "Iterative reweighted algorithms for sparse signal recovery with temporally correlated source vectors," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 3932–3935.
- [47] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [48] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.
- [49] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J Mach Learn Res*, vol. 1, no. Jun, pp. 211–244, 2001.
- [50] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.
- [51] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [52] —, "Message passing algorithms for compressed sensing: Ii. analysis and validation," in *Proc. IEEE Inform. Theory Workshop*, 2010, pp. 1–5.
- [53] T. Goldstein, C. Studer, and R. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," *arXiv preprint arXiv:1411.3406*, 2014.
- [54] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Inform. Theory Workshop*, 2010, pp. 1–5.
- [55] J. P. Vila and P. Schniter, "Expectation-maximization gaussian-mixture approximate message passing," *IEEE Trans. Signal Processing*, vol. 61, no. 19, pp. 4658–4672, 2013.
- [56] C. Rush and R. Venkataramanan, "Finite-sample analysis of approximate message passing," in *Proc. IEEE Int. Symp. Inform. Theory*, 2016, pp. 755–759.
- [57] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. Inform. Theory*, vol. 65, no. 9, pp. 5339–5351, 2019.

- [58] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Acc.*, vol. 5, pp. 2020–2033, 2017.
- [59] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inform. Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.
- [60] S. Haghhighatshoar, P. Jung, and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inform. Theory*, 2018, pp. 381–385.
- [61] A. Fengler, S. Haghhighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inform. Theory*, vol. 67, no. 5, pp. 2925–2951, 2021.
- [62] Z. Chen, F. Sotiraki, Y.-F. Liu, and W. Yu, "Covariance based joint activity and data detection for massive random access with massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, 2019, pp. 1–6.
- [63] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," in *Proc. Int. Symp. Wireless Commun. Systems*, 2013, pp. 1–5.
- [64] C. Bockelmann, H. F. Schepker, and A. Dekorsy, "Compressive sensing based multi-user detection for machine-to-machine communication," *European Trans. Telecommun.*, vol. 24, no. 4, pp. 389–400, 2013.
- [65] Y. Du, C. Cheng, B. Dong, Z. Chen, X. Wang, J. Fang, and S. Li, "Block-sparsity-based multiuser detection for uplink grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7894–7909, 2018.
- [66] W. Zhang, J. Li, X. Zhang, and S. Zhou, "A joint user activity detection and channel estimation scheme for packet-asynchronous grant-free access," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 338–342, 2022.
- [67] X. Miao, D. Guo, and X. Li, "Grant-free NOMA with device activity learning using long short-term memory," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 981–984, 2020.
- [68] B. Wang, L. Dai, T. Mir, and Z. Wang, "Joint user activity and data detection based on structured compressive sensing for NOMA," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1473–1476, 2016.
- [69] Q. He, T. Q. Quek, Z. Chen, Q. Zhang, and S. Li, "Compressive channel estimation and multi-user detection in C-RAN with low-complexity methods," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3931–3944, 2018.
- [70] Y. Jiang, J. Su, Y. Shi, and B. Houska, "Distributed optimization for massive connectivity," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1412–1416, 2020.
- [71] Q. Lin, Y. Li, and Y.-C. Wu, "Sparsity constrained joint activity and data detection for massive access: A difference-of-norms penalty framework," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1480–1494, 2023.
- [72] P. Gao, Z. Liu, P. Xiao, C. H. Foh, and J. Zhang, "Low-complexity block coordinate descend based multiuser detection for uplink grant-free NOMA," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9532–9543, 2022.
- [73] Y. Li, M. Xia, and Y.-C. Wu, "Activity detection for massive connectivity under frequency offsets via first-order algorithms," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1988–2002, 2019.
- [74] J. Yuan, Q. He, M. Matthaiou, T. Q. Quek, and S. Jin, "Toward massive connectivity for IoT in mixed-ADC distributed massive MIMO," *IEEE Internet of Things J.*, vol. 7, no. 3, pp. 1841–1856, 2019.
- [75] L. Zhu, K.-H. Liu, L. Wan, and L. Sun, "Active user detection and channel estimation via fast ADMM," in *Proc. IEEE Wireless Commun. and Networking Conf.*, 2023, pp. 1–6.

- [76] T. Jiang, Y. Shi, J. Zhang, and K. B. Letaief, “Joint activity detection and channel estimation for IoT networks: Phase transition and computation-estimation tradeoff,” *IEEE Internet of Things J.*, vol. 6, no. 4, pp. 6212–6225, 2019.
- [77] Z. Wang, Z. Chen, Y.-F. Liu, F. Sohrab, and W. Yu, “An efficient active set algorithm for covariance based joint data and activity detection for massive random access with massive MIMO,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing.* IEEE, 2021, pp. 4840–4844.
- [78] J. Dong, J. Zhang, Y. Shi, and J. H. Wang, “Faster activity and data detection in massive random access: A multiarmed bandit approach,” *IEEE Internet of Things J.*, vol. 9, no. 15, pp. 13 664–13 678, 2022.
- [79] X. Shao, X. Chen, D. W. K. Ng, C. Zhong, and Z. Zhang, “Covariance-based cooperative activity detection for massive grant-free random access,” in *Proc. IEEE Global Telecommun. Conf.*, 2020, pp. 1–6.
- [80] Z. Wang, Y.-F. Liu, Z. Chen, and W. Yu, “Accelerating coordinate descent via active set selection for device activity detection for multi-cell massive random access,” in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.*, 2021, pp. 366–370.
- [81] D. Jiang and Y. Cui, “ML and MAP device activity detections for grant-free massive access in multi-cell networks,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3893–3908, 2022.
- [82] X. Xu, X. Rao, and V. K. Lau, “Active user detection and channel estimation in uplink cran systems,” in *Proc. IEEE Int. Conf. Commun.* IEEE, 2015, pp. 2727–2732.
- [83] X. Zhang, Y.-C. Liang, and J. Fang, “Novel Bayesian inference algorithms for multiuser detection in M2M communications,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7833–7848, 2017.
- [84] J. Fang, Y. Shen, H. Li, and P. Wang, “Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals,” *IEEE Trans. Signal Processing*, vol. 63, no. 2, pp. 360–372, 2015.
- [85] L. Liu and W. Yu, “Massive connectivity with massive MIMO—part I: Device activity detection and channel estimation,” *IEEE Trans. Signal Processing*, vol. 66, no. 11, pp. 2933–2946, 2018.
- [86] —, “Massive connectivity with massive MIMO—part II: Achievable rate characterization,” *IEEE Trans. Signal Processing*, vol. 66, no. 11, pp. 2947–2959, 2018.
- [87] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, “Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO,” *IEEE Trans. Signal Processing*, vol. 68, pp. 764–779, 2020.
- [88] K. Senel and E. G. Larsson, “Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach,” *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, 2018.
- [89] S. Xia, Y. Shi, Y. Zhou, and X. Yuan, “Reconfigurable intelligent surface for massive connectivity: Joint activity detection and channel estimation,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 5693–5707, 2021.
- [90] S. Jiang, J. Dang, Z. Zhang, L. Wu, B. Zhu, and L. Wang, “EM-AMP-based joint active user detection and channel estimation in cell-free system,” *IEEE Sensors J.*, vol. 17, no. 3, pp. 4026–4037, 2023.
- [91] S. Li, W. Zhang, Y. Cui, H. V. Cheng, and W. Yu, “Joint design of measurement matrix and sparse support recovery method via deep auto-encoder,” *IEEE Signal Processing Lett.*, vol. 26, no. 12, pp. 1778–1782, 2019.

- [92] W. Zhang, S. Li, and Y. Cui, “Jointly sparse support recovery via deep auto-encoder with applications in MIMO-based grant-free random access for mMTC,” in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.*, 2020, pp. 1–5.
- [93] Z. Mao, X. Liu, M. Peng, Z. Chen, and G. Wei, “Joint channel estimation and active-user detection for massive access in internet of things—a deep learning approach,” *IEEE Internet of Things J.*, vol. 9, no. 4, pp. 2870–2881, 2021.
- [94] Y. Qiang, X. Shao, and X. Chen, “A model-driven deep learning algorithm for joint activity detection and channel estimation,” *IEEE Wireless Commun. Lett.*, vol. 24, no. 11, pp. 2508–2512, 2020.
- [95] X. Shao, X. Chen, Y. Qiang, C. Zhong, and Z. Zhang, “Feature-aided adaptive-tuning deep learning for massive device detection,” *IEEE J. Select. Areas Commun.*, vol. 39, no. 7, pp. 1899–1914, 2021.
- [96] H. Yu, Z. Fei, Z. Zheng, N. Ye, and Z. Han, “Deep learning-based user activity detection and channel estimation in grant-free NOMA,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2202–2214, 2023.
- [97] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, “DNN-aided block sparse bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12 000–12 012, 2019.
- [98] Y. Cui, S. Li, and W. Zhang, “Jointly sparse signal recovery and support recovery via deep learning with applications in MIMO-based grant-free random access,” *IEEE J. Select. Areas Commun.*, vol. 39, no. 3, pp. 788–803, 2021.
- [99] L. Sanguinetti, E. Björnson, and J. Hoydis, “Towards massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination,” *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, 2020.
- [100] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017. [Online]. Available: <http://dx.doi.org/10.1561/20000000093>
- [101] H. Djelouat, M. Leinonen, L. Ribeiro, and M. Juntti, “Joint user identification and channel estimation via exploiting spatial channel covariance in mMTC,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 887–891, 2021.
- [102] H. Djelouat, M. Leinonen, and M. Juntti, “Iterative reweighted algorithms for joint user identification and channel estimation in spatially correlated massive MTC,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2021, pp. 4805–4809.
- [103] —, “Spatial correlation aware compressed sensing for user activity detection and channel estimation in massive MTC,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6402–6416, 2022.
- [104] —, “Joint estimation of clustered user activity and correlated channels with unknown covariance in mMTC,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.
- [105] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.
- [106] T. P. Minka, “Expectation propagation for approximate Bayesian inference,” in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI’01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, p. 362–369.
- [107] E. J. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J. FOURIER Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

- [108] H. Djelouat, L. Marata, M. Leinonen, H. Alves, and M. Juntti, "User activity detection and channel estimation of spatially correlated channels via AMP in massive MTC," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1200–1204.
- [109] H. Djelouat, M. J. Sillanpää, M. Leinonen, and M. Juntti, "Hierarchical MTC user activity detection and channel estimation with unknown spatial covariance," *arXiv preprint arXiv:2310.10204*, 2023.
- [110] H. Djelouat, M. J. Sillanpää, and M. Juntti, "Device detection and channel estimation in MTC with correlated activity pattern," *arXiv preprint arXiv:2310.14578*, 2023.
- [111] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, 2016, pp. 974–978.
- [112] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast MIMO channels: Curse or blessing?" in *Proc. IEEE Global Telecommun. Conf.*, 2008, pp. 1–5.
- [113] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 3352–3366, 06 2015.
- [114] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 1, pp. 4–16, 2013.
- [115] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 1985–2001, 2011.
- [116] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.
- [117] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [118] B. K. Sriperumbudur, D. A. Torres, and G. R. Lanckriet, "A majorization-minimization approach to the sparse generalized eigenvalue problem," *Mach. Learn.*, vol. 85, no. 1-2, pp. 3–39, 2011.
- [119] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," in *Wavelets and Sparsity XV*, vol. 8858. International Society for Optics and Photonics, 2013, p. 88580R.
- [120] H. Lu, X. Long, and J. Lv, "A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods," in *Int. Conf. Artif. Intell. Statist.*, 2011, pp. 461–469.
- [121] A. C. Cirik, N. M. Balasubramanya, and L. Lampe, "Multi-user detection using ADMM-based compressive sensing for uplink grant-free NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 46–49, 2017.
- [122] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing," *SIAM J. Imaging Sci.*, vol. 1, no. 1, pp. 143–168, 2008.
- [123] Y. Cheng, L. Liu, and L. Ping, "Orthogonal AMP for massive access in channels with spatial and temporal correlations," *IEEE J. Select. Areas Commun.*, vol. 39, no. 3, pp. 726–740, 2021.
- [124] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.

- [125] A. Rajoriya and R. Budhiraja, "Joint AMP-SBL algorithms for device activity detection and channel estimation in massive MIMO mMTC systems," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2136–2152, 2023.
- [126] J. Bai and E. G. Larsson, "Activity detection in distributed MIMO: Distributed AMP via likelihood ratio fusion," *IEEE Commun. Lett.*, vol. 11, no. 10, pp. 2200–2204, 2022.
- [127] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Processing*, vol. 19, no. 1, pp. 53–63, 2009.
- [128] J.-C. Jiang, H.-M. Wang, and H. V. Poor, "Performance analysis of joint active user detection and channel estimation for massive connectivity," *IEEE Trans. Signal Processing*, vol. 70, pp. 3647–3662, 2022.
- [129] W. Liu, Y. Cui, F. Yang, L. Ding, and J. Sun, "Joint optimization of preamble selection and access barring for random access in MTC with general device activities," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3930–3946, 2022.
- [130] E. Becirovic, E. Björnson, and E. G. Larsson, "Activity detection in distributed massive MIMO with pilot-hopping and activity correlation," *IEEE Commun. Lett.*, vol. 12, no. 2, pp. 272–276, 2023.
- [131] L. Chetot, M. Egan, and J.-M. Gorce, "Joint identification and channel estimation for fault detection in industrial IoT with correlated sensors," *IEEE Acc.*, vol. 9, pp. 116 692–116 701, 2021.
- [132] L. Marata, O. Luis Alcaraz López, H. Djelouat, M. Leinonen, H. Alves, and M. Juntti, "Joint coherent and non-coherent detection and decoding techniques for heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1730–1744, 2023.
- [133] M. R. Andersen, O. Winther, and L. K. Hansen, "Bayesian inference for structured spike and slab priors," *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [134] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Acc.*, vol. 6, pp. 47 682–47 698, 2018.
- [135] A. Vehtari, A. Gelman, T. Sivula, P. Jylänki, D. Tran, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich, and C. P. Robert, "Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data," *J Mach Learn Res*, vol. 21, no. 1, pp. 577–629, 2020.
- [136] P. Schniter, "A simple derivation of AMP and its state evolution via first-order cancellation," *IEEE J. Select. Topics Signal Processing*, vol. 68, pp. 4283–4292, 2020.
- [137] J. Kim, W. Chang, B. Jung, D. Baron, and J. C. Ye, "Belief propagation for joint sparse recovery," *arXiv preprint arXiv:1102.3289*, 2011.
- [138] A. M. Mathai and S. B. Provost, *Quadratic forms in random variables: theory and applications*. Dekker, 1992.
- [139] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

Appendix 1 Joint MMSE Estimator

The received signal in (12) can be rewritten as

$$\mathbf{y} = \Theta_{\mathcal{A}} \mathbf{x}_{\mathcal{A}} + \mathbf{w}, \quad (135)$$

where $\mathbf{y} = \text{vec}(\mathbf{Y}^T) \in \mathbb{C}^{\tau_p M}$, $\mathbf{w} = \text{vec}(\mathbf{W}^T) \in \mathbb{C}^{\tau_p M}$, and $\Theta_{\mathcal{A}} = \Phi_{\mathcal{A}} \otimes \mathbf{I}_M \in \mathbb{C}^{M \tau_p \times KM}$. The vectorization in (135) transforms the matrix estimation into a classical vector estimation. Thus, we utilize the MMSE estimator [139] to *jointly* estimate the channels of the active devices, as

$$\mathbf{x}^{\text{J-MMSE}} = \text{vec}(\mathbf{X}^{\text{J-MMSE}}) = \bar{\mathbf{x}} + (\mathbf{R}_{\text{diag}} \Theta_{\mathcal{A}}^H \Theta_{\mathcal{A}} \mathbf{R}_{\text{diag}} \Theta_{\mathcal{A}}^H + \sigma^2 \mathbf{I}_{\tau_p M})^{-1} (\mathbf{y} - \Theta_{\mathcal{A}} \bar{\mathbf{x}}), \quad (136)$$

where $\bar{\mathbf{x}}$ denotes the mean of \mathbf{x} , and \mathbf{R}_{diag} denotes the covariance matrix of $\mathbf{x}_{\mathcal{A}}$ given as a block diagonal matrix whose main diagonal blocks are given by the scaled covariance matrices $\tilde{\mathbf{R}}_i$ corresponding to the active devices $i \in \mathcal{A}$.

Appendix 2 EP derivation

2.1 Product and quotient rules

This section derives the product and quotient rules of two Gaussian random vectors θ_1 and θ_2 . These rules are useful for the derivation of EP updates, as well as the AMP denoiser function, as we will see later.

First, let us denote the two Gaussian vectors $\mathbf{a} \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $b \sim \mathcal{N}(\mu_2, \Sigma_2)$. This give us the following two results

Product

Let the vector θ_p denotes the random variable obtained by $\theta_p = \mathbf{a}b$. The PDF of θ_p follows

$$\mathcal{N}(\mathbf{a}; \mu_1, \Sigma_1) \mathcal{N}(b; \mu_2, \Sigma_2) = K_p \mathcal{N}(\theta_p; \mu_p, \Sigma_p), \quad (137)$$

where $\Sigma_p = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$, $\mu_p = \Sigma^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$, and $K_p = \mathcal{N}(\mu_1; \mu_2, \Sigma_2 + \Sigma_1)$.

Fraction

Let the vector θ_q denotes the random variable obtained by $\theta_q = \frac{\mathbf{a}}{b}$. The PDF of θ_q follows

$$\frac{\mathcal{N}(\theta_1; \mu_1, \Sigma_1)}{\mathcal{N}(\theta_2; \mu_2, \Sigma_2)} = K_q \mathcal{N}(\theta_q; \mu_q, \Sigma_q), \quad (138)$$

where $\Sigma_q = (\Sigma_1^{-1} - \Sigma_2^{-1})^{-1}$, $\mu = \Sigma^{-1}(\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2)$, and $K_q = \frac{|\Sigma_2 \Sigma_1^{-1}|}{\mathcal{N}(\mu_1; \mu_2, \Sigma_2 - \Sigma_1)}$.

Thus, by using (137) and (138) and few simple manipulations, we straightforwardly obtain the terms in (83), (89), and (93).

2.2 Derivation of $q_2(\cdot)$ update

We start by computing the normalizing constant for $f_2(\mathbf{X}_{\mathcal{C}_l}, c_l) Q^{2,l}(\mathbf{X}_{\mathcal{C}_l}, c_l)$

$$\begin{aligned}
G_{l,0} &= \sum_{c_l} \int_{\mathbf{X}_{\mathcal{C}_l}} f_2(\mathbf{X}_{\mathcal{C}_l}, c_l) Q^{\setminus 2,l}(\mathbf{X}_{\mathcal{C}_l}, c_l) d\mathbf{X}_{\mathcal{C}_l} dc_l \\
&\stackrel{(a)}{=} \sum_{c_l} \int_{\mathbf{X}_{\mathcal{C}_l}} \left[(1-c_l) \prod_{i \in \mathcal{C}_l} \delta(\mathbf{x}_i) + c_l \prod_{i \in \mathcal{C}_l} \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}_i) \right] \\
&\quad \mathcal{C}\mathcal{N}(\mathbf{x}_i; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l}) \mathcal{B}(c_l; p_c) d\mathbf{X}_{\mathcal{C}_l} \\
&= (1-p_c) \int_{\mathbf{X}_{\mathcal{C}_l}} \prod_{i \in \mathcal{C}_l} \delta(\mathbf{x}_i) \mathcal{C}\mathcal{N}(\mathbf{x}_i; \hat{\mathbf{m}}_i^{\setminus 2,l}, \Sigma_i^{\setminus 2,l}) d\mathbf{X}_{\mathcal{C}_l} \\
&\quad + p_c \int_{\mathbf{x}_i} \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}_i) \mathcal{C}\mathcal{N}(\mathbf{x}_i; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l}) d\mathbf{x}_i \\
&\quad \prod_{j \notin \mathcal{C}_l} \int_{\mathbf{x}_j} \mathcal{C}\mathcal{N}(\mathbf{x}_j; \mathbf{0}, \mathbf{R}_j) \mathcal{C}\mathcal{N}(\mathbf{x}_j; \hat{\mathbf{m}}_j^{\setminus 2,l}, \hat{\Sigma}_j^{\setminus 2,l}) d\mathbf{x}_j \tag{139} \\
&\stackrel{(b)}{=} (1-p_c) \prod_{i \in \mathcal{C}_l} \mathcal{C}\mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l}) \\
&\quad + p_c \underbrace{\int_{\mathbf{x}_i} \prod_{i \in \mathcal{C}_l} \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{R}_i (\mathbf{R}_i + \hat{\Sigma}_i^{\setminus 2,l})^{-1} \hat{\mathbf{m}}_i^{\setminus 2,l}, (\mathbf{R}_i^{-1} + \hat{\Sigma}_i^{\setminus 2,l^{-1}})^{-1}) d\mathbf{x}_i}_{\stackrel{(c)}{=} 1} \\
&\quad \prod_{i \in \mathcal{C}_l} \mathcal{C}\mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l} + \mathbf{R}_i) \\
&\stackrel{(d)}{=} b_l + a_l,
\end{aligned}$$

where (a) is obtained by plugging the densities, (b) by straightforward application of (137), (c) by integrating all the involved Gaussian distributions to unity, and (d) by setting $b_l = (1-p_c) \prod_{i \in \mathcal{C}_l} \mathcal{C}\mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l})$ and $a_l = p_c \prod_{i \in \mathcal{C}_l} \mathcal{C}\mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{\setminus 2,l}, \hat{\Sigma}_i^{\setminus 2,l} + \mathbf{R}_i)$.

Subsequently, by following the same steps, the n th moment with respect to \mathbf{x}_i is computed as

$$\begin{aligned}
\mathbb{E}_{f_2 Q^{\setminus 2,l}}[\mathbf{x}_i^n] &= \frac{1}{G_{l,0}} \sum_{c_l} \int_{\mathbf{X}_{\mathcal{C}_l}} \mathbf{x}_i^n f_2(\mathbf{X}_{\mathcal{C}_l}, c_l) Q^{\setminus 2,l}(\mathbf{X}_{\mathcal{C}_l}, c_l) d\mathbf{X}_{\mathcal{C}_l} dc_l \\
&= \frac{a_l}{G_{l,0}} \int_{\mathbf{x}_i} \mathbf{x}_i^n \mathcal{C}\mathcal{N}(\mathbf{x}_i; \mathbf{R}_i (\mathbf{R}_i + \hat{\Sigma}_i^{\setminus 2,l})^{-1} \hat{\mathbf{m}}_i^{\setminus 2,l}, (\mathbf{R}_i^{-1} + \hat{\Sigma}_i^{\setminus 2,l^{-1}})^{-1}) d\mathbf{x}_i. \tag{140}
\end{aligned}$$

Note that the integral term represents the n th moment of a multivariate Gaussian random vector. Thus, the mean $\mathbb{E}_{f_2 Q^{\setminus 2,l}}[\mathbf{x}_i]$, and the variance $\text{Var}_{f_2 Q^{\setminus 2,l}}[\mathbf{x}_i]$ are deduced directly from (140) and they are given in (92). Finally, the posterior mean with respect to c_l is

computed as

$$\begin{aligned}
\mathbb{E}_{f_2 Q^{2,l}}[c_l] &= \frac{1}{G_{l,0}} \sum_{\mathbf{c}} \int c_l f_2(\mathbf{X}_{\mathcal{G}_l}, c_l) Q^{2,l}(\mathbf{X}_{\mathcal{G}_l}, \mathbf{c}) d\mathbf{X} d\mathbf{c} \\
&= \sum_{c_l} c_l (1 - c_l) \mathcal{B}(c_l; p_{\mathbf{c}}) dc_l \int \prod_{i \in \mathcal{G}_l} \delta(\mathbf{x}_i) \mathcal{C} \mathcal{N}(\mathbf{x}_i; \hat{\mathbf{m}}_i^{2,l}, \hat{\Sigma}_i^{2,l}) \\
&\quad + \sum_{c_l} c_l^2 \mathcal{B}(c_l; p_{\mathbf{c}}) dc_l \int \prod_{i \in \mathcal{G}_l} \mathcal{C} \mathcal{N}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}_i) \mathcal{C} \mathcal{N}(\mathbf{x}_i; \hat{\mathbf{m}}_i^{2,l}, \hat{\Sigma}_i^{2,l}) d\mathbf{X}_{\mathcal{G}_l} \\
&= \frac{1}{G_{l,0}} \mathbb{E}[c_l - c_l^2] \prod_{i \in \mathcal{G}_l} \mathcal{C} \mathcal{N}(\mathbf{0}; \hat{\mathbf{m}}_i^{2,l}, \hat{\Sigma}_i^{2,l}) + \frac{1}{G_{l,0}} \mathbb{E}[c_l^2] \prod_{i \in \mathcal{G}_l} \mathcal{C} \mathcal{N} \\
&= \frac{a_l}{G_{l,0}}.
\end{aligned} \tag{141}$$

Appendix 3 MMSE-based denoiser

Recall the linear signal model given by (118)

$$\boldsymbol{\theta}_i^{(k)} = \mathbf{x}_i + \mathbf{e}^{(k)}, \quad \forall i \in \mathcal{N}.$$

In this section, we provide a detailed derivation for the MMSE denoiser function $\boldsymbol{\eta}(\cdot, \cdot) = \mathbb{E}[\mathbf{x}|\boldsymbol{\theta}]$. For clarity of presentation, we drop both the device index i and the iteration index (t) .

Let us consider the following linear model

$$\boldsymbol{\theta} = \mathbf{x} + \mathbf{e}, \quad \forall i \in \mathcal{N},$$

where $\mathbf{e} \sim \mathcal{CN}(\mathbf{e}; \mathbf{0}, \Sigma_{\theta})$, and $p(\mathbf{x}) = (1 - p_a)\delta(\mathbf{x}_i) + p_a\mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \mathbf{R})$. The MMSE-based denoising function $\boldsymbol{\eta}(\cdot, \cdot)$ estimator is given by $\boldsymbol{\eta} = \mathbb{E}[\mathbf{x}|\boldsymbol{\theta}]$

$$\begin{aligned} \boldsymbol{\eta}(\cdot) &= \mathbb{E}[\mathbf{x}|\boldsymbol{\theta}] \\ &= \int_{\mathbf{x}} \mathbf{x} p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} \\ &= \int_{\mathbf{x}} \frac{p(\boldsymbol{\theta}|\mathbf{x})((1 - p_a)\delta(\mathbf{h}_i) + p_a\mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}))}{p(\boldsymbol{\theta})} \\ &= \frac{(1 - p_a)}{p(\boldsymbol{\theta})} \underbrace{\int_{\mathbf{x}} \mathbf{x} p(\boldsymbol{\theta}|\mathbf{x}) \delta(\mathbf{x}) d\mathbf{x}}_{\eta_1} + \underbrace{\int_{\mathbf{x}} \mathbf{x} \frac{p_a}{p(\boldsymbol{\theta})} p(\boldsymbol{\theta}|\mathbf{x}) \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \mathbf{R}) d\mathbf{x}}_{\eta_2}. \end{aligned} \quad (142)$$

Since $\int_{\mathbf{x}} f(\mathbf{x})\delta(\mathbf{x})d\mathbf{x} = f(\mathbf{x} = \mathbf{0}) = \mathbf{0}$, $\eta_1 = 0$ and, hence, it does not contribute to $\mathbb{E}[\mathbf{x}|\boldsymbol{\theta}]$.

Thus, in order to evaluate $\mathbb{E}[\mathbf{x}|\boldsymbol{\theta}]$, we have to evaluate only $\int_{\mathbf{x}} \mathbf{x}\eta_2 d\mathbf{x}$ in (142).

Next, we express η_2 as follows

$$\begin{aligned} p_a \frac{p(\boldsymbol{\theta}|\mathbf{x})\mathcal{CN}(\mathbf{x}; \mathbf{0}, \mathbf{R})}{p(\boldsymbol{\theta})} &\stackrel{(a)}{=} \frac{\mathcal{CN}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{R} + \Sigma_{\theta})\mathcal{CN}(\mathbf{x}; (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})\Sigma_{\theta}^{-1}\boldsymbol{\theta}, (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})^{-1})}{(1 - p_a)\mathcal{CN}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{R} + \Sigma_{\theta}) + p_a\mathcal{CN}(\boldsymbol{\theta}; \mathbf{0}, \Sigma_{\theta})} \\ &\stackrel{(b)}{=} \frac{\mathcal{CN}(\mathbf{x}; (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})\Sigma_{\theta}^{-1}\boldsymbol{\theta}, (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})^{-1})}{1 + \frac{1 - p_a}{p}\mathcal{CN}(\boldsymbol{\theta}; \mathbf{0}, \Sigma_{\theta}^{-1} - (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})^{-1}) \det(\mathbf{R} + \mathbf{I}_M)} \\ &\stackrel{(c)}{=} \frac{\mathcal{CN}(\mathbf{x}; (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})\Sigma_{\theta}^{-1}\boldsymbol{\theta}, (\Sigma_{\theta}^{-1} + \mathbf{R}^{-1})^{-1})}{1 + \frac{1 - p_a}{p_a} \det(\mathbf{I}_M + \mathbf{R}) e^{-\boldsymbol{\theta}^H (\Sigma_{\theta}^{-1} - (\Sigma_{\theta} + \mathbf{R})^{-1}) \boldsymbol{\theta}}}, \end{aligned} \quad (143)$$

where

- (a) follows from the fact that the nominator is obtained by applying the rule of product for two Gaussian random vectors with respect to \mathbf{x} as shown in Appendix 2.1.

- (b) is accomplished by dividing both the denominator and numerator by the scaled Gaussian term $p_a \mathcal{C} \mathcal{N}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{R} + \Sigma_\theta)$ and then applying the rule for the division of two Gaussian random vectors with respect to $\boldsymbol{\theta}$.
- (c) is achieved through straightforward manipulation of the Gaussian term in the denominator.

Finally, note that the numerator in (143) identifies as a circularly symmetric complex Gaussian distribution with mean $(\Sigma_\theta^{-1} + \mathbf{R}^{-1})^{-1} \Sigma_\theta^{-1} \boldsymbol{\theta} = \mathbf{R}(\Sigma_\theta^{-1} + \mathbf{R}^{-1})^{-1} \boldsymbol{\theta}$ and covariance matrix $(\Sigma_\theta^{-1} + \mathbf{R}^{-1})^{-1}$. Thus, by definition,

$$\eta(\boldsymbol{\theta}, \Sigma_\theta) = \mathbb{E}[\mathbf{x}|\boldsymbol{\theta}] = \int_{\mathbf{x}} \mathbf{x} \eta_2 d\mathbf{x} = \boldsymbol{\psi}(\boldsymbol{\theta}, \Sigma_\theta) \mathbf{R}(\Sigma_\theta^{-1} + \mathbf{R}^{-1})^{-1} \boldsymbol{\theta} \quad (144)$$

where we used the matrix inverse lemma [100, Lemma B.3] to get $(\Sigma_\theta^{-1} + \mathbf{R}^{-1})^{-1} \Sigma_\theta^{-1} \boldsymbol{\theta} = \mathbf{R}(\Sigma_\theta^{-1} + \mathbf{R}^{-1})^{-1} \boldsymbol{\theta}$ and $\boldsymbol{\psi}(\boldsymbol{\theta}, \Sigma_\theta) = \frac{1}{1 + \frac{1-p_a}{p_a} \frac{\det(\Sigma_\theta + \mathbf{R})}{\det(\Sigma_\theta)} e^{-\theta^H (\Sigma_\theta^{-1} - (\Sigma_\theta + \mathbf{R})^{-1}) \boldsymbol{\theta}}}$. This completes the proof.

920. Hatami, Mohammad (2023) Information freshness optimization in energy harvesting IoT networks
921. Montaseri, Mohammad Hassan (2023) Analysis and design of stacked MOS mm-wave power amplifiers
922. Pauna, Tommi (2023) Development of a collaborative delivery model for industrial engineering projects
923. Heikkilä, Marjo (2024) Field trials and test methods in vertical driven 5G and beyond mobile network research
924. Gong, Zhengya (2024) Understanding cultural influence on creativity in ideation
925. Tran, Thuong-Khanh (2024) Micro-expression spotting based on supervised learning
926. Javed, Zunera (2024) Crowdsourcing driven truthful data collection methods for wireless nodes in enterprise networks
927. Qiao, Yang (2024) An experimental study on fracture process zone of rock
928. Marata, Leatile (2024) Advanced signal processing techniques for machine type communications
929. Susarla, Praneeth (2024) Learning-based strategies for millimeter wave radio beamforming and sensing
930. Salomäki, Marko (2024) On the connections between occupational safety, health management, and safety culture development in construction : insights from a large-scale infrastructure alliance project
931. Kokkonen, Mikko (2024) Lens structures from novel materials for fixed and steerable beams antenna systems at sub-THz range
932. Bello, Adedayo (2024) Valorization of tannin from spruce bark and its application as biocoagulant in water treatment
933. Ayati, Sayed Mohammad (2024) Toward circular business models and supply chains in manufacturing : building circular dynamic capabilities
934. Kursula, Kalle (2024) Recycling and reusing of waste concrete fines as granulated aggregates
935. Christophliemk, Mika (2024) Porous alkali-activated manganese and copper composites for wastewater treatment : catalytic wet air oxidation of bisphenol A and pharmaceutical residues

S E R I E S E D I T O R S

A
SCIENTIAE RERUM NATURALIUM
University Lecturer Mahmoud Filali

B
HUMANIORA
University Lecturer Santeri Palviainen

C
TECHNICA
Senior Research Fellow Antti Kajjalainen

D
MEDICA
University Lecturer Pirjo Kaakinen

E
SCIENTIAE RERUM SOCIALIUM
University Lecturer Henri Pettersson

E
SCRIPTA ACADEMICA
Strategy Officer Mari Katvala

G
OECONOMICA
University Researcher Marko Korhonen

H
ARCHITECTONICA
Associate Professor Anu Soikkeli

EDITOR IN CHIEF
University Lecturer Santeri Palviainen

