Sparse Bayesian Learning for Joint Channel Estimation and Data Detection in OFDM Systems

A thesis

submitted in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

in the Faculty of Engineering

by

Ranjitha Prasad

Electrical Communication Engineering
Indian Institute of Science, Bangalore
Bangalore – 560 012 (INDIA)

Oct. 2014

TO

My parents
Smt. Padmini and Sri. Ramaprasad,
my husband
Harshan and
my parents-in-law
Smt. Lalitha and Sri. Jagadeesh

Acknowledgements

My gratitude goes to my advisor Dr. Chandra R. Murthy for his tremendous patience, timely advice and support throughout my stint as a graduate student. He has given me invaluable insights into our research work and has taught me the art of writing papers and presenting the results, along with helping me to strive for a good work-life balance. His pursuit for excellence and perfection has inspired me to be a better researcher and a better person.

My special thanks goes out to the members of the ECE office, especially Mr. Srinivasa Murthy and Mr. C. T. Nagaraj. I would also like to thank IISc for providing me an opportunity to pursue graduate studies along with good hostel and mess facilities. I appreciate the financial support from VIMSS for pursuing an internship at UCSD, and TCS research scholar program for supporting my Ph.D. and conference travels.

I would like to express my gratitude to Prof. B. D. Rao for providing me financial assistance and an opportunity to interact with him at Dept. of ECE, UCSD. I am grateful to Prof. E. Viterbo who hosted me as an intern for 4 months at Monash university.

Life as a PhD student is incomplete without a technically interactive and intellectually vibrant lab. I take this opportunity to thank my colleagues Abhay, Sanjeev, Venu, Bharath, Gana, Partha, and Karthik who have been a part of my Ph.D. journey since the beginning. The wonderful experiences that we have shared in our numerous treks and trips will be cherished all my life. Special thanks to Nagananda for frequently taking us out in his car. I will always be thankful to Nanda for the car ride that helped me pay my thesis fees. I would also like to thank Santhosh, Nirmal, Srinivas, KC, Chandrashekar, Saurabh, Geethu, Mohit, Vinuthna and Vinnu with whom I have enjoyed several discussions. My stay at IISc would not have been as colorful without the girls of signal processing west wing- Shilpa, Suma, Ananya, Geethu, Priyanka and Vinuthna. I thank

my friends Haricharan and Shubha for all the interesting times together.

I am indebted to my family who have supported me throughout my Ph.D. I begin by thanking my parents Padmini Ramaprasad and B. S. Ramaprasad, and my husband Harshan, who have been supportive in all my endeavors. I would like to express my sincere gratitude to my parents-in-law Lalitha Jagadeesh and G. D. Jagadeesh, and brother-in-law Kiran, for their encouragement. I am grateful to my sister Rachana and brother-in-law Arvind, who have helped me through my Ph.D. program.

Abstract

Bayesian approaches for sparse signal recovery have enjoyed a long-standing history in signal processing and machine learning literature. Among the Bayesian techniques, the expectation maximization based Sparse Bayesian Learning (SBL) approach is an iterative procedure with global convergence guarantee to a local optimum, which uses a parameterized prior that encourages sparsity under an evidence maximization framework. SBL has been successfully employed in a wide range of applications ranging from image processing to communications. In this thesis, we propose novel, efficient and low-complexity SBL-based algorithms that exploit structured sparsity in the presence of fully/partially known measurement matrices. We apply the proposed algorithms to the problem of channel estimation and data detection in Orthogonal Frequency Division Multiplexing (OFDM) systems. Further, we derive Cramér Rao type lower Bounds (CRB) for the single and multiple measurement vector SBL problem of estimating compressible vectors and their prior distribution parameters. The main contributions of the thesis are as follows:

- We derive Hybrid, Bayesian and Marginalized Cramér Rao lower bounds for the problem of estimating compressible vectors drawn from a Student-t prior distribution. We derive CRBs that encompass the deterministic or random nature of the unknown parameters of the prior distribution and the regression noise variance. We use the derived bounds to uncover the relationship between the compressibility and Mean Square Error (MSE) in the estimates. Through simulations, we demonstrate the dependence of the MSE performance of SBL based estimators on the compressibility of the vector.

- OFDM is a well-known multi-carrier modulation technique that provides high spectral efficiency and resilience to multi-path distortion of the wireless channel.

It is well-known that the impulse response of a wideband wireless channel is approximately sparse, in the sense that it has a small number of significant components relative to the channel delay spread. In this thesis, we consider the estimation of the unknown channel coefficients and its support in SISO-OFDM systems using a SBL framework. We propose novel pilot-only and joint channel estimation and data detection algorithms in block-fading and time-varying scenarios. In the latter case, we use a first order auto-regressive model for the time-variations, and propose recursive, low-complexity Kalman filtering based algorithms for channel estimation. Monte Carlo simulations illustrate the efficacy of the proposed techniques in terms of the MSE and coded bit error rate performance.

- Multiple Input Multiple Output (MIMO) combined with OFDM harnesses the inherent advantages of OFDM along with the diversity and multiplexing advantages of a MIMO system. The impulse response of wireless channels between the N_t transmit and N_r receive antennas of a MIMO-OFDM system are group approximately sparse (ga-sparse), i.e., the $N_t N_r$ channels have a small number of significant paths relative to the channel delay spread, and the time-lags of the significant paths between transmit and receive antenna pairs coincide. Often, wireless channels are also group approximately-cluster sparse (ga-csparse), i.e., every ga-sparse channel consists of clusters, where a few clusters have all strong components while most clusters have all weak components. In this thesis, we cast the problem of estimating the ga-sparse and ga-csparse block-fading and time-varying channels using a multiple measurement SBL framework. We propose a bouquet of novel algorithms for MIMO-OFDM systems that generalize the algorithms proposed in the context of SISO-OFDM systems. The efficacy of the proposed techniques are demonstrated in terms of MSE and coded bit error rate performance.

Glossary

3GPP	:	Third Generation Partnership Project
AR	:	Autoregressive
AWGN	:	Additive White Gaussian Noise
BCRB	:	Bayesian Cramér Rao Bound
BER	:	Bit Error Rate
BIM	:	Bayesian Information matrix
BPSK	:	Binary Phase Shift Keying
FIM	:	Fisher Information Matrix
CP	:	Cyclic Prefix
CRLB	:	Cramér Rao Lower Bound
CS	:	Compressed Sensing
EM	:	Expectation Maximization
HCRB	:	Hybrid Cramér Rao Bounds
HIM	:	Hybrid Information Matrix
KFS	:	Kalman Filter and Smoother
LLR	:	Log-Likelihood Ratio
LOS	:	Line-Of-Sight
LTE-A	:	Long Term Evolution - Advanced
MAP	:	Maximum a Posteriori
MCRB	:	Marginalized Cramér Rao Bounds
MIMO	:	Multiple-Input Multiple-Output
MIP	:	Multipath Intensity Profile
MISO	:	Multiple-Input Single-Output
ML	:	Maximum Likelihood
MMSE	:	Minimum Mean Square Error
MMV	:	Multiple Measurement Vector
MSE	:	Mean Squared Error
OFDM	:	Orthogonal Frequency Division Multiplexing
OMP	:	Orthogonal Matching Pursuit
QPSK	:	Quadrature Phase Shift Keying
QAM	:	Quadrature Amplitude Modulation
SBL	:	Sparse Bayesian Learning
SMV	:	Single Measurement Vector
SNR	:	Signal-to-Noise Ratio
SISO	:	Single-Input Single-Output

Notation

Boldface lower case letters	:	Vectors
$\lfloor c \rfloor$	:	Largest integer less than c
$\mathcal{C}$	:	Field of complex numbers
$\mathcal{CN}(\mu, \sigma^2)$	:	Circularly symmetric complex Gaussian distribution with mean μ and σ^2 variance
$\mathbb{E}_X[\cdot]$	:	Expectation with respect to the random variable X
$\mathbf{I}_M$	:	$M \times M$ Identity matrix
$\Im(\cdot)$	:	Imaginary part of the complex argument
K	:	Number of OFDM symbols in an OFDM frame
L	:	Length of the channel
N	:	Number of subcarriers in an OFDM symbol
N_t	:	Number of Transmit Antenna in a MIMO-OFDM system
N_r	:	Number of Receive Antenna in a MIMO-OFDM system
$p(x)$	:	probability density function of the random variable X
$\mathcal{R}$	:	Field of real numbers
$\mathcal{R}^+$	:	Field of non-negative real numbers
$\text{Tr}(\cdot)$	:	Trace of a matrix
$\otimes$	:	Kronecker Product
$\Re(\cdot)$	:	Real part of the complex argument
$(\cdot)^T$	:	Transposition
$(\cdot)^H$	:	Hermitian transposition
$(\cdot)^*$	:	Complex conjugation
$\lvert \cdot \rvert$	:	Absolute value of a complex number or the determinant of a matrix or the cardinality of a set, depending on the context
$\lVert \cdot \rVert_2$	:	Euclidean norm of a vector
$\lVert \cdot \rVert_F$	:	Frobenius norm of a matrix
Upper case letters	:	Matrices

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the recent years, Sparse Signal Recovery (SSR) has received a lot of interest in the signal processing and machine learning communities as it facilitates reliable recovery of underlying sparse signals using compressed domain samples by exploiting the inherent redundancy in the sparse signal [1–4]. The objective of SSR is to accurately reconstruct a sparse signal from an underdetermined set[1] of noisy linear measurements. In practice, the signal may not be exactly sparse but *compressible*. Compressible signals are signals whose ordered coefficients decay quickly, for e.g., they may obey a power law: for some $p > 1$, the kth largest coefficient in magnitude is at most $c_p k^{-p}$, where $c_p > 0$ is some constant. In this context, one seeks to find computationally efficient algorithms that recover sparse/compressible vectors from an underdetermined set of linear equations. In several applications such as audio and video processing [6], medical imaging [7], and baseband signal processing in communication systems [8], [9], the signal acquisition process involves an unknown sparse vector $\mathbf{x} \in \mathbb{C}^{L \times 1}$ projected into a lower dimensional space and corrupted by additive noise, resulting in the observations

---

[1]This is also referred to as an overcomplete set in the literature [5]

$\mathbf{y} \in \mathbb{C}^{L \times 1}$ given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}, \tag{1.1}$$

where $\mathbf{\Phi} \in \mathbb{C}^{N \times L}$ denotes the measurement matrix with $N < L$ and $\mathbf{v} \in \mathbb{C}^{N \times 1}$ denotes the ambient noise. The difficulty in reconstructing a sparse signal arises from the fact that neither the locations of the non-zero entries of $\mathbf{x}$ nor the corresponding values at those locations are known.

In general, the problem of recovering $\mathbf{x}$ from $\mathbf{y}$ is ill-posed, as measurement matrix $\mathbf{\Phi}$ has a non-trivial null space ($N < L$). In other words, any non-zero vector $\mathbf{x}_n$ in the null-space of $\mathbf{\Phi}$ contributes to a new solution $\mathbf{x} + \mathbf{x}_n$ of (1.1), thus, making it impossible to recover a unique solution. In [5] and [10], the authors have independently shown that sparse vectors can be uniquely recovered from a set of N ($N < M$) linear and non-adaptive measurements using the knowledge of sparsity, S, which allows one to restrict the search for $\mathbf{x}$ to the set consisting of all possible S-sparse vectors. Furthermore, these studies also elucidate the crucial role of $\mathbf{\Phi}$ in admitting sparse representations along with recovery guarantees. These sufficient conditions for unique and reliable recovery of sparse vectors typically take the form of a requirement on the so-called mutual coherence or restricted isometry property of the measurement matrix.

In practice, sensing devices record the measurements $\mathbf{y}$ as a function of the sparse signal $\mathbf{x}$, but the measurement matrix is system-dependent. It is often hard to verify that these matrices possess the required properties to guarantee reliable recovery of $\mathbf{x}$ [5, 10]. Moreover, many of the known sufficient conditions for reliable recovery tend to be overly pessimistic. The presence of measurement noise in the sensing process further complicates the problem. Despite these and several other issues that render

the SSR seemingly hard, the idea that the sparsity structure in the signal can be used to obtain unique solutions from underdetermined works in principle, and a number of recovery algorithms are available in the literature. The recovery algorithms can be broadly classified into two categories, namely, Compressed Sensing (CS) and Bayesian methods. In the rest of this section, we discuss the various CS and Bayesian techniques that have been proposed and studied in the literature.

1.1   Compressed Sensing Techniques

In this subsection, we first introduce the fundamental problem of sparse recovery as an optimization problem with ℓ_0 norm based sparsity constraint. We later discuss the different relaxations of the fundamental problem, such as the ℓ_1 and the ℓ_p norm based optimization problems. Furthermore, we discuss the algorithms proposed in the literature in order to solve the fundamental sparse recovery problem and its variants.

In general, a Minimum Mean Square Error (MMSE) or Least Squares (LS) solution to the noisy linear equations in (1.1) can be obtained by solving for $\mathbf{x}$ that minimizes $\|\mathbf{y} - \mathbf{\Phi x}\|_2$. However, in scenarios where $N < L$, the MMSE/LS problem is ill-posed. If $\mathbf{x}$ is sparse, one can exploit the sparse structure and obtain the solution to (1.1) by solving the constrained optimization problem given by

$$\arg\min_{\mathbf{x}} |\mathbf{x}|_0 \quad \text{subject to } \|\mathbf{y} - \mathbf{\Phi x}\|_2 \leq \epsilon, \tag{1.2}$$

where $\epsilon > 0$ is a user-defined parameter, $|\cdot|_0$ denotes the ℓ_0 norm and $\|\cdot\|_2$ denotes the

ℓ_2 norm. The implication of sparsity is clear when we rewrite (1.1) as follows:

$$\mathbf{y} = \sum_{i=1}^{L} \boldsymbol{\Phi}_i x(i) + \mathbf{v}, \tag{1.3}$$

where $\boldsymbol{\Phi}_i$ represents the ith column of $\boldsymbol{\Phi}$. If $\mathbf{x}$ is S-sparse, only S columns of $\boldsymbol{\Phi}$ contribute to $\mathbf{y}$. Hence, in the noiseless scenario, if the locations of the non-zero entries of $\mathbf{x}$ are known, then only S rows of $\boldsymbol{\Phi}$ are sufficient in order to recover $\mathbf{x}$. However, since the locations of non-zero entries are not known, intuitively, we require $N > S$ measurements, especially in the presence of noise. However, for sufficiently small S, it is possible that $N < L$, i.e., it is possible to uniquely recover $\mathbf{x}$ even when the linear equations given in (1.1) are underdetermined. Notice that in (1.2), the ℓ_0 norm of $\mathbf{x}$ counts the number of non-zero components of the vector, i.e., one seeks to find the *sparsest* solution. Hence, by imposing the sparsity constraint, it is possible to recover $\mathbf{x}$ in (1.1).

Greedy algorithms such as Matching Pursuit [11], Orthogonal Matching Pursuit (OMP) [12] and Compressive Sampling Matching Pursuit (CoSAMP) [13] have been proposed for the recovery of $\mathbf{x}$ in (1.2). In [14], iterative thresholding algorithms using a variational formulation for hard thresholding has been proposed as a solution to the ℓ_0 problem. In [15], the authors provide two low-complexity iterative algorithms which employ hard thresholding and show that they minimize a cost function similar to the ℓ_0 norm.

Since solving the ℓ_0 norm problem is known to be NP-hard, algorithms based on convex optimization techniques have been studied in the literature. Convex optimization based techniques relax the ℓ_0 norm on $\mathbf{x}$ to a convex ℓ_1 norm, thereby converting the

problem in (1.2) to an optimization problem given by

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to } \|\mathbf{y} - \mathbf{\Phi x}\|_2 \leq \epsilon. \tag{1.4}$$

where $\|\mathbf{x}\|_1$ denotes the ℓ_1 norm of $\mathbf{x}$, defined as $\|\mathbf{x}\|_1 \triangleq \sum_{i=1}^{L} |\mathbf{x}|$. It turns out, rather surprisingly, that when the measurement matrix $\mathbf{\Phi}$ satisfies the so-called Restricted Isometry Property (RIP) property with an appropriate constant, (1.2) and (1.4) can be guaranteed to have the same unique globally optimal solution [10]. We reiterate that (1.4) is a convex optimization problem, and many numerically efficient algorithms such as the Basis Pursuit (BP) and BP denoising/LASSO [16] have been proposed for finding sparse solutions.

An unconstrained form of the optimization problem in (1.4) can be written as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 + \lambda(\|\mathbf{y} - \mathbf{\Phi x}\|_2), \tag{1.5}$$

where the scalar λ appropriately weighs the relative importance given to the sparsity of the solution and the Euclidean fit error term. In [16], a recovery technique known as the Least Absolute Shrinkage and Selection Operator (LASSO) is proposed for solving (1.5). Simple and efficient methods for solving the basis pursuit problem (1.4) and the unconstrained problem (1.5), based on Bregman iterative regularization are proposed in [17]. Several efficient methods such as the iterative shrinkage and thresholding [18], gradient projection for sparse reconstruction [19] have been proposed to solve (1.5). These methods are computationally fast, with complexity being linear in the size of the product of $\mathbf{\Phi}$ and $\mathbf{x}$, and in some special cases the complexity is linear in the size of $\mathbf{x}$. As a result, such algorithms are a natural choice in solving large-scale sparse recovery

problems. An accelerated projected gradient method for regularization of (1.5) via iterative soft-thresholding algorithms and ℓ_1 penalization methods have been proposed for cases where the vector is almost sparse [20].

A non-convex relaxation of (1.2) leads to a ℓ_p $(0 < p < 1)$ norm minimization problem, whose unconstrained variant is proposed in [21] given by,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} |\mathbf{x}|_p + \lambda(\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2), \tag{1.6}$$

where $0 < p \le 1$. The authors propose the FOcal Underdetermined System Solver (FOCUSS) algorithm as a solution to the above optimization problem. Exact reconstruction of sparse signals in (1.6) via nonconvex minimization methods is proposed in [22], where the authors show that exact recovery is possible with substantially fewer measurements compared to the ℓ_1 based solution. Iteratively reweighted algorithms for computing local minima of such nonconvex problems is proposed in [23].

Several other studies in the literature have sought to provide a coding-theoretic perspective, and leverage algorithms from coding theory, for sparse signal recovery. In [24], the authors perform fast approximate Bayesian inference using belief propagation (BP) decoding, which is based on representing the CS measurement matrix as a graphical model. Using the theory of codes on graphs, joint design of sensing matrices and low complexity reconstruction algorithms using a new family of list decoding and multiple-basis belief propagation algorithms are proposed in [25].

Several alternate perspectives on CS based recovery techniques have been proposed in the literature. In [26], the authors seek to bridge the two major algorithmic approaches

to SSR - ℓ_1 minimization methods and iterative methods such as matching pursuit algorithms. The authors propose a simple, Regularized version of Orthogonal Matching Pursuit (ROMP) which has advantages of both ℓ_1 minimization methods and iterative methods.

In scenarios where the observations arrive sequentially, several methods have been proposed for online ℓ_1 learning [27,28]. A recursive perspective in terms of Kalman and RLS filters to the ℓ_1 optimization problem is provided in [29] and [30], respectively. A recursive ℓ_1 regularized least squares algorithm, also known as SPARLS for the estimation of a sparse vector has been proposed using the expectation maximization approach in [31]. In the next subsection, we review discuss the Bayesian formalism used in SSR and several algorithms that are popularly employed.

1.2 Bayesian Techniques

Bayesian techniques impose a sparsity promoting prior distribution on the sparse vector, and perform MAP estimation, variational Bayes' minimization or evidence maximization. Unlike the CS based algorithms that provide point estimates of the sparse vector, the Bayesian techniques evaluate the posterior distribution of **x** conditioned on **y**, or the posterior statistics such as the mean, variance and other higher order moments. A variety of Bayesian methods have been employed for finding sparse representations from a linear underdetermined set of equations.

Bayesian approaches for recovering a sparse vectors have a long history in the signal processing and machine learning literature. In [32], the author proposed Bayesian

regularization and pruning techniques using Laplacian priors for neural network applications. In [33] and [34], the authors proposed a Sparse Bayesian Learning (SBL) technique using an evidence maximization framework, largely based on approaches proposed in [35] and [36]. The relationship between LASSO and quadratic penalization was discussed in [37], along with proposing an expectation maximization based algorithm for obtaining the LASSO solution. In [38], the authors propose Bayesian techniques based on the cost functions that emerge from different priors, rather than justifying the choice of the prior based on the physical process that resulted in the sparse vectors, and postulate conditions that any sparse learning objective should satisfy. Bayesian inference using Markov chain Monte Carlo (MCMC) methods are proposed in [39]. The relationship between the compressed sensing and Bayesian experimental design framework is explored in [40]. Also, novel algorithms are proposed, based on expectation propagation in a sequential setup for large scale problems.

Several papers on Bayesian sparse recovery are based on Bernoulli-Gaussian priors, also known as spike-and-slab priors. In [41], the authors use the spike-and-slab prior [42] for sparse unsupervised learning. Further, approximate sparse inference techniques which employ the Bernoulli Gaussian prior are proposed in [43, 44].

It turns out that CS based approaches such as BP and OMP can also be viewed as a problem in the Bayesian framework, where the goal is to obtain a Maximum A-posteriori (MAP) estimate of $\mathbf{x}$ using a fixed sparsity inducing prior distribution. As an example, consider a Laplacian prior on $\mathbf{x}$, with distribution $p(\mathbf{x}; \lambda) \propto \exp^{-\lambda \|\mathbf{x}\|_1}$, where $\lambda > 0$. Let $\mathbf{v}$ be the additive Gaussian noise distributed as $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The MAP estimation problem using such a Laplacian prior is the same as solving (1.5), where λ is

determined by the parameters of the prior. An approach for estimating x using Laplacian priors has been proposed in [45].

A low-complexity recursive procedure for model selection and minimum mean squared error (MMSE) estimation in ill-posed linear regression problems using a fast Bayesian matching pursuit algorithm was proposed in [46, 47]. In [48], the authors present a sparse reconstruction technique using a distribution agnostic Bayesian matching pursuit algorithm. Here, the authors leverage the fast matching pursuit method and obtain the Bayesian estimates of sparse signals even when the signal prior is non-Gaussian or unknown using greedy approach.

In this thesis, we focus on SBL, which uses a two-stage hierarchical model on the unknown vector, as shown in Fig. 1.1. Here, we assume that $\mathbf{x} \sim \mathcal{N}(0, \boldsymbol{\Gamma})$, where the diagonal matrix $\boldsymbol{\Gamma}$ contains the *hyperparameters* $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L]^T$ as its diagonal elements. Further, an Inverse Gamma (IG) *hyperprior* is assumed for $\boldsymbol{\gamma}$, as it leads to a Student-t prior on the vector $\mathbf{x}$, which, in-turn, is known to be sparsity/compressible vector-promoting [34]. The significance of using the IG hyperprior arises from the fact that it is a conjugate to the Gaussian distribution. As a result, it leads to closed-form posterior and marginalized distributions on $\mathbf{y}$ [34]. In scenarios where the noise variance is unknown and random, an IG prior is used for the distribution of the noise variance as well.

SBL involves joint estimation of the sparse vector $\mathbf{x}$ and the hyperparameters, $\boldsymbol{\gamma}$. Different updates have been proposed for SBL [34, 35], among which the exact inference Expectation Maximization (EM) based updates are widely used. In addition to the monotonicity property of SBL by virtue of the EM framework, SBL offers guarantees

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\Phi\mathbf{x}, \sigma^2\mathbf{I}_{N\times N})$$
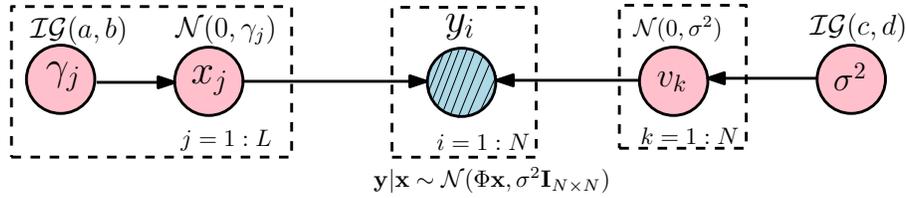
Figure 1.1: Graphical model of SBL: Two stage hierarchical model with the unknown sparse vector taking a conditional Gaussian distribution and the hyperparameters taking an Inverse Gamma distribution.

such as convergence to the sparsest solution when the noise variance is zero, and convergence to a sparse local minimum irrespective of the noise variance [49]. In contrast, approximate inference methods [44], although lower in computational complexity, do not offer such rigorous convergence guarantees. In Fig. 1.2 and Fig. 1.3, we plot the performance of BP, OMP and the SBL techniques to recover a $L = 150$ length vector of sparsity $S = 30$. The plots show that the SBL algorithm can recover the sparse signal with lower number of measurements as compared to conventional techniques. Moreover, we observe that SBL consistently performs well both when sparse vectors $\mathbf{x}$ have components $x(i) \in [0, 1]$ and when $\mathbf{x}$ is sparse and consists of highly scaled components.

Although the experiment shows that SBL is a very promising technique, the choice of employing CS or Bayesian methods primarily depends on the system, its dynamics, the nature of signals encountered, and the performance metric to be optimized. For e.g., it has been shown that CS based techniques are effective for MRI applications [7], video sensing [50], surveillance video processing [51] and coding [6], image processing [52], and medical imaging [53]. On the other hand, Bayesian techniques have been successfully employed in the fields of bioinformatics [54], genome analysis [55], image/visual tracking [56], neuro-imaging [57, 58] and beamforming [59]. In this thesis, we focus on
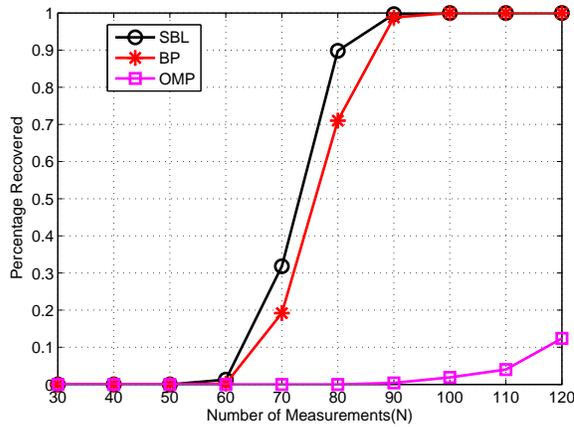
Figure 1.2: Performance comparison of the percentage successful recovery performance of SBL, BP and OMP algorithms, for $S = 30$ sparse signals with coefficients being 0 or 1 with $L = 150$
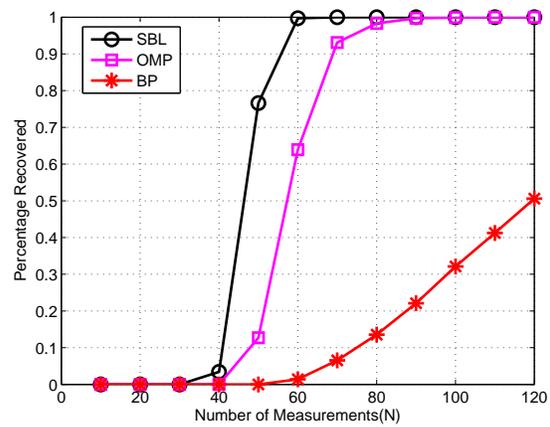
Figure 1.3: Performance comparison of the percentage successful recovery performance of SBL, BP and OMP algorithms, for $S = 30$ sparse signals with highly scaled coefficients with $L = 150$

the application of SSR techniques to baseband signal processing for wireless communications; in particular for the sparse/compressible wireless channel estimation problem in the framework of a well-known multiplexing technique called Orthogonal Frequency Division Multiplexing (OFDM).

Until now, we have introduced the optimization problems for SSR, its advantages, algorithms proposed in literature and various applications that feature a linear regression model with a sparse vector. In the rest of the chapter, we discuss the issues regarding channel estimation in OFDM systems and throw light on the SBL technique that we adopt for exploiting sparsity in OFDM channel estimation.

1.3 SSR Techniques for Wireless Communications

In this section, we address the problem of channel estimation in OFDM systems incorporating the inherent sparse and approximately sparse nature of the wireless channel.

Some of the key questions which we address are as follows:

1. $Q1$ - Why does sparsity arise in wideband wireless channels? (See Sec. 1.3.1)

2. $Q2$ - When do we encounter an under-determined system of linear equations in channel estimation problems? (See Sec. 1.3.2)

3. $Q3$ - What are the conventional channel estimation and the existing sparse channel recovery methods? (See Sec. 1.3.3)

4. $Q4$ - What are the contributions of the thesis to the problem of wideband channel estimation exploiting the sparsity of the channel? (See Sec. 1.3.4)

We focus specifically on OFDM as the underlying air interface technology due to its ubiquitous use in present-day wireless standards [60, 61].

1.3.1   Sparsity of Wireless Channels in OFDM Systems

In this subsection, we focus on $Q1$ and present a brief discussion on the nature of OFDM channels. Here, the sparse signal of interest is the impulse response of wireless channels. In the complex baseband representation, the impulse response $\tilde{h}[t], t \in \mathbb{R}$ of a wireless channel can be modeled as a stationary tapped delay line filter in the lag-domain, given by

$$\tilde{h}[t] = \sum_{l=1}^{\tilde{L}} \tilde{h}_l \delta[t - \tau_l], \tag{1.7}$$

where $\delta[t]$ is the Dirac delta function, $\tilde{h}_l$ and τ_l represent the attenuation and propagation delay for the lth path, respectively, and $\tilde{L}$ denotes the total number of resolvable paths [62].

Typically, a communication system employs filters at the transmitter and receiver in order to reduce the adjacent channel power radiation and interference from other transmitters. The transmit and receive filtering operations performed at the respective RF front ends lead to an equivalent channel given by

$$\mathbf{h}(t) = \mathbf{g}_t(t) * \tilde{\mathbf{h}}(t) * \mathbf{g}_r(t), \tag{1.8}$$

where $\tilde{\mathbf{h}}(t)$ is given in (1.7), $\mathbf{g}_t(t)$ and $\mathbf{g}_r(t)$ are the baseband transmit and receive filters, respectively, and $*$ represents the convolution operation. The effect of transmit and receive filtering can be represented by the convolution with an equivalent Raised Cosine (RC) filter, with an impulse response given by

$$\mathbf{r}(t) = \text{sinc}\left(\frac{t}{T}\right) \frac{\cos\left(\frac{\pi t \beta}{T}\right)}{\left(1 - \left(\frac{2\beta t}{T}\right)^2\right)}, \tag{1.9}$$

where T is the sampling period and β is the roll-off factor of the RC filter [63]. Hence, $\mathbf{h}(t)$ in (1.8) can be computed as

$$\mathbf{h}(t) = \int_{-\infty}^{t} \tilde{\mathbf{h}}(\tau) \text{sinc}\left(\frac{t-\tau}{T}\right) \frac{\cos\left(\frac{\pi(t-\tau)\beta}{T}\right)}{\left(1 - \left(\frac{2\beta(t-\tau)}{T}\right)^2\right)} d\tau. \tag{1.10}$$

Sampling $\mathbf{h}(t)$ at $t = nT$, we obtain the discrete-time channel impulse response as

$$\mathbf{h}(nT) = \int_{-\infty}^{t} \tilde{\mathbf{h}}(\tau) \text{sinc}\left(n - \frac{\tau}{T}\right) \frac{\cos\left(\frac{\pi(nT-\tau)\beta}{T}\right)}{\left(1 - \left(\frac{2\beta(nT-\tau)}{T}\right)^2\right)} d\tau. \tag{1.11}$$

In the ideal case of *sample-spaced* channel, $\mathbf{g}_t(t) = \mathbf{g}_r(t) = \delta(t)$ and the sampling instants $\{0, T, 2T, \ldots\}$ coincide with the arrival instants τ_l, and as a result, the discrete

version of the channel $\tilde{h}[nT]$ will have exactly $\tilde{L}$ non-zero elements. Thus, the sample spaced impulse response of a wireless channel exhibits sparsity which, in-turn, motivates us to investigate the use of SSR techniques for the channel estimation problem. However, in this thesis, we address the larger issue of non-ideal non-sample spaced channels, as elaborated in the next section. In the sequel, we describe the problem of channel estimation in an OFDM system and the conventional channel estimation techniques, followed by the issues addressed in this thesis.

1.3.2 OFDM System and the Channel Estimation Problem

Two important technologies for energy-and-bandwidth-efficient wireless communications that have been developed over the past couple of decades are OFDM and MIMO. OFDM is known to effectively mitigate the inter-symbol interference caused by the frequency selective fading channel [64]. Multiple-Input Multiple-Output (MIMO) systems provide spatial diversity, leading to significant gains in the capacity of the time-varying wireless channel. To reap the advantages of the above two systems, MIMO-OFDM systems have emerged as the clear choice in nearly all wireless standards [60]. In this section, we describe the OFDM system model and the problem of channel estimation in order to address $Q2$.

 The goal of a MIMO-OFDM system is high-rate reliable communication along with accurate data detection at the output of the decoder. The transmission between the N_t transmit and the N_r receive antennas takes place through OFDM frames, where every frame consists of K OFDM symbols. Fig. 1.4 shows the block diagram of a typical
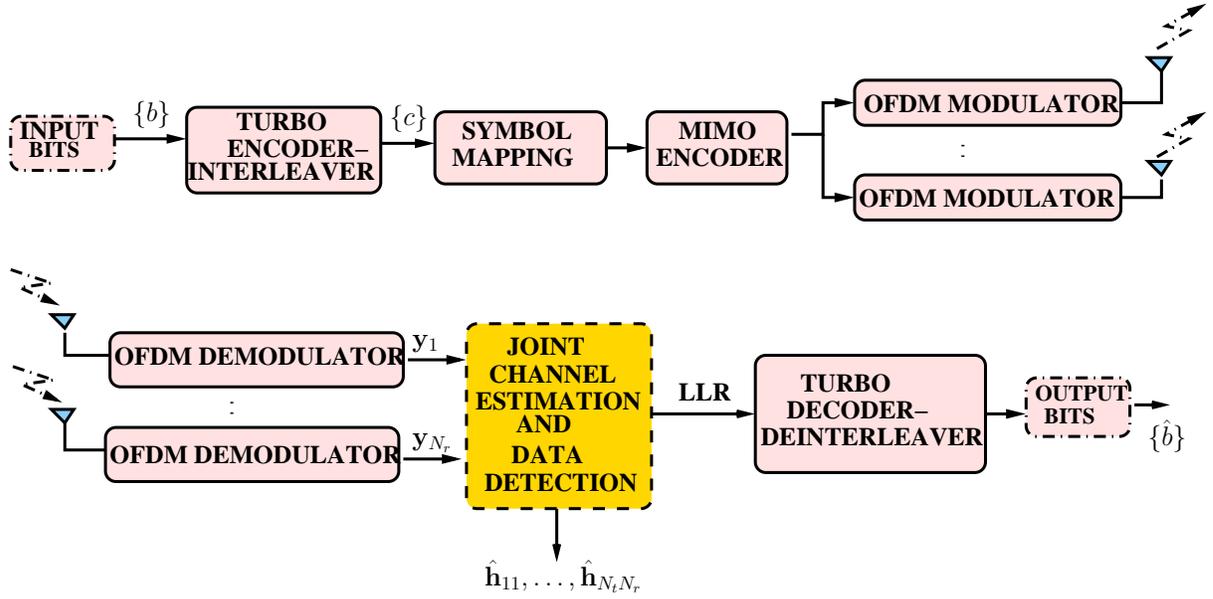
Figure 1.4: Turbo encoded/decoded transmit and receive chain of a MIMO-OFDM system. The dashed box (block shaded in yellow) highlights the proposed algorithms. Note that the quantities of interest are the channel estimates and output bits $\{\hat{b}\}$.

MIMO-OFDM system with N subcarriers, N_t transmit antennas and N_r receive antennas. At the transmitter of the discrete-time MIMO-OFDM system, $\{b\}$ input bits are first encoded and interleaved into a new sequence of coded bits, $\{c\}$. The coded bits $\{c\}$ are mapped into an M-ary complex symbol sequence, which is further divided into N_t streams. At every transmit antenna, P pilots are inserted in an OFDM frame. The pilot symbols along with coded data symbols $\{c\}$ are OFDM modulated and transmitted over the multipath fading channel denoted by $\mathbf{h}_{n_t n_r} \in \mathbb{C}^{L \times 1}$ (1.11). Here, n_t (n_r) denotes the transmit (receive) antenna index, and L is the length of the channel. After OFDM demodulation, the signal received at the n_r^{th} receive antenna (for simplicity, we

consider $K = 1$ frame here) given by

$$\mathbf{y}_{n_r} = \sum_{n_t=1}^{N_t} \mathbf{X}_{n_t} \mathbf{F} \mathbf{h}_{n_t n_r} + \mathbf{v}_{n_r}, \quad n_r = 1, \ldots, N_r, \tag{1.12}$$

where the diagonal matrix $\mathbf{X}_{n_t} \in \mathbb{C}^{N \times N}$ consists of the pilot as well as the data transmitted over the n_t^{th} transmit antenna, and $\mathbf{F} \in \mathbb{C}^{N \times L}$ represents the matrix consisting of the first L columns of the $N \times N$ DFT matrix. Each component of $\mathbf{v}_{n_r} \in \mathbb{C}^{N \times 1}$ is a circularly symmetric additive white Gaussian noise with probability distribution $\mathcal{CN}(0, \sigma^2)$. For the case where $N_t = N_r = 1$, the system model given in (1.12) simplifies to that of a SISO-OFDM system given by

$$\mathbf{y} = \mathbf{X} \mathbf{F} \mathbf{h} + \mathbf{v}, \tag{1.13}$$

where $\mathbf{X} \in \mathbb{C}^{N \times N}$ is a diagonal matrix consisting of P pilot symbols and $N - P$ data symbols, $\mathbf{F} \in \mathbb{C}^{N \times L}$ consists of the first L columns of the $N \times N$ Discrete Fourier Transform (DFT) matrix, and $\mathbf{v} \in \mathbb{C}^{N \times 1}$ is the additive white Gaussian noise, distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$. Note that we drop the subscripts n_t and n_r since we set $N_t = N_r = 1$.

Typically, the observations corresponding to the pilot subcarriers in SISO and MIMO-OFDM systems are used to estimate the channel, this is referred to as pilot-based estimation. Note that, at the receiver, the diagonal entries of $\mathbf{X}$ corresponding to the pilot subcarriers are known. Simplifying (1.13) further, we sample $\mathbf{y}$ at the pilot subcarriers to obtain $\mathbf{y}_p$, given by

$$\mathbf{y}_p = \mathbf{X}_p \mathbf{F}_p \mathbf{h} + \mathbf{v}_p, \tag{1.14}$$

By arriving at the signal model in (1.14), we are close to providing an answer to $Q2$. If $\mathbf{h}$ is sparse, then the model in (1.14) resembles that of (1.1). As a one-to-one comparison,

note that the observation at the pilot subcarriers, $\mathbf{y}_p$ corresponds to the $\mathbf{y}$, and the matrix $\mathbf{X}_p \mathbf{F}_p$ corresponds to the known measurement matrix $\mathbf{\Phi}$ in (1.1). The channel estimation problem in (1.14) is ill-posed whenever the number of pilots is smaller than the length of the channel ($P < L$). If the Multipath Intensity Profile (MIP), defined as the average multipath power profile measured at a particular location on a measurement grid [65] of the wireless channel is known, then having $P \geq S$ is sufficient for estimating the channel without the need for SSR techniques. However, the MIP depends not only on the location, but also on a variety of other factors such as the speed of the user (in mobile communications), the presence of obstacles, environmental fluctuations, etc. Moreover, measuring the MIP at all possible locations is impractical. Hence, there is a need to develop methods that can work without requiring the knowledge of the MIP. When the MIP is unknown, SSR techniques are useful for recovering the channel for the following reasons:

- The problem of recovering $\mathbf{h}$ is ill-posed since L can be significantly larger than P. In such scenarios, SSR techniques can be used to estimate the sparse channel.

- In order to use conventional techniques, L pilots are required. However, SSR techniques can recover the sparse vector using $P < L$ pilots. Hence, they allow larger number of data symbols per OFDM symbol, thereby resulting in better spectral efficiency compared to conventional systems.

We note that, upon obtaining an estimate of the sparse channel given by $\hat{\mathbf{h}}$, the information bits $\{\hat{b}\}$ can be decoded by feeding the log-likelihood values to the channel decoder.

1.3.3 Conventional and Sparsity-exploiting OFDM Channel Estimation Techniques

In this subsection, we discuss the conventional channel estimation techniques for OFDM systems, followed by existing sparse channel recovery techniques. Hence, we address $Q3$. We note that the goal here is to provide an introductory overview of the main approaches that have been adopted for OFDM channel estimation in the literature. We refer the readers to later chapters in the thesis for an exhaustive survey of the relevant literature.

Least Squares (LS) and Minimum Mean Square Error (MMSE) techniques are the two well-known lag-domain channel estimation techniques for OFDM systems. The least squares estimate of $\mathbf{h}$ is given by [66]

$$\hat{\mathbf{h}}_{\text{LS}} = (\mathbf{F}_p^H \mathbf{X}_p^H \mathbf{X}_p \mathbf{F}_p)^{-1} \mathbf{F}_p^H \mathbf{X}_p \mathbf{y}_p, \tag{1.15}$$

where $\hat{\mathbf{h}}_{\text{LS}}$ is the solution to an ℓ_2-norm minimization problem. Due to the inverse operation, the LS estimator works only when $P \geq L$, and hence, this method cannot be used in our regime of interest, i.e., when $P < L$. On the other hand, the MMSE estimate is given by [66]

$$\hat{\mathbf{h}}_{\text{MMSE}} = (\mathbf{F}_p^H \mathbf{X}_p^H \mathbf{X}_p \mathbf{F}_p + \mathbf{\Gamma}^{-1})^{-1} \mathbf{F}_p^H \mathbf{X}_p \mathbf{y}_p, \tag{1.16}$$

where $\mathbf{\Gamma} \triangleq \mathbb{E}[\mathbf{h}\mathbf{h}^H]$ denotes the covariance matrix of the channel, also referred to as the *prior information*. In the practical wireless channel scenario, this prior information represents the MIP. Although we can obtain the MMSE estimate when $P < L$, the accuracy of the prior information influences the MMSE estimate. Since the support

information is crucial to the structure of the covariance matrix, reliable information about the support is required to obtain accurate estimates.

LS and MMSE techniques can also be applied in the frequency domain, where the goal is to directly estimate the channel $\mathbf{F}_p\mathbf{h}$ [67]. Such techniques also either require the knowledge of the MIP or require a large number of pilots ($P > L$), when the MIP is not available. Further, even though $\mathbf{h}$ is sparse, $\mathbf{F}_p\mathbf{h}$ is not sparse, and, as a result, frequency domain techniques are not useful if one is interested in SSR-based methods for channel estimation.

CS based sparse OFDM channel estimation using the basis pursuit technique has been proposed in [8]. Approximate inference Bayesian methods have been used to solve the problem of joint channel estimation and data decoding in a Bit Interleaved Coded Modulation (BICM) OFDM system, where the time-varying sparse channel is modeled using a Bernoulli-Gaussian prior [63, 68]. In [69], the authors propose variational algorithms for pilot-based channel estimation. However, none of the proposed techniques are Bayesian exact inference techniques.

In the next subsection, we list some practical issues and challenges in using SBL based techniques for OFDM channel estimation.

1.3.4   Problems Addressed in this Thesis

The work presented in this thesis is motivated by the sparse nature of OFDM channels and the versatility of the SBL based framework, which makes it applicable to a variety of practical scenarios in OFDM. The overall theme of this work is to develop methods that incorporate any available knowledge to efficiently incorporate the prior knowledge

about OFDM system and devise novel SBL based sparse channel recovery algorithms. In this section, we present several practically relevant scenarios for OFDM systems, and state the problems addressed in this thesis. In essence, we provide an answer to question $Q4$ raised earlier.

Approximately Sparse Channels

In practice, the impulse response of a wireless channel is not exactly sparse. The channel models obtained using channel sounding experiments are known to exhibit *approximate* sparsity in the lag-domain [70]. This behavior is due to the non-ideal low-pass and pulse shaping filters employed at the transmitter and the receiver for reducing the spectral leakage and meeting the spectral mask. Further, even without the filtering, the channel is only approximately sparse in practice because it is, in general, non-sample spaced: the nonzero values in $\mathbf{h}(t)$ do not necessarily occur at precisely the sampling instants.

A sample instantiation of the lag-domain channel is approximately sparse (a-sparse) channel used in the simulations and the filtered MIP are depicted in Fig. 1.5. The figure captures the leakage effect due to finite bandwidth sampling and practical filtering. To generate the plot, we have used the Pedestrian B channel model [71] with Rayleigh fading. We have also used RC filtering at the receiver and transmitter with a roll-off factor of 0.5 [61].

We see that at the sampling frequencies considered, the number of significant channel taps are far fewer than the weak channel taps in the filtered impulse response, as seen in Fig. 1.5. Thus, the channel is a-sparse in the lag-domain, and the primary focus of this thesis is to develop SBL-based techniques for such a-sparse channel estimation and
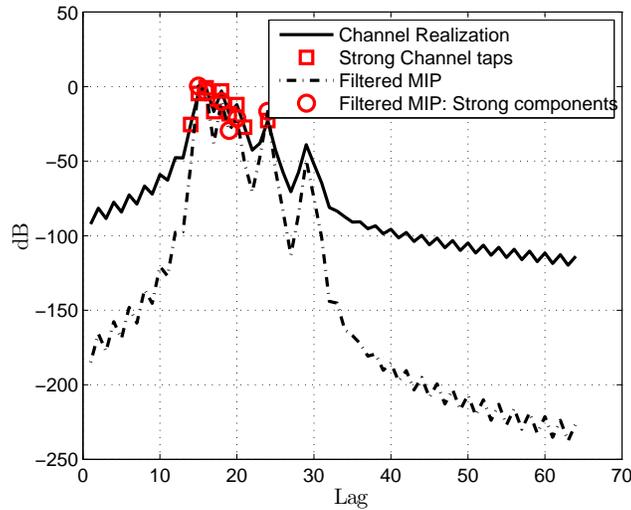
Figure 1.5: One sample channel realization of the a-sparse channel, along with the fil-tered MIP, i.e., the MIP when RC filters are employed at the transmitter and receiver. The plot shows the strong (> -30 dB) channel taps and filtered-MIP components, to illustrate that the channel can indeed be modeled as being approximately sparse.

joint channel estimation and data detection.

Joint Channel Estimation and Data Detection

In this thesis, we not only develop methods for pilot-based a-sparse channel estima-tion, but also for joint channel estimation and data detection. The motivation for jointly estimating the data and the channel comes from [72, 73], where the authors show that the MIP-aware EM based joint channel estimation and data detection techniques for a SISO-OFDM system are far superior compared to the pilot-only LS and the MMSE algo-rithms. The EM algorithm iteratively estimates the channel $\mathbf{h}$ and the unknown trans-mit data in $\mathbf{X}$, i.e., the decisions on the unknown transmit data $\mathbf{X}$ are coalesced with the pilot symbols and used in the channel estimation process. Naturally, if the decisions on the data subcarriers are accurate, then the decisions act as pilots, and hence, improve

the quality of channel estimates. Approximate inference methods have been used to solve the problem of joint channel estimation and data decoding in a BICM-OFDM system, where the time-varying sparse channel is modeled using a Bernoulli-Gaussian prior [63, 68].

In this thesis, we explore such techniques for a-sparse (SISO-OFDM), ga-sparse (MIMO-OFDM) and ga-csparse (SISO and MIMO-OFDM) channels using exact-inference based SBL methods. We provide details regarding the nature of ga-sparse and ga-csparse channels later in the thesis. In particular, the challenge lies in combining the data detection along with channel estimation and obtain an SBL based iterative algorithm with convergence guarantees.

Block-fading and Time-varying Channels

In this thesis, we first design SBL based algorithms for estimating block-fading channels, where the channel coefficients remain fixed across the OFDM frame duration and vary in an i.i.d. fashion from frame to frame. We also address the time-varying nature of the a-sparse channels caused due to the mobility of the receiver. When the channel is time-varying, the nonzero channel coefficients vary slowly and are temporally correlated, but the MIP of the channel remains constant for several OFDM frames [74], and hence, the locations of the significant components coincide in successive channel instantiations. Such channels are *group approximately-sparse* (ga-sparse).

CS techniques have been proposed for the estimation of the time-varying channel over all the symbols in a frame when the channel consists of a few significant nonzero entries but the path delays are unknown [75–77]. Approximate inference techniques for estimating the time-varying sparse vector and support have been proposed in [78]. In

the context of SBL, block-based methods such as Block SBL (BSBL) and Temporal SBL (TSBL) algorithms [79] have been proposed to estimate the time-varying correlated sparse vectors when the correlation among the group-sparse vectors is modeled using a general correlation structure. In this thesis, we use an autoregressive (AR) model to capture the time-varying nature of the channel, and design novel, exact-inference based low complexity Kalman-SBL channel tracking algorithms that exploit the temporally ga-sparse nature of the channel.

Group Approximately-sparse MIMO-OFDM Channels

In MIMO-OFDM systems, identical transmit and receive filters are employed at the N_t transmit antennas and the N_r receive antennas, which leads to a-sparse channel between every pair of transmit and receive antenna. Moreover, such channels are *group sparse*, i.e., the locations of significant paths of the $N_t N_r$ channels coincide. In scenarios where the MIP is not known, blind methods [80] and techniques based on Compressed Sensing (CS) using group-sparse based formulation [81,82] are employed. Specifically, CS based Simultaneous Orthogonal Matching Pursuit (OMP) [83], modified OMP [84], subspace OMP [85], group basis Pursuit Denoising and group OMP [86] have been proposed for ga-sparse pilot-assisted channel estimation in MIMO-OFDM systems.

Although the SISO-OFDM system model given in (1.14) has a one-to-one correspondence with the sparse recovery problem given in (1.1), it is not straightforward to formulate the problem of sparse channel recovery in the MIMO-OFDM framework given by (1.12). In this work, we first formulate the problem of pilot-assisted and joint ga-sparse channel estimation and data detection in MIMO-OFDM systems using the SBL

framework. Later, we design algorithms for both block-fading and time-varying ga-sparse channels.

Cluster-sparse MIMO and SISO-OFDM Channels

In this thesis, we model the MIMO and SISO-OFDM channel as ga-cluster sparse (ga-csparse), i.e., every channel consists of clusters where a few clusters have all strong while most clusters have all weak components. CS based Block OMP (BOMP) has been proposed for pilot-assisted ga-csparse MIMO-OFDM channel estimation [87].

In this thesis, we focus on recovering block-fading and time-varying ga-csparse channels using pilot symbols only, and jointly with the transmit data symbols. Since ga-csparse channels need to estimate a smaller number of hyperparameters compared to ga-sparse channels, we obtain a low-complexity methods for joint channel estimation and data detection.

Thus far, we addressed the several generalizations of the basic OFDM system. In the following section, we list the contents of each chapter along with our contributions, not only to the literature on OFDM channel estimation, but to the theory of SBL as well.

1.4 Contributions and Outline of the Thesis

In this section, we summarize the thesis by providing a detailed outline of the contents of the thesis.

In [88], we considered sparse baseband channel where the discrete multipath components of the channel are located at the sampling instants, and an ideal low pass filter is used. We simulated the SBL and the joint-SBL (J-SBL) algorithm for channel estimation

in an uncoded QPSK modulated SISO-OFDM system, in the presence of a sparse baseband channel. Further, we used BCRB for both the algorithms in order to benchmark the MSE performance of the algorithms. From Fig. 1.6, we see that the BCRB averaged
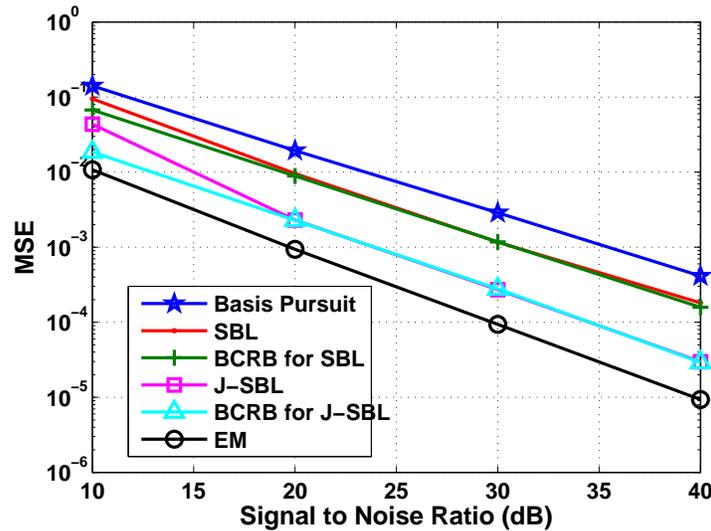
Figure 1.6: SISO-OFDM System: MSE vs. BCRB for SBL and J-SBL.

over all the channel instantiations forms a tight lower bound for both the SBL and the J-SBL algorithms. This motivated us to derive a general class of bounds including the Hybrid, Bayesian and Marginalized Cramér Rao lower bounds for the single and multiple measurement vector SBL problem of estimating compressible vectors and their prior distribution parameters in chapter 2. Through simulations, we demonstrate the dependence of the MSE performance of SBL based estimators on the compressibility of the vector for several values of the number of observations and at different signal powers. The contents of this chapter have been published in parts in [89–91].

Figure 1.7 provides a pictorial overview of the rest of the thesis. In chapter 3 we propose SBL based algorithms for joint a-sparse channel estimation and data detection in
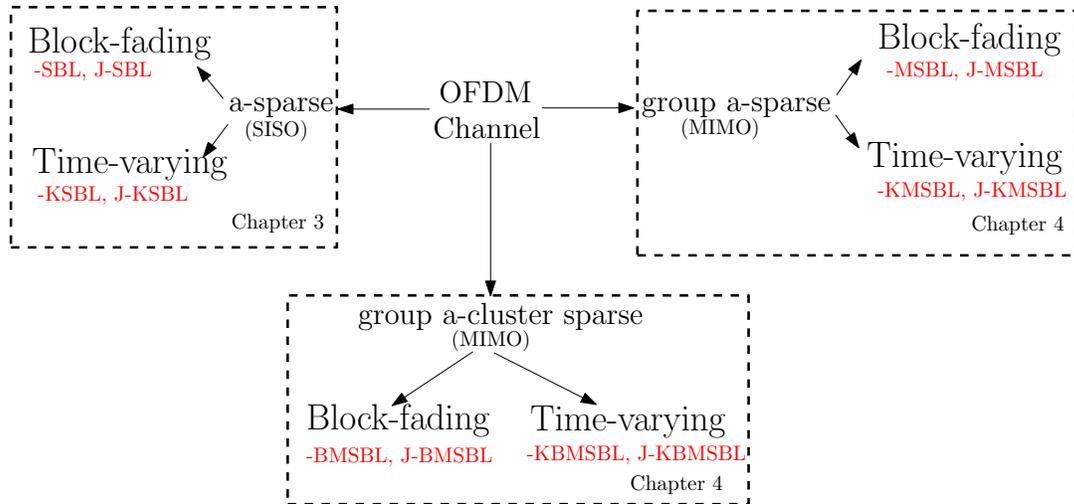
Figure 1.7: OFDM Systems: different scenarios under which pilot-based channel estimation and joint channel estimation and data detection algorithms are proposed in this work.

SISO-OFDM systems. In a quasi-static, block-fading scenario, we employ the SBL algorithm for channel estimation, and propose a Joint SBL (J-SBL) and a low-complexity recursive J-SBL algorithm for joint channel estimation and data detection. In a time-varying scenario, we use a first order auto-regressive model for the wireless channel, and propose a novel, recursive, low-complexity Kalman filtering-based SBL (KSBL) algorithm for channel estimation. We generalize the KSBL algorithm to obtain the recursive Joint KSBL algorithm that performs joint channel estimation and data detection. Monte Carlo simulations illustrate the efficacy of the proposed techniques in terms of the MSE and coded bit error rate performance. The contents of this chapter have been published in [90].

In chapter 4 we propose SBL based algorithms for estimating the ga-sparse and ga-csparse block-fading and time-varying channels. We propose a bouquet of novel algorithms for pilot-based channel estimation and joint channel estimation and data detection in MIMO-OFDM systems. In the time-varying scenario, we employ a first order AR based model for the temporal variation of the wireless ga-sparse and ga-csparse channels and propose a recursive, low-complexity Kalman filtering and smoothing framework for joint channel estimation, tracking and data detection. Monte Carlo simulations illustrate the efficacy of proposed techniques in terms of MSE and coded bit error rate performance. We demonstrate that the algorithms that exploit the ga-csparse nature of the wireless channel offers improvement in the MSE and coded BER performance. The contents of this chapter have been published in [91, 92].

In chapter 5 of this thesis, we provide a compendium of the algorithms proposed in this thesis in a generic SBL framework. We demonstrate the MSE and support recovery performance of the proposed algorithms as compared to the existing CS and SBL based algorithms using Monte Carlo simulations.

1.5 List of Publications

Journal Papers

1. R. Prasad and C. R. Murthy, "Cramér-Rao-Type Bounds for Sparse Bayesian Learning", *IEEE Trans. on Sig. Proc.*, vol. 61, no. 3, pp. 622-632, Mar. 2013.

2. R. Prasad, C. R. Murthy, and B. D. Rao, "Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems using Sparse Bayesian Learning," *IEEE Trans. on Sig. Proc.*, vol. 62, no. 14, pp. 3591-3603, July 2014.

3. R. Prasad, C. R. Murthy, and B. D. Rao, "Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach," *submitted to IEEE Trans. on Sig. Proc.*, Oct. 2014.

Conference Papers

1. R. Prasad and C. R. Murthy, "Bayesian Learning for Joint Sparse OFDM Channel Estimation and Data Detection," *Proc. Global Communications Conference (Globecom)*, Dec. 2010.

2. R. Prasad, B. N. Bharath, and C. R. Murthy, "Joint Data Detection and Dominant Singular Mode Estimation in Time Varying Reciprocal MIMO Systems," Proc. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, Prague, Czech Republic, May 2011, pp. 3240-3243.

3. R. Prasad and C. R. Murthy, "Joint Approximately Group Sparse Channel Estimation and Data Detection in MIMO-OFDM Systems Using Sparse Bayesian Learning," *Proc. National Conference on Communications (NCC)*, IIT Kanpur, India, Feb.

2014. (NCC 2014 best paper award in the communications track.)

4. R. Prasad, C. R. Murthy, and B. Rao, "Nested Sparse Bayesian Learning for Block-Sparse Signals with Intra-Block Correlation", *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

Chapter 2

Cramér Rao-Type Bounds for Sparse Bayesian Learning

2.1 Introduction

Recent results in the theory of compressed sensing have generated immense interest in sparse vector estimation problems, resulting in a multitude of successful practical signal recovery algorithms. In several applications, such as the processing of natural images, audio, and speech, signals are not exactly sparse, but *compressible*, i.e., the magnitudes of the sorted coefficients of the vector follow a power law decay [13]. In [93] and [94], the authors show that random vectors drawn from a special class of probability distribution functions (pdf) known as *compressible priors* result in compressible vectors. Assuming that the vector to be estimated (henceforth referred to as the unknown vector) has a compressible prior distribution enables one to formulate the compressible vector recovery problem in the Bayesian framework, thus allowing the use of Sparse Bayesian Learning (SBL) techniques [34]. In his seminal work, Tipping proposed an SBL algorithm for estimating the unknown vector, based on the Expectation

Maximization (EM) and McKay updates [34]. Since these update rules are known to be slow, fast update techniques are proposed in [95]. A duality based algorithm for solving the SBL cost function is proposed in [96], and $\ell_1 - \ell_2$ based reweighting schemes are explored in [97]. Such algorithms have been successfully employed for image/visual tracking [56], neuro-imaging [57, 58], beamforming [59], and joint channel estimation and data detection for OFDM systems [98].

Many of the aforementioned papers study the complexity, convergence and support recovery properties of SBL based estimators (e.g., [95, 96]). In [94], the general conditions required for the so-called instance optimality of such estimators are derived. However, it is not known whether these recovery algorithms are optimal in terms of the Mean Square Error (MSE) in the estimate or by how much their performance can be improved. In the context of estimating *sparse* signals, Cramér Rao lower bounds on the MSE performance are derived in [99–101]. However, to the best of our knowledge, none of the existing works provide a lower bound on the MSE performance of *compressible* vector estimation. Such bounds are necessary, as they provide absolute yardsticks for comparative analysis of estimators, and may also be used as a criterion for minimization of MSE in certain problems [102]. In this chapter, we close this gap in theory by providing Cramér Rao type lower bounds on the MSE performance of estimators in the SBL framework.

As our starting point, we consider a linear Single Measurement Vector (SMV) SBL model given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{2.1}$$

where the observations $\mathbf{y} \in \mathbb{R}^N$ and the measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times L}$ are known,

and $\mathbf{x} \in \mathbb{R}^L$ is the unknown sparse/compressible vector to be estimated [103]. Each component of the additive noise $\mathbf{n} \in \mathbb{R}^N$ is white Gaussian, distributed as $\mathcal{N}(0, \sigma^2)$, where the variance σ^2 may be known or unknown. The SMV-SBL system model in (2.1) can be generalized to a linear Multiple Measurement Vector (MMV) SBL model given by

$$\mathbf{T} = \boldsymbol{\Phi}\mathbf{W} + \mathbf{V}. \tag{2.2}$$

Here, $\mathbf{T} \in \mathbb{R}^{N \times M}$ represents the M observation vectors, the columns of $\mathbf{W} \in \mathbb{R}^{L \times M}$ are the M sparse/compressible vectors, and each column of $\mathbf{V} \in \mathbb{R}^{N \times M}$ is modeled similar to $\mathbf{n}$ in (2.1) [104]. Since the M vectors in $\mathbf{W}$ have a common underlying compressible distribution, (2.1) is a special case of (2.2) for $M = 1$.

In typical compressible vector estimation problems, $\boldsymbol{\Phi}$ is underdetermined ($N < L$), rendering the problem ill-posed. Bayesian techniques circumvent this problem by using a prior distribution on the compressible vector as a regularization, and computing the corresponding posterior estimate. To incorporate a compressible prior in (2.1) and (2.2), SBL uses a two-stage hierarchical model on the unknown vector, as shown in Fig. 2.1. Here, $\mathbf{x} \sim \mathcal{N}(0, \boldsymbol{\Upsilon})$, where the diagonal matrix $\boldsymbol{\Upsilon}$ contains the *hyperparameters* $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L]^T$ as its diagonal elements. Further, an Inverse Gamma (IG) *hyperprior* is assumed for $\boldsymbol{\gamma}$ itself, because it leads to a Student-t prior on the vector $\mathbf{x}$, which is known to be compressible [34].[1] In scenarios where the noise variance is unknown and random, an IG prior is used for the distribution of the noise variance as well. For the system model in (2.2), every compressible vector $\mathbf{w}_i \sim \mathcal{N}(0, \boldsymbol{\Upsilon})$, i.e., the M compressible vectors are governed by a common $\boldsymbol{\Upsilon}$.

---

[1]The IG hyperprior is conjugate to the Gaussian pdf [34].
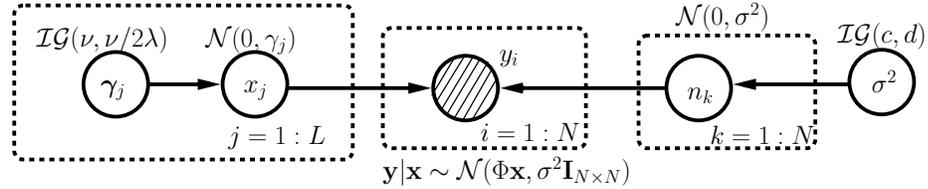
Figure 2.1: Graphical model for SBL: Two stage hierarchical model with the compressible vector taking a conditional Gaussian distribution and the hyperparameters taking an Inverse Gamma distribution. The noise is modeled as white Gaussian distributed, with the noise variance modeled as deterministic/random and known or unknown.

It is well known that the Cramér Rao Lower Bound (CRLB) provides a fundamental limit on the MSE performance of unbiased estimators [105] for deterministic parameter estimation. For the estimation problem in SBL, an analogous bound known as the Bayesian Cramér Rao Bound (BCRB) is used to obtain lower bounds [106], by incorporating the prior distribution on the unknown vector. If the unknown vector consists of both deterministic and random components, Hybrid Cramér Rao Bounds (HCRB) are derived [107].

In SBL, the unknown vector estimation problem can also be viewed as a problem involving nuisance parameters. Since the assumed hyperpriors are conjugate to the Gaussian likelihood, the marginalized distributions have a closed form and the Marginalized Cramér Rao Bounds (MCRB) [108] can be derived. For example, in the SBL hyperparameter estimation problem, $\mathbf{x}$ itself can be considered a nuisance variable and marginalized from the joint distribution, $p_{\mathbf{Y},\mathbf{X}|\boldsymbol{\gamma}}(\mathbf{y}, \mathbf{x}|\boldsymbol{\gamma})$, to obtain the log-likelihood as

$$\mathbf{L}(\boldsymbol{\gamma}) = \log \int_{\mathbf{x}} p_{\mathbf{Y},\mathbf{X}|\Gamma}(\mathbf{y}, \mathbf{x}|\boldsymbol{\gamma})\mathrm{d}\mathbf{x} = \frac{-(\log |\boldsymbol{\Sigma}_y| + \mathbf{y}^T\boldsymbol{\Sigma}_y^{-1}\mathbf{y})}{2}, \tag{2.3}$$
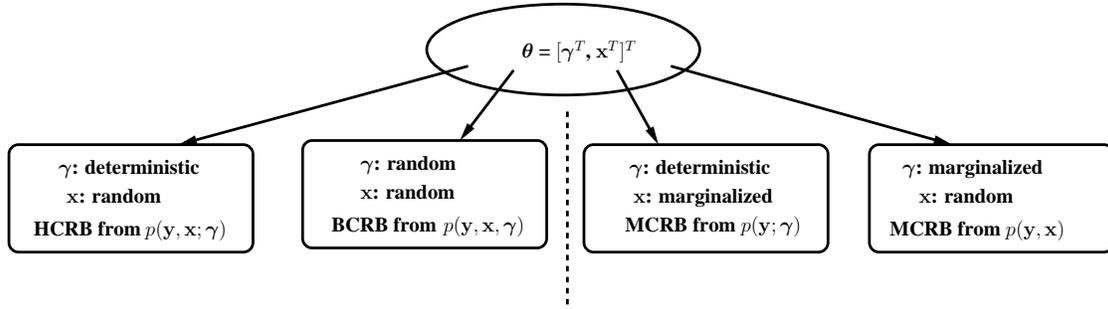
Figure 2.2: Summary of the lower bounds derived in this work when the noise variance is known.

where $\mathbf{\Sigma}_y = \sigma^2 I + \mathbf{\Phi}\mathbf{\Upsilon}\mathbf{\Phi}^T$ [109]. Henceforth, in this chapter, the function $p_X(x)$ represents the pdf of the random variable X evaluated at its realization x.

The goal of this chapter is to derive Cramér Rao type lower bounds on the MSE performance of estimators based on the SBL framework. Our contributions are as follows:

- Under the assumption of known noise variance, we derive the HCRB and the BCRB for the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, as indicated in the left half of Fig. 2.2.

- When the noise variance is known, we marginalize nuisance variables ($\boldsymbol{\gamma}$ or $\mathbf{x}$) and derive the corresponding MCRB, as indicated in the right half of Fig. 2.2. Since the MCRB is a function of the parameters of the hyperprior (and hence is an offline bound), it yields insights into the relationship between the MSE performance of the estimators and the compressibility of $\mathbf{x}$.

- In the unknown noise variance case, we derive the BCRB, HCRB and MCRB for the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T, \sigma^2]^T$, as indicated in Fig. 2.3.

- We derive the MCRB for a general parametric form of the compressible prior [94] and deduce lower bounds for two (Student-t and Generalized double Pareto) of

$$\boldsymbol{\theta} = [\boldsymbol{\gamma}^T, \mathbf{x}^T, \sigma^2]^T$$

$\boldsymbol{\gamma}$: **deterministic** **x: random** σ^2: **random** **HCRB from** $p(\mathbf{y}, \mathbf{x}, \sigma^2; \boldsymbol{\gamma})$	$\boldsymbol{\gamma}$: **random** **x: random** σ^2: **random** **BCRB from** $p(\mathbf{y}, \mathbf{x}, \sigma^2, \boldsymbol{\gamma})$	$\boldsymbol{\gamma}$: **deterministic** **x: marginalized** σ^2: **deterministic** **MCRB from** $p(\mathbf{y}; \boldsymbol{\gamma}, \sigma^2)$
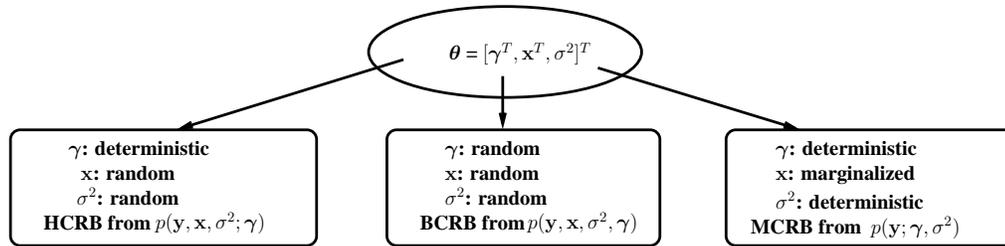
Figure 2.3: Different modeling assumptions and the corresponding bounds derived in this work when the noise variance is unknown.

the well-known compressible priors.

- Similar to the SMV-SBL case, we derive the BCRB, HCRB and MCRB for the MMV-SBL model in (2.2).

Through numerical simulations, we show that the MCRB on the compressible vector x is the tightest lower bound [108], and that the MSE performance of the EM algorithm achieves this bound at high SNR and as $N \to L$. The techniques used to derive the bounds can be extended to handle different compressible prior pdfs used in literature [93]. These results provide a convenient and easy-to-compute benchmark for comparing the performance of the existing estimators, and in some cases, for establishing their optimality in terms of the MSE performance.

The rest of this chapter is organized as follows. In Sec. 2.2, we provide the basic definitions and describe the problem set up. In Secs 2.3 and 2.4, we derive the lower bounds for the cases shown in Figs. 2.2 and 2.3, respectively. The bounds are extended to the MMV-SBL signal model in Sec. 2.5. The efficacy of the lower bounds is graphically illustrated through simulation results in Sec. 2.6. We provide some concluding remarks in Sec. 2.7. In the Appendix, we provide proofs for the Propositions and Theorems stated

in the chapter.

2.2   Preliminaries

As a precursor to the sections that follow, we define the MSE matrix and the Fisher Information Matrix (FIM) [105], and state the assumptions under which we derive the lower bounds in this chapter. Consider a general estimation problem where the unknown vector $\boldsymbol{\theta} \in \mathbb{R}^n$ can be split into sub-vectors $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_d^T]^T$, where $\boldsymbol{\theta}_r \in \mathbb{R}^m$ consists of *random* parameters distributed according to a known pdf, and $\boldsymbol{\theta}_d \in \mathbb{R}^{n-m}$ consists of *deterministic* parameters. Let $\hat{\boldsymbol{\theta}}(\mathbf{y})$ denote the estimator of $\boldsymbol{\theta}$ as a function of the observations $\mathbf{y}$. The MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ is defined as

$$\mathbf{E}^{\boldsymbol{\theta}} \triangleq \mathbb{E}_{\mathbf{Y},\Theta_r} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}))(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}))^T \right], \tag{2.4}$$

where Θ_r denotes the random parameters to be estimated, whose realization is given by $\boldsymbol{\theta}_r$. The first step in obtaining Cramér Rao type lower bounds is to derive the FIM $\mathbf{I}^{\boldsymbol{\theta}}$ [105]. Typically, $\mathbf{I}^{\boldsymbol{\theta}}$ is expressed in terms of the individual blocks of submatrices, where the $(ij)^{th}$ block is given by

$$\mathbf{I}_{ij}^{\boldsymbol{\theta}} \triangleq -\mathbb{E}_{\mathbf{Y},\Theta_r}[\nabla_{\boldsymbol{\theta}_i} \nabla_{\boldsymbol{\theta}_j}^T \log p_{\mathbf{Y},\Theta_r;\Theta_d}(\mathbf{y}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)]. \tag{2.5}$$

In this chapter, we use the notation $\mathbf{I}^{\boldsymbol{\theta}}$ to represent the FIM under the different modeling assumptions. For example, when $\boldsymbol{\theta}_r \neq \emptyset$ and $\boldsymbol{\theta}_d \neq \emptyset$, $\mathbf{I}^{\boldsymbol{\theta}}$ represents a Hybrid Information Matrix (HIM). When $\boldsymbol{\theta}_r \neq \emptyset$ and $\boldsymbol{\theta}_d = \emptyset$, $\mathbf{I}^{\boldsymbol{\theta}}$ represents a Bayesian Information matrix (BIM). Assuming that the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ exists and the FIM is non-singular, a lower

bound on the MSE matrix $\mathbf{E}^{\theta}$ is given by the inverse of the FIM:

$$\mathbf{E}^{\theta} \succeq \left(\mathbf{I}^{\theta}\right)^{-1}. \tag{2.6}$$

It is easy to verify that the underlying pdfs considered in the SBL model satisfy the regularity conditions required for computing the FIM (see Sec. 5.2.3 in [108]).

We conclude this section by making one useful observation about the FIM in the SBL problem. An assumption in the SMV-SBL framework is that $\mathbf{x}$ and $\mathbf{n}$ are independent of each other (for the MMV-SBL model, $\mathbf{T}$ and $\mathbf{W}$ are independent). This assumption is reflected in the graphical model in Fig. 2.1, where the compressible vector $\mathbf{x}$ (and its attribute $\boldsymbol{\gamma}$) and the noise component $\mathbf{n}$ (and its attribute σ^2) are on unconnected branches. Due to this, a submatrix of the FIM is of the form

$$\mathbf{I}_{\gamma\xi}^{\theta} = -\mathbb{E}_{\mathbf{X},\mathbf{Y},\Gamma,\Xi} \left[\nabla_{\gamma}\nabla_{\xi} \left\{\log p_{\mathbf{Y}|\mathbf{X},\Xi}(\mathbf{y}|\mathbf{x},\xi) + \log p_{\mathbf{X},\Gamma}(\mathbf{x},\boldsymbol{\gamma}) + \log p_{\Xi}(\xi)\right\}\right], \tag{2.7}$$

where there are no terms in which both $\boldsymbol{\gamma}$ and $\xi = \sigma^2$ are jointly present. Hence, the corresponding terms in the above mentioned submatrix are always zero. This is formally stated in the following Lemma.

Lemma 1. *When $\boldsymbol{\theta}_i = \boldsymbol{\gamma}$ and $\boldsymbol{\theta}_j = \sigma^2$, the ij^{th} block matrix of the information matrix $\mathbf{I}^{\theta}$, given by (2.5), simplifies to $\mathbf{I}_{ij}^{\theta} = \mathbf{0}$, i.e., to an all zero matrix.*

2.3   SMV-SBL: Lower Bounds when σ^2 is Known

In this section, we derive lower bounds for the system model in (2.1) for the scenarios in Fig. 2.2, where the unknown vector is $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$. We examine different modeling

assumptions on γ and derive the corresponding lower bounds.

2.3.1 Bounds from the Joint pdf

HCRB for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$

In this subsection, we consider the unknown variables as a hybrid of a deterministic vector $\boldsymbol{\gamma}$ and a random vector $\mathbf{x}$ distributed according to a Gaussian distribution parameterized by $\boldsymbol{\gamma}$. Using the assumptions and notation in the previous section, we obtain the following proposition.

Proposition 1. *For the signal model in* (2.1), *the HCRB on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ of the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$ with the parameterized distribution of the compressible signal $\mathbf{x}$ given by $\mathcal{N}(0, \boldsymbol{\Upsilon})$, and with $\boldsymbol{\gamma}$ modeled as unknown and deterministic, is given by $\mathbf{E}^{\boldsymbol{\theta}} \succeq (\mathbf{H}^{\boldsymbol{\theta}})^{-1}$, where*

$$\mathbf{H}^{\boldsymbol{\theta}} \triangleq \begin{bmatrix} \mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x}) & \mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}) \\ (\mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}))^T & \mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} \left(\frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right) & \mathbf{0}_{L \times L} \\ \mathbf{0}_{L \times L} & \mathrm{diag}(2\gamma_1^2,\ 2\gamma_2^2,\ \ldots,\ 2\gamma_L^2)^{-1} \end{bmatrix}. \quad (2.8)$$

Proof: See Appendix A.

Note that the lower bound on the estimate of $\mathbf{x}$ depends on the prior information through the diagonal matrix $\boldsymbol{\Upsilon}$. In the SBL problem, the realization of the random parameter $\boldsymbol{\gamma}$ has to be used to compute the bound above, and hence, it is referred to as an online bound. Note that the lower bound on the MSE matrix of $\mathbf{x}$ is $\mathbf{E}^{\boldsymbol{\theta}} \succeq \left(\frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right)^{-1}$, which is the same as the lower bound on the error covariance of the Bayes vector estimator for a linear model (see Theorems 10.2 and 10.3 in [105]), and is achievable by the MMSE estimator when $\boldsymbol{\Upsilon} = \mathrm{diag}(\gamma_1, \ldots, \gamma_L)$ is known.

BCRB for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$

For deriving the BCRB, a hyperprior distribution is considered on $\boldsymbol{\gamma}$, and the resulting $\mathbf{x}$ is viewed as being drawn from a compressible prior distribution. The most commonly used hyperprior distribution in the literature is the IG distribution [34], where $\gamma_i, i = 1, 2, \ldots, L$ is distributed as $\mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, given by

$$p_\Gamma(\gamma_i) \triangleq \left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1} \left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}} \gamma_i^{\left(-\frac{\nu}{2}-1\right)} \exp\left\{-\frac{\nu}{2\lambda\gamma_i}\right\}; \quad \gamma_i \in (0, \infty), \ \nu, \lambda > 0. \quad (2.9)$$

Using the definitions and notation in the previous section, we state the following proposition.

Proposition 2. *For the signal model in* (2.1), *the BCRB on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ of the unknown random vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, where the conditional distribution of the compressible signal $\mathbf{x}|\boldsymbol{\gamma}$ is $\mathcal{N}(0, \boldsymbol{\Upsilon})$, and the hyperprior distribution on $\boldsymbol{\gamma}$ is $\prod_{i=1}^{L} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, is given by $\mathbf{E}^{\boldsymbol{\theta}} \succeq (\mathbf{B}^{\boldsymbol{\theta}})^{-1}$, where*

$$\mathbf{B}^{\boldsymbol{\theta}} \triangleq \begin{bmatrix} \mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) & \mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}) \\ (\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}))^T & \mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} \left(\frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2} + \lambda \mathbf{I}_{L\times L}\right) & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \frac{\lambda^2(\nu+2)(\nu+7)}{2\nu} \mathbf{I}_{L\times L} \end{bmatrix}. \quad (2.10)$$

Proof: See Appendix B.

It can be seen from $\mathbf{B}^{\boldsymbol{\theta}}$ that the lower bound on the MSE of $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ is a function of the parameters of the IG prior on $\boldsymbol{\gamma}$, i.e., a function of ν and λ, and it can be computed without the knowledge of realization of $\boldsymbol{\gamma}$. Thus, it is an offline bound.

2.3.2 Bounds from Marginalized Distributions

MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$

Here, we derive the MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$, where $\boldsymbol{\gamma}$ is an unknown deterministic parameter. This requires the marginalized distribution $p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$, which is obtained by considering $\mathbf{x}$ as a nuisance variable and marginalizing it out of the joint distribution $p_{\mathbf{X},\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{x},\mathbf{y};\boldsymbol{\gamma})$ to obtain (2.3). Since $\boldsymbol{\gamma}$ is a deterministic parameter, the pdf $p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$ must satisfy the regularity condition in [105]. Using the definitions and notations of the previous sections, we state the following theorem to obtain the MCRB.

Theorem 1. *For the signal model in (2.1), the log likelihood function $\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$ satisfies the regularity conditions. Further, the MCRB on the MSE matrix $\mathbf{E}^{\gamma}$ of the unknown deterministic vector $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$ is given by $\mathbf{E}^{\gamma} \succeq (\mathbf{M}^{\gamma})^{-1}$, where the ij^{th} element of $\mathbf{M}^{\gamma}$ is given by*

$$\mathbf{M}_{ij}^{\gamma} = \frac{1}{2}(\Phi_j^T \boldsymbol{\Sigma}_y^{-1} \Phi_i)^2, \tag{2.11}$$

for $1 \leq i,j \leq L$, where Φ_i is the ith column of $\boldsymbol{\Phi}$, and $\boldsymbol{\Sigma}_y = \sigma^2 \mathbf{I}_{N \times N} + \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^T$, as defined earlier.

Proof: See Appendix C.

To intuitively understand (2.11), we consider a special case of $\boldsymbol{\Phi}^T \boldsymbol{\Phi} = N\mathbf{I}_{N \times N}$, and use the Woodbury's identity to simplify $\boldsymbol{\Sigma}_y^{-1}$, to obtain the $(ii)^{th}$ entry of the matrix $\mathbf{M}^{\gamma}$ as

$$\mathbf{M}_{ii}^{\gamma} = 2\left(\frac{\sigma^2}{N} + \gamma_i\right)^{-2}. \tag{2.12}$$

Hence, the error in γ_i is bounded as $\mathbf{E}_{ii}^{\gamma} \geq 2\left(\frac{\sigma^2}{N} + \gamma_i\right)^2$. As $N \to \infty$, the bound reduces

to $2\gamma_i^2$, which is the same as the lower bound on the estimate of γ obtained as the lower-right submatrix in (2.8). For finite N, the MCRB is tighter than the HCRB.

MCRB for $\boldsymbol{\theta} = [\mathbf{x}]$

In this subsection, we assume a hyperprior on γ, which leads to a joint distribution of x and γ, from which γ can be marginalized. Further, assuming specific forms for the hyperprior distribution can lead to a compressible prior on x. For example, assuming an IG hyperprior on γ leads to an x with a Student-t distribution. In [93], the authors show that sampling from a Student-t distribution with parameters ν and λ results in a ν-compressible x. The Student-t prior is given by

$$p_{\mathbf{X}}(\mathbf{x}) \triangleq \left(\frac{\boldsymbol{\Gamma}((\nu+1)/2)}{\Gamma(\nu/2)}\right)^L \left(\frac{\lambda}{\pi\nu}\right)^{L/2} \prod_{i=1}^{L} \left(1 + \frac{\lambda x_i^2}{\nu}\right)^{-(\nu+1)/2} \; ; \quad x_i \in (-\infty, \infty), \; \nu, \lambda > 0,$$

$$(2.13)$$

where ν represents the number of degrees of freedom and λ represents the inverse variance of the distribution. Using the notation developed so far, we state the following theorem.

Theorem 2. *For the signal model in (2.1), the MCRB on the MSE matrix $\mathbf{E}^{\mathbf{x}}$ of the unknown compressible random vector $\boldsymbol{\theta} = [\mathbf{x}]$ distributed as (2.13), is given by $\mathbf{E}^{\mathbf{x}} \succeq (\mathbf{M}^{\mathbf{x}})^{-1}$, where*

$$\mathbf{M}^{\mathbf{x}} = \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2} + \frac{\lambda(\nu+1)}{(\nu+3)}\mathbf{I}_{L \times L}. \qquad (2.14)$$

Proof: See Appendix D.

We see that the bound derived depends on the parameters of the Student-t pdf. From [94], the prior is *"somewhat"* compressible for $2 < \nu < 4$, and (2.14) is nonnegative and bounded for $2 < \nu < 4$, i.e., the bound is meaningful in the range of ν used in practice.

Note that by choosing λ to be large (or the variance of $\mathbf{x}$ to be small), the bound is dominated by the prior information, rather than the information from the observations, as expected in Bayesian bounds [105].

It is conjectured in [108] that, in general, the MCRB is tighter than the BCRB. Analytically comparing the MCRB (2.14) with the BCRB (2.8), we see that for the SBL problem of estimating a compressible vector, the MCRB is indeed tighter than the BCRB, since

$$\left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \frac{\lambda(\nu+1)}{(\nu+3)}\mathbf{I}_{L\times L}\right)^{-1} \succeq \left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \lambda\mathbf{I}_{L\times L}\right)^{-1}.$$

The techniques used to derive the bounds in this subsection can be applied to any family of compressible distributions. In [94], the authors propose a parametric form of the Generalized Compressible Prior (GCP) and prove that such a prior is compressible for certain values of ν. In the following subsection, we derive the MCRB for the GCP.

2.3.3 General Marginalized Bounds

In this subsection, we derive MCRBs for the parametric form of the GCP. The GCP encompasses the double Pareto shrinkage type prior [110] and the Student-t prior (2.13) as its special cases. We consider the GCP on $\mathbf{x}$ as follows

$$p_{\mathbf{x}}(\mathbf{x}) \triangleq K^L \prod_{i=1}^{L} \left(1 + \frac{\lambda|x_i|^\tau}{\nu}\right)^{-(\nu+1)/\tau} \; ; \;\; x_i \in (-\infty, \infty), \;\; \tau, \nu, \lambda > 0, \tag{2.15}$$

where the normalizing constant $K \triangleq \frac{\tau}{2}\left(\frac{\lambda}{\nu}\right)^{1/\tau}\frac{\Gamma((\nu+1)/\tau)}{\Gamma(1/\tau)\Gamma(\nu/\tau)}$. When $\tau = 2$, the above distribution reduces to the Student-t prior as given in (2.13), and when $\tau = 1$, it reduces to a generalized double Pareto shrinkage prior [110]. Note that the expression for the GCP in [94] can be obtained from (2.15) by setting $\lambda = 1$, and defining $\nu \triangleq s - 1$. The following theorem provides the MCRB for the GCP.

Theorem 3. *For the signal model in (2.1), the MCRB on the MSE matrix $\mathbf{E}_\tau^\theta$ of the unknown random vector $\boldsymbol{\theta} = [\mathbf{x}]$, where $\mathbf{x}$ is distributed by a GCP in (2.15) is given by $\mathbf{E}_\tau^\theta \succeq (\mathbf{M}_\tau^\theta)^{-1}$, where*

$$\mathbf{M}_\tau^\theta = \frac{\mathbf{\Phi}^T \mathbf{\Phi}}{\sigma^2} + T_\tau, \tag{2.16}$$

where $T_\tau = \frac{\tau^2(\nu+1)}{(\nu+\tau+1)} \left(\frac{\lambda}{\nu}\right)^{2/\tau} \frac{\Gamma\left(\frac{\nu+2}{\tau}\right)\Gamma\left(2-\frac{1}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)\Gamma\left(\frac{v}{\tau}\right)} \mathbf{I}_{L\times L}$.

Proof: See Appendix E.

It is straightforward to verify that for $\tau = 2$, (2.16) reduces to the MCRB derived in (2.14) for the Student-t distribution. For $\tau = 1$, the inverse of the MCRB can be reduced to

$$\mathbf{M}_\tau^\theta = \frac{\mathbf{\Phi}^T \mathbf{\Phi}}{\sigma^2} + \frac{\lambda^2(\nu+1)^2}{\nu(\nu+2)} \mathbf{I}_{L\times L}. \tag{2.17}$$

Hence, this method is useful in obtaining a Cramér Rao type lower bound for the estimators based on the double Pareto shrinkage prior, which uses the generalized prior with $\tau = 1$ [110, 111].

Further, we plot the expression (2.16) in Fig. 2.4 and observe that, in general, the bounds predict an increase in MSE for higher values of τ. Also, the lower bounds at different signal to noise ratios (SNRs) converge as the value of τ increases at a given value of N, indicating that increasing τ renders the bound insensitive to the SNR. The lower bounds also predict a smaller value of MSE for a lower value of ν.

Thus far, we have presented the lower bounds on the MSE in estimating the unknown parameters of the SBL problem when the noise variance is known. In the next section, we extend the results to the case of unknown noise variance.
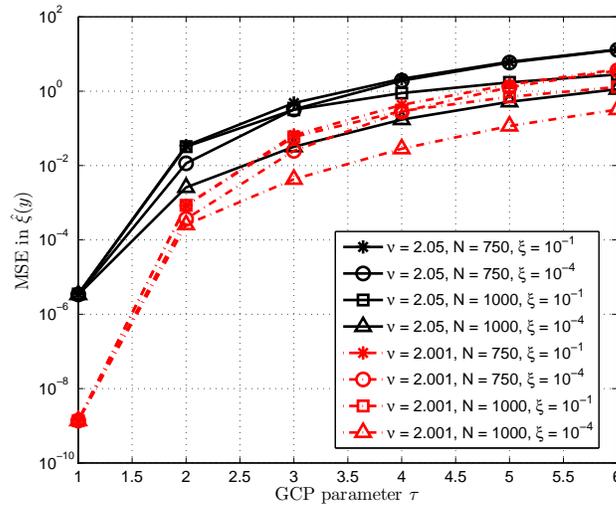
Figure 2.4: Behavior of the MCRB (2.16) for the parametric form of the GCP, as a function of τ, ν, N and noise variance ξ.

2.4 SMV-SBL: Lower Bounds when σ^2 is Unknown

Let us denote the unknown noise variance as $\xi = \sigma^2$. In the Bayesian formulation, the noise variance is associated with a prior, and since the IG prior is conjugate to the Gaussian likelihood $p_{\mathbf{Y}|\mathbf{X},\Xi}(\mathbf{y}|\mathbf{x},\xi)$, it is assumed that $\sigma^2 \sim \mathcal{IG}(c,d)$ [34], i.e., $\xi = \sigma^2$ is distributed as

$$p_\Xi(\xi) \triangleq \frac{d^c}{\Gamma(c)} \xi^{(-c-1)} \exp\left\{-\frac{d}{\xi}\right\}; \quad \xi \in (0,\infty),\ c,d > 0. \tag{2.18}$$

Under this assumption, one can marginalize the unknown noise variance and obtain the likelihood $p(\mathbf{y}|\mathbf{x})$ as

$$p(\mathbf{y}|\mathbf{x}) \triangleq \int_{\xi=0}^{\infty} p(\mathbf{y},\xi|\mathbf{x})\mathrm{d}\xi = \frac{(2d)^c \Gamma\left(\frac{N}{2}+c\right)}{\Gamma(c)(\pi)^{N/2}} \left((\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})^T(\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})+2d\right)^{-\left(\frac{N}{2}+c\right)}, \tag{2.19}$$

which is a multivariate Student-t distribution. It turns out that the straightforward approach of using the above multivariate likelihood to directly compute lower bounds for the various cases given in the previous section is analytically intractable, and that the lower bounds cannot be computed in closed form. Hence, we compute lower bounds from the *joint* pdf, i.e., we derive the HCRB and BCRBs for the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T, \xi]^T$ with the MSE matrix $\mathbf{E}_\xi^\theta$ defined by (2.4).[2] Using the assumptions and notation from the previous sections, we obtain the following proposition.

Proposition 3. *For the signal model in (2.1), the HCRB on the MSE matrix $\mathbf{E}_\xi^\theta$ of the unknown vector $\boldsymbol{\theta} = [\boldsymbol{\theta}'^T, \xi]^T$, where $\boldsymbol{\theta}' = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, with the distribution of the compressible vector $\mathbf{x}$ given by $\mathcal{N}(0, \boldsymbol{\Upsilon})$, where $\boldsymbol{\gamma}$ is modeled as a deterministic or as a random parameter distributed as $\prod_{i=1}^{L} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, and ξ is modeled as a deterministic parameter, is given by $(\mathbf{H}_\xi^\theta)^{-1}$, where*

$$\mathbf{H}_\xi^\theta = \begin{bmatrix} \mathbf{H}^{\theta'} & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{1\times L} & \frac{N}{2\xi^2} \end{bmatrix}. \tag{2.20}$$

In the above expression, with a slight abuse of notation, $\mathbf{H}^{\theta'}$ is the FIM given by (2.8) when $\boldsymbol{\gamma}$ is unknown deterministic and by (2.10) when $\boldsymbol{\gamma}$ is random.

Proof: See Appendix F.

The lower bound on the estimation of ξ matches with the well known lower bounds on noise variance estimation (see Sec. 3.5 in [105]). One disadvantage of such a bound on $\hat{\xi}(\mathbf{y})$ is that the knowledge of the noise variance is essential to compute the bound, and hence, it cannot be computed offline. Instead, assigning a hyperprior to ξ would result in a lower bound that only depends on the parameters of the hyperprior, which

---

[2]We use the subscript ξ to indicate that the error matrices and bounds are obtained for the case of unknown noise variance.

are assumed to be known, allowing the bound to be computed offline. We state the following proposition in this context.

Proposition 4. *For the signal model in (2.1), the HCRB on the MSE matrix $\mathbf{E}_\xi^\theta$ of the unknown vector $\boldsymbol{\theta} = [\boldsymbol{\theta'}^T, \xi]^T$, where $\boldsymbol{\theta'} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, with the distribution of the vector $\mathbf{x}$ given by $\mathcal{N}(0, \boldsymbol{\Upsilon})$, where $\boldsymbol{\gamma}$ is modeled as a deterministic or as a random parameter distributed as $\prod_{i=1}^{L} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, and the random parameter ξ is distributed as $\mathcal{IG}(c, d)$, is given by $(\mathbf{H}_\xi^\theta)^{-1}$, where*

$$\mathbf{H}_\xi^\theta = \begin{bmatrix} \mathbf{H}^{\theta'} & \mathbf{0}_{L \times 1} \\ \mathbf{0}_{1 \times L} & \frac{c(c+1)(N/2+c+3)}{d^2} \end{bmatrix}. \tag{2.21}$$

In (2.21), $\mathbf{H}^{\theta'}$ is the FIM given in (2.8) when $\boldsymbol{\gamma}$ is unknown deterministic and by (2.10) when $\boldsymbol{\gamma}$ is random.

Proof: See Appendix G.

In SBL problems, a non-informative prior on ξ is typically preferred, i.e., the distribution of the noise variance is modeled to be as flat as possible. In [34], it was observed that a non-informative prior is obtained when $c, d \to 0$. However, as $c, d \to 0$, the bound in (2.21) is indeterminate. In Sec. 2.6, we illustrate the performance of the lower bound in (2.21) for practical values of c and d.

2.4.1 Marginalized Bounds

In this subsection, we obtain lower bounds on the MSE of the estimator $\hat{\xi}(\mathbf{y})$, in the presence of nuisance variables in the joint distribution. To start with, we consider the marginalized distributions of $\boldsymbol{\gamma}$ and ξ, i.e., $p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y};\boldsymbol{\gamma},\xi)$ where both, $\boldsymbol{\gamma}$ and ξ are deterministic variables. Since the unknowns are deterministic, the regularity condition has

to be satisfied for $\boldsymbol{\theta} = [\boldsymbol{\gamma}^T, \xi]^T$. We state the following theorem.

Theorem 4. *For the signal model in (2.1), the log likelihood function* $\log p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y};\boldsymbol{\gamma},\xi)$ *satisfies the regularity condition [105]. Further, the MCRB on the MSE matrix* $\mathbf{E}_\xi^{\boldsymbol{\theta}}$ *of the unknown deterministic vector* $\boldsymbol{\theta} = [\boldsymbol{\gamma}^T, \xi]^T$ *is given by* $\mathbf{E}_\xi^{\boldsymbol{\theta}} \succeq (\mathbf{M}_\xi^{\boldsymbol{\theta}})^{-1}$, *where*

$$\mathbf{M}_\xi^{\boldsymbol{\theta}} \triangleq \begin{bmatrix} \mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) & \mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\xi) \\ \mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi,\boldsymbol{\gamma}) & \mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi) \end{bmatrix}, \tag{2.22}$$

where the ij^{th} *entry of the matrix* $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ *is given by* $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}))_{ij} = \frac{1}{2}\left\{(\Phi_j^T\Sigma_y^{-1}\Phi_i)^2\right\}$, *and* $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi) = \frac{1}{2}Tr(\Sigma_y^{-2})$. *Further,* $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\xi))_i = (\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi,\boldsymbol{\gamma}))_i = \frac{\Phi_i^T\Sigma_y^{-2}\Phi_i}{2}$, $i,j = 1, 2, \ldots, L$.

Proof: See Appendix H

Remark: From the graphical model in Fig. 2.1, it can be seen that the branches consisting of γ_i and ξ are independent conditioned on $\mathbf{x}$. However, when $\mathbf{x}$ is marginalized, the nodes ξ and γ_i are connected, and hence, Lemma 2 is no longer valid. Due to this, the lower bound on $\boldsymbol{\gamma}$ depends on ξ and vice versa, i.e., $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ and $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi)$ depend on both ξ and $\boldsymbol{\Upsilon} = \text{diag}(\boldsymbol{\gamma})$ through $\Sigma_y = \xi\mathbf{I}_{N\times N} + \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^T$.

Thus far, we have presented several bounds for the MSE performance of the estimators $\hat{\mathbf{x}}(\mathbf{y})$, $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ and $\hat{\xi}(\mathbf{y})$ in the SMV-SBL framework. In the next section, we derive Cramér Rao type lower bounds for the MMV-SBL signal model.

2.5 Lower Bounds for the MMV-SBL

In this section, we provide Cramér Rao type lower bounds for the estimation of unknown parameters in the MMV-SBL model given in (2.2). We consider the estimation of the compressible vector $\mathbf{w}$ from the vector of observations $\mathbf{t}$, which contain the

stacked columns of $\mathbf{W}$ and $\mathbf{T}$, respectively. In the MMV-SBL model, each column of

$\mathbf{W}$ is distributed as $\mathbf{w}_i \sim \mathcal{N}(0, \mathbf{\Upsilon})$, for $i = 1, \ldots M$, and the likelihood is given by

$\prod_{i=1}^{M} p_{\mathbf{T}|\mathbf{w}_i, \Xi}(\mathbf{t}_i|\mathbf{w}_i, \xi)$, where $p_{\mathbf{T}|\mathbf{w}_i \Xi}(\mathbf{t}_i|\mathbf{w}_i, \xi) = \mathcal{N}(\mathbf{\Phi}\mathbf{w}_i, \xi)$ and $\xi = \sigma^2$. The modeling

assumptions on γ and ξ are the same as in the SMV-SBL case, given by (2.9) and (2.18),

respectively [104].

Using the notation developed in Sec. 2.2, we derive the bounds for the MMV SBL case

similar to the SMV-SBL cases considered in Secs. 2.3 and 2.4. Since the derivation of

these bounds follow along the same lines as in the previous sections, we simply state

results in Table 2.1.

Bound Derived	Expression
HCRB on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$	$\mathbf{H}_M^{\boldsymbol{\theta}} = \text{diag}\left(\frac{M}{2\gamma_i^2}\right), i = 1, 2 \ldots, L$
BCRB on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$	$\mathbf{B}_M^{\boldsymbol{\theta}} = \frac{\lambda^2(\nu+2)(M+\nu+6)}{2\nu}\mathbf{I}_{L \times L}$
MCRB on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$	$\mathbf{M}_M^{\boldsymbol{\theta}} = [\mathbf{M}_{ij}^{\boldsymbol{\theta}}],$ where $\mathbf{M}_{ij}^{\boldsymbol{\theta}} = \frac{M}{2}(\Phi_j^T \boldsymbol{\Sigma}_y^{-1} \Phi_i)^2$
HCRB on $\hat{\mathbf{w}}(\mathbf{y})$	$\mathbf{H}_M^{\boldsymbol{\theta}} = \left(\frac{\mathbf{\Phi}^T\mathbf{\Phi}}{\sigma^2} + \mathbf{\Upsilon}^{-1}\right) \otimes \mathbf{I}_{M \times M}$
BCRB on $\hat{\mathbf{w}}(\mathbf{y})$	$\mathbf{B}_M^{\boldsymbol{\theta}} = \left(\frac{\mathbf{\Phi}^T\mathbf{\Phi}}{\sigma^2} + \lambda\mathbf{I}_{L \times L}\right) \otimes \mathbf{I}_{M \times M}$
HCRB on $\hat{\xi}(\mathbf{y})$	$\mathbf{H}_{M,\xi}^{\boldsymbol{\theta}} = \left(\frac{MN}{2\xi^2}\right)$
BCRB on $\hat{\xi}(\mathbf{y})$	$\mathbf{B}_{M,\xi}^{\boldsymbol{\theta}} = \frac{c\left(\frac{MN}{2}+c+3\right)(c+1)}{d^2}$
MCRB on $[\hat{\boldsymbol{\gamma}}(\mathbf{y})^T, \hat{\xi}(\mathbf{y})]^T$	$\mathbf{M}_{M,\xi}^{\boldsymbol{\theta}} = M \times \mathbf{M}_{\xi}^{\boldsymbol{\theta}}$

Table 2.1: Cramér Rao Type Bounds for the MMV-SBL Case.

We see that the lower bounds on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ and $\hat{\xi}(\mathbf{y})$ are reduced by a factor of M compared

to the SMV case. This is intuitively satisfying, since a higher number of observations are

available for the estimation of the parameters γ and ξ. It turns out that it is not possible

to obtain the MCRB on $\mathbf{w}$ in the MMV-SBL setting, since closed form expressions for

the FIM are not available.

In the next section, we consider two algorithms for SBL, namely the EM algorithm and the ARD based reweighted ℓ_1 algorithm, and numerically illustrate the utility of the lower bounds.

2.6   Simulations and Discussion

The vector estimation problem in the SBL framework typically involves the joint estimation of the hyperparameter and the unknown compressible vector **x**. Since the hyperparameter estimation problem cannot be solved in closed form, iterative estimators are employed [34]. In this section, we consider the iterative updates based on the EM algorithm first proposed in [34]. We also consider the algorithm proposed in [96] based on the Automatic Relevance Determination (ARD) framework. We plot the MSE performance in estimating **x**, γ and ξ with the linear model in (2.1) and (2.2), for the EM algorithm, labeled EM, and the ARD based Reweighted ℓ_1 algorithm, labeled ARD-SBL. We compare the performance of the estimators against the derived lower bounds.

We simulate the lower bounds for a random underdetermined ($N < L$) measurement matrix $\mathbf{\Phi}$, whose entries are *i.i.d.* and standard Bernoulli ($\{+1, -1\}$) distributed. A compressible signal of dimension L is generated by sampling from a Student-t distribution with the value of ν ranging from 2.01 to 2.05, which is the range in which the signal is *"somewhat"* compressible, for high dimensional signals [94]. Figure 2.5 shows the decay profile of the sorted magnitudes of $L = 1024$ *i.i.d.* samples drawn from a Student-t distribution for different degrees of freedom and with the value of $\mathbb{E}(x_i^2)$ fixed at 10^{-3}.
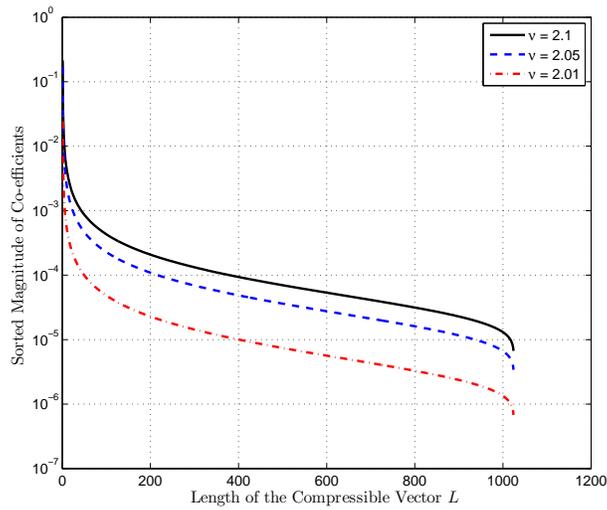
Figure 2.5: Decay profile of the sorted magnitudes of *i.i.d.* samples drawn from a Student-*t* distribution.
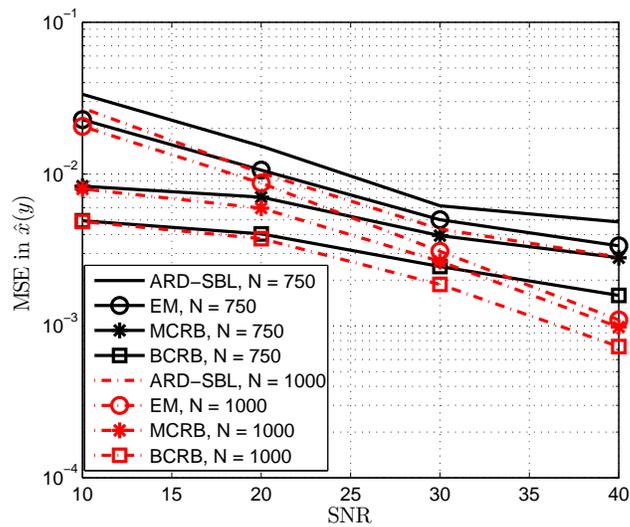
Figure 2.6: Plot of the MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$, the corresponding MCRB and BCRB as a function of SNR, where $\nu = 2.01$.

2.6.1 Lower Bounds on the MSE Performance of $\hat{\mathbf{x}}(\mathbf{y})$

In this subsection, we compare the MSE performance of the ARD-SBL estimator and the EM based estimator $\hat{\mathbf{x}}(\mathbf{y})$. Figure 2.6 depicts the MSE performance of $\hat{x}(y)$ for different SNRs and $N = 750$ and 1000, with $\nu = 2.01$. We compare it with the HCRB/BCRB derived in (2.8), which is obtained by assuming the knowledge of the realization of the hyperparameters $\boldsymbol{\gamma}$. We see that the MCRB derived in (2.14) is a tight lower bound on the MSE performance at high SNR and N.
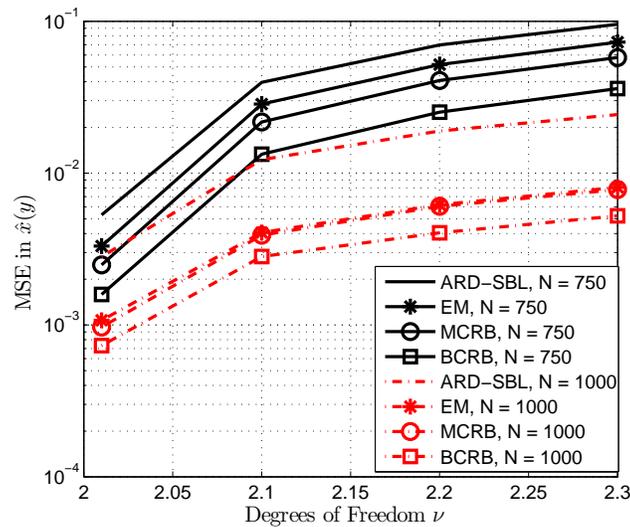
Figure 2.7: Plot of the MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$, the corresponding MCRB and BCRB as a function of ν, where SNR = 40dB.

Figure 2.7 shows the comparative MSE performance of the ARD-SBL estimator and EM based estimator as a function of varying degrees of freedom ν, at an SNR of 40dB and $N = 1000$ and 750. As expected, the MSE performance of the algorithms is better at low values of ν since the signal is more compressible, and the MCRB and BCRB also reflect this behavior. The MCRB is a very tight lower bound, especially for high values
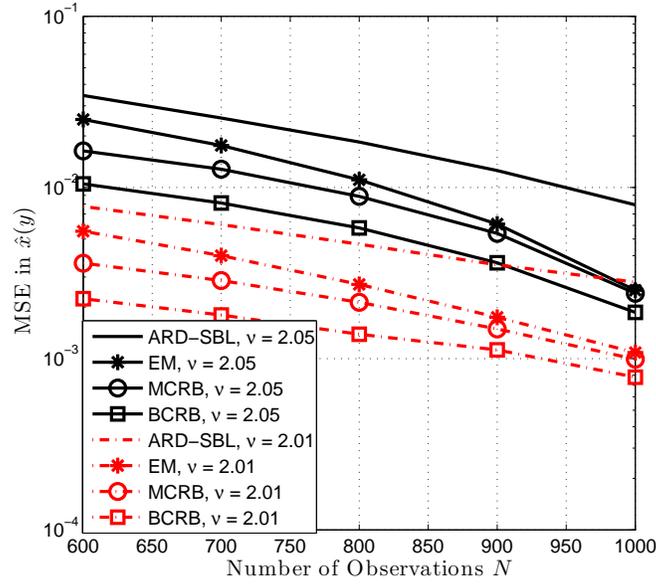
Figure 2.8: Plot of the MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$, the corresponding MCRB and BCRB as a function of N, where SNR = 40dB.

of N. Figure 2.8 shows the comparative MSE performance of the ARD-SBL estimator and EM based estimator as a function of N, at an SNR of 40dB and for two different values of ν. The MSE performance of the EM algorithm converges to that of the MCRB at higher N.

2.6.2   Lower Bounds on the MSE Performance of $\hat{\boldsymbol{\gamma}}(\mathbf{y})$

In this subsection, we compare the different lower bounds for the MSE of the estimator $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ for the SMV and MMV-SBL system model. Figure 2.9 shows the MSE performance of $\hat{\boldsymbol{\gamma}}(y)$ as a function of SNR and M, when γ is a random parameter, $N = 1000$ and $\nu = 2.01$. In this case, it turns out that there is a large gap between the performance of the EM based estimate and the lower bound.
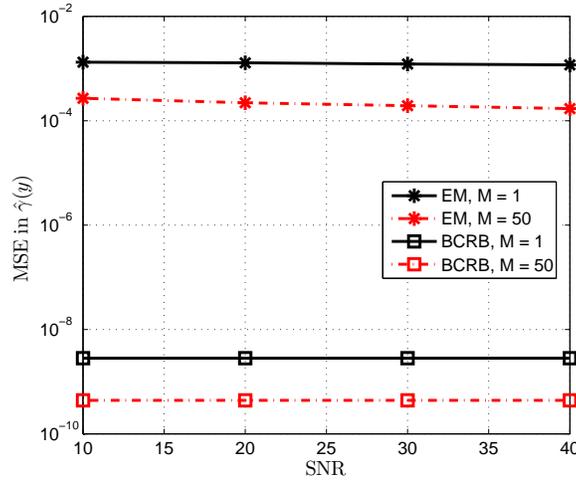
Figure 2.9: Plot of the MSE performance of $\hat{\gamma}(\mathbf{y})$ and the corresponding HCRB as a function of SNR, where $N = 1000$.

When γ is deterministic, we first note that the EM based ML estimator for γ is asymptotically optimal and the lower bounds are practical for large data samples [105]. The results are listed in Table 2.2. We see that for $L = 2048$ and $N = 1500$, the MCRB and BCRB are tight lower bounds, with MCRB being marginally tighter than the BCRB. However, as M increases, the gap between the MSE and the lower bounds increases.

SNR(dB)		10	20	30	40
$M = 1$	MSE	0.054	0.053	0.051	0.050
	MCRB	0.052	0.051	0.050	0.049
	HCRB	0.049	0.049	0.049	0.049
$M = 50$	MSE	0.0450	0.039	0.035	0.030
	MCRB $\times 10^{-2}$	0.120	0.110	0.100	0.090
	HCRB $\times 10^{-3}$	0.977	0.977	0.977	0.977

Table 2.2: Values of the MSE of the Estimator $\hat{\gamma}(\mathbf{y})$, the MCRB and the HCRB, for $\boldsymbol{\theta}_d = [\gamma]$ as a Function of SNR, for $N = 1500$.

2.6.3 Lower Bounds on the MSE Performance of $\hat{\xi}(\mathbf{y})$

In this subsection, we compare the lower bounds on the MSE of the estimator $\hat{\xi}(\mathbf{y})$ in the SMV and MMV-SBL setting. Figure 2.10 shows the MSE performance of $\hat{\xi}(\mathbf{y})$ and the corresponding HCRB for different values of N and M. Here, ξ is sampled from the IG pdf (2.18), with parameters $c = 3$ and $d = 0.2$.
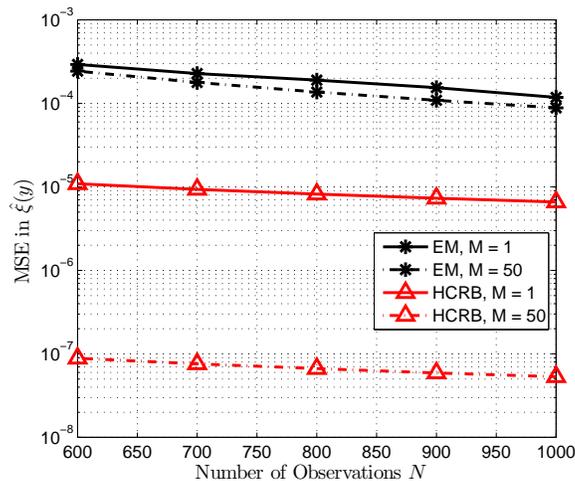
Figure 2.10: Plot of MSE performance of $\hat{\gamma}(\mathbf{y})$ along with the HCRB as a function of N.

When ξ is deterministic, the EM based ML estimator for $\boldsymbol{\gamma}$ is asymptotically optimal and the lower bounds are practical for large data samples [105]. Table 2.3 lists the MSE values of $\hat{\xi}(\mathbf{y})$, the corresponding HCRB and MCRB for deterministic but unknown noise variance, while the true noise variance is fixed at 10^{-3}. We see that for $L = 2048$ and $N = 1500$, the MCRB is marginally tighter than the HCRB. However, when the noise variance is random, we see from Fig. 2.10 that there is a large gap between the MSE performance and the HCRB.

	N	1500	1600	1700	1800
$M = 1$	MSE $\times 10^{-8}$	0.736	0.663	0.636	0.592
	MCRB$\times 10^{-8}$	0.380	0.340	0.307	0.279
	HCRB$\times 10^{-8}$	0.133	0.125	0.118	0.111
$M = 50$	MSE $\times 10^{-9}$	0.930	0.892	0.866	0.847
	MCRB$\times 10^{-10}$	0.680	0.652	0.614	0.573
	HCRB$\times 10^{-10}$	0.267	0.250	0.235	0.222

Table 2.3: Values of the MSE of the Estimator $\hat{\xi}(\mathbf{y})$, the MCRB and the HCRB for $\theta_d = [\xi]$, as a Function of N.

2.7 Conclusions

In this work, we derived Cramér Rao type lower bounds on the MSE, namely, the HCRB, BCRB and MCRB, for the SMV-SBL and the MMV-SBL problem of estimating compressible signals. We used a hierarchical model for the compressible priors to obtain the bounds under various assumptions on the unknown parameters. The bounds derived by assuming a hyperprior distribution on the hyperparameters themselves provided key insights into the MSE performance of SBL and the values of the parameters that govern these hyperpriors. We derived the MCRB for the generalized compressible prior distribution, of which the Student-t and Generalized Pareto prior distribution are special cases. We showed that the MCRB is tighter than the BCRB. We compared the lower bounds with the MSE performance of the ARD-SBL and the EM algorithm using Monte Carlo simulations. The numerical results illustrated the near-optimality of EM based updates for SBL, which makes it interesting for practical implementations.

Chapter 3

Joint Sparse Channel Estimation and Data Detection in SISO-OFDM Systems using Sparse Bayesian Learning

3.1 Introduction and System Model

In practice, wireless channels have a large delay spread with a few significant channel tap coefficients, and therefore, the channel is *approximately sparse* (a-sparse) in the lag domain. Several papers in literature have proposed sparse channel estimation techniques (see [112–114] and references therein). In the context of channel estimation for OFDM systems, spectrally efficient techniques (for which $P < L$, where P is the number of pilots and L is the length of the channel) that leverage this approximate sparsity using Compressed Sensing (CS) [5] have been proposed [8,63,70,114–116]. In this work, we propose to formulate the problem of channel estimation in a Sparse Bayesian Learning (SBL) framework [34, 49]. Specifically, we design novel SBL algorithms for OFDM systems in the following scenarios: (i) The block-fading case, where the channel coefficients remain fixed across the OFDM frame duration and vary in an i.i.d. fashion from

frame to frame; and (ii) the time-varying case, where the channel coefficients across successive OFDM symbols are temporally correlated but have a common support.

3.1.1 Problem Formulation and Contributions

In this subsection, we cast the channel estimation problem in the SBL framework and describe the contributions of this work. In an OFDM system with N subcarriers, the instantaneous received signal, denoted by $\mathbf{y} \in \mathbb{C}^{N \times 1}$, is mathematically represented as [64]

$$\mathbf{y} = \mathbf{X}\mathbf{F}\mathbf{h} + \mathbf{v}. \tag{3.1}$$

Here, $\mathbf{F} \in \mathbb{C}^{N \times L}$ ($N > L$) contains the first L columns of the $N \times N$ Discrete Fourier Transform (DFT) matrix, $\mathbf{h} \in \mathbb{C}^{L}$ is the channel impulse response. The dictionary matrix is given by $\boldsymbol{\Phi} = \mathbf{X}\mathbf{F}$, where the diagonal matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$ contains the N transmitted symbols comprising both known pilot symbols and unknown M-PSK/M-QAM modulated data along the diagonal. Each component of $\mathbf{v} \in \mathbb{C}^{N \times 1}$ is a zero mean circularly symmetric additive white Gaussian noise with pdf denoted by $\mathcal{CN}(0, \sigma^2)$, where σ^2 is the noise variance. Typically, the communication between the transmitter and the receiver occurs in frames consisting of K consecutive OFDM symbols. Suppose that, in a given OFDM symbol, P of the N subcarrier locations are pilot subcarriers and the remaining $(N - P)$ subcarriers carry unknown data symbols. The system model pertaining to the *pilot* subcarriers can be written as

$$\mathbf{y}_p = \mathbf{X}_p \mathbf{F}_p \mathbf{h} + \mathbf{v}_p, \tag{3.2}$$

where $\mathbf{y}_p$ is a $P \times 1$ vector containing the entries of $\mathbf{y}$ sampled at pilot locations, $\mathbf{X}_p$ is a $P \times P$ diagonal matrix with the known pilot symbols along its diagonal, $\mathbf{F}_p$ is the $P \times L$ ($P < L$) submatrix of $\mathbf{F}$ consisting of the rows corresponding to the pilot locations and $\mathbf{v}_p$ is a $P \times 1$ vector, again consisting of components of $\mathbf{v}$ sampled at pilot locations.

As elucidated in chapter 1, the complex baseband representation of the scalar channel impulse response $\tilde{h}[t], t \in \mathbb{R}$ can be modeled as a stationary tapped delay line filter in the lag-domain:

$$\tilde{h}[t] = \sum_{l=1}^{\tilde{L}} \tilde{h}_l \delta[t - \tau_l], \tag{3.3}$$

where $\delta[t]$ is the Dirac delta function, $\tilde{h}_l$ and τ_l represent the attenuation and propagation delay between the transmitter and the receiver path l, respectively, and $\tilde{L}$ is the number of paths [62]. It is known that the wireless channel models obtained using channel sounding experiments exhibit approximate sparsity in the lag-domain (for e.g., due to non-perfect low-pass filtering using raised cosine filtering), as the communication bandwidth and sampling frequency increase [70]. Hence, based on these practical considerations, we consider the lag-domain filtered channel impulse response, which can be represented as $h[t] = g_t[t] * \tilde{h}[t] * g_r[t]$, where $g_t[t]$ and $g_r[t]$ represent the baseband transmit and receive filters and $*$ represents the convolution operation [63]. Then, the discrete-time channel can be represented as, $h(l) = h[(l-1)T]$, where T is the baud interval. The overall channel is represented as $\mathbf{h} = (h(1), h(2), \ldots, h(L))^T$. Further, in an SBL framework, we model the channel as $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{\Gamma})$, where $\mathbf{\Gamma} = \text{diag}(\gamma(1), \ldots, \gamma(L))$. Note that if $\gamma(l) \to 0$, then the corresponding $h(l) \to 0$ [33, 49].

The traditional methods for channel estimation in OFDM systems assume knowledge of the MIP and use pilots for channel estimation and tracking [66], or employ iterative

techniques based on the Expectation Maximization (EM) algorithms for joint channel estimation/tracking and data detection [72,73]. CS techniques have been proposed for the estimation of the time-varying channel over all the symbols in a frame when the channel consists of a few significant nonzero entries but the path delays are unknown [75–77]. Further, approximate inference methods have been used to solve the problem of joint channel estimation and decoding in a BICM-OFDM system, where the time-varying sparse channel is modeled using a Bernoulli-Gaussian prior [63,68]. In [69], the authors design variational message-passing algorithms based on hierarchical Bayesian prior models for pilot-assisted OFDM channel estimation.

In this chapter, we propose SBL algorithms for exact inference[1] based channel estimation, channel tracking, and data detection. In addition to the monotonicity property of SBL by virtue of the EM framework, SBL offers guarantees such as convergence to the sparsest solution when the noise variance is zero, and converging to a sparse local minimum irrespective of the noise variance [49]. In contrast, approximate inference methods [44], although lower in computational complexity, do not offer such rigorous convergence guarantees. Given the prior distributions of the noise $\mathbf{v}_p$ and the channel $\mathbf{h}$ in (3.2), the a-sparse channel estimation problem is given by

$$(P1) \quad \hat{\mathbf{h}} = \underset{\mathbf{h}, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}}{\arg \min} \frac{\|\mathbf{y}_p - \mathbf{X}_p \mathbf{F}_p \mathbf{h}\|_2^2}{\sigma^2} + \log |\boldsymbol{\Gamma}| + \mathbf{h}^H \boldsymbol{\Gamma}^{-1} \mathbf{h}, \tag{3.4}$$

where[2] $\boldsymbol{\Gamma} \triangleq \mathrm{diag}(\gamma(1), \ldots, \gamma(L))$ and $|\cdot|$ denotes the determinant of a matrix. In the objective function above, the first term originates from the data likelihood and the other

---

[1]In the machine learning literature (e.g., [117–119]), "exact inference" is an attribute associated with algorithms that obtain the exact posterior distribution of the hidden/missing variable.

[2]Due to the one-to-one correspondence between the vector γ and the diagonal matrix $\boldsymbol{\Gamma}$, we use them interchangeably.

terms are from the Gaussian prior (conditioned on γ) assumed on the wireless channel.

In this work, we specifically address the problem of OFDM channel estimation.

Note that, the above problem addresses the estimation of the wireless channel us-
ing pilot subcarriers only. However, in the OFDM scenario, several subcarriers carry
unknown data as well. In this work, we also consider the problem of joint channel
estimation and data detection, which can be stated as

$$(P2) \quad \hat{\mathbf{h}}, \hat{\mathbf{X}} = \underset{\mathbf{h}, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}, \mathbf{X} \in \mathcal{S}}{\arg \min} \frac{\|\mathbf{y} - \mathbf{X}\mathbf{F}\mathbf{h}\|_2^2}{\sigma^2} + \log |\boldsymbol{\Gamma}| + \mathbf{h}^H \boldsymbol{\Gamma}^{-1} \mathbf{h}. \tag{3.5}$$

where $\mathcal{S} \subset \mathbb{C}$ denotes M-QAM/M-PSK constellation from which the symbol is trans-
mitted.

Depending on the mobility of the receiver, the channel may remain essentially con-
stant over the frame duration, or may be slowly time-varying. If the channel is con-
stant, the a-sparse channel estimate can be obtained from the pilot subcarriers by solv-
ing $(P1)$. When the channel is time-varying, typically, the nonzero channel coefficients
vary slowly and are temporally correlated, but the hyperparameters of the channel re-
main constant for several OFDM frames [74]. Consequently, the locations of the sig-
nificant components coincide in successive channel instantiations, i.e., the channels
are approximately *group-sparse* (a-group-sparse). In this work, we cast the channel es-
timation problem as a a-group-sparse channel estimation problem and devise exact
Bayesian inference based solutions. Approximate inference techniques for estimating
the time-varying sparse vector and support have been proposed in [78]. In the context
of SBL, block-based methods such as Block SBL (BSBL) and Temporal SBL (TSBL) algo-
rithms [79] have been proposed to estimate the time-varying correlated sparse vectors

when the correlation among the group-sparse vectors is modeled using a general correlation structure. In contrast to the above-mentioned works, the autoregressive (AR) state space model has been employed to model the correlation among the group sparse vectors and approximate Kalman filtering techniques have been proposed [44]. Further, CS based Kalman filtering has been proposed in the context of sparse correlated vector estimation [29].

In this work, we adopt the Kalman Filter (KF) based exact inference, where the temporal variations of the channel are captured by an AR model. Moreover, it is known that the first order AR model accurately captures the local behavior of fading wireless channels [120]. The first order AR model for the kth channel tap is given by

$$\mathbf{h}_k = \rho \mathbf{h}_{k-1} + \mathbf{u}_k, \tag{3.6}$$

where the Jakes' Doppler spectrum leads to $\rho = J_0(2\pi f_d T_s) \in \mathbb{R}$ where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, f_d is the Doppler frequency, and T_s is the OFDM symbol duration [121]. The driving noise $\mathbf{u}_k$ consists of independent components $\mathbf{u}_k(i) \sim \mathcal{CN}(0, (1 - \rho^2)\gamma(i))$. The initial condition for the a-sparse channel is given by $\mathbf{h}_1 \sim \mathcal{CN}(0, \mathbf{\Gamma})$.

When the hyperparameters are known, a KF approach has been used for channel tracking using the pilot symbols [122]. The EM based KF has also been proposed for joint channel tracking and data detection in OFDM systems [121, 123]. However, these algorithms are not applicable in scenarios where the hyperparameters are unknown and need to be estimated along with the channel tap coefficients and the data symbols.

In contrast, we use the exact inference techniques employed for linear dynamical systems [117, 124] to exploit the known correlation structure of the channel. We note that by using an AR state space model, it is possible to significantly reduce the computational complexity compared to the block-based a-sparse estimation techniques such as the ARSBL [125].

Since the unknown channels have a common hyperparameter vector, the joint pdf of the K received OFDM signals and the a-group-sparse temporally correlated channels is given by

$$p(\mathbf{Y}_{p,K}, \mathbf{h}_1, \ldots, \mathbf{h}_K; \boldsymbol{\gamma}) = \prod_{m=1}^{K} p(\mathbf{y}_{p,m}|\mathbf{h}_m)p(\mathbf{h}_m|\mathbf{h}_{m-1}; \boldsymbol{\gamma}), \qquad (3.7)$$

where $\mathbf{Y}_{p,K} = [\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,K}]$, and, by convention, we use $p(\mathbf{h}_1|\mathbf{h}_0; \boldsymbol{\gamma}) \triangleq p(\mathbf{h}_1; \boldsymbol{\gamma})$ where $\mathbf{h}_1 \sim \mathcal{CN}(0, \boldsymbol{\Gamma})$. To obtain the optimization problem, we consider $-\log p(\mathbf{Y}_{p,K}, \mathbf{h}_1, \ldots, \mathbf{h}_K; \boldsymbol{\gamma})$ and neglect the terms that are constant w.r.t. $\mathbf{h}$ and $\boldsymbol{\gamma}$, to obtain

$$f(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}) = \sum_{m=1}^{K} \frac{\|\mathbf{y}_{p,m} - \mathbf{X}_{p,m}\mathbf{F}_{p,m}\mathbf{h}_m\|_2^2}{\sigma^2} + K \log |\boldsymbol{\Gamma}| + \sum_{m=2}^{K} \frac{(\mathbf{h}_m - \rho\mathbf{h}_{m-1})^H \boldsymbol{\Gamma}^{-1}(\mathbf{h}_m - \rho\mathbf{h}_{m-1})}{(1-\rho^2)}$$
$$+ \mathbf{h}_1^H \boldsymbol{\Gamma}^{-1}\mathbf{h}_1. \qquad (3.8)$$

From the equation above, the pilot-based channel estimation problem for K OFDM symbols can be written as

$$(P3) \quad \hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K = \underset{\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}}{\arg \min} f(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}). \qquad (3.9)$$

Problem $(P3)$ addresses the estimation of *time-varying* wireless channels using only the pilot subcarriers. However, as mentioned earlier, several subcarriers in each of the OFDM symbols carry unknown data. Hence, we can also consider the problem of joint

time-varying channel estimation and data detection, by modifying (3.9) as follows:

$$(P4) \quad \hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K, \hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_K = \underset{\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}, \mathbf{X}_1, \ldots, \mathbf{X}_K \in \mathcal{S}}{\arg \min} g(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}, \mathbf{X}_1, \ldots, \mathbf{X}_K)$$

(3.10)

where

$$g(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}, \mathbf{X}_1, \ldots, \mathbf{X}_K) = \sum_{m=1}^{K} \frac{\|\mathbf{y}_m - \mathbf{X}_m \mathbf{F} \mathbf{h}_m\|_2^2}{\sigma^2} + \sum_{m=2}^{K} \frac{(\mathbf{h}_m - \rho \mathbf{h}_{m-1})^H \boldsymbol{\Gamma}^{-1} (\mathbf{h}_m - \rho \mathbf{h}_{m-1})}{(1-\rho^2)}$$

$$+ K \log |\boldsymbol{\Gamma}| + \mathbf{h}_1^H \boldsymbol{\Gamma}^{-1} \mathbf{h}_1. \tag{3.11}$$

Contributions

In this work, we propose a practical and principled approach for joint a-group-sparse channel estimation and data detection in SISO-OFDM systems, that evaluates well in Monte-Carlo simulations. First, we show that the problem in $(P1)$ can be solved using the SBL framework of [49]. We next generalize the SBL framework to obtain the J-SBL algorithm as a solution to $(P2)$. A key feature of the J-SBL algorithm is that the observations from both the data and the pilot subcarriers are incorporated to jointly estimate the a-sparse channel as well as the unknown data. We also propose a low complexity, recursive J-SBL (RJ-SBL) algorithm to solve $(P2)$. We show that the joint estimation procedure leads to a significant improvement in the Mean Square Error (MSE) of the channel estimate at SNRs of practical interest. Further, we propose a novel, low-complexity K-SBL algorithm as a recursive solution to $(P3)$. We enhance the K-SBL algorithm to obtain the JK-SBL algorithm, which is a recursive solution to $(P4)$. The results are summarized in the Table 3.1.

Although our work focuses on a-sparse channel estimation for OFDM systems using

Table 3.1: The main contributions of this chapter.

Sl. no.	Novel algorithms proposed and Section number	Goal	Applicability
1	SBL in Sec. 3.2	Joint channel, hyperparameter estimation	Block-fading channels ($P1$)
2	J-SBL in Sec. 3.3 and Recursive J-SBL	Joint channel, hyperparameter estimation and data detection	Block-fading channels ($P2$)
3	K-SBL in Sec. 3.2	Recursive joint channel and hyperparameter estimation	Time-varying and Block-fading channels ($P3$)
4	JK-SBL in Sec. 3.3	Recursive joint channel, hyperparameter estimation and data detection	Time-varying and Block-fading channels ($P4$)

the SBL framework, the algorithms we develop are important in their own right as described in 2. This is the first work in the literature that proposes recursive techniques for exact inference in sparse signal recovery. We show that the joint problems of hyperparameter estimation and data detection separate out in the M-step. This leads to a simple maximization procedure in the M-step, with no loss of optimality. The joint algorithms involve estimation of the unknown data symbols, which necessitates the development of techniques that are capable of handling partially unknown dictionary matrices.[3] Finally, the recursive versions of the algorithms have the advantage of computational simplicity compared to other exact inference methods, while retaining the performance advantages of SBL estimators.

The rest of this chapter is organized as follows. In Sec. 3.2, we propose algorithms for a-sparse channel estimation using pilots. In Sec. 3.3, the joint channel estimation and data detection algorithms are proposed and the implementation issues are discussed. The efficacy of the proposed techniques is demonstrated through simulation results in Sec. 3.4. We offer some concluding remarks in Sec. 3.5.

---

[3]That is, the algorithms are capable of handling the fact that, due to the $N - P$ unknown data symbols in $\mathbf{X}$, the measurement matrix $\mathbf{\Phi} = \mathbf{XF}$ is partially unknown.

3.2 Channel Estimation and Tracking Using Pilot Subcarriers

In this section, we propose SBL algorithms for a-group-sparse channel estimation in OFDM systems using pilot symbols, for both block-fading and time-varying channels. First, we discuss the SBL algorithm to solve $(P1)$, i.e., the problem of a-sparse channel estimation using P_b pilots in the entire OFDM frame when the channel is block-fading. Subsequently, we consider the time-varying channel using P_t pilots in every symbol, and propose a novel, recursive approach for a-group-sparse channel estimation, i.e., a solution to $(P3)$.

3.2.1 The SBL Algorithm: Block-fading Case

Here, we propose the SBL algorithm for channel estimation using pilot subcarriers in a single OFDM symbol; this forms the basis for the algorithms developed in the sequel. The observation model is given by (2). SBL uses a parametrized prior to obtain sparse solutions, given by

$$p(\mathbf{h}; \boldsymbol{\gamma}) = \prod_{i=1}^{L} (\pi \gamma(i))^{-1} \exp\left(-\frac{|h(i)|^2}{\gamma(i)} \right). \tag{3.12}$$

Typically, the hyperparameters $\boldsymbol{\gamma}$ can be estimated using the type-II ML procedure [34], i.e., by maximizing the marginalized pdf $p(\mathbf{y}_p; \boldsymbol{\gamma})$ as

$$\hat{\boldsymbol{\gamma}}_{ML} = \arg \max_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} p(\mathbf{y}_p; \boldsymbol{\gamma}). \tag{3.13}$$

Since the above problem cannot be solved in closed form, iterative estimators such as the EM based SBL algorithm [49] have to be employed. The sparse channel $\mathbf{h}$ is

considered as the hidden variable and the ML estimate of $\boldsymbol{\gamma}$ is obtained in the M-step. The steps of the algorithm can be given as

$$\text{E-step}: Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}_p, \mathbf{h}; \boldsymbol{\gamma})] \tag{3.14}$$

$$\text{M-step}: \boldsymbol{\gamma}^{(r+1)} = \arg\max_{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}} Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \tag{3.15}$$

The E-step above requires the posterior density of the sparse vector with the hyperparameter $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(r)}$, which can be expressed as

$$p\left(\mathbf{h}|\mathbf{y}_p; \boldsymbol{\gamma}^{(r)}\right) = \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.16}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^{(r)} - \boldsymbol{\Gamma}^{(r)}\boldsymbol{\Phi}_p^H\left(\sigma^2\mathbf{I}_{P_b} + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}^{(r)}\boldsymbol{\Phi}_p^H\right)^{-1}\boldsymbol{\Phi}_p\boldsymbol{\Gamma}^{(r)}$, and $\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_p^H\mathbf{y}_p$, where $\boldsymbol{\Phi}_p = \mathbf{X}_p\mathbf{F}_p$. Notice that the EM algorithm given by the steps in (3.14), (3.15) also solves $(P1)$, where we obtain a MAP estimate of the a-sparse channel, i.e., $\hat{\mathbf{h}} = \boldsymbol{\mu}$ with $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma}^{(r)})$. The M-step in (3.15) can be simplified, to obtain

$$\gamma^{(r+1)}(i) = \arg\max_{\gamma(i)\in\mathbb{R}_+} \mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{h}; \boldsymbol{\gamma})] \tag{3.17}$$

$$= \mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}\left[|h(i)|^2\right] = \Sigma(i,i) + |\mu(i)|^2. \tag{3.18}$$

In (3.17), the term $\mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}_p|\mathbf{h}; \boldsymbol{\gamma})]$ has been dropped, as it is not a function of $\gamma(i)$. Note that, since all the algorithms proposed in this work use the EM updates, they have monotonicity property, i.e., the likelihood is guaranteed to increase at each iteration [126, 127].[4]

In the case of multiple OFDM symbols in a block-fading channel, since the channel

---

[4]We have found, empirically, that the straightforward initialization such as $\boldsymbol{\Gamma}^{(0)} = \mathbf{I}_L$ leads to accurate solutions.

remains constant for the K OFDM symbols, the system model in (3.2) is modified as

$$\mathbf{y}_{p,m} = \mathbf{X}_{p,m}\mathbf{F}_{p,m}\mathbf{h} + \mathbf{v}_{p,m}, \quad m = 1, \ldots, K. \tag{3.19}$$

The equation above has a one-to-one correspondence with (3.2), since $\mathbf{y}_{p,m}$ denotes the observations corresponding to pilot subcarriers in the mth OFDM symbol and $\mathbf{\Phi}_{p,m} = \mathbf{X}_{p,m}\mathbf{F}_{p,m}$ denotes the matrix consisting of measurements corresponding to pilot sub-carriers in the mth OFDM symbol.

We note that the SBL algorithm proposed in this section is not equipped to use the correlations between the channel across successive OFDM symbols in a time-varying channel. A straightforward approach to exploit the correlation is to use a block-based method, where the estimates of all the K channel vectors are obtained jointly using the observations for the K OFDM symbols [79, 125]. However, this joint processing of all K OFDM symbols is computationally expensive, as it requires inverting matrices of the size $KP_t \times KP_t$. In the next subsection, we propose a recursive approach that is not only low-complexity compared to the block-based techniques, but also exploits the temporal channel correlation across symbols, resulting in an enhanced channel tracking performance.

3.2.2 The K-SBL Algorithm: Time-varying Case

In this subsection, we derive algorithms for tracking the *slowly time-varying* channel using an SBL framework to learn the hyperparameters along with the channel coeffi-cients, i.e., we solve $(P3)$. We derive recursive techniques based on the Kalman Filter

and Smoother (KFS), with an AR model for the temporal evolution of the channel. Interestingly, the framework developed in this section can also be used to accommodate detection of the unknown data (i.e., a solution to $(P4)$), as we show in the next section.

In the time-varying case, the measurement equation given by the OFDM system model, and the state equation given by the first order AR channel model, for K consecutive symbols, are as follows:

$$\mathbf{y}_{p,m} = \mathbf{\Phi}_{p,m}\mathbf{h}_m + \mathbf{v}_{p,m}, \tag{3.20}$$

$$\mathbf{h}_{m+1} = \rho\mathbf{h}_m + \mathbf{u}_{m+1}, \quad m = 1, 2, \ldots, K, \tag{3.21}$$

where $\mathbf{\Phi}_{p,m} = \mathbf{X}_{p,m}\mathbf{F}_{p,m}$. Typically, in a KF approach to $(P3)$, the goal is to recursively estimate the channel state and its covariance matrix using forward and backward recursions, given the observations $\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,K}$ sampled at the P_t pilot subcarriers. In the forward recursion, for each OFDM symbol, the KF operates on the received symbol to obtain the estimates of the a-sparse channel as a weighted average of the previous estimate and the current received symbol. These weights are given by the the Kalman gain matrix, and are updated for each OFDM symbol. In the backward recursion, the Kalman *smoother* ensures that the observations until the Kth OFDM symbol are included in the estimation of the a-sparse channel corresponding to the mth symbol for $1 \leq m < K$. Hence, it improves the accuracy of the estimates of the previous channel states in every recursion.

For the moment, if we assume that $\mathbf{\Gamma}$ is known, and if we denote the posterior mean

and the covariance matrix of channel in the mth OFDM symbol by $\hat{\mathbf{h}}_{m|m}$ and $\mathbf{P}_{m|m}$, respectively, for $1 \leq m \leq K$, then the KFS update equations are as follows [124, 128]:

for $m = 1, \ldots, K$ **do**

Prediction: $\hat{\mathbf{h}}_{m|m-1} = \rho \hat{\mathbf{h}}_{m-1|m-1}$ \hfill (3.22)

$\mathbf{P}_{m|m-1} = \rho^2 \mathbf{P}_{m-1|m-1} + (1 - \rho^2)\mathbf{\Gamma}$ \hfill (3.23)

Filtering:

$\mathbf{G}_m = \mathbf{P}_{m|m-1} \mathbf{\Phi}_{p,m}^H \left(\sigma^2 \mathbf{I}_{P_t} + \mathbf{\Phi}_{p,m} \mathbf{P}_{m|m-1} \mathbf{\Phi}_{p,m}^H \right)^{-1}$ \hfill (3.24)

$\hat{\mathbf{h}}_{m|m} = \hat{\mathbf{h}}_{m|m-1} + \mathbf{G}_m (\mathbf{y}_{p,m} - \mathbf{\Phi}_{p,m} \hat{\mathbf{h}}_{m|m-1})$ \hfill (3.25)

$\mathbf{P}_{m|m} = (\mathbf{I}_L - \mathbf{G}_m \mathbf{\Phi}_{p,m}) \mathbf{P}_{m|m-1}$ \hfill (3.26)

end

for $j = K, K-1, \ldots, 2$ **do**

Smoothing: $\hat{\mathbf{h}}_{j-1|K} = \hat{\mathbf{h}}_{j-1|j-1} + \mathbf{J}_{j-1} (\hat{\mathbf{h}}_{j|K} - \hat{\mathbf{h}}_{j|j-1})$ \hfill (3.27)

$\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1} (\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1}) \mathbf{J}_{j-1}^H,$ \hfill (3.28)

end

where $\mathbf{J}_{j-1} \triangleq \rho \mathbf{P}_{j-1|j-1} \mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{G}_m$ is the Kalman gain. In the above, the symbols $\hat{\mathbf{h}}_{m|m-1}$, $\mathbf{P}_{m|m-1}$, etc. have their usual meanings as in the KF literature [124]. For example, $\hat{\mathbf{h}}_{m|m-1}$ is the channel estimate at the mth OFDM symbol given the observations $\mathbf{Y}_{p,m-1} = [\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,m-1}]$ and $\mathbf{P}_{m|m-1}$ is the covariance of the mth channel estimate given $\mathbf{Y}_{p,m-1}$. The above KFS equations are initialized by setting $\hat{\mathbf{h}}_{0|0} = \mathbf{0}$ and $\mathbf{P}_{0|0} = \mathbf{\Gamma}$. They track the channel in the forward direction using the prediction and the filtering equations in (3.22)-(3.26) and *smooth* the obtained channel estimates using the backward

recursions in (3.27)-(3.28). However, in the a-sparse channel tracking problem, $\mathbf{\Gamma}$ is unknown. Hence, we propose the K-SBL algorithm, which simultaneously estimates the channel coefficients and also learns the unknown $\mathbf{\Gamma}$.

Recall that the a-group-sparse channel has a common hyperparameter set. The joint pdf of the received signals and the a-group-sparse channel for K OFDM symbols is given by (3.7), which leads to the optimization problem as given by $(P3)$. We propose the K-SBL algorithm using the EM updates, as follows:

$$\text{E-step}: Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}_{p,K},\mathbf{h}_1,\ldots,\mathbf{h}_K;\boldsymbol{\gamma})] \tag{3.29}$$

$$\text{M-step}: \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max} Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \tag{3.30}$$

To compute the E-step given above, we require the posterior distribution of the unknown a-sparse channel, which is obtained using the recursive update equations given by (3.22)-(3.28). In order to obtain an ML estimate of $\boldsymbol{\gamma}$, K-SBL incorporates an M-step, which, in turn, utilizes the mean and covariance of the posterior distribution from the E-step. From (3.7), the M-step results in the following optimization problem:

$$\boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max} \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}\left[c - K\log|\mathbf{\Gamma}|\right.$$
$$\left. -\sum_{j=2}^{K} \frac{(\mathbf{h}_j - \rho\mathbf{h}_{j-1})^H\mathbf{\Gamma}^{-1}(\mathbf{h}_j - \rho\mathbf{h}_{j-1})}{(1-\rho^2)} - \mathbf{h}_1^H\mathbf{\Gamma}^{-1}\mathbf{h}_1\right], \tag{3.31}$$

where c is a constant independent of $\boldsymbol{\gamma}$. As mentioned earlier, we see that the M-step requires the computation of $\hat{\mathbf{h}}_{j|K} \triangleq \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j]$, and covariance $\mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j\mathbf{h}_j^H] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H$ for $j = 1,\ldots,K$, which is obtained from (3.22)-(3.28). The M-step also requires the computation of $\mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j\mathbf{h}_{j-1}^H] \triangleq \mathbf{P}_{j,j-1|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j-1|K}^H$ for

$j = K, K - 1, \ldots, 2$, which we obtain from [124] as follows:

$$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1}\mathbf{J}_{j-2}^H + \mathbf{J}_{j-1}^H(\mathbf{P}_{j,j-1|K} - \rho\mathbf{P}_{j-1|j-1})\mathbf{J}_{j-2}. \tag{3.32}$$

The above recursion is initialized using $\mathbf{P}_{K,K-1|K} = \rho(\mathbf{I}_L - \mathbf{G}_K\mathbf{\Phi}_{p,K})\mathbf{P}_{K-1|K-1}$. Using the above expressions, (3.31) simplifies as

$$\boldsymbol{\gamma}^{(r+1)} = \arg\max_{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times1}} \left\{ c' - K\log|\mathbf{\Gamma}| - \text{Trace}(\mathbf{\Gamma}^{-1}\mathbf{M}_{1|K}) - \frac{1}{(1-\rho^2)}\sum_{j=2}^{K}\text{Trace}(\mathbf{\Gamma}^{-1}\mathbf{M}_{j|K}) \right\},$$

$$\tag{3.33}$$

where c' is a constant independent of $\boldsymbol{\gamma}$, $\mathbf{M}_{j|K} \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H + \rho^2(\mathbf{P}_{j-1|K} + \hat{\mathbf{h}}_{j-1|K}\hat{\mathbf{h}}_{j-1|K}^H) - 2\rho\text{Re}(\mathbf{P}_{j,j-1|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j-1|K}^H)$ and $\mathbf{M}_{1|K} \triangleq \mathbf{P}_{1|K} + \hat{\mathbf{h}}_{1|K}\hat{\mathbf{h}}_{1|K}^H$. Differentiating (3.33) w.r.t. $\gamma(i)$ and setting the resulting equation to zero gives the update for the ith hyperparameter as follows:

$$\gamma^{(r+1)}(i) = \frac{1}{K}\left(\sum_{j=2}^{K}\frac{M_{j|K}(i,i)}{(1-\rho^2)} + M_{1|K}(i,i) \right), \tag{3.34}$$

for $i = 1, \ldots, L$. Thus the K-SBL algorithm learns $\boldsymbol{\gamma}$ in the M-step and provides low-complexity and recursive estimates of the a-sparse channel in the E-step. This completes the EM based solution to $(P3)$.

Remarks: When $\rho = 1$, the AR model simplifies to $\mathbf{h} = \mathbf{h}_1 = \ldots = \mathbf{h}_K$, and hence, it reduces to the block-fading channel scenario. The recursive updates in the E-step are given by the KFS equations (3.22)-(3.28), and the M-step is given by

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r+1)}) = \mathbb{E}_{\mathbf{h}|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[c' - (\mathbf{h}^H\mathbf{\Gamma}^{-1}\mathbf{h} + \log|\mathbf{\Gamma}|)], \tag{3.35}$$

which results in the same M-step as that of the SBL algorithm in the block-fading case.

Hence, this algorithm provides a low-complexity recursive solution to the SBL problem in the block-fading scenario, which we discuss in detail in Sec. 3.3.2. At the other extreme, when $\rho = 0$, the AR model simplifies to $\mathbf{h}_m = \mathbf{u}_m$ for $m = 1, \ldots, K$, i.e., the channels for OFDM symbols are mutually independent of each other. In this case, the prediction equations of the KFS equations simplify as $\hat{\mathbf{h}}_{m|m-1} = \mathbf{0}$ and $\mathbf{P}_{m|m-1} = \mathbf{\Gamma}$, and the expressions for $\hat{\mathbf{h}}_{m|m}$ and $\mathbf{P}_{m|m}$ simplify to the mean and covariance matrix, as obtained in the SBL algorithm for a single OFDM symbol. The smoothing equations simplify to $\hat{\mathbf{h}}_{m-1|m} = \hat{\mathbf{h}}_{m-1|m-1}$ and $\mathbf{P}_{m-1|m} = \mathbf{P}_{m-1|m-1}$, i.e., the smoothed mean and covariance at the $(m-1)^{\text{th}}$ symbol depend only on observations of the $(m-1)^{\text{th}}$ OFDM symbol, as expected.

Although the algorithms proposed in this section are easy to implement and computationally simple due to their recursive nature, they do not utilize all the information available from the observation vectors $\mathbf{y}_1, \ldots, \mathbf{y}_K$. Only the pilot subcarriers are used for channel estimation. Hence, in the next section, we extend the SBL framework developed in this section to detect the unknown data. We show how these decisions can be coalesced into the EM iterations, leading to joint channel estimation and data detection.

3.3 Joint Channel Estimation and Data Detection Using Pilot and Data Subcarriers

In this section, we start by deriving the J-SBL and the RJ-SBL algorithm for joint estimation of the unknown a-sparse channel and transmit data in a block-fading OFDM system. Subsequently, we consider the time-varying channel, and generalize the K-SBL

to obtain the JK-SBL for jointly estimating the unknown data and tracking the a-group-sparse channel. Our proposed algorithms solve the problems $(P2)$ and $(P4)$ using an SBL framework.

3.3.1 The J-SBL Algorithm: Block-fading Case

To derive the algorithm for an OFDM frame consisting of K OFDM symbols, we consider $\mathbf{h}$ as a hidden variable and $[\boldsymbol{\gamma}, \mathbf{X}_1, \ldots, \mathbf{X}_K]$ as the parameters to be estimated. The E and the M-steps of the J-SBL algorithm can be given as

$$\text{E-step}: \ Q(\mathbf{X}, \boldsymbol{\gamma} | \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}) = \mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}} \left[\log p(\mathbf{y}, \mathbf{h}; \mathbf{X}, \boldsymbol{\gamma}) \right] \tag{3.36}$$

$$\text{M-step}: \ \left(\mathbf{X}^{(r+1)}, \boldsymbol{\gamma}^{(r+1)} \right) = \underset{\mathbf{X}, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}}{\arg \max} \, Q \left(\mathbf{X}, \boldsymbol{\gamma} | \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)} \right), \tag{3.37}$$

where $\mathbf{X} \in \mathbb{C}^{NK \times NK}$ is a block diagonal matrix consisting of the matrices $\mathbf{X}_1, \ldots, \mathbf{X}_K$ whose diagonal entries consist of symbols from the transmit constellation, and $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T$. The posterior density computed in the E-step is $p\left(\mathbf{h}|\mathbf{y}; \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$
\begin{aligned}
\boldsymbol{\mu} &= \sigma^{-2} \boldsymbol{\Sigma} \mathbf{F}_b^H \mathbf{X}^{(r)H} \mathbf{y} \\
\boldsymbol{\Sigma} &= \left(\sigma^{-2} \mathbf{F}_b^H \mathbf{X}^{(r)H} \mathbf{X}^{(r)} \mathbf{F}_b + \boldsymbol{\Gamma}^{(r)-1} \right)^{-1},
\end{aligned}
\tag{3.38}
$$

where $\mathbf{F}_b \in \mathbb{C}^{NK \times L}$ with $\mathbf{F}_b = \mathbf{1}_K \otimes \mathbf{F}$, where $\mathbf{1}_K$ is a vector of all ones. Notice that (3.38) and (3.16) are different since the former uses the known pilot symbols, $\mathbf{X}_p \in \mathbb{C}^{P \times P}$, whereas the latter uses the pilot symbols along with the estimated transmit data, together given by $\mathbf{X}^{(r)}$ in the rth iteration. The proposed algorithm is pictorially depicted in Fig. 3.1.
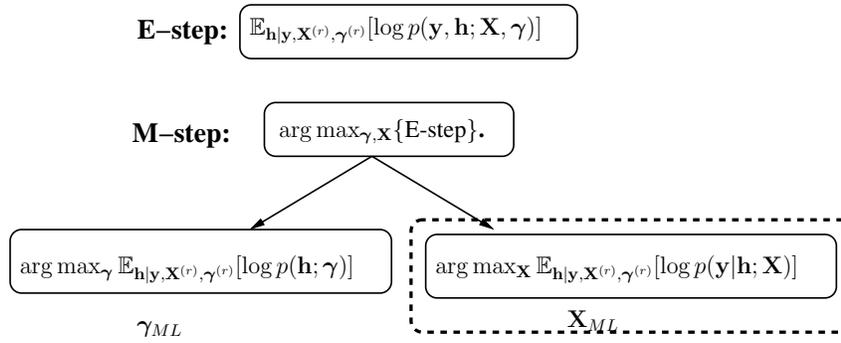
Figure 3.1: The J-SBL algorithm: the E-step computes the expectation over the posterior density of $\mathbf{h}$. The joint maximization in the M-step simplifies into two independent maximizations over $\gamma(i)$ and $\mathbf{X}$. The step inside the dashed box indicates the new ingredient in the J-SBL algorithm.

The objective function in the M-step given in (3.37) can be written as

$$Q\left(\mathbf{X}, \boldsymbol{\gamma}|\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = c'' - \mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}\left[\frac{\|\mathbf{y} - \mathbf{X}\mathbf{F}_b\mathbf{h}\|_2^2}{\sigma^2} + \log|\boldsymbol{\Gamma}| + \mathbf{h}^H\boldsymbol{\Gamma}^{-1}\mathbf{h}\right] \quad (3.39)$$

where c'' is a constant independent of $\boldsymbol{\gamma}$ and $\mathbf{X}$. The objective function given above is the sum of two independent functions, $\mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}|\mathbf{h};\mathbf{X})]$ and $\mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{h};\boldsymbol{\gamma})]$. The key aspect of the M-step below is that the function $\mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}|\mathbf{h};\mathbf{X})]$ is maximized over $\mathbf{X}$, which incorporates the information discarded in the M-step of the SBL algorithm presented in Sec. 3.2.1. Now, since the first term does not depend on $\boldsymbol{\gamma}$, we optimize the second function with respect to $\gamma(i)$ to obtain $\gamma^{(r+1)}(i)$ as in the SBL algorithm, given by (3.18). On the other hand, the first function can be optimized by solving the following problem:

$$X^{(r+1)}(i,i) = \arg\min_{x_i \in \mathcal{S}}\left\{C_b(i,i)|x_i|^2 + |y(i) - x_i\mathbf{F}_b(i,:)\boldsymbol{\mu}|^2\right\} \quad (3.40)$$

where $i \in \mathcal{D}$, $\mathcal{D}$ is an index set consisting of the data subcarrier locations, $\mathbf{C}_b = \mathbf{F}_b\boldsymbol{\Sigma}\mathbf{F}_b^H$,

$\mathbf{F}_b(i,:)$ is the ith row of the $\mathbf{F}_b$ matrix, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given in (3.38) and $\mathcal{S}$ is the constellation from which the symbol is transmitted. Due to the above maximization, $Q(\mathbf{X}, \boldsymbol{\gamma}|\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)})$ increases monotonically for $1 \leq r \leq r_{\text{max}}$, where r_{max} is the maximum number of iterations. That is,

$$Q(\mathbf{X}^{(r+1)}, \boldsymbol{\gamma}^{(r+1)}|\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}) \geq Q(\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}|\mathbf{X}^{(r-1)}, \boldsymbol{\gamma}^{(r-1)}), \;\; \text{for} \;\; 1 \leq r \leq r_{\text{max}}. \quad (3.41)$$

Note that the above function $Q(.)$ monotonically (in $\mathbf{X}$ and $\boldsymbol{\gamma}$) approaches the likelihood function, which in turn is bounded. This guarantees the convergence of the proposed J-SBL algorithm. Further, by the same reasoning, the convergence guarantee holds good for the JK-SBL algorithm which will be presented in the sequel.

The J-SBL requires initial estimates of the unknown parameters $\boldsymbol{\gamma}$ and $\mathbf{X}$. The initial estimate of $\boldsymbol{\Gamma}$ is taken to be the identity matrix, as in the previous section. The initialization of the $(KN - P_b)$ non-pilot data in turn requires an initial channel estimate. Channel estimates using methods like LS and MMSE cannot be used here, as they require knowledge of the support and the hyperparameters, respectively. Hence, the initialization of $\mathbf{X}$ can be obtained from the channel estimate obtained from a few iterations of the SBL algorithm from the $P_b = P$ pilots (denoted as $\hat{\mathbf{h}}_{SBL}$). The ML data detection problem is given by

$$X^{(0)}(i,i) = \arg\min_{x_i \in \mathcal{S}} |y(i) - x_i \mathbf{F}_b(i,:)\hat{\mathbf{h}}_{SBL}|^2, \quad i \in \mathcal{D}. \quad (3.42)$$

J-SBL algorithm is a block-based algorithm, and hence, the complexity of the algorithm is dominated by the E-step, which incurs a complexity of $\mathcal{O}(N^2 L K^3)$ [104]. In the next subsection, we derive a low-complexity, recursive version of the J-SBL algorithm,

using the K-SBL algorithm with $\rho = 1$.

3.3.2   Recursive J-SBL Algorithm: Block-fading Case

In this subsection, we derive the recursive joint SBL algorithm which is mathematically equivalent to the J-SBL algorithm proposed in Sec. 3.3.1, using the framework of the K-SBL algorithm with $\rho = 1$, i.e., for the block-fading channel. Hence, we solve the problem ($P3$), using a low-complexity RJ-SBL algorithm.

Consider the state space model in (3.20) and (3.21) in the block-fading case, where the channel remains constant for K OFDM symbols, with $\mathbf{h} = \mathbf{h}_1 = \cdots = \mathbf{h}_K$. The prediction equations of the KFS update equations in (3.22)-(3.26) simplify as $\hat{\mathbf{h}}_{m|m-1} = \hat{\mathbf{h}}_{m-1|m-1} \triangleq \hat{\mathbf{h}}_{m-1}$ and $\mathbf{P}_{m|m-1} = \mathbf{P}_{m-1|m-1} \triangleq \mathbf{P}_{m-1}$, for $m = 1, \ldots, K$. Moreover, for $\rho = 1$, the smoothing equations in (3.27)-(3.28) simplify as $\mathbf{h}_{j-1|K} = \mathbf{h}_{j|K}$ and $\mathbf{P}_{j-1|K} = \mathbf{P}_{j|K}$ for $j = K, \ldots, 1$. Hence, the filtering equations of the KFS updates suffice to describe the recursions, as follows. For $m = 1, \ldots, K$, the E-step of the J-SBL algorithm can be replaced by

$$\mathbf{G}_m = \mathbf{P}_{m-1}\mathbf{\Phi}_m^H \left(\sigma^2\mathbf{I}_N + \mathbf{\Phi}_m\mathbf{P}_{m-1}\mathbf{\Phi}_m^H\right)^{-1} \tag{3.43}$$

$$\hat{\mathbf{h}}_m = \hat{\mathbf{h}}_{m-1} + \mathbf{G}_m(\mathbf{y}_m - \mathbf{\Phi}_m\hat{\mathbf{h}}_{m-1}) \tag{3.44}$$

$$\mathbf{P}_m = (\mathbf{I}_L - \mathbf{G}_m\mathbf{\Phi}_m)\mathbf{P}_{m-1}, \tag{3.45}$$

where $\mathbf{\Phi}_k$ denotes the measurement matrix of the kth OFDM symbol given by $\mathbf{\Phi}_k = \mathbf{X}_k\mathbf{F}$. However, since $\mathbf{\Gamma}$ is unknown and $\mathbf{X}_k$ is known only at pilot locations, the SBL framework is incorporated to learn the unknown $\mathbf{\Gamma}$ and unknown data in $\mathbf{X}_k$. Hence, the update equations given above form the E-step, while the M-step is the same as that

of the J-SBL algorithm, given by (3.37). The update for γ is given by,

$$\gamma^{(r+1)}(i) = P_K(i,i) + \left|\hat{h}_K(i)\right|^2, \tag{3.46}$$

where $\hat{\mathbf{h}}_K$ and $\mathbf{P}_K$ are given by (3.44) and (3.45), respectively. The unknown data can be detected by solving the following optimization problem:

$$X^{(r+1)}(i,i) = \arg\min_{x_i \in \mathcal{S}} \left\{ |x_i|^2 C(i,i) + |y(i) - x_i \mathbf{F}_b(i,:)\hat{\mathbf{h}}_K|^2 \right\} \tag{3.47}$$

where $i \in \mathcal{D}$, $\mathbf{C} = \mathbf{F}_b \mathbf{P}_K \mathbf{F}_b^H$. The initialization of γ and $\mathbf{X}^{(0)}$ is the same as the J-SBL algorithm of Sec. 3.3.1. The complexity of the RJ-SBL algorithm is dominated by the computation of $\mathbf{G}_k$, and is given by $\mathcal{O}(NL^2K)$. Hence, for large K, the RJ-SBL algorithm is computationally significantly cheaper than the J-SBL algorithm.

The E-step of the RJ-SBL is a recursive implementation of the E-step of the J-SBL algorithm, and the M-steps of the algorithms are the same. Hence, the algorithms are mathematically equivalent if the same initializations are employed. This is illustrated via simulations in Sec. 3.4 (see Fig. 3.4).

3.3.3 The JK-SBL Algorithm: Time-varying Case

In this section, we generalize the K-SBL algorithm of Sec. 3.2.2 to obtain the JK-SBL algorithm, which utilizes the observations available at all the N subcarriers and performs data detection at the $(N - P_t)$ data subcarriers of the OFDM symbol. The algorithm is recursive in nature, and the channel estimates for K OFDM symbols are used to jointly estimate the a-sparse channel and the unknown data of the $m^{th}, 1 \leq m \leq K$ OFDM symbol. In essence, we solve the problem given by $(P4)$.

Our starting point, again, is the state space model given by (3.20) and (3.21). Using the observations $\mathbf{Y}_K = [\mathbf{y}_1, \ldots, \mathbf{y}_K]$, the recursive updates of the mean and the covariance of the posterior distribution are given by (3.22)-(3.28), with $\mathbf{y}_{p,m}$ and $\mathbf{\Phi}_{p,m}$ replaced by $\mathbf{y}_m$ and $\mathbf{\Phi}_m$, respectively. Thus, the JK-SBL algorithm uses the observations available at all the N subcarriers. Further, since $\mathbf{\Gamma}$ and data at the non-pilot subcarriers are unknown, the SBL framework leads to the objective function for the M-step given by

$$Q\left(\mathbf{X}_1, \ldots, \mathbf{X}_K, \boldsymbol{\gamma} | \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = c''' - \mathbb{E}_{\mathbf{h}_1, \ldots, \mathbf{h}_K | \mathbf{Y}_K; \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}, \boldsymbol{\gamma}^{(r)}} \left[\sum_{j=1}^{K} \frac{\|\mathbf{y}_j - \mathbf{X}_j \mathbf{F} \mathbf{h}_j\|^2}{\sigma^2} \right.$$

$$\left. + K \log |\mathbf{\Gamma}| + \sum_{j=2}^{K} \frac{(\mathbf{h}_j - \rho \mathbf{h}_{j-1})^H \mathbf{\Gamma}^{-1} (\mathbf{h}_j - \rho \mathbf{h}_{j-1})}{(1-\rho^2)} + \mathbf{h}_1^H \mathbf{\Gamma}^{-1} \mathbf{h}_1 \right], \tag{3.48}$$

where c''' is a constant independent of $\boldsymbol{\gamma}$ and $\mathbf{X}_1, \ldots, \mathbf{X}_K$. The expression above is a sum of terms which are independent functions of $\boldsymbol{\gamma}$ and $\mathbf{X}_K$, denoted as $Q\left(\boldsymbol{\gamma} | \boldsymbol{\gamma}^{(r)}\right)$ and $Q\left(\mathbf{X}_1, \ldots, \mathbf{X}_K | \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}\right)$, respectively. Further, we see that $Q\left(\boldsymbol{\gamma} | \boldsymbol{\gamma}^{(r)}\right)$ is the same as the expression in (3.33). Hence, the learning rule for $\boldsymbol{\gamma}$ follows from the M-step of the K-SBL algorithm, and is given by (3.34). The expression for $Q\left(\mathbf{X}_1, \ldots, \mathbf{X}_K | \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}\right)$ is given by

$$Q\left(\mathbf{X}_1, \ldots, \mathbf{X}_K | \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}\right) = c - \mathbb{E}_{\mathbf{h}_1, \ldots \mathbf{h}_K | \mathbf{Y}_K; \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}, \boldsymbol{\gamma}^{(r)}} \left[\sum_{m=1}^{K} \frac{\|\mathbf{y}_m - \mathbf{X}_m \mathbf{F} \mathbf{h}_m\|^2}{\sigma^2} \right]. \tag{3.49}$$

As mentioned earlier, the M-step requires the computation of $\hat{\mathbf{h}}_{j|K} \triangleq \mathbb{E}_{\mathbf{h}_1, \ldots, \mathbf{h}_K | \mathbf{Y}_{p,K}; \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}, \boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j]$, and covariance $\mathbb{E}_{\mathbf{h}_1, \ldots, \mathbf{h}_K | \mathbf{Y}_{p,K}; \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}, \boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j \mathbf{h}_j^H] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K} \hat{\mathbf{h}}_{j|K}^H$ for $j = 1, \ldots, K$, which are given by the KFS equations of the E-step.

The maximization of $Q\left(\mathbf{X}_1, \ldots, \mathbf{X}_K | \mathbf{X}_1^{(r)}, \ldots, \mathbf{X}_K^{(r)}\right)$ in (3.49) leads to the following optimization problem for $\mathbf{X}_m$:

$$X_m^{(r+1)}(i,i) = \arg\min_{x_i \in \mathcal{S}} \left\{ |x_i|^2 C_m(i,i) + |y_m(i) - x_i \mathbf{F}(i,:)\hat{\mathbf{h}}_{m|K}|^2 \right\}, \ 0 \leq m \leq K, \ i \in \mathcal{D}$$

(3.50)

where $\mathbf{C}_m = \mathbf{F}\mathbf{P}_{m|K}\mathbf{F}^H$ and $\mathbf{F}(i,:)$ represents the ith row of the matrix $\mathbf{F}$. The iterations of the JK-SBL proceed similar to the K-SBL algorithm, except for the additional M-step to estimate the unknown data. Also, the measurement matrix is given by $\mathbf{\Phi}_m^{(r)}$ in the rth iteration of the mth OFDM symbol, instead of the $\mathbf{\Phi}_{p,m}$ used in the K-SBL algorithm, which consisted of pilot subcarriers only. We provide a pictorial representation of the overall JK-SBL algorithm in Fig. 3.2. We use the channel estimate obtained from a few iterations of the K-SBL algorithm using P_t pilots (denoted as $\hat{\mathbf{h}}_{KSBL}$) to obtain the initial estimate $\mathbf{X}_m^{(0)}$ for $0 \leq m \leq K$ as

$$X_m^{(0)}(i,i) = \arg\min_{x_i \in \mathcal{S}} |y_m(i) - x_i \mathbf{F}(i,:)\hat{\mathbf{h}}_{KSBL}|^2, \ i \in \mathcal{D}.$$

(3.51)
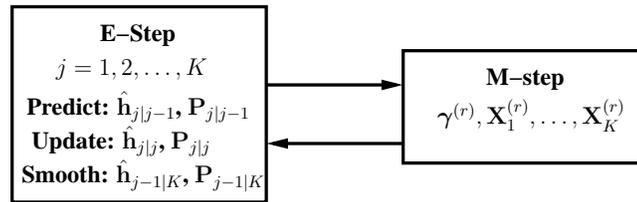
Figure 3.2: Block diagram depicting the JK-SBL algorithm. The a-sparse channel is estimated and tracked in the E-step, while the M-step learns the unknown hyperparameters $\boldsymbol{\gamma}$ and detects the unknown transmit data $\mathbf{X}_1, \ldots, \mathbf{X}_K$.

Thus far, we proposed algorithms for joint a-sparse channel estimation and data detection in block-fading and time-varying channels in OFDM systems. We now discuss

some implementation aspects of the proposed algorithms.

3.3.4 Discussion

In this subsection, we discuss the implementation of the proposed exact inference algorithms, and contrast their complexity with the block-based Autoregressive-SBL (AR-SBL) algorithm [125] and the approximate inference algorithm [44].

Consider the estimation of the wireless channels when the data is observed up to the Kth OFDM symbol. First, in the forward recursion, (3.22)-(3.26) are applied recursively until we reach the Kth OFDM symbol. Hence, in the forward recursion, we store the values of $\hat{\mathbf{h}}_{j|j}$, $\hat{\mathbf{h}}_{j|j-1}$, $\mathbf{P}_{j|j}$ and $\mathbf{P}_{j|j-1}$ for $j = 1, 2, \ldots, K$. Next, we apply the backward recursion using the Kalman smoother given by (3.27)-(3.28), i.e., KFS is applied to the whole sequence of observations before updating γ. The Kalman smoother helps to utilize all the information available in both the past and future symbols, and hence improves the channel estimates. For the K-SBL and JK-SBL algorithms, the smoothed mean and covariance are required for the computation of the M-step.

The K-SBL and JK-SBL algorithms are iterative in nature, and the filtering and smoothing equations are executed in the E-step of every iteration using the hyperparameters obtained in the M-step of the previous iteration and the unknown data for K symbols. Hence, the E-step performs exact inference, by obtaining the exact posterior distribution of the a-sparse channel, given the estimate of the hyperparameters. Exact inference ensures that the likelihood function increases at each EM iteration. However, the price paid for the exact inference methods is their higher complexity, as has been well-demonstrated by the simulation results in [44].

Using a flop-count analysis [129], for K ($K > 1$) OFDM symbols, the computations of the K-SBL and JK-SBL algorithms are dominated by the computation of the $\mathbf{J}_{K-1}$ term in the smoothing step, which has a complexity of $\mathcal{O}(KL^3)$ per iteration. In a block-based method such as the ARSBL, the computation of the covariance matrix Σ incurs a complexity of $\mathcal{O}(K^3P_t^2L)$ per iteration. Hence, we see that if the number of OFDM symbols to be tracked are such that $KP_t > L$, the complexity of the ARSBL algorithm is larger than the K-SBL algorithm. In other words, the K-SBL algorithm is a good choice among the exact inference techniques when the number of OFDM symbols to be tracked is large.

The proposed recursive algorithms are very flexible. For example, a pruning step, where small channel coefficients or hyperparameters are set to zero, can be incorporated between iterations. This leads to a reduced support set, which in turn results in faster convergence and lower complexity [9]. However, pruning may eliminate some of the basis vectors of the measurement matrix before achieving convergence and result in support recovery errors.

The improved channel estimation accuracy achieved by using the SBL techniques can lead to performance enhancements in different ways. As will be demonstrated in the next section, the BER performance can be improved, in both uncoded and coded systems. An additional approach could be to reduce, or optimize, the pilot density, with the aim of maximizing the outage capacity [70,114].

3.4 Simulation Results

In this section, we demonstrate the performance of the proposed channel estimation algorithms through Monte Carlo simulations. We consider the parameters in the 3GPP/LTE broadband standard [60,61]. We use a 3MHz OFDM system with 256 subcarriers, with a sampling frequency of $f_s = 3.84$MHz, resulting in an OFDM symbol duration of $\sim 83.3\mu s$ with Cyclic Prefix (CP) of $16.67\mu s$ (64 subcarriers). The length of the a-sparse channel (L) is taken to be equal to the length of the CP. Each OFDM frame consists of $K = 7$ OFDM symbols, which is also known as an OFDM slot. The data is transmitted using a rate $1/2$ Turbo code with QPSK modulation. For the Turbo code generation, we use the publicly available software [130], which uses a maximum of 10 Turbo iterations.

A sample instantiation of the a-sparse channel used in the simulations and the filtered MIP are depicted in Fig. 3.3. The figure captures the leakage effect due to finite bandwidth sampling and practical filtering. To generate the plot, we have used the Pedestrian B channel model [71] with Rayleigh fading. We have also used raised cosine filtering at the receiver and transmitter with a roll-off factor of 0.5 [61]. At the sampling frequencies considered, the number of significant channel taps are far fewer than the weak channel taps in the filtered impulse response, as seen in Fig. 3.3. In the following subsections, we present the simulation results for the block fading and time varying scenarios.

3.4.1 Block-fading Channel

In this subsection, we consider a block-fading channel and use $P_b = 44$ pilot subcarriers, uniformly placed in each OFDM symbol. Each OFDM frame consists of $K = 7$
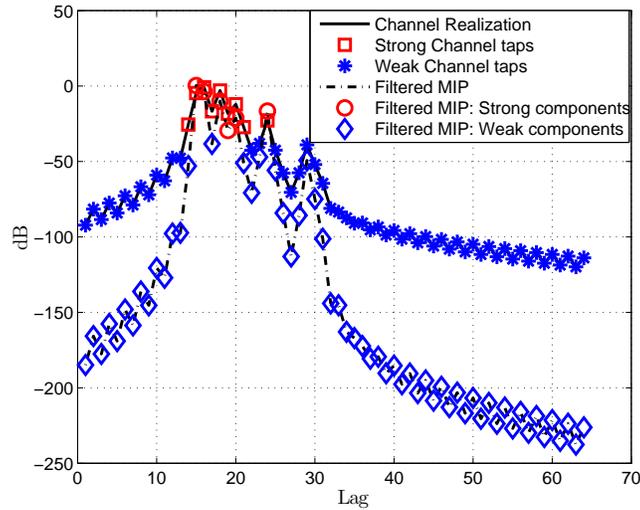
Figure 3.3: One sample channel realization of the a-sparse channel, along with the filtered MIP, i.e., the MIP when raised cosine filters are employed at the transmitter and receiver. The plot also shows the strong (> -30 dB) and weak (< -30 dB) channel taps and filtered-MIP components, to illustrate that the channel can indeed be modeled as being approximately sparse.

OFDM symbols. We implement the SBL and the J-SBL algorithm and plot the MSE performance of both the algorithms in Fig. 3.4, using a convergence criteria of $\epsilon = 10^{-9}$ and $r_{max} = 200$ for both the algorithms. We compare the MSE performance of the proposed algorithms with the CS based channel estimation technique [8], and the MIP-aware methods: pilot-only MIP-aware estimation [66] and the MIP-aware joint data and channel estimation algorithm, which we refer to as the EM-OFDM algorithm [72]. From Fig. 3.4, we observe that the SBL algorithms perform better than the MIP-unaware, non-iterative schemes such as the Frequency Domain Interpolation (FDI) technique. Among the iterative methods, the J-SBL algorithm performs an order of magnitude better than the SBL algorithm, especially at higher values of SNR, while being within 3 dB from the

MIP-aware EM-OFDM algorithm. The J-SBL jointly detects the $(KN - P_b)$ data symbols along with the estimating channel, resulting in a significantly lower overall MSE. As mentioned earlier, the RJ-SBL is mathematically equivalent to, and computationally simpler than, the J-SBL algorithm. Hence, they have the same performance.
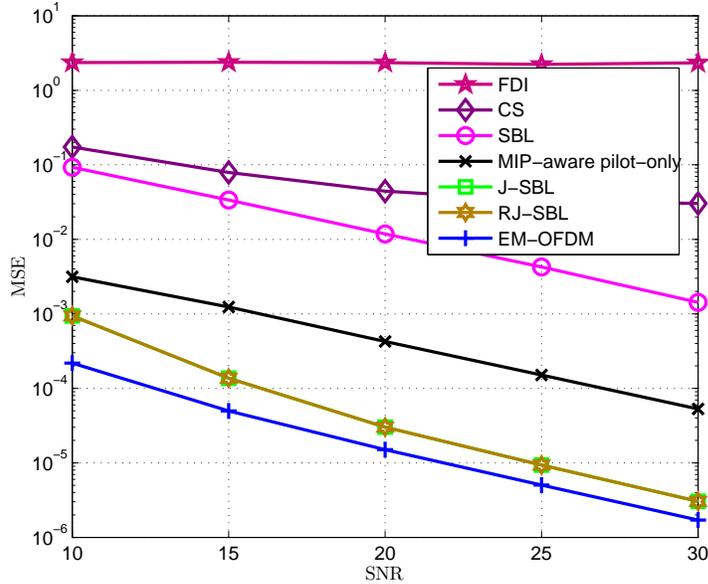
Figure 3.4: MSE performance of SBL, J-SBL/RJ-SBL algorithms compared to FDI [67], CS [8], MIP-aware pilot-only [66] and EM [72] schemes in a block-fading channel, with $P_b = 44$ pilot subcarriers, as a function of SNR in dB.

The coded and the uncoded BER performance of the EM, J-SBL and a genie receiver, i.e., a receiver with perfect knowledge of the channel (labeled as `Genie`), is shown in Fig. 3.5. We also compare the performance with SBL and MIP-aware pilot-only channel estimation followed by data detection. The BER performance of the RJ-SBL is superior that of the SBL and CS algorithms in both coded and uncoded cases. The MIP-aware pilot-only estimation method has a better BER performance compared to RJ-SBL for SNRs < 15 dB, in both coded and the uncoded cases. Also, the MIP-aware EM-OFDM

algorithm outperforms the proposed RJ-SBL algorithm by 3 dB. This is because, in the block-fading case, J-SBL algorithm suffers due to error propagation from the large number of data symbols that are simultaneously detected.
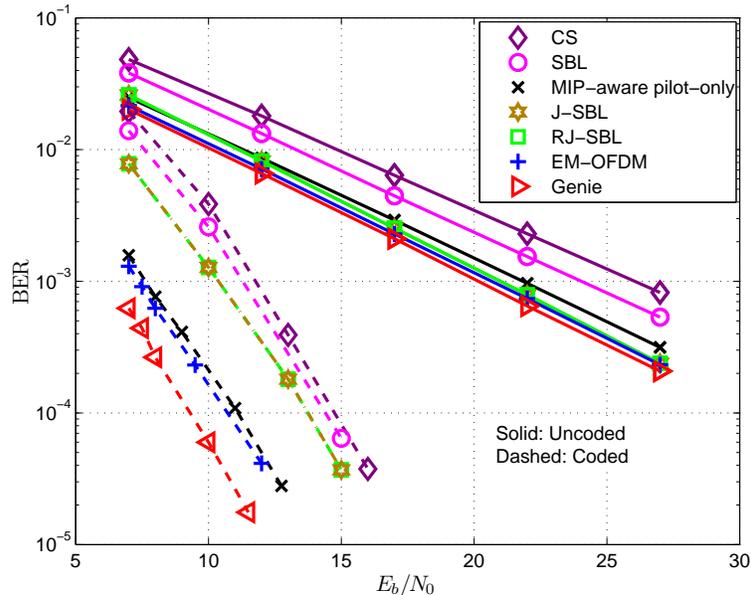
Figure 3.5: BER performance of the proposed algorithms in a block-fading channel, with $P_b = 44$ pilot subcarriers, as a function of E_b/N_0.

3.4.2 Slowly Time-varying Channel

In this section, we consider a slowly time-varying channel, simulated according to a Jakes' model [131] with a normalized fade rate of $f_d T_s = 0.001$ and $P_t = 44$ pilot subcarriers in every OFDM symbol. The MSE performance of the K-SBL and the JK-SBL algorithms are plotted against SNR in Fig. 3.6 and compared with the per-symbol MIP-unaware FDI [67], and the per-symbol J-SBL and the SBL algorithm. Figure 3.6 also shows the performance of the optimal MIP-aware Kalman tracking algorithm [121]

which considers all the subcarriers as carrying pilot symbols. The SBL and the J-SBL algorithms are not designed to exploit the temporal correlation in the channel, and hence, they perform 7-8 dB poorer than their recursive counterparts, the K-SBL and the JK-SBL algorithms. At higher SNR, we observe that the performance of the JK-SBL algorithm is only 2 dB worse than the MIP-aware Kalman tracking algorithm with all subcarriers being pilot subcarriers.
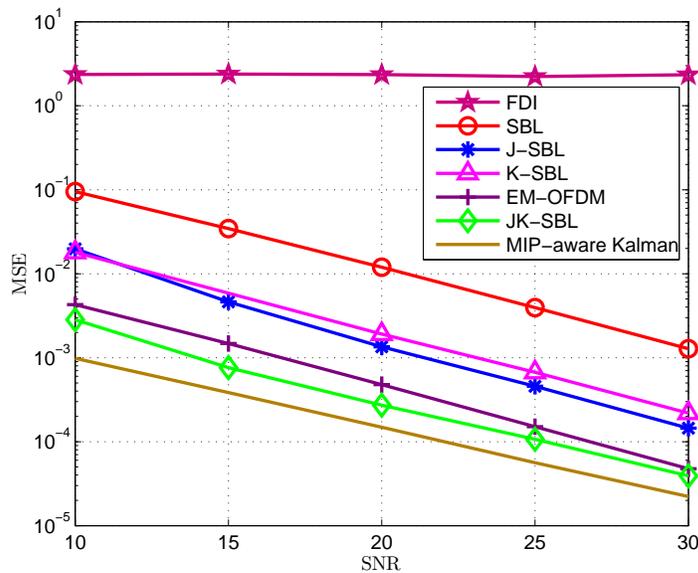
Figure 3.6: MSE performance of different schemes in a time-varying channel, compared to the optimal Kalman tracker [121] with $f_d T_s = 0.001$ and $P_t = 44$, as a function of SNR in dB.

In Fig. 3.7, we depict the BER performance of the proposed algorithms. We see that, in the coded case, while the JK-SBL performs about 2 dB better than the J-SBL algorithm, it is only a fraction of a dB away from performance of the genie receiver which has perfect channel knowledge. The JK-SBL outperforms pilots-only based channel

estimation using the K-SBL and the SBL algorithms by a large (4-5 dB) margin. Further, it outperforms the MIP-aware EM-OFDM algorithm, since the latter is unaware of the channel correlation, and performs channel estimation on a per-OFDM symbol basis; while the JK-SBL algorithm exploits its knowledge of the channel correlation to improve the channel estimates.
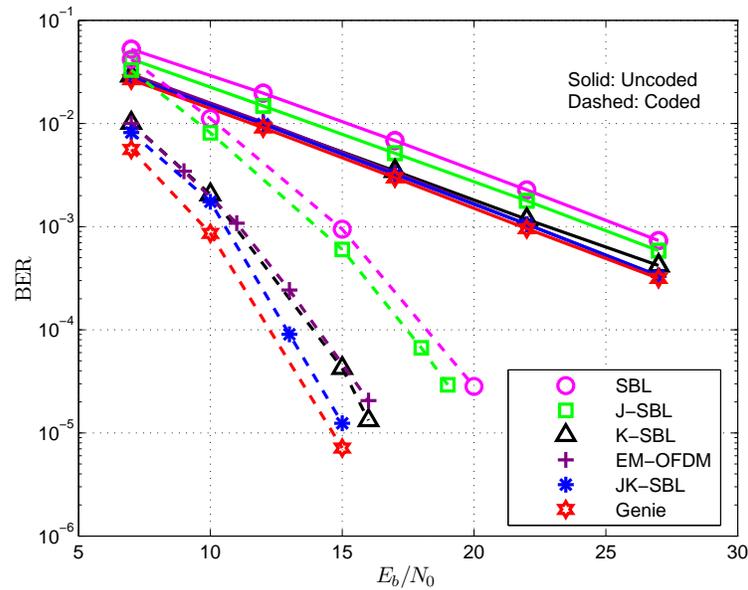
Figure 3.7: BER performance of different schemes in a time-varying channel with $f_d T_s = 0.001$ and $P_t = 44$, as a function of E_b/N_0.

In Fig. 3.8, we study the MSE performance of K-SBL and the JK-SBL algorithm across the OFDM frame as a function of the OFDM symbol index for SNRs of 10 and 30 dB. It is observed that after an initial reduction in the MSE, the MSE tends to remain more or less unchanged throughout the frame, especially at an SNR of 30 dB, indicating that the algorithms learn the hyperparameters within the first few OFDM symbols. Hence, this study shows that at a given SNR, it is possible to restrict the number of OFDM symbols over which the proposed algorithms need to learn the hyperparameters. After the

hyperparameters are estimated, channel tracking can be accomplished using the con-

ventional MIP-aware Kalman tracking algorithm. This can lead to additional reduction

in the computational complexity of the algorithms.
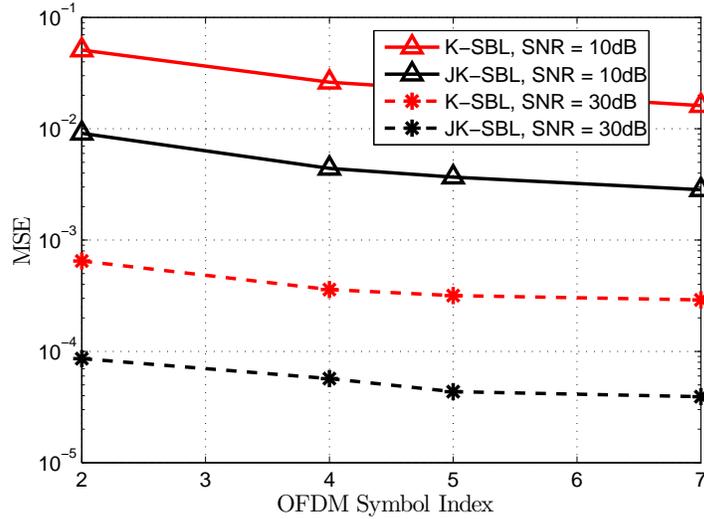
Figure 3.8: MSE performance of the K-SBL and the JK-SBL algorithms, as a function of the OFDM symbol index with $f_d T_s = 0.001$ and $P_t = 44$.

3.5 Conclusions

In this work, we considered the joint approximately sparse channel estimation and data

detection for block-fading and time-varying channels in SISO-OFDM systems, from the

perspective of SBL. To estimate the a-sparse block-fading channel, we proposed the SBL

algorithm and generalized it to obtain the J-SBL algorithm for joint a-sparse channel es-

timation and data detection. Furthermore, we obtained a mathematically equivalent

low-complexity RJ-SBL algorithm. For the time-varying channels, we used a first order

AR model to capture the temporal correlation of the a-sparse channel and proposed a

novel K-SBL algorithm, using which we tracked the a-sparse channel. We generalized the K-SBL algorithm to obtain the JK-SBL algorithm for joint channel estimation and data detection. We discussed the implementation issues of the recursive algorithms and showed that the proposed algorithms entail a significantly lower computational complexity compared to the previously known SBL techniques. Simulation results showed that the proposed recursive techniques exploit the temporal correlation of the channel, leading to an enhanced channel estimation and data detection capability compared to the per-symbol SBL and J-SBL algorithms, and also learn the hyperparameters within a few OFDM symbols.

Chapter 4

Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach

4.1 Introduction

Multiple Input Multiple Output (MIMO) technology combined with Orthogonal Frequency Division Multiplexing (OFDM) is the air-interface solution for next-generation broadband wireless systems and standards [64]. Most MIMO-OFDM wireless standards such as DVB-T, IEEE 802.11a, IEEE 802.16e etc., employ pilot-based channel estimation for data decoding. However, this approach necessitates the transmission of known pilots symbols on a set of anchor subcarriers per transmit antenna, leading to severe overheads on the spectral efficiency. In this chapter, we propose novel MIMO-OFDM channel estimation techniques using far fewer pilots compared to the conventional methods [66,67], by exploiting the inherent, approximate sparsity of the physical wireless channel.

In this chapter, we model the $N_t N_r$ spatially uncorrelated MIMO-OFDM wireless
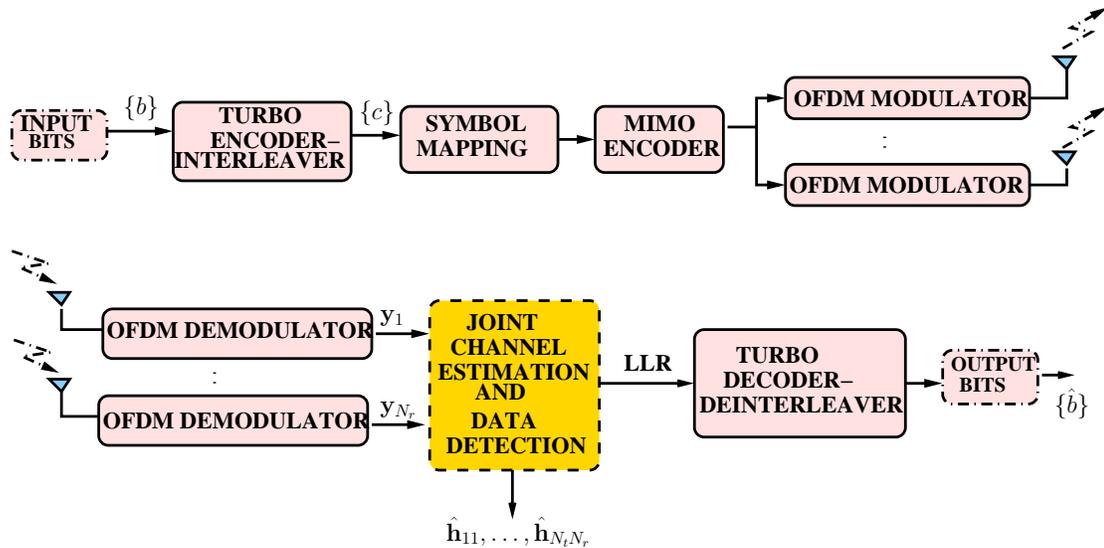
Figure 4.1: Turbo encoded/decoded transmitter and receiver chain of a MIMO-OFDM system. The dashed box (block shaded in yellow) highlights the proposed algorithms. Note that, the quantities of interest are the channel estimates $\hat{\mathbf{h}}_{11}, \ldots, \hat{\mathbf{h}}_{N_t N_r}$ and output bits $\{\hat{b}\}$.

channels as (a) group approximately-sparse (ga-sparse), and (b) group approximately-cluster sparse (ga-csparse). Further, we formulate the channel estimation problem in block-fading and time-varying channels and investigate both pilot-only channel estimation and joint channel estimation and data detection. Our goal is to design novel algorithms exploiting the structure in ga-sparse and ga-csparse channels for channel estimation and data detection in MIMO-OFDM systems.

4.1.1 Background and Literature Survey

In this subsection, we present the basic set-up of the coded MIMO-OFDM system considered in this work and formulate the problem of pilot-based channel estimation and joint channel estimation and data detection in a MIMO-OFDM system using the Multiple Measurement Vector (MMV) framework.

The goal of a MIMO-OFDM system is accurate data detection at the output of the decoder. We assume that the transmissions between the N_t transmit and the N_r receive antennas take place through OFDM frames, where every frame consists of K OFDM symbols. In this work, we consider both the block-fading channel, where the channel coefficients remain fixed across the OFDM frame duration and vary in an i.i.d. fashion from frame to frame, and the slowly time-varying channel, where the channel coefficients can vary across the OFDM frame duration. Fig. 4.1 shows the block diagram of a typical MIMO-OFDM system with N subcarriers, N_t transmit antennas and N_r receive antennas. At the transmitter of the discrete-time MIMO-OFDM system, $\{b\}$ input bits are first encoded and interleaved into a sequence of coded bits, $\{c\}$. The coded bits $\{c\}$ are mapped into an M-ary complex symbol sequence, which is split into N_t streams. At each transmit antenna, P_b pilots are inserted in an OFDM frame in the case of block-fading channels, and P_t pilots are inserted in every OFDM symbol of an OFDM frame in the case of time-varying channels. The pilot symbols along with coded data symbols $\{c\}$ are OFDM modulated and transmitted over the multipath fading channel of the kth OFDM symbol, denoted by $\mathbf{h}_{n_t n_r, k} \in \mathbb{C}^{L \times 1}$. Here, n_t (n_r) denotes the transmit (receive) antenna index, and L is the length of the channel. After OFDM demodulation, the signal received at the n_r^{th} receive antenna of the kth OFDM symbol is given by

$$\mathbf{y}_{n_r, k} = \sum_{n_t=1}^{N_t} \mathbf{X}_{n_t, k} \mathbf{F} \mathbf{h}_{n_t n_r, k} + \mathbf{v}_{n_r, k}, \quad n_r = 1, \ldots, N_r, \tag{4.1}$$

where the diagonal matrix $\mathbf{X}_{n_t, k} \in \mathbb{C}^{N \times N}$ consists of the pilot and data transmitted over the n_t^{th} transmit antenna and k^{th} OFDM symbol, and the matrix $\mathbf{F} \in \mathbb{C}^{N \times L}$ consists of the first L columns of the $N \times N$ DFT matrix. Each component of $\mathbf{v}_{n_r, k} \in \mathbb{C}^{N \times 1}$

is an circularly symmetric additive white Gaussian noise with probability distribution $\mathcal{CN}(0, \sigma^2)$.

In the complex baseband representation, the time domain channel impulse response, denoted by $\tilde{h}_{n_t n_r, k}[t], t \in \mathbb{R}$, can be modeled as a stationary tapped delay-line filter in the lag domain:

$$\tilde{h}_{n_t n_r, k}[t] = \sum_{l=1}^{\tilde{L}} \tilde{h}_{n_t n_r, k, l} \delta[t - \tau_l], \tag{4.2}$$

where $\delta[t]$ is the Dirac delta function, $\tilde{h}_{n_t n_r, k, l}$ and τ_l represent the attenuation and propagation delay, respectively, on the path l, and $\tilde{L}$ is the number of resolvable paths [62]. Wireless channel models obtained using channel sounding experiments, on the other hand, exhibit *approximate* sparsity in the lag domain, for e.g., due to non-perfect low-pass filtering using raised cosine filtering [70]. Based on these practical considerations, we model the lag domain *filtered* channel impulse response as, $h_{n_t n_r, k}[t] = g_t[t] * \tilde{h}_{n_t n_r, k}[t] * g_r[t]$, where $g_t[t]$ and $g_r[t]$ represent the baseband transmit and receive filters employed at the transmit and receive antennas, and $*$ represents the convolution operation. The corresponding discrete-time channel can be represented as $h_{n_t n_r, k}(l) = h_{n_t n_r, k}[(l-1)T]$, where T is the baud interval. The overall channel is represented as $\mathbf{h}_{n_t n_r, k} = [h_{n_t n_r, k}(1), h_{n_t n_r, k}(2), \ldots, h_{n_t n_r, k}(L)]^T$. In addition, it is known that the sample-spaced representation of $\tilde{h}_{n_t n_r, k}[t]$ between different transmit and receive antenna pairs are group-sparse [81, 82], i.e., the locations of non-zero elements of the sparse vectors coincide. Since $g_t[t]$ and $g_r[t]$ are identical for every transmit and receive antenna, we deduce that the locations of the significant components in $\mathbf{h}_{n_t n_r, k}$ also coincide across the entire MIMO-OFDM system. In this work, we consider the following scenarios:

- $\mathbf{h}_{n_t n_r,k}$ is *group approximately-sparse* (ga-sparse), i.e., the $N_t N_r$ wireless channels consists of a few strong components and several weak components, and the time-lags of strong and weak components between transmit and receive antenna pairs coincide.

- $\mathbf{h}_{n_t n_r,k}$ is *group approximately-cluster sparse* (ga-csparse), i.e., the $N_t N_r$ ga-sparse wireless channels consists of clusters such that the components of a given cluster are all strong or all weak. In addition, there are only a few clusters consisting of strong components.

Now, to estimate the ga-sparse and ga-csparse channels, we rewrite (4.1) in an MMV framework [104, 132]. Here, in the kth OFDM symbol, the observations from the N_r receivers form an observation matrix, $\mathbf{Y}_k$, which is related to the vectors in the channel matrix, $\mathbf{H}_k$, through a common dictionary $\mathbf{\Phi}_k$, as follows:

$$\underbrace{[\mathbf{y}_{1,k},\ldots,\mathbf{y}_{N_r,k}]}_{\mathbf{Y}_k \in \mathbb{C}^{N \times N_r}} = \underbrace{\mathbf{X}_k(\mathbf{I}_{N_t} \otimes \mathbf{F})}_{\mathbf{\Phi}_k \in \mathbb{C}^{N \times LN_t}} \underbrace{\begin{bmatrix} \mathbf{h}_{11,k} & \ldots & \mathbf{h}_{1N_r,k} \\ \vdots & \vdots & \\ \mathbf{h}_{N_t1,k} & \ldots & \mathbf{h}_{N_t N_r,k} \end{bmatrix}}_{\mathbf{H}_k \in \mathbb{C}^{LN_t \times N_r}} + \underbrace{[\mathbf{v}_{1,k},\ldots,\mathbf{v}_{N_r,k}]}_{\mathbf{V}_k \in \mathbb{C}^{N \times N_r}}, \qquad (4.3)$$

where the overall transmit data matrix $\mathbf{X}_k \in \mathbb{C}^{N \times NN_t}$ is given by $\mathbf{X}_k \triangleq [\mathbf{X}_{1,k}, \mathbf{X}_{2,k}, \ldots, \mathbf{X}_{N_t,k}]$. At the *P pilot subcarriers*, the MIMO-OFDM system model can be written as

$$\mathbf{Y}_{p,k} = \mathbf{\Phi}_{p,k}\mathbf{H}_k + \mathbf{V}_{p,k}, \qquad (4.4)$$

where $\mathbf{Y}_{p,k} \in \mathbb{C}^{P \times N_r}$, $\mathbf{\Phi}_{p,k} \in \mathbb{C}^{P \times LN_t}$ and $\mathbf{V}_{p,k} \in \mathbb{C}^{P \times N_r}$ are obtained by sampling $\mathbf{Y}_k$, $\mathbf{\Phi}_k$ and $\mathbf{V}_k$ at the pilot subcarriers, respectively.

Several channel estimation techniques for MIMO-OFDM systems have been proposed

in literature. Conventional pilot-based interpolation techniques based on frequency domain Least Squares (LS) or Minimum Mean Square Error (MMSE) based methods [66,67] and lag domain LS and MMSE [66]. These methods do not provide reliable estimates when $P_b < L$ unless the prior knowledge of the the average multipath power profile measured at a particular location, also known as the Multipath Intensity Profile (MIP) of the channel [65], is known. In scenarios where the MIP is not known, blind methods [80] and techniques based on Compressed Sensing (CS) [81, 82] have been employed. Specifically, CS based simultaneous Orthogonal Matching Pursuit (OMP) [83], Modified OMP [84], Simultaneous Basis Pursuit Denoising and Simultaneous OMP [86] have been proposed for pilot-assisted ga-sparse channel estimation in MIMO-OFDM systems. Further, CS based Block OMP (BOMP) has been proposed for pilot-assisted ga-csparse MIMO-OFDM channel estimation [87]. In general, CS based methods recover an approximately sparse vector by recovering the s significant non-zero coefficients [133]: a large value of s guarantees recovery accuracy, but requires a correspondingly large number of measurements. Bayesian algorithms such as the Temporal SBL (TSBL) [79] have been proposed for recovery of temporally correlated group-sparse vectors, by modeling the correlation among the group-sparse vectors using a general correlation structure. However, due to the generality of the correlation structure assumed, the complexity of such algorithms quickly becomes prohibitive as the time-window over which estimation is performed increases, making these algorithms unsuitable for OFDM channel tracking.

If the MIP is known, it is well-known that incorporating the observations available

at the data subcarriers into channel estimation using joint data detection and channel estimation techniques enhance the quality of channel estimates in MIMO-OFDM systems [132]. Recently, we showed that such joint approximately sparse channel estimation and data detection schemes enhance the quality of channel estimates in SISO-OFDM systems [90]. However, using multiple SISO-OFDM estimators in parallel to obtain estimators in the MIMO-OFDM context does not benefit from the spatial ga-sparse and the ga-csparse nature of the channel. The novelty of this work lies in formulating the problem such that the proposed Bayesian techniques for joint channel estimation and data detection harness the group and cluster-sparse nature of the wireless channels. To the best of the authors' knowledge, this is the first study in which such schemes are proposed for MIMO-OFDM systems.

4.1.2 Problem Formulation and Contributions

In this work, we propose Sparse Bayesian Learning (SBL)-based methods [34,49] for ga-sparse and ga-csparse channel estimation and data detection in MIMO-OFDM systems. Among the known Bayesian sparse signal recovery techniques [44,45], SBL exhibits the Expectation Maximization (EM) based monotonicity property, and offers guarantees such as convergence to the sparsest solution when the noise variance is zero, and convergence to a sparse local minimum, irrespective of the noise variance [49]. The SBL framework models the sparse channel vector to be estimated as follows:

- The ga-sparse channel is modeled as $\mathbf{h}_{n_t n_r,k} \sim \mathcal{CN}(0, \mathbf{\Gamma})$, where the so-called hyperparameters $\mathbf{\Gamma} = \text{diag}(\gamma(1), \ldots, \gamma(L))$ are *common* for the $N_t N_r$ channels over $1 \leq k \leq K$, i.e., the channels are spatially and temporally ga-sparse. Note that, if $\gamma(l) \to 0$, then the corresponding $h_{n_t n_r,k}(l) \to 0$ for all the $N_t N_r K$ channels [33,49].

- The ga-csparse channel $\mathbf{h}_{n_t n_r,k}$ is modeled as consisting of B clusters, each of length M, as follows:

$$\mathbf{h}_{n_t n_r,k} = [\underbrace{h_{n_t n_r,k}(1), \ldots, h_{n_t n_r,k}(M)}_{\mathbf{h}_{n_t n_r,1k} \in \mathbb{C}^{1 \times M}}; \ldots; \underbrace{h_{n_t n_r,k}((M-1)B+1), \ldots, h_{n_t n_r,k}(MB)}_{\mathbf{h}_{n_t n_r,Bk} \in \mathbb{C}^{1 \times M}}],$$

(4.5)

for $1 \leq k \leq K$. Here, the ga-csparse structure is exploited by modeling the bth cluster of the channel, $\mathbf{h}_{n_t n_r,bk}$, as $\mathcal{N}(0, \gamma(b)\mathbf{I}_M)$, where $\gamma(b)$ is an unknown hyper-parameter, such that, when $\gamma(b) = 0$, the bth block of $\mathbf{h}_{n_t n_r,k}$ is zero [134]. In addition, the different clusters of the ga-csparse channel are mutually uncorrelated, and hence, the overall covariance matrix of $\mathbf{h}_{n_t n_r,k}$ is a block-diagonal matrix with principal blocks given by $\gamma(b)\mathbf{I}_M$, $1 \leq b \leq B$.

The above prior models facilitate the estimation of the channel under the SBL framework, as they lead to closed-form solutions to the E and M steps of the EM algorithm. We note that the model need not match the true channel statistics: The goal here is to develop novel channel estimation and data detection algorithms that offer a coded BER performance comparable to that of a genie-aided MIP-aware receiver that has exact knowledge of the channel support. In particular, the above prior model results in sparse solutions to the channel vector, accurate channel estimation, and superior coded BER performance, while requiring far fewer pilot symbols compared to conventional methods that do not exploit the structured sparsity of the channel.

Depending on the mobility of the receiver, the ga-sparse and the ga-csparse channels may remain constant over the frame duration (block-fading), or may be slowly time-varying. When the channel is time-varying, the nonzero channel coefficients vary
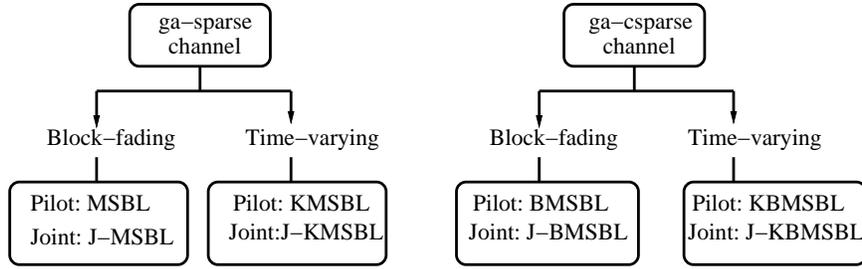
Figure 4.2: Pictorial representation of the algorithms proposed in this work.

slowly and are temporally correlated, but the locations of significant components of the channel remain constant for several OFDM frames [74]. In addition, it is known that a first order autoregressive (AR) model accurately captures the temporal behavior of fading wireless channels [120]. The channel evolution under the first order AR model is given by

$$\mathbf{h}_{n_t n_r, k} = \rho \mathbf{h}_{n_t n_r, k-1} + \mathbf{u}_{n_t n_r, k}, \tag{4.6}$$

where $\rho = J_0(2\pi f_d T_s)$, $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, f_d is the Doppler frequency, and T_s is the OFDM symbol duration [121]. The driving noise $\mathbf{u}_{n_t n_r, k}$ is distributed as $\mathcal{CN}(0, (1 - \rho^2)\gamma(i)\mathbf{I}_M)$.

The main contributions of this chapter are pictorially depicted in Fig. 4.2, and are as follows. First, for recovering ga-sparse channels in the MMV framework (4.3), we propose the following algorithms:

- In Sec. 4.2, we adapt the multiple response MSBL algorithm [104] and propose the novel Kalman MSBL (KMSBL) for pilot-based channel estimation in block-fading channels and time-varying channels, respectively.

- In Sec. 4.3, we propose the joint-MSBL (J-MSBL) and joint-KMSBL (J-KMSBL) algorithms for joint channel estimation/tracking and data detection in the block-fading and time-varying scenarios, respectively.

Next, in the context of recovering ga-csparse channels, we propose the following algorithms:

- In Sec. 4.4, we propose the Block MSBL (BMSBL) and Kalman BMSBL (KBMSBL) algorithms for pilot-based ga-csparse channel estimation in block-fading and time-varying channels, respectively.

- In Sec. 4.5, we propose the and Joint-BMSBL (J-BMSBL) and Joint-KBMSBL (J-KBMSBL) algorithms for joint channel tracking and data detection in the block-fading and time-varying scenario, respectively.

The joint counterparts of the proposed algorithms, wherein a joint ML estimation of both the hyperparameters and the data is performed, leads to significant enhancement in the quality of channel estimates. In the M-step, this joint estimation problem separates as independent optimization problems, leading to a simple, computationally inexpensive maximization procedure, with no loss of optimality. This, in turn, leads to significant improvement in the coded Bit Error Rate (BER) performance compared to the pilot-based and conventional methods. The algorithms proposed to handle the time-varying channel conditions fully exploit the correlation structure of the channel, resulting in gains of $1-2$ dB in the coded BER performance, as illustrated using Monte Carlo simulations. Further, we propose novel, implementation-friendly structures which lead to a lower computational load for ga-csparse channels.

4.2 Channel Estimation and Tracking Using Pilot Subcarriers for Ga-sparse Channels

In this section, we propose algorithms for ga-sparse channel estimation and tracking, using the pilot-subcarriers $\mathbf{Y}_{p,k}$ in (4.4), in both block-fading (the MSBL algorithm) and time-varying (the KMSBL algorithm) scenarios.

4.2.1 The MSBL Algorithm

In the MSBL framework, multiple group-sparse vectors are recovered from multiple observation vectors [104] with a parameterized prior incorporated to obtain group-sparse solutions. The prior density is given by

$$p(\mathbf{H}; \boldsymbol{\Gamma}) = \prod_{n_r=1}^{N_r} p(\mathbf{h}_{n_r}; \boldsymbol{\Gamma}), \tag{4.7}$$

where $\mathbf{h}_{n_r}$ represents the n_r^{th} column of $\mathbf{H}$, given by $\mathbf{h}_{n_r} = [\mathbf{h}_{1n_r}^T, \ldots, \mathbf{h}_{N_t n_r}^T]^T$, with a prior pdf of $\mathbf{h}_{n_r} \sim \mathcal{CN}(0, \boldsymbol{\Gamma}_b)$, $\boldsymbol{\Gamma}_b = \mathbf{I}_{N_t} \otimes \boldsymbol{\Gamma}$ which control the variances of elements in $\mathbf{H}$. The hyperparameters in $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = [\gamma(1), \gamma(2), \ldots, \gamma(L)]^T$, can be estimated using the type-II ML procedure [34], i.e., by maximizing the marginalized pdf $p(\mathbf{y}_{p,n_r}; \boldsymbol{\gamma})^1$ at the n_r^{th} receive antenna, as follows:

$$\gamma_{ML}(i) = \arg\max_{\gamma(i)\in\mathbb{R}_+} p(\mathbf{y}_{p,n_r}; \boldsymbol{\gamma}); \ 1 \leq i \leq L. \tag{4.8}$$

---

[1]Here, we describe the MSBL algorithm for $k = 1$, and hence, we drop the subscript k in $\mathbf{y}_{p,n_r}$ and $\mathbf{Y}_p$.

Since the above problem cannot be solved in closed form, iterative estimators such as the EM based[2] MSBL algorithm [104] are employed. In this approach, $\mathbf{H}$ is treated as the hidden variable, and the posterior distribution of $\mathbf{H}$ is obtained in the E-step and the ML estimate of $\boldsymbol{\gamma}$ is obtained in the M-step. The steps of the algorithm are given as

$$\text{E-step} : Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \mathbb{E}_{\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}_p, \mathbf{H}; \boldsymbol{\gamma})] \tag{4.9}$$

$$\text{M-step} : \gamma^{(r+1)}(i) = \underset{\gamma(i)\in\mathbb{R}_+}{\arg\max} Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}), \tag{4.10}$$

for $1 \leq i \leq L$, and the E and M steps are iterated until convergence. The E-step requires the posterior distribution $p(\mathbf{H}|\mathbf{Y}_p; \boldsymbol{\gamma}^{(r)})$, which can be obtained from the likelihood at the n_r^{th} receiver, as follows:

$$p(\mathbf{y}_{p,n_r}|\mathbf{h}_{n_r}) = \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left(-\frac{\|\mathbf{y}_{p,n_r} - \boldsymbol{\Phi}_p\mathbf{h}_{n_r}\|_2^2}{\sigma^2}\right). \tag{4.11}$$

Combining the likelihood and the prior distribution, the posterior distribution of $\mathbf{h}_{n_r}$ is given by $p\left(\mathbf{h}_{n_r}|\mathbf{y}_{p,n_r}; \boldsymbol{\gamma}^{(r)}\right) \sim \mathcal{CN}(\boldsymbol{\mu}_{n_r}, \boldsymbol{\Sigma})$, with mean and covariance given by

$$\boldsymbol{\mu}_{n_r} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_p^H\mathbf{y}_{p,n_r}, \ \boldsymbol{\Sigma} = \left(\frac{\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p}{\sigma^2} + \boldsymbol{\Gamma}_b^{(r)^{-1}}\right)^{-1}. \tag{4.12}$$

Here, $\boldsymbol{\Gamma}_b^{(r)}$ is the hyperparameter value in the rth iteration and $\boldsymbol{\Sigma}$ is common to all receive antenna, and hence, independent of the subscript n_r.

The M-step, given by (4.10), can be simplified to obtain the update equation for γ as

$$\gamma^{(r+1)}(i) = \frac{1}{N_t N_r} \sum_{n_r=1}^{N_r} \sum_{n_t=0}^{N_t-1} \left(|\boldsymbol{\mu}_{n_r}(i + n_t L)|^2 + \boldsymbol{\Sigma}(i + n_t L, i + n_t L)\right). \tag{4.13}$$

---

[2]Note that, all the algorithms proposed in the chapter use EM-based updates, and hence, they have a convergence guarantee to a local optima, with the likelihood increasing in each iteration [127].

Note that, in the above equation, the ga-sparse nature of the channel results in the update of γ which is *averaged* over the $N_t N_r$ channels of the MIMO-OFDM system. For a SISO-OFDM system, $N_t = N_r = 1$, and the above expression simplifies to the one obtained in [90].

The E-step involves computing the posterior mean and variance of the ga-sparse MIMO-OFDM channel as given in (4.12), incurring a computational complexity given by $\mathcal{O}(P_b^2 L)$ [104], while M-step computes the hyperparameter update as given in (4.13), incurring a computational complexity of $\mathcal{O}(N_t N_r L)$. In practice, it is found that an initial estimate for Γ given by $\Gamma^{(0)} = \mathbf{I}_{L \times L}$ is sufficient for the MSBL algorithm.

In the case of multiple OFDM symbols in a block-fading channel, the channel remains constant for the K OFDM symbols. The system model in (4.4) can be used for channel estimation, such that the number of observations corresponding to pilot subcarriers is P_b.

The MSBL algorithm, in the current form, cannot benefit from the correlation that exists in time-varying channels across OFDM symbols. In the following subsection, we extend MSBL algorithm to obtain the recursive KMSBL algorithm which exploits the temporal correlation across OFDM symbols, resulting in a significant performance improvement when the channel is slowly time-varying.

4.2.2 The KMSBL Algorithm

In this subsection, we describe the KMSBL algorithm which tracks the $N_t N_r$ ga-sparse MIMO-OFDM channels by exploiting both the group-sparsity and the temporal channel correlation using a Kalman filter and smoother (KFS) based recursive framework.

In the time-varying scenario, the state space equations for $k = 1, 2, \ldots, K - 1$ are as

follows:

$$\mathbf{Y}_{p,k} = \boldsymbol{\Phi}_{p,k}\mathbf{H}_k + \mathbf{V}_{p,k}, \tag{4.14}$$

$$\mathbf{H}_{k+1} = \rho\mathbf{H}_k + \mathbf{U}_{k+1}, \tag{4.15}$$

where $\boldsymbol{\Phi}_{p,k} = [\boldsymbol{\Phi}_{p,1,k}, \ldots, \boldsymbol{\Phi}_{p,N_t,k}]$, $\boldsymbol{\Phi}_{p,n_t,k} \triangleq \mathbf{X}_{p,n_t,k}\mathbf{F}_{p,n_t} \in \mathbb{C}^{P_t \times L}$, $\mathbf{X}_{p,n_t,k} \in \mathbb{C}^{P_t \times P_t}$ is a diagonal matrix consisting of pilots symbols transmitted from the n_t^{th} antenna in the kth OFDM symbol, and $\mathbf{F}_{p,n_t} \in \mathbb{C}^{P_t \times L}$ is a truncated DFT matrix consisting of the first L columns and the P_t rows corresponding to the pilot subcarriers of the n_t^{th} transmit antenna. Further, $\mathbf{H}_k$ consists of the N_r channels corresponding to the kth OFDM symbol, i.e., $\mathbf{H}_k = [\mathbf{h}_{1,k}, \ldots, \mathbf{h}_{N_r,k}]$ where $\mathbf{h}_{n_r,k} = [\mathbf{h}_{1n_r,k}^T, \ldots, \mathbf{h}_{N_tn_r,k}^T]^T$. In the above equation, we define $\mathbf{H}_0 \triangleq \mathbf{0}_{N_tL \times N_r}$, where $\mathbf{0}_{N_tL \times N_r}$ is an $N_tL \times N_r$ matrix of zeros. Note that the columns of the matrix $\mathbf{U}_{k+1}$ contain the driving noise vectors, $\mathbf{u}_{n_r,k+1}$ which consists of independent components $\mathbf{u}_{n_r,k+1}(i) \sim \mathcal{CN}(0, (1-\rho^2)\gamma(i))$. The initial condition for the ga-sparse channel is given by $\mathbf{h}_1 \sim \mathcal{CN}(0, \boldsymbol{\Gamma})$.

The EM steps of the KMSBL algorithm are as follows:

$$\text{E-step} : Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K|\mathbf{Y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}_p, \mathbf{H}_1, \ldots, \mathbf{H}_K; \boldsymbol{\gamma})]$$

$$\text{M-step} : \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \tag{4.16}$$

In the above expression, $\mathbf{Y}_p \triangleq [\mathbf{Y}_{p,1}, \ldots, \mathbf{Y}_{p,K}]$ represents the overall observation matrix.

To compute the E-step given above, we require the posterior distribution of the unknown ga-sparse channel $\mathbf{H}_k$. For this, we employ the Kalman based recursive update equations. The Kalman Filtering and Smoothing (KFS) equations for K OFDM symbols

are as follows [90, 124, 128]:

for $k = 1, \ldots, K$ **do**

Prediction: $\hat{\mathbf{H}}_{k|k-1} = \rho \hat{\mathbf{H}}_{k-1|k-1}$ (4.17)

$\mathbf{P}_{k|k-1} = \rho^2 \mathbf{P}_{k-1|k-1} + (1 - \rho^2)\mathbf{\Gamma}_b$ (4.18)

Filtering:

$\mathbf{G}_k = \mathbf{P}_{k|k-1}\mathbf{\Phi}_{p,k}^H \left(\sigma^2 \mathbf{I}_{P_t} + \mathbf{\Phi}_{p,k}\mathbf{P}_{k|k-1}\mathbf{\Phi}_{p,k}^H\right)^{-1}$ (4.19)

$\hat{\mathbf{H}}_{k|k} = \hat{\mathbf{H}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}_{p,k} - \mathbf{\Phi}_{p,k}\hat{\mathbf{H}}_{k|k-1})$ (4.20)

$\mathbf{P}_{k|k} = (\mathbf{I}_{N_t L} - \mathbf{G}_k\mathbf{\Phi}_{p,k})\mathbf{P}_{k|k-1}$ (4.21)

end (4.22)

for $j = K, K-1, \ldots, 2$ **do**

Smoothing:

$\hat{\mathbf{H}}_{j-1|K} = \hat{\mathbf{H}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{H}}_{j|K} - \hat{\mathbf{H}}_{j|j-1})$ (4.23)

$\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^H$ (4.24)

end (4.25)

where $\mathbf{J}_{j-1} \triangleq \rho \mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{G}_k$ is the Kalman gain matrix. In the above, the symbols $\hat{\mathbf{H}}_{k|k-1}$, $\mathbf{P}_{k|k-1}$, etc. have their usual meanings as in the KF literature [124]. For example, $\hat{\mathbf{H}}_{k|k-1}$ is the channel estimate at the kth OFDM symbol given the observations $\mathbf{Y}_{p,k-1}$; $\mathbf{P}_{k|k-1}$ is the covariance of the kth channel estimate given $\mathbf{Y}_{p,k-1}$, etc. The above KFS equations are initialized by $\hat{\mathbf{H}}_{0|0} = \mathbf{0}$ and $\mathbf{P}_{0|0} = \mathbf{\Gamma}_b = \mathbf{\Gamma} \otimes \mathbf{I}_{N_t}$.

In order to simplify (4.16), we use the joint pdf of the observations $\mathbf{Y}_p$ and $\mathbf{H}_1, \ldots, \mathbf{H}_K$,

given by

$$p(\mathbf{Y}_p, \mathbf{H}_1, \dots, \mathbf{H}_K; \boldsymbol{\gamma}) = \prod_{k=1}^{K} p(\mathbf{Y}_p | \mathbf{H}_1, \dots, \mathbf{H}_K) p(\mathbf{H}_k | \mathbf{H}_{k-1}; \boldsymbol{\gamma}). \tag{4.26}$$

Since $\mathbf{H}_k$ consists of columns $\mathbf{h}_{n_r,k}$ for $1 \leq n_r \leq N_r$, the M-step results in the following optimization problem:

$$\boldsymbol{\gamma}^{(r+1)} = \arg\max_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} \mathbb{E}_{\mathbf{H}_1, \dots, \mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}} \Big[K N_r \log |\boldsymbol{\Gamma}_b| + \frac{1}{(1-\rho^2)} \sum_{n_r=1}^{N_r} \sum_{k=2}^{K} \big[(\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1})^H \boldsymbol{\Gamma}_b^{-1} $$
$$(\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1}) + \mathbf{h}_{n_r,1}^H \boldsymbol{\Gamma}_b^{-1} \mathbf{h}_{n_r,1} \big] \Big] . \tag{4.27}$$

We see that the M-step requires the computation of

$\hat{\mathbf{H}}_{j|K} \triangleq \mathbb{E}_{\mathbf{H}_1, \dots, \mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}}[\mathbf{H}_j]$, and the covariance $\mathbb{E}_{\mathbf{H}_1, \dots, \mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}}[\mathbf{H}_j \mathbf{H}_j^H] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j|K}^H$

for $j = 1, \dots, K$, which is obtained from (4.17)-(4.25). The M-step also requires the computation of $\mathbb{E}_{\mathbf{H}_1, \dots, \mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}}[\mathbf{H}_j \mathbf{H}_{j-1}^H] \triangleq \mathbf{P}_{j,j-1|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j-1|K}^H$ for $j = K, K-1, \dots, 2$, which we obtain from [124] as follows:

$$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1} \mathbf{J}_{j-2}^H + \mathbf{J}_{j-1}^H (\mathbf{P}_{j,j-1|K} - \rho \mathbf{P}_{j-1|j-1}) \mathbf{J}_{j-2}. \tag{4.28}$$

The above recursion is initialized using $\mathbf{P}_{K,K-1|k} = \rho(\mathbf{I}_{N_t L} - \mathbf{G}_K \boldsymbol{\Phi}_{p,K}) \mathbf{P}_{K-1|K-1}$. Using the above expressions, the optimization problem in (4.27) can be written as

$$\boldsymbol{\gamma}^{(r+1)} = \arg\min_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} \Big\{ K N_t N_r \log |\boldsymbol{\Gamma}| + \sum_{n_t=1}^{N_t} \mathrm{Tr}(\boldsymbol{\Gamma}^{-1} \mathbf{M}_{n_t,1|K}) + \tfrac{1}{(1-\rho^2)} \sum_{j=2}^{K} \sum_{n_t=1}^{N_t} \mathrm{Tr}(\boldsymbol{\Gamma}^{-1} \mathbf{M}_{n_t,j|K}) \Big\}, \tag{4.29}$$

where $\mathbf{M}_{n_t,1|K} \in \mathbb{C}^{L \times L}$ is the submatrix consisting of rows and columns $(n_t - 1)L$ through $n_t L$ from the matrix $\mathbf{M}_{1|K} \triangleq N_r \mathbf{P}_{j|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j|K}^H + \rho^2 (N_r \mathbf{P}_{j-1|K} + \hat{\mathbf{H}}_{j-1|K} \hat{\mathbf{H}}_{j-1|K}^H) -$

$2\rho\text{Re}(N_r\mathbf{P}_{j,j-1|K} + \hat{\mathbf{H}}_{j|K}\hat{\mathbf{H}}^H_{j-1|K})$. Similarly, $\mathbf{M}_{n_t,j|K} \in \mathbb{C}^{L \times L}$ is the submatrix of $\mathbf{M}_{j|K} \triangleq$

$N_r\mathbf{P}_{1|K} + \hat{\mathbf{H}}_{1|K}\hat{\mathbf{H}}^H_{1|K}$, consisting of rows and columns $(n_t - 1)L$ through $n_t L$. Since the

individual channel components of $\mathbf{h}_{n_r,k}$ given by $\mathbf{h}_{n_t n_r,k}$ for $1 \leq n_t \leq N_t$ are governed

by $\boldsymbol{\gamma}$, we note that the update of $\boldsymbol{\gamma}$ is averaged over the N_t components via the summa-

tion over n_t. Differentiating (4.29) w.r.t. $\gamma(i)$ and setting the resulting expression to zero

and solving for $\boldsymbol{\gamma}$ gives the update for the ith hyperparameter as follows:

$$\gamma^{(r+1)}(i) = \left[\frac{1}{KN_tN_r} \left(\sum_{j=2}^{K} \sum_{n_t=1}^{N_t} \frac{M_{n_t,j|K}(i,i)}{(1-\rho^2)} + M_{n_t,1|K}(i,i) \right) \right]^+ , \tag{4.30}$$

for $i = 1,\ldots,L$. Thus the KMSBL algorithm learns $\boldsymbol{\gamma}$ in the M-step and provides low-

complexity and recursive estimates of the ga-sparse channel in the E-step.

Remarks: When $\rho = 1$, $\mathbf{H}_1 = \ldots = \mathbf{H}_K$, and hence, the channel is constant across the

OFDM frame, i.e., the channel is block-fading. Substituting $\rho = 1$ in (4.17)-(4.25), the

KFS update equations collapse to the following three equations:

$$\mathbf{G}_k = \mathbf{P}_{k-1|k-1}\mathbf{\Phi}^H_{p,k}(\sigma^2\mathbf{I}_{P_t} + \mathbf{\Phi}_{p,k}\mathbf{P}_{k-1|k-1}\mathbf{\Phi}^H_{p,k})^{-1} \tag{4.31}$$

$$\hat{\mathbf{H}}_{k|k} = \hat{\mathbf{H}}_{k-1|k-1} + \mathbf{G}_k(\mathbf{Y}_{p,k} - \mathbf{\Phi}_{p,k}\hat{\mathbf{H}}_{k-1|k-1}) \tag{4.32}$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_{N_tL} - \mathbf{G}_k\mathbf{\Phi}_{p,k})\mathbf{P}_{k-1|k-1}. \tag{4.33}$$

Further, when $\rho = 1$, the M-step of (4.27) simplifies to the M-step of MSBL given in

(4.10).

The KMSBL algorithm proposed in this section is a generalized version of the KSBL

algorithm proposed in [90] for pilot-based SISO-OFDM channel estimation, i.e., setting

$N_t = N_r = 1$ in the KMSBL algorithm leads to the KSBL algorithm. However, in con-

trast to the KSBL algorithm, the KMSBL algorithm is capable of harnessing the spatial

sparsity that exists in the MIMO-OFDM channel, and hence track N_r correlated channel vectors governed by a common γ.

In order to estimate the wireless channel when the data is observed up to the Kth OFDM symbol, (4.17)-(4.21) are applied recursively until we reach the Kth OFDM symbol in the forward recursion. We store the values of $\hat{\mathbf{H}}_{j|j}$, $\hat{\mathbf{H}}_{j|j-1}$, $\mathbf{P}_{j|j}$ and $\mathbf{P}_{j|j-1}$ for $j = 0, \ldots, K$ in the forward recursion. Next, we apply the backward recursion using the Kalman smoother given by (4.23)-(4.25), i.e., KFS is applied to the whole sequence of observations before updating γ. The Kalman smoother helps to utilize all the information available in both the past and future symbols, and hence improves the channel estimates.

Using a flop-count analysis [129], the complexity of the KMSBL algorithm is dominated by the computation of $\mathbf{J}_{k-1}$ in the smoothing step, which has a complexity of $\mathcal{O}(KL^3)$ per iteration per receive antenna. Hence, if $KP_t > L$, the complexity of the block-based ARSBL algorithm [79] is higher than the KMSBL algorithm. Thus, the KMSBL algorithm is a good choice among the exact inference techniques when the number of OFDM symbols to be tracked is large [90].

Although the algorithms proposed in this section are simple to implement, they are based on the pilot subcarriers only, and do not utilize the data subcarriers. In the following section, we propose such joint channel estimation and data detection schemes for ga-sparse channel estimation.

$$\boxed{\text{\textbf{E-step:} } Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}) = \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}}[\log p(\mathbf{H}, \mathbf{Y}; \boldsymbol{\gamma}, \mathbf{X})]}$$

$$\boxed{\text{\textbf{M-step:} } \arg\max_{\boldsymbol{\gamma}, \mathbf{X}} Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)})}$$

$$\boxed{\boldsymbol{\gamma}^{(r+1)} = \arg\max_{\boldsymbol{\gamma}} Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})} \qquad \boxed{\mathbf{X}^{(r+1)} = \arg\max_{\mathbf{X}} Q(\mathbf{X}|\mathbf{X}^{(r)})}$$
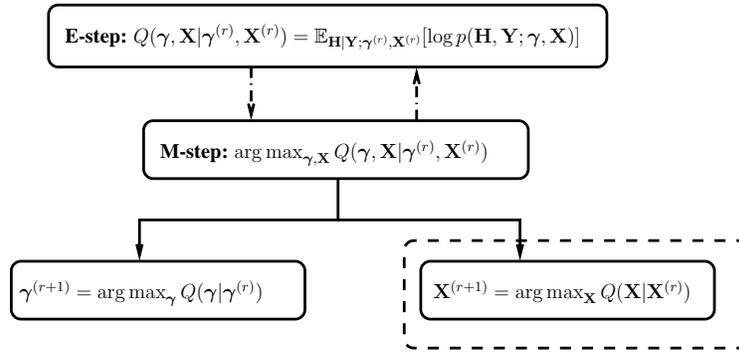
Figure 4.3: The J-MSBL algorithm: E-step computes the expectation over the posterior density of $\mathbf{H}$. The joint maximization in the M-step simplifies into two independent maximizations over $\boldsymbol{\gamma}$ and $\mathbf{X}$. The dashed box indicates the novelty in the J-MSBL approach.

4.3 Joint Channel Estimation/Tracking and Data Detection Using Pilot and Data Subcarriers for Ga-sparse Channels

In this section, we derive the novel J-MSBL and J-KMSBL algorithms that generalize the pilot-based MSBL and KMSBL algorithms for joint ga-sparse channel estimation and data detection in MIMO-OFDM systems. Further, using the recursive J-KMSBL algorithm, we show that a low-complexity recursive variant of J-MSBL can be derived using the KFS update equations given in (4.17)-(4.25).

4.3.1 The J-MSBL Algorithm

In this subsection, we derive the J-MSBL algorithm for joint estimation of the ga-sparse channels and the transmit data. To derive this algorithm, we modify the MSBL framework to include both the hyperparameters and the data transmitted in K OFDM symbols as unknown variables. That is, we consider $\mathbf{H}$ in (4.3) as the hidden variable and

$\left(\boldsymbol{\gamma}, \mathbf{X} \triangleq [\mathbf{X}_{11}, \ldots, \mathbf{X}_{n_t k}, \ldots, \mathbf{X}_{N_t K}]\right)$ as the parameters to be estimated. Here, $\mathbf{X}_{n_t k}$ consists of the data corresponding to the n_t^{th} antenna in the kth OFDM symbol. The E and the M-steps of the J-MSBL algorithm can be given as

$$\text{E-step} : Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}) = \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}, \mathbf{H}; \boldsymbol{\gamma}, \mathbf{X})]$$

$$\text{M-step} : \left(\boldsymbol{\gamma}^{(r+1)}, \mathbf{X}^{(r+1)}\right) = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}, \mathbf{X}:x_i\in\mathcal{S}}{\arg\max} \ Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}), \tag{4.34}$$

where x_i is an element in $\mathbf{X}$, and $\mathcal{S}$ is the constellation from which the symbols are transmitted. The E-step computes the posterior distribution at every receive antenna, and is given as $p(\mathbf{h}_{n_r}|\mathbf{y}_{n_r}; \boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}) \sim \mathcal{CN}(\boldsymbol{\mu}_{n_r}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu}_{n_r} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_b^H \mathbf{y}_{n_r k}, \ \boldsymbol{\Sigma} = \left(\sigma^{-2}\boldsymbol{\Phi}_b^H\boldsymbol{\Phi}_b + \boldsymbol{\Gamma}^{(r)^{-1}}\right)^{-1}, \tag{4.35}$$

for the K OFDM symbols in a frame. In (4.35), $\boldsymbol{\Phi}_b = [\boldsymbol{\Phi}_1^T, \ldots, \boldsymbol{\Phi}_K^T]^T$, and for $1 \leq k \leq K$, $\mathbf{F}_b = \mathbf{1}_{N_t} \otimes \mathbf{F}$, $\boldsymbol{\Phi}_k = \mathbf{F}_b\text{blkdiag}(\mathbf{X}_{1k}^{(r)}, \ldots, \mathbf{X}_{N_t k}^{(r)})$ and $\mathbf{y}_{n_r,k} = [\mathbf{y}_{1,k}^T, \ldots, \mathbf{y}_{N_r,k}^T]^T$.

At the outset, solving the optimization problem in the M-step in (4.34) seems an uphill task, as it involves joint optimization over $\mathbf{X}$ and $\boldsymbol{\gamma}$. However, in (4.34), the optimization problem w.r.t. to $\boldsymbol{\gamma}$ and $\mathbf{X}$ can be decoupled as the sum of two independent functions, $Q(\mathbf{X}|\mathbf{X}^{(r)}) \triangleq \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}}[\log p(\mathbf{Y}|\mathbf{H}; \mathbf{X})]$ and $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) \triangleq \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}}[\log p(\mathbf{H}; \boldsymbol{\gamma})]$. This is schematically illustrated in Fig. 4.3.[3] Further, we see that $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$ of the MSBL algorithm and the J-MSBL algorithm are identical, and hence, upon optimizing $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$ with respect to $\gamma(i)$, we obtain the expression for $\gamma^{(r+1)}(i)$ as in the MSBL algorithm,

---

[3]Notice that (4.10) and (4.34) are different, since the former uses the measurement matrix containing only the known pilot symbols, $\boldsymbol{\Phi}_p$, whereas the latter uses the measurement matrix consisting of pilot symbols along with the estimated data, together given by $\boldsymbol{\Phi}^{(r)}$.

given by (4.10). Further, the objective function to obtain $\mathbf{X}$, i.e., $Q(\mathbf{X}|\mathbf{X}^{(r)})$, can be derived as follows:

$$Q(\mathbf{X}|\mathbf{X}^{(r)}) = \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}} \left[\log \prod_{n_r=1}^{N_r} p(\mathbf{y}_{n_r,k}|\mathbf{h}_{n_r};\mathbf{X}) \right]$$

$$= -\mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}} \left[\sum_{n_r=1}^{N_r} \|\mathbf{y}_{n_r,k} - \boldsymbol{\Phi}_b \mathbf{h}_{n_r}\|_2^2 \right]. \tag{4.36}$$

and hence, the optimization problem for $\mathbf{X}$ is given by

$$X_{11}^{(r+1)}(i,i),\ldots,X_{N_t K}^{(r+1)}(i,i) = \underset{x_1,\ldots,x_{N_t}\in\mathcal{S}}{\arg\min}\ C(i,i) + \sum_{n_r=1}^{N_r} |y_{n_r,k}(i) - \sum_{n_t=1}^{N_t} x_{n_t,k} \mathbf{F}_b(i,:)\boldsymbol{\mu}_{n_r}|^2,$$

$$\tag{4.37}$$

where $i \in \mathcal{D}$, $\mathcal{D}$ is an index set consisting of the data subcarrier locations, $\mathbf{C} = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^H$, $\mathbf{F}_b(i,:)$ is the ith row of the $\mathbf{F}_b$ matrix, $\boldsymbol{\mu}_{n_r}$ and $\boldsymbol{\Sigma}$ are given in (4.35). The computational complexity of this algorithm is dominated by the inverse operation in (4.35), and is $\mathcal{O}(K^2 N^2 L N_t)$.

As stated in the previous section, the initial estimate of $\boldsymbol{\Gamma}$ is taken to be the identity matrix. The initialization of the $(KNN_t - P_b N_t)$ data symbols in turn requires an initial channel estimate. Hence, the initialization of $\mathbf{X}$ is obtained from the channel estimate obtained from a few iterations of the MSBL algorithm from the P_b pilots (denoted as $\hat{\mathbf{h}}_{MSBL}$). The ML data detection problem for obtaining the initial data estimates is given by

$$X_1^{(0)}(i,i),\ldots,X_{N_t}^{(0)}(i,i) = \underset{x_1,\ldots,x_{N_t}\in\mathcal{S}}{\arg\min}\ |y_{n_r,k}(i) - \sum_{n_t=1}^{N_t} x_{n_t} \mathbf{F}_b(i,:)\hat{\mathbf{h}}_{MSBL}|^2, \quad i \in \mathcal{D}. \tag{4.38}$$

In order to obtain the solution for both (4.37) and (4.38), we need to find the vector $[x_1,\ldots,x_{N_t}]$ that jointly minimizes (4.37). Although we can solve this problem with

moderate complexity for MIMO-OFDM systems with N_t up to 4 [60], the complexity of this problem is high for large values of N_t. In such scenarios, one can use sphere decoding [132].

In the following section, we discuss the pilot-based and joint channel estimation and data detection for time-varying ga-sparse MIMO-OFDM channels.

4.3.2 The J-KMSBL Algorithm

In this section, we generalize the KMSBL algorithm of Sec. 4.2.2 to obtain the J-KMSBL algorithm, which utilizes the observations available at the N subcarriers of the K OFDM symbols, and performs data detection at the $(N - P_t)$ data subcarriers of the OFDM symbol. Generalizing the J-MSBL to the J-KMSBL algorithm involves incorporating an E-step that exploits the correlation in the time-varying channels such that the algorithm is recursive in nature, and the smoothed channel estimates obtained for the K OFDM symbols are used to jointly estimate the ga-sparse channel and the unknown data of the K OFDM symbols.

Our starting point, again, is the state space model given by (4.15). The EM update equations in this context are given by

$$\text{E-step}: Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}) = \mathbb{E}_{\mathbf{H}_1, \ldots, \mathbf{H}_K|\mathbf{Y}; \boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}, \mathbf{H}_1, \ldots, \mathbf{H}_K; \boldsymbol{\gamma}, \mathbf{X})]$$

$$\text{M-step}: \left(\boldsymbol{\gamma}^{(r+1)}, \mathbf{X}^{(r+1)}\right) = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}, \mathbf{X}: x_i \in \mathcal{S}}{\arg\max} Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}), \tag{4.39}$$

where $\mathbf{X}$ comprises of the data transmitted on the K OFDM symbols, as defined in the previous subsection. Since the J-KMSBL algorithm uses the observations available at all the N subcarriers of an OFDM symbol, the recursive updates of the mean and

the covariance of the posterior distribution are given by (4.17)-(4.25), with $\mathbf{Y}_p$ and $\mathbf{\Phi}_p$ replaced by $\mathbf{Y}$ and $\mathbf{\Phi}$, respectively. Further, since $\mathbf{\Gamma}$ and data at the non-pilot subcarriers are unknown, the objective function in the M-step given by

$$
\begin{aligned}
Q\left(\mathbf{X}, \boldsymbol{\gamma} \mid \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = c - K N_r \log |\mathbf{\Gamma}_b| - \mathbb{E}_{\mathbf{H}_1, \ldots, \mathbf{H}_K \mid \mathbf{Y}; \mathbf{X}, \boldsymbol{\gamma}^{(r)}} \\
\left[\sum_{j=1}^{K} \sum_{n_r=1}^{N_r} \sigma^{-2} \left\| \mathbf{y}_{n_r, j} - \sum_{n_t=1}^{N_t} \mathbf{X}_{n_t, j} \mathbf{F} \mathbf{h}_{n_t n_r, j} \right\|^2 \right. \\
\left. - \sum_{j=2}^{K} \sum_{n_r=1}^{N_r} \frac{(\mathbf{h}_{n_r, j} - \rho \mathbf{h}_{n_r, j-1})^H \mathbf{\Gamma}_b^{-1} (\mathbf{h}_{n_r, j} - \rho \mathbf{h}_{n_r, j-1})}{(1 - \rho^2)} - \mathbf{h}_{n_r, 1}^H \mathbf{\Gamma}_b^{-1} \mathbf{h}_{n_r, 1} \right],
\end{aligned}
\tag{4.40}
$$

where c is a constant independent of $\boldsymbol{\gamma}$ and $\mathbf{X}$. The expression above is a sum of terms which are independent functions of $\boldsymbol{\gamma}$ and $\mathbf{X}_k \triangleq [\mathbf{X}_{1,k}, \ldots, \mathbf{X}_{N_t, k}]$ for $1 \leq k \leq K$, denoted as $Q\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(r)}\right)$ and $Q\left(\mathbf{X}_k \mid \mathbf{X}_k^{(r)}\right), 1 \leq k \leq K$, respectively. Further, we see that $Q\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(r)}\right)$ is the same as (4.29). Hence, the learning rule for $\boldsymbol{\gamma}$ follows from the M-step of the KMSBL algorithm, and is given by (4.30). The expression for $Q\left(\mathbf{X}_k \mid \mathbf{X}_k^{(r)}\right)$ is given by

$$
Q\left(\mathbf{X}_k \mid \mathbf{X}_k^{(r)}\right) = \mathbb{E}_{\mathbf{H}_k \mid \mathbf{Y}; \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}} \left[c - \sum_{n_r=1}^{N_r} \sigma^{-2} \left\| \mathbf{y}_{n_r, k} - \sum_{n_t=1}^{N_t} \mathbf{X}_{n_t, k} \mathbf{F} \mathbf{h}_{n_t n_r, k} \right\|^2 \right].
\tag{4.41}
$$

The M-step requires $\hat{\mathbf{H}}_{k \mid K} \triangleq \mathbb{E}_{\mathbf{H}_k \mid \mathbf{Y}; \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}}[\mathbf{H}_k]$ and $\mathbf{P}_{k \mid K} \triangleq \mathbb{E}_{\mathbf{H}_k \mid \mathbf{Y}; \mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}}[\mathbf{H}_k \mathbf{H}_k^H]$, which are given by the KFS equations of the E-step. The maximization of $Q\left(\mathbf{X}_k \mid \mathbf{X}_k^{(r)}\right)$ in (4.41) leads to the following optimization problem for $\mathbf{X}_k$:

$$
X_{1,k}^{(r+1)}(i, i), \ldots, X_{N_t, k}^{(r+1)}(i, i) = \underset{x_1, \ldots, x_{N_t} \in \mathcal{S}}{\arg \min} \, C(i, i) + \sum_{n_r=1}^{N_r} \left| y_{n_r}(i) - \sum_{n_t=1}^{N_t} x_{n_t} \mathbf{F}(i, :) \hat{\mathbf{h}}_{n_r, k \mid K} \right|^2,
\tag{4.42}
$$

where $i \in \mathcal{D}$, $\mathcal{D}$ is an index set consisting of the data subcarrier locations, $\mathbf{C} = \mathbf{\Phi} \mathbf{P}_{k \mid K} \mathbf{\Phi}^H$,

$\mathbf{F}(i,:)$ is the ith row of $\mathbf{F}$ and $\hat{\mathbf{h}}_{n_r,k|K}$ is the n_r^{th} column of $\hat{\mathbf{H}}_{k|K}$. Note that, in contrast to the expression for $\mathbf{C}$ in (4.37), the above expression is a function of $\mathbf{P}_{k|k}$ since the co-variance is computed recursively.

Data detection in the M-step results in the measurement matrix $\boldsymbol{\Phi}_k^{(r)}$ in the rth iteration and k^{th} OFDM symbol. Hence, the iterations of the J-KMSBL comprise the KFS update equations that incorporate $\boldsymbol{\Phi}_k^{(r)}$ instead of the pilot-only $\boldsymbol{\Phi}_{p,k}$ used in the KSBL algorithm. Further, the data detection in the M-step necessitates the initialization of transmit data, $\mathbf{X}_k^{(0)}$ for $0 \leq k \leq K$. We use the channel estimate obtained from a few iterations of the KMSBL algorithm from the P_t pilots (denoted as $\hat{\mathbf{h}}_{KMSBL}$) to obtain the initial estimate $\mathbf{X}_k^{(0)}$ for $0 \leq k \leq K$ and $i \in \mathcal{D}$ as

$$X_{1,k}^{(0)}(i,i), \dots, X_{N_t,k}^{(0)}(i,i) = \underset{x_1, \dots, x_{N_t} \in \mathcal{S}}{\arg \min} \ |y_{n_r}(i) - \sum_{n_t=1}^{N_t} x_{n_t} \mathbf{F}(i,:) \hat{\mathbf{h}}_{KMSBL}|^2. \qquad (4.43)$$

As mentioned in Sec. 4.2.2, when $\rho = 1$, the channel is block-fading in nature. Employing P_b pilots in an OFDM frame, we can emulate the block-fading scenario described in Sec. 4.3.1, and hence implement the J-MSBL algorithm recursively using KFS equations given by (4.33). Further, the M-step of the J-KMSBL algorithm is given by (4.30) and (4.43).

Thus far, we focussed on recovering the block-fading and time-varying ga-sparse channels using pilot-only and joint techniques. Next, we design pilot-only and joint channel estimation and data detection algorithms for group approximately *cluster-sparse* block-fading and time-varying channels.

4.4   Channel Estimation and Tracking Using Pilot Subcarriers for Ga-csparse Channels

In this section, we model the channel as ga-csparse, i.e., the entries of the approximately sparse channel are constrained to lie in a few clusters. The ga-csparse channel $\mathbf{h}_{n_t n_r}$ consists of B blocks each of length M, with each block containing all strong or all weak components, and the strong component clusters are few in number. The parametric prior modeling in SBL can be extended to the ga-csparse channels by assigning a hyperparameter $\gamma_c(i)$ to the ith cluster, $1 \leq i \leq B$, instead of the ith component, as given in Sec. 4.2.1. That is, a B length hyperparameter vector $\boldsymbol{\gamma}_c$ is associated with the pdf of $\mathbf{h}_{n_t n_r}$, such that every M length cluster of the channel is distributed as $\mathcal{CN}(0, \gamma_c(i))$.

First, we propose the Block MSBL (BMSBL) for pilot-based ga-csparse block-fading channel estimation in a MIMO-OFDM framework and propose to implement the BMSBL algorithm using the parallel cluster MSBL (PCMSBL) approach [135], which is same as the BMSBL in performance but has the advantage of computational complexity as it allows for the parallel implementation of the algorithm. Thereafter, we develop the Kalman-BMSBL (KBMSBL) algorithm for pilot-based ga-csparse time-varying channel estimation, and propose to implement the algorithm using the low-complexity Nested MSBL (NMSBL) approach [135].

4.4.1   The BMSBL Algorithm

In this subsection, we propose the Block MSBL (BMSBL) algorithm for pilot-based estimation of the block-fading ga-csparse channels, by generalizing the BSBL algorithm [134] to the multiple measurement scenario. That is, we recover the N_r ga-csparse

channels, $\mathbf{h}_{n_r,m}$ from N_r observation vectors, $\mathbf{y}_{n_r,m}$. Note that setting $N_r = N_t = 1$ leads to the SISO-OFDM problem, making the proposed algorithm backward compatible for the SISO-OFDM ga-csparse channel estimation.

The EM algorithm for ML estimation of the parameter $\boldsymbol{\gamma}_c$ in the BMSBL framework is as follows:

$$\text{E-step}: Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r)}\right) = \mathbb{E}_{\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}_c^{(r)}}[\log p(\mathbf{Y}_p, \mathbf{H}; \boldsymbol{\gamma}_c)]$$

$$\text{M-step}: \boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg\max} Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r)}\right). \tag{4.44}$$

The posterior distribution in the E-step above can be derived as $p(\mathbf{h}_{n_r}|\mathbf{y}_{p,n_r}; \boldsymbol{\gamma}_c^{(r)}) \sim \mathcal{CN}(\boldsymbol{\mu}_{c,n_r}, \boldsymbol{\Sigma}_c)$, where

$$\boldsymbol{\mu}_{c,n_r} = \boldsymbol{\Sigma}_c \boldsymbol{\Phi}_p^H \mathbf{y}_{p,n_r}, \quad \boldsymbol{\Sigma}_c = \sigma^{-2} \left(\frac{\boldsymbol{\Phi}_p^H \boldsymbol{\Phi}_p}{\sigma^2} + (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)^{-1} \right)^{-1} \tag{4.45}$$

Observe that the MSBL Sec. 4.2.1 and the BMSBL algorithms differ in the prior distribution of $\mathbf{H}$. The log-likelihood of the ga-csparse channel $\mathbf{H}$ is given by

$$\log p(\mathbf{H}; \boldsymbol{\gamma}_c) = c' - N_t N_r \log |(\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)| - \sum_{n_t=1}^{N_t} \sum_{n_r=1}^{N_r} \mathbf{h}_{n_t n_r}^H (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)^{-1} \mathbf{h}_{n_t n_r}, \tag{4.46}$$

where c' is a constant independent of $\boldsymbol{\gamma}_c$. Maximizing $Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r)}\right)$ in (4.44) w.r.t. $\boldsymbol{\gamma}_c$, we obtain the following

$$\boldsymbol{\gamma}_c^{(r+1)}(i) = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+}{\arg\min} MN_t N_r \log |\boldsymbol{\Gamma}_c| + \mathbb{E}_{\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}_c^{(r)}} \sum_{n_r=1}^{N_r} \left[\sum_{n_t=1}^{N_t} (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)^{-1} \text{Tr}[\mathbf{h}_{n_t n_r} \mathbf{h}_{n_t n_r}^H] \right].$$

$$\tag{4.47}$$

Simplifying the above, we obtain

$$\boldsymbol{\gamma}_c^{(r+1)}(i) = \frac{1}{MN_tN_r}\sum_{m=1}^{M}\sum_{n_r=1}^{N_r}\sum_{n_t=1}^{N_t}\boldsymbol{\Sigma}_{c,n_tn_r}(m,m) + |\boldsymbol{\mu}_{c,n_tn_r}(m)|^2. \qquad (4.48)$$

Note that, in contrast to (4.13), we obtain the averaging over the size of the cluster, since $\boldsymbol{\gamma}_c^{(r+1)}(i)$ is common to the entries of the cluster. Further, since the vectors are ga-csparse over N_t transmit and N_r receive antenna, we obtain the update, $\boldsymbol{\gamma}_c^{(r+1)}(i)$ which is averaged over N_tN_r channels of the MIMO-OFDM system.

Implementation of BMSBL

Here, we discuss the implementation of the BMSBL algorithm. We employ the PCSBL approach [135] which can significantly decrease the complexity of the proposed BMSBL approach.

The complexity of the BMSBL algorithm is dominated by the computation of the posterior covariance matrix $\boldsymbol{\Sigma}_c$, which incurs a computational load of $\mathcal{O}(N^2MB)$. In [135], we proposed an approach for estimating cluster-sparse signals and showed that the block-based algorithm in [134] is amenable to a parallel cluster SBL (PCSBL) implementation.

We employ the PCSBL approach to handle multiple measurements, as depicted in Fig. 4.4, where the ga-csparse channel is recovered by solving M parallel problems. The M-step is simply the average of the hyperparameter updates obtained from the M parallel problems per receive antenna. This approach incurs a maximum computational load of $\mathcal{O}(P_b^3)$, i.e., the complexity does not scale with $L = MB$.

The BMSBL algorithm is designed for block-fading channels, and cannot avail itself
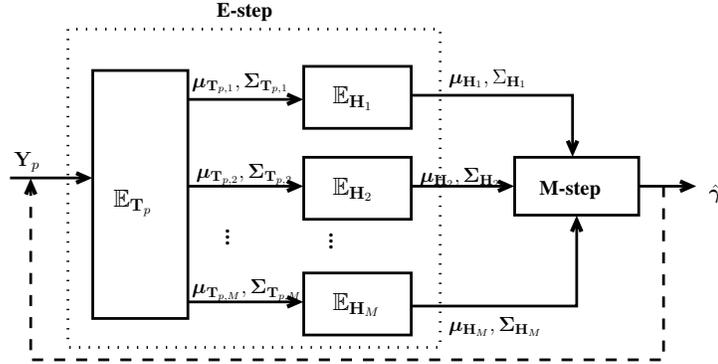
Figure 4.4: Block Diagram of the PCMSBL algorithm depicting M parallel branches.

of the temporal correlation in time-varying channels. In the following subsection, we propose a recursive KBMSBL algorithm for time-varying ga-csparse channel estimation that exploits its temporal correlation.

4.4.2 The KBMSBL Algorithm

In this subsection, we derive an algorithm for tracking the *slowly time-varying* ga-csparse MIMO-OFDM channel using the SBL framework. As in Sec. 4.2.2, we employ an AR model for the temporal evolution of the ga-csparse channel and derive recursive KFS based techniques. In addition, we propose a nested SBL approach [135] which facilitates the implementation of the proposed algorithm using M parallel Kalman filters/smoothers.

We formulate the ga-csparse channel estimation problem by modeling the channel corresponding to the kth OFDM symbol as $\mathbf{h}_{n_r,k} \sim \mathcal{CN}(0, (\mathbf{\Gamma}_c \otimes \mathbf{I}_M))$. We model the temporal variation of every cluster in the ga-csparse channel using the first order AR model as given in (4.15), i.e.,

$$\mathbf{h}_{n_t n_r, mk} = \rho \mathbf{h}_{n_t n_r, mk-1} + \mathbf{u}_{n_t n_r, mk}, \tag{4.49}$$

for $1 \leq m \leq M$, where $\mathbf{u}_{n_t n_r, mk}$ is spatio-temporally white, i.e., $\mathbf{u}_{n_t n_r, mk} \sim \mathcal{CN}(0, (1 - \rho^2)\gamma(m)\mathbf{I}_M)$. The E and M steps for pilot-based ga-csparse channel estimation are given by

$$\text{E-step} : Q\left(\boldsymbol{\gamma}_c | \boldsymbol{\gamma}_c^{(r)}\right) = \mathbb{E}_{\mathbf{H}_1, \ldots, \mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}_c^{(r)}} [\log p(\mathbf{Y}_p, \mathbf{H}_1, \ldots, \mathbf{H}_K; \boldsymbol{\gamma}_c)] \tag{4.50}$$

$$\text{M-step} : \boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg \max} \, Q\left(\boldsymbol{\gamma}_c | \boldsymbol{\gamma}_c^{(r)}\right). \tag{4.51}$$

The posterior distribution of $\mathbf{H}_1, \ldots, \mathbf{H}_K$ can be efficiently evaluated using the KFS equations given in (4.17) - (4.25), by replacing $\boldsymbol{\Gamma}_b$ by $\boldsymbol{\Gamma}_{cb} \triangleq (\mathbf{I}_{N_t} \otimes (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M))$. The logarithm of the conditional prior distribution is given by

$$\log p(\mathbf{H}_k | \mathbf{H}_{k-1}; \boldsymbol{\Gamma}_c) = KN_r \log |\boldsymbol{\Gamma}_{cb}| - \frac{\sum_{n_r=1}^{N_r} \sum_{k=2}^{K} (\mathbf{h}_{n_r,k} - \rho\mathbf{h}_{n_r,k-1})^H \boldsymbol{\Gamma}_{cb}^{-1}(\mathbf{h}_{n_r,k} - \rho\mathbf{h}_{n_r,k-1})}{(1 - \rho^2)}$$

$$- \sum_{n_r=1}^{N_r} \mathbf{h}_{n_r,1}^H \boldsymbol{\Gamma}_{cb}^{-1} \mathbf{h}_{n_r,1}, \tag{4.52}$$

Note that, in the above expression, $\boldsymbol{\Gamma}_{cb}^{-1}$ imposes the ga-csparse structure on the channel for K OFDM symbols. The M-step of KBMSBL can be simplified as follows:

$$\boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg \min} \, \mathbb{E}_{\mathbf{H}_1, \ldots, \mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\Gamma}_c^{(r)}} [KN_r \log |\boldsymbol{\Gamma}_{cb}|$$

$$+ \sum_{k=2}^{K} \sum_{n_r=1}^{N_r} \frac{(\mathbf{h}_{n_r,k} - \rho\mathbf{h}_{n_r,k-1})^H \boldsymbol{\Gamma}_{cb}^{-1}(\mathbf{h}_{n_r,k} - \rho\mathbf{h}_{n_r,k-1})}{(1 - \rho^2)} + \sum_{n_r=1}^{N_r} \mathbf{h}_{n_r,1}^H \boldsymbol{\Gamma}_{cb}^{-1} \mathbf{h}_{n_r,1}]. \tag{4.53}$$

Using (4.52), and invoking the fact that $\boldsymbol{\Gamma}_{cb} = (\mathbf{I}_{N_t} \otimes (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M))$, we can simplify (4.53)

as

$$\boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg \min} KMN_rN_t \log |\boldsymbol{\Gamma}_c| + \sum_{n_t=1}^{N_t} \mathrm{Tr}(\boldsymbol{\Gamma}_{cb}^{-1} \mathbf{M}_{cn_t,1|K})$$

$$+ \frac{1}{(1-\rho^2)} \sum_{k=2}^{K} \sum_{n_t=1}^{N_t} \mathrm{Tr}(\boldsymbol{\Gamma}_{cb}^{-1} \mathbf{M}_{cn_t,k|K}), \tag{4.54}$$

where $\mathbf{M}_{cn_t,j|K}$ consists of rows and columns $(n_t - 1)L$ through n_tL from the matrix

$$\mathbf{M}_{c,j|K} \triangleq N_r\mathbf{P}_{j|K} + \hat{\mathbf{H}}_{j|K}\hat{\mathbf{H}}_{j|K}^H + \rho^2(N_r\mathbf{P}_{j-1|K} + \hat{\mathbf{H}}_{j-1|K}\hat{\mathbf{H}}_{j-1|K}^H) - 2\rho\mathrm{Re}(N_r\mathbf{P}_{j,j-1|K} + \hat{\mathbf{H}}_{j|K}\hat{\mathbf{H}}_{j-1|K}^H).$$

Likewise, $\mathbf{M}_{cn_t,1|k}$ consists of rows and columns $(n_t - 1)L$ through n_tL from the matrix

$\mathbf{M}_{c,1|k} \triangleq N_r\mathbf{P}_{1|k} + \hat{\mathbf{H}}_{1|k}\hat{\mathbf{H}}_{1|k}^H$. Note that, in the above expressions, $\mathbf{P}_{j|k}$ and $\hat{\mathbf{H}}_{j|k}$ are a

function of $\boldsymbol{\gamma}_c$, unlike (4.29), where the expressions are a function of $\boldsymbol{\gamma}$. Differentiating

(4.54) w.r.t. $\gamma_c(i)$ and setting the resulting equation to zero gives the update for the ith

hyperparameter as follows:

$$\gamma_c^{(r+1)}(i) = \frac{1}{MKN_tN_r} \left(\sum_{n_t=1}^{N_t} \sum_{k=2}^{K} \sum_{m=1}^{M} \frac{\mathbf{M}_{m,cn_t,j|K}}{(1-\rho^2)} + \sum_{n_t=1}^{N_t} \sum_{m=1}^{M} \mathbf{M}_{m,cn_t,1|K} \right), \tag{4.55}$$

where, $\mathbf{M}_{m,cn_t,j|K}$ consists of rows and columns $(B-1)M$ through BM from the matrix

$\mathbf{M}_{cn_t,j|K}$ and $\mathbf{M}_{m,cn_t,1|K}$ consists of rows and columns $(B-1)M$ through BM from the

matrix $\mathbf{M}_{m,cn_t,1|K}$. Thus the KBMSBL algorithm learns $\boldsymbol{\gamma}_c$ in the M-step and provides

low-complexity and recursive estimates of the time-varying ga-csparse channel in the

E-step using the KFS framework.

Implementation of KBMSBL

The complexity of the KBMSBL algorithm is dominated by the term $\mathbf{J}_{k-1}$, whose com-

putational complexity is given by $\mathcal{O}(KL^3)$. In [135], we proposed a low-complexity,

Nested SBL (NSBL) approach for estimating cluster-sparse signals. In this approach,

we introduce auxiliary variables $\mathbf{t}_{n_t n_r, k} \in \mathbb{C}^{N \times 1}$, such that

$$\mathbf{t}_{n_t n_r, k} = \mathbf{\Phi}_{n_t, k} \mathbf{h}_{n_t n_r, kc} + \mathbf{z}_{n_t n_r, kc}. \tag{4.56}$$

The structuring of the vectors $\mathbf{h}_{n_t n_r, kc}$ is crucial for the NSBL algorithm since it directly affects the computational complexity. Here, we construct a vector $\mathbf{h}_{n_t n_r, kc}$, such that it consists of sub-vectors governed by a common hyperparameter vector $\boldsymbol{\gamma}_c$, i.e.,

$$\mathbf{h}_{n_t n_r, kc} = [h_{n_t n_r, 1k}(1), h_{n_t n_r, 2k}(1), \ldots, h_{n_t n_r, Bk}(1), \ldots, h_{n_t n_r, 1k}(M),$$
$$h_{n_t n_r, 2k}(M), \ldots, h_{n_t n_r, Bk}(M)]. \tag{4.57}$$

Accordingly, $\mathbf{\Phi}_{n_t, k}$ consists of the columns of $\mathbf{\Phi}_k$ corresponding to entries of $\mathbf{h}_{n_t n_r, kc}$. Although $\mathbf{z}_{n_t n_r, kc}$ cannot explicitly obtained, we note that its covariance can be written as $\mathbf{z}_{n_t n_r, kc} \sim \mathcal{CN}(0, \beta_m \sigma^2 \mathbf{I}_N)$ where, $0 \leq \beta_m \leq 1$ and $\sum_{m=1}^{M} \beta_m = 1$. [135]. Further, using $\mathbf{t}_{n_r, k} = \left[\mathbf{t}_{1n_r, k}^T, \ldots, \mathbf{t}_{N_t n_r, k}^T\right]$, we construct the matrix $\mathbf{T}_k \in \mathbb{C}^{N_t ML \times N_r}$ by stacking $\mathbf{t}_{1, k}, \ldots, \mathbf{t}_{N_r, k}$ as its columns. The auxiliary variable matrix $\mathbf{T}_k$ decomposes the problem of tracking ga-csparse channels into a problem of tracking M length ga-sparse channel component vectors.

The NSBL technique is implemented using two EM loops, one nested within the other, as depicted in Fig. 4.5. The outer EM loop consists of updating the posterior distribution of $\mathbf{T}_k$ for $1 \leq k \leq K$, and the inner EM loop consists of updating the posterior distribution of the ga-csparse channel using the KFS framework across the K OFDM symbols.

The inner EM loop in the NSBL algorithm is amenable to parallel implementation as M parallel KFS chains. Each KFS chain incurs a computational load of $\mathcal{O}(KB^3)$,
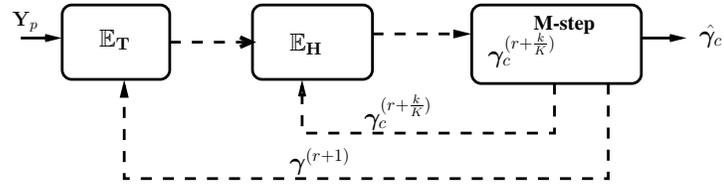
Figure 4.5: Block Diagram of the NSBL approach.

since the parallel chains track a vector in a lower dimension (B). The computational complexity of the outer loop of the NSBL approach dominates the overall complexity of the algorithm, and hence, the complexity of NSBL is given by $\mathcal{O}(K^3 M^2 P_t^2 L)$. Note that, in contrast to KBMSBL which incurs a computational complexity of $\mathcal{O}(KL^3)$, the complexity of the NSBL approach scales linearly in L. Hence, the NSBL approach leads to an efficient implementation of the proposed KBMSBL algorithm for large L.

In the following section, we briefly discuss the generalization of the BMSBL and KBMSBL approaches to joint channel estimation and data detection in time-varying ga-csparse MIMO-OFDM channels, before presenting our simulation results.

4.5 Joint Channel Estimation/Tracking and Data Detection Using Pilot and Data Subcarriers for Ga-csparse Channels

In this section, we develop the novel J-BMSBL and J-KBMSBL algorithm that generalize the pilot-based BMSBL and KBMSBL algorithms for joint ga-cparse channel estimation and data detection in MIMO-OFDM systems.

Here, the unknown variables are not only the hyperparameters but also the unknown transmit data symbols in the entire OFDM frame. We consider $\mathbf{H}$ as the hidden variable,

and, in contrast to BMSBL and KBMSBL, we consider $[\boldsymbol{\gamma}_c, \mathbf{X}]$, where $\mathbf{X} \triangleq [\mathbf{X}_1, \ldots, \mathbf{X}_{N_t}]$, as the parameters to be estimated.

Note that the BMSBL/KBMSBL algorithms differ from the ga-sparse based MSBL/KMSBL algorithms in the cluster-based hierarchical channel model, which, in turn, affects the posterior mean and variance of the channel. These posterior statistics affect the updates of $\boldsymbol{\gamma}_c$ as shown in (4.47) and (4.53). The updates of the transmit data $[\mathbf{X}_1, \ldots, \mathbf{X}_{N_t}]$ in the case of J-BMSBL, and $[\boldsymbol{\gamma}_c, \mathbf{X}_1, \ldots, \mathbf{X}_k]$ in the case of J-KBMSBL, can be obtained from the posterior estimates of the ga-csparse channel, from the E-step. Hence, the update equation for the transmit data remains the same as (4.37) and (4.42).

4.6 Simulation Results

In this section, we demonstrate the performance of the proposed channel estimation algorithms using Monte Carlo simulations. We consider the parameters in the 3GPP/LTE broadband standard [60,61]. We use a 3MHz 2×2 MIMO-OFDM system with 256 subcarriers, with a sampling frequency of $f_s = 3.84$MHz, resulting in an OFDM symbol duration of $\sim 83.3 \mu s$ with Cyclic Prefix (CP) of $16.67 \mu s$ (64 subcarriers). The length of ga-sparse channel (L) is taken to be equal to the length of the CP. Each frame of the MIMO-OFDM system consists of $K = 7$ OFDM symbols. The data is transmitted using a rate $1/2$ turbo code with QPSK modulation. For the turbo code generation, we use publicly available software [130], which uses a maximum of 10 turbo iterations. We use a convergence criterion of $\epsilon = 10^{-9}$ and $r_{max} = 200$ for all the algorithms. Further, we use a raised cosine filtering in every receive and transmit antenna chain with a roll-off factor of 0.5 [61] and the Pedestrian B channel model [71], which leads to approximately

sparse channels at the sampling frequencies considered [90].

4.6.1 Block-fading Ga-sparse and Ga-csparse Channels

In this subsection, we consider the pilot-only channel estimation and joint channel estimation and data detection in block-fading ga-sparse and ga-csparse channels. Each OFDM frame consists of $K = 7$ OFDM symbols, with $P_b = 44$ uniformly placed pilots in an OFDM frame of each transmitter. We implement the MSBL and the J-MSBL algorithm for ga-sparse and BMSBL and J-BMSBL algorithms (with block sizes of 4 and 6) in the case of ga-csparse block-fading channels, and plot the MSE and the coded BER performance of the algorithms in Fig. 4.6 and Fig. 4.7, respectively. We compare the performance of the proposed algorithms with the CS based Simultaneous OMP (SOMP) [12] using 50 pilots, MIP-aware methods: pilot-only MIP-aware estimation [66] and the MIP-aware joint data and channel estimation algorithm, which we refer to as the EM-OFDM algorithm [132].

From the top half of Fig. 4.6, we observe that the MSBL algorithms performs at least 1 dB better than the CS based SOMP technique. Since the proposed MSBL technique exploits spatial sparsity, it performs 5 dB better than the symbol-by-symbol SBL algorithm in [90]. We also observe that since BMSBL exploits the cluster-sparse structure, it outperforms the MSBL by $2 - 2.5$ dB. The bottom half of Fig. 4.6 depicts the MSE performance of joint data detection techniques that detect the $(KN - P_b)$ data symbols along with estimating the channel, resulting in a significantly lower overall MSE compared to pilot-only schemes. We see that among the joint SBL based iterative methods, the J-MSBL algorithm performs an order of magnitude better than the MSBL algorithm,

especially at higher values of SNR. Further, we see that J-BMSBL has a superior performance compared to J-MSBL and the symbol-by-symbol J-SBL [90]. Note that, J-BMSBL is less than a dB from the MIP-aware (support-aware) EM-OFDM algorithm.
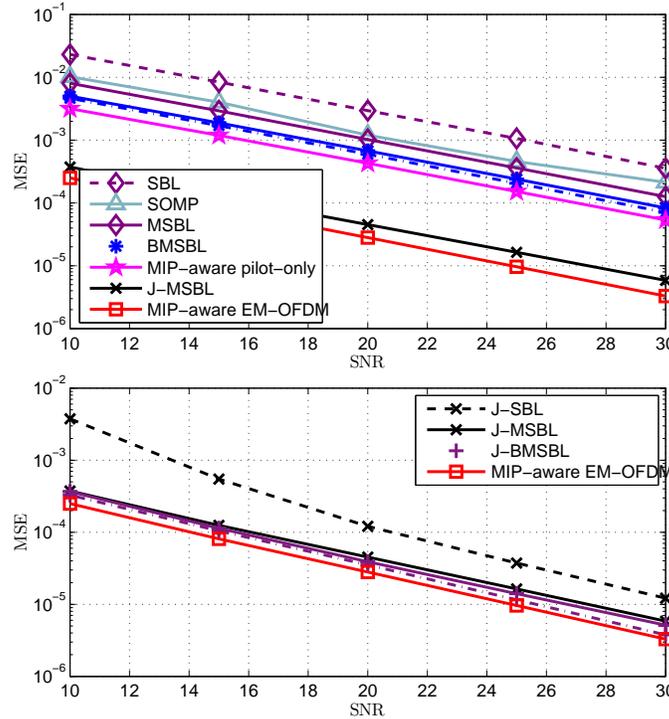
Figure 4.6: MSE performance in block-fading channels as a function of SNR in dB: Top: symbol-by-symbol SBL vs. MSBL vs. BMSBL. Bottom: J-SBL vs. J-MSBL vs. J-BMSBL. $P_b = 44$, ga-csparse: Solid curves - block size = 4, Dashed-dot curves - block size = 6.

The coded BER performance of the proposed schemes is compared with the MIP-aware EM-OFDM and with a genie receiver, i.e., a receiver with perfect knowledge of the channel (labeled as Genie), in Fig. 4.7. We also compare the performance with MSBL, BMSBL and MIP-aware pilot-only channel estimation followed by data detection. First, we observe that the MSBL algorithm performs 2 dB better than the SOMP scheme, while being more than a dB poorer than the BMSBL scheme. Further, the J-BMSBL technique, performs 1 dB better than the BMSBL scheme and 0.5 dB better

than the J-MSBL scheme and only 0.5 dB away from the MIP-aware pilot-only tech-
nique. Since the MIP-aware pilot-only technique estimates the channel from an overde-
termined system of equations, it outperforms the MIP-unaware pilot-only techniques.
Moreover, at lower SNRs (between $0 - 10$ dB), the joint channel estimation and data
detection techniques are prone to errors in the detected transmit data, and hence, they
are outperformed by MIP-unaware pilot-only techniques.
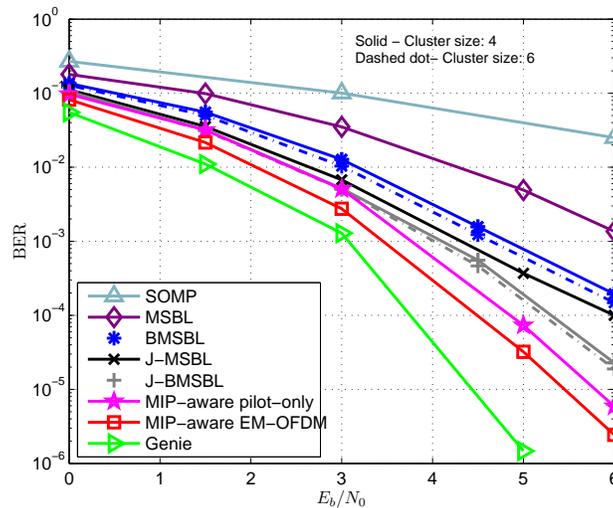
Figure 4.7: Coded BER performance of the proposed algorithms in a block-fading chan-
nel, with $P_b = 44$ pilot subcarriers, as a function of E_b/N_0.

4.6.2 Time-varying Ga-sparse and Ga-csparse Channels

In this section, we consider a slowly time-varying channel, simulated according to a
Jakes' model [131] with a normalized fade rate of $f_d T_s = 0.001$ and $P_t = 44$ pilot sub-
carriers in every OFDM symbol.

  The MSE performance of the proposed algorithms as a function of SNR are depicted

in Fig. 4.8. In the top half of the plot, we demonstrate that the pilot-only KMSBL algorithm performs $5 - 7$ dB better than the MSBL and symbol-by-symbol J-SBL algorithms, since the KMSBL algorithm exploits the temporal correlation and joint sparsity in time-varying channels. Further, we demonstrate that the KBMSBL technique which exploits the approximate cluster-sparsity performs $1.5 - 2.5$ dB better than the KMSBL algorithm, while being $5 - 6$ dB away from the MIP-aware Kalman tracking algorithm [121]. However, the J-KBMSBL algorithm performs 5 dB better than its pilot-only counterpart, i.e., the KBMSBL algorithm, while being less than a dB away from the MIP-aware Kalman tracking algorithm. The MIP-aware algorithm performs joint channel estimation and data detection, i.e., uses an MIP-aware EM algorithm which implements the channel estimation in the E-step using a Kalman tracker, and detects the transmit data in the M-step.

In the bottom half of Fig. 4.8, we demonstrate the performance of joint channel estimation and data detection schemes in time-varying channels. First, we observe that the symbol-by-symbol J-SBL algorithm that is not designed to exploit the temporal correlation performs 5-6 dB poorer than the recursive KMSBL and JKMSBL algorithms. At higher SNR, we observe that the performance of the JKMSBL algorithm is 2 dB away from the MIP-aware Kalman tracking algorithm. In contrast to pilot-only schemes, J-KMSBL and J-KBMSBL have the same performance, while being 1 dB away from the MIP-aware Kalman tracking algorithm. The gains due to modeling the channel as being cluster-sparse diminishes as the algorithm effectively has a higher number of known symbols to work with due to accurate detection of transmit data, especially at higher SNRs.
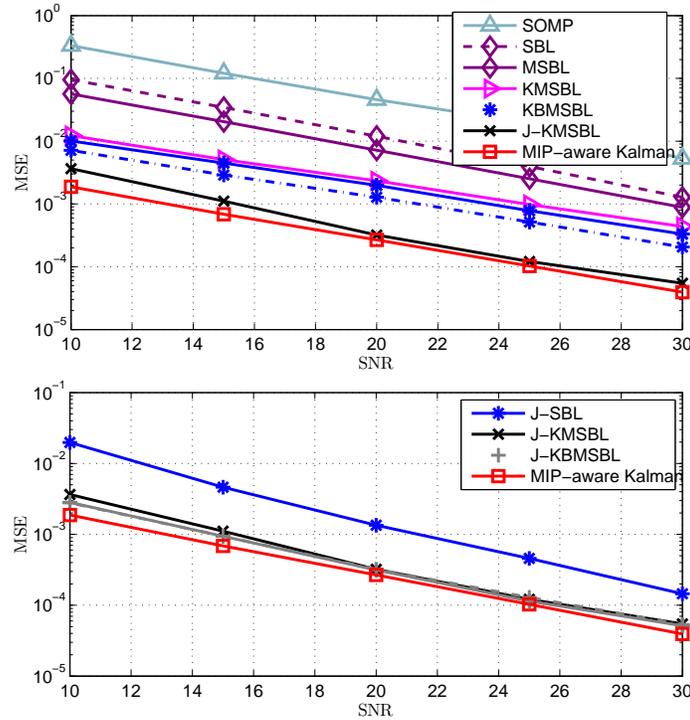
Figure 4.8: MSE performance in time-varying channels as a function of SNR in dB: Top: symbol-by-symbol SBL vs. MSBL vs. KMSBL vs. KBMSBL, Bottom: J-SBL vs. J-MSBL vs. J-KMSBL vs. J-KBMSBL, compared with the MIP-aware Kalman tracker [121]. $f_d T_s = 0.001$ and $P_t = 44$. Cluster-sparse: Solid curves - block size = 4, Dashed-dot curves - block size = 6.

In Fig. 4.9, we depict the coded BER performance of the proposed algorithms. We see that, while the proposed algorithms perform better than the SOMP algorithm by a margin larger than 2.5 dB, the J-KBMSBL is only a fraction of a dB away from performance of the MIP-aware Kalman receiver and the genie receiver which has perfect channel knowledge. The J-KSBL outperforms the pilots-only KMSBL by a margin of 0.5 dB. Further, the ga-csparse KBMSBL and J-KBMSBL algorithms perform better than their ga-sparse counterparts, i.e., KMSBL and J-KMSBL, by a margin of 0.5dB.
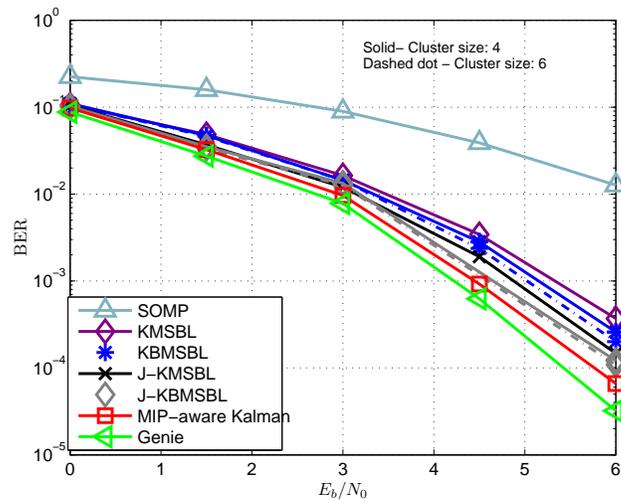
Figure 4.9: Coded BER performance of different schemes in a time-varying channel with $f_d T_s = 0.001$ and $P_t = 44$, as a function of E_b/N_0.

In both block-fading and time-varying channel conditions, the algorithms proposed for ga-csparse channels outperform their ga-sparse counterparts, affirming that exploiting the cluster-sparse nature of the channel does offer better performance in MIMO-OFDM channels. Intuitively, modeling the channel using $\gamma \in \mathbb{R}_+^{L \times 1}$ leads to over-fitting, which can be overcome by modeling the channel using $\gamma_c \in \mathbb{R}_+^{B \times 1}$, where $B < L$. We also observe that the performance is better for a block size of $M = 6$ (i.e., $B = 10$ blocks)

compared to $M = 4$ (i.e., $B = 15$ blocks), demonstrating that $M = 6$ is more appropriate for modeling the MIMO-OFDM channel, as it reduces the over-fitting compared to $M = 4$. We have seen, empirically, that $M = 10$ (i.e., $B = 6$ blocks) performs worse than $M = 6$. This is because $M = 10$ is perhaps an overly parsimonious model for the channel, leading to model mismatch and associated estimation errors. In general, the value of M needs to be tuned depending on the channel behavior, the transmit-receive filtering used, the sampling frequency, etc. A deeper study of the choice of M is beyond the scope of this chapter.

4.7   Conclusions

In this chapter, we considered the pilot-only channel estimation and joint channel estimation and data detection for block-fading and time-varying ga-sparse and ga-csparse channels in MIMO-OFDM systems, using the SBL framework. In the block-fading scenario, to estimate the ga-sparse and ga-csparse channels, we presented an adaptation of the MSBL and BMSBL algorithms. We also generalized them to obtain the J-MSBL and J-BMSBL algorithms for joint ga-sparse and ga-csparse channel estimation and data detection, respectively.

 In the time-varying scenario, we used a first order AR model to capture the temporal correlation of the channel and proposed the pilot-only based KMSBL and KBMSL algorithms, respectively. We generalized these algorithms to obtain the J-KMSBL and J-KBMSBL algorithms, respectively, for joint channel estimation and data detection. We discussed the computational aspects of the proposed algorithms and showed that the proposed recursive algorithms entail a significantly lower computational complexity

compared to the previously known SBL based techniques. Further, we discussed efficient implementation structures for ga-csparse channels in both block-fading and time-varying cases.

Simulation results showed that (i) joint algorithms outperformed their pilot-only counterparts, especially at higher SNRs, (ii) recursive techniques outperformed the symbol-by-symbol algorithms, and (iii) algorithms proposed in the context of ga-csparse channels outperformed their ga-sparse counterparts. Thus, it is, in general, beneficial to exploit any available structure in the sparsity of the signal being estimated. Future work can study the use of the algorithms developed in this chapter for other applications besides channel estimation and data detection in wireless communications.

Chapter 5

On the Recovery of Sparse and Structured Sparse Signals Using Sparse Bayesian Learning

5.1 Introduction

In problems of sparse learning and regression, in order to obtain sparse estimates, Bayesian inference is performed by incorporating a prior distribution on the vector to be recovered. Typically, the prior density contains a set of free parameters that are capable of inducing sparsity: the sparse vector estimation task involves inferring the values of such parameters based on the observed data. Lately, Bayesian methods for sparse recovery have found widespread applications as they often lead to simple, convergent algorithms accompanied by efficient implementations. Typically, Bayesian methods involve deriving the full posterior distribution of the missing variables conditioned on the observations, hence providing valuable statistical information as compared to CS based techniques that provide point estimates. In dynamic, time-varying scenarios that

131

involve recovering sparse vectors, the statistical information obtained from the posterior distribution is used to obtain predictive densities, which potentially lead to future states that are sparse [34]. Furthermore, in the context of structured sparsity, it has been found that it is hard to adapt non-Bayesian sparse recovery algorithms to scenarios where sparse signals exhibit inter and intra-vector correlations. In contrast, it is simple to incorporate structured sparsity using Bayesian methods by employing these constraints into the hierarchy of prior distributions or hierarchy of mixture distributions, which intuitively behave as a constraint on the space of all possible structured sparse solutions [44,79]. In addition, Bayesian methods infer the number of non-sparse components automatically, and hence, do not necessitate the information regarding the level of sparsity.

In this thesis, we focus on a popular Bayesian technique known as Sparse Bayesian Learning (SBL), which uses a parameterized prior to promote sparsity through a process of evidence maximization [34]. The prior distribution used in SBL regularizes the overcomplete problem given in (5.5), thereby circumventing its ill-posed nature. SBL uses a two-stage hierarchical distribution on the unknown vector, as shown in Fig. 1.1. Here, $\mathbf{x} \sim \mathcal{N}(0, \mathbf{\Gamma})$, where the diagonal matrix $\mathbf{\Gamma}$ contains the *hyperparameters* $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L]^T$ as its diagonal elements. In order to obtain an explicit posterior density function and yet promote sparsity, a conjugate Inverse Gamma (IG) *hyperprior* parameterized by a and b is assumed for $\boldsymbol{\gamma}$, i.e.

$$\boldsymbol{\gamma} \sim \mathcal{IG}(a, b), \quad a > 0, \ b > 0. \tag{5.1}$$

By marginalizing over $\boldsymbol{\gamma}$, the distribution of sparse vector $\mathbf{x}$ parameterized by a and b

can be obtained as

$$p(\mathbf{x}|a, b) = \int p(\mathbf{x}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}; a, b)\mathrm{d}\boldsymbol{\gamma} \tag{5.2}$$

$$\propto \prod_{l=1}^{L} \left(b + \frac{x(l)^2}{2} \right)^{\left(a + \frac{1}{2}\right)}, \tag{5.3}$$

which is a Student-t prior on the vector $\mathbf{x}$ [34]. Further, note that as $a, b \to 0$, the Student-t distribution is given by

$$p(\mathbf{x}|a = 0, b = 0) \propto \prod_{l=1}^{L} \frac{1}{|x(l)|}. \tag{5.4}$$

Such a prior is strongly peaked around $x(i) = 0$, and consequently, the overall prior $p(\mathbf{x})$ is sparsity promoting.

An empirical SBL technique which incorporates a Student-t prior with $a = b = 0$ was first proposed by Tipping for learning the sparse vector, based on the Expectation Maximization (EM) and McKay updates [34]. Several results that elucidate the general behavior of the nonconvex SBL cost function and solid theoretical justification for using the EM-based update equations are provided in [49]. Specifically, the authors prove that the global minima of the SBL cost function is always achieved at the maximally sparse solution in the absence of noise, and, irrespective of noise, the local minima are sparse. SBL based algorithms have been successfully employed for image/visual tracking [56], neuro-imaging [57, 58], beamforming [59], and joint channel estimation and data detection for OFDM systems [90].

The Single Measurement Vector (SMV) SBL system model given by

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n}, \tag{5.5}$$

as described in (1.1). The model given above can be generalized to a Multiple Measurement Vector (MMV) SBL model given by

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{V}, \tag{5.6}$$

where, $\mathbf{Y} \in \mathbb{C}^{N \times K}$ represents the K observation vectors stacked as its columns. The columns of $\mathbf{X} \in \mathbb{C}^{L \times K}$ represent the K sparse/compressible vectors, and each column of $\mathbf{V} \in \mathbb{C}^{N \times K}$ is modeled as i.i.d. and with the same distribution as $\mathbf{n}$ in (1.1) [104]. Albeit related, a more general MMV system model is given by

$$\mathbf{y}_k = \mathbf{\Phi}_k \mathbf{x}_k + \mathbf{v}_k, \quad \text{for} \quad 1 \leq k \leq K. \tag{5.7}$$

Note that compared to (5.6), the equation given above involves K distinct matrices $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K$ for different instantiations of the sparse vector, i.e., the system model given above collapses to (5.6) when $\mathbf{\Phi} = \mathbf{\Phi}_1 = \ldots = \mathbf{\Phi}_K$, and to (5.5) for $K = 1$.

Several extensions of SBL have been proposed that handle group sparsity, cluster sparsity and correlation constraints in (5.6). In [104], the authors have proposed the MSBL algorithm for simultaneous recovery of sparse columns of $\mathbf{X}$ from measurements in $\mathbf{Y}$. Consider a scenario where the columns of $\mathbf{X}$ are correlated, i.e., for $i \neq j$,

$$\mathbb{E}[\mathbf{x}_i \mathbf{x}_j^H] = \mathbf{B}, \tag{5.8}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ represents two distinct columns of $\mathbf{X}$ and $\mathbf{B}$ is a covariance matrix such that $B(i, j) \neq 0$ for $i \neq j$, i.e., $\mathbf{B}$ is not a diagonal matrix. In order to recover columns of $\mathbf{X}$ in such scenarios, the Temporal SBL (TSBL) algorithm has been proposed [79]. TSBL converts the problem in the MMV framework to a problem in the SMV framework

by vectorizing $\mathbf{X}^T$, i.e., $\mathbf{x}_c = \text{vec}(\mathbf{X}^T)$. Then, the MMV recovery problem in (5.6) can be reformulated as a problem of estimating the vector $\mathbf{x}_c$ in an SMV formulation, as follows:

$$\mathbf{y}_c = \mathbf{\Phi}_c \mathbf{x}_c + \mathbf{v}_c, \tag{5.9}$$

where $\mathbf{\Phi}_c \in \mathbb{C}^{N \times MB}$ is given as $\mathbf{\Phi}_c = \mathbf{\Phi} \otimes \mathbf{I}_K$ is the restructured measurement matrix. Further, $\mathbf{v}_c$ is the ambient noise given by $\mathbf{v}_c = \text{vec}(\mathbf{V}^T)$. By construction, the vector $\mathbf{x}_c$ consists of correlated non-zero entries which occur in clusters, and hence, such vectors are called as *block-sparse* or *cluster-sparse*. From (5.9), we see that the problem of recovering correlated MMV vectors can be formulated as a cluster-sparse vector recovery problem, where the *inter-vector* correlation is manifested as *intra-vector* correlation, and in particular *intra-cluster* correlation. There are several applications where cluster sparsity and intra-vector correlation arise naturally (see [136] and references therein). In particular, strong intra-cluster correlation has been observed in EEG, ECG and several physiological signals [137]. Popular CS based approaches include those which exploit cluster sparsity in linear models using mixed penalty such as the $\ell_1 - \ell_2$ and $\ell_1 - \ell_\infty$ [138–140] and techniques such as block matching pursuit, block orthogonal matching pursuit [139], and block-CoSamp [141]. However, none of the techniques based on CS exploit the intra-cluster correlation in the cluster-sparse signal. In the SBL framework, a cluster-sparse vector recovery algorithm known as the BSBL algorithm [134, 142] is proposed, which, in addition to incorporating the cluster-sparse

structure into the Bayesian framework, also exploits the intra-cluster correlation.[1] Furthermore, when the intra-block correlation is not known, the BSBL framework uses an approximate heuristic to estimate the intra-block correlation from the observations themselves, in an MMV setup.

5.1.1 Proposed Algorithms

In this subsection, we describe the novel SBL-based algorithms proposed in this thesis. These algorithms are proposed in the context of group-sparse and cluster-sparse signals, and they offer efficient and low-complexity solutions. We also provide a variant of SBL that addresses recovery of partially known measurement matrices jointly with the recovery of sparse vectors.

First, we consider the problem of recovering correlated vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ which have a common support in an generalized MMV framework, as given in (5.7). We formulate the recovery of correlated vectors using a recursive framework by modeling the correlation using a first order AR model leading to the state space model given by

$$\mathbf{y}_k = \boldsymbol{\Phi}_k \mathbf{x}_k + \mathbf{v}_k, \tag{5.10}$$

$$\mathbf{x}_k = \rho \mathbf{x}_{k-1} + \mathbf{z}_k, \quad k = 1, 2, \ldots, K, \tag{5.11}$$

where, by definition, $\mathbf{x}_0 \triangleq \mathbf{0}_L$, where $\mathbf{0}_L$ represents an L-length vector of zeros, ρ represents the AR co-efficient. Also, $\mathbf{z}_k$ is the the the noise driving the state sequence, consisting

---

[1]A subtle difference between the TSBL and the BSBL algorithm lies in the structure of the measurement matrix. In the TSBL algorithm, the measurement matrix is restructured in order to facilitate cluster-sparse vector recovery, and hence, the measurement matrix is of the form $\boldsymbol{\Phi} \otimes \mathbf{I}_K$. However, the BSBL algorithm recovers the cluster-sparse vectors in the presence of generic measurement matrices.

of independent components given by $z_k(i) \sim \mathcal{CN}(0, (1 - \rho^2)\mathbf{\Gamma})$. The recursive formulation along with the nature of the prior distributions allows us to use the Kalman filtering and smoothing (KFS) framework. Such a KFS framework is amenable to an efficient implementation of the proposed algorithm, which we refer to as the Kalman SBL (KSBL) algorithm. We show that when K is large, the proposed algorithm is of significantly lower complexity compared to the TSBL algorithm.

In the context of cluster-sparse signals with intra-cluster correlation, we propose the Nested SBL (NSBL) algorithm. Unlike the BSBL and the ARSBL algorithm [134], we obtain closed form EM based updates for estimating the correlation coefficient. Further, when intra-block correlation is absent, we simplify the NSBL approach and propose the PCSBL algorithm. While the PCSBL algorithm offers the same performance as the BSBL algorithm, we show that PCSBL is based on a divide and conquer approach and hence amenable to parallel implementation.

Another novel algorithm proposed in this thesis learns the partially unknown dictionary matrix jointly with recovery of the sparse vector in an SBL framework. Consider the system model given by

$$\mathbf{y} = \mathbf{A}\boldsymbol{\psi}\mathbf{x} + \mathbf{n}. \tag{5.12}$$

The expression given above is identical to (1.1), with $\mathbf{\Phi} = \mathbf{A}\boldsymbol{\psi}$. Let $\boldsymbol{\psi}$ be a known matrix such as the Fourier or the wavelet basis. We consider a scenario where $\mathbf{A}$ is diagonal with some missing entries. We elegantly generalize the conventional SBL based EM framework in order to incorporate the estimation of diagonal entries of $\mathbf{A}$ in the M-step of the SBL algorithm. We show that such an algorithm continues to enjoy the likelihood monotonicity property of the conventional SBL framework. Further, using Monte Carlo

simulations, we also show that incorporating the missing entries of $\mathbf{A}$ into the update equations leads to superior MSE and support recovery efficiency as compared to the SBL framework where only known entries of $\mathbf{A}$ are included. Until now, we described the problem setup and provided an overview of the algorithms proposed in this thesis. In the next section, we describe the SBL algorithm which is the basis of the novel SSR algorithms proposed in this thesis.

5.2 Sparse Bayesian Learning

Here, we describe the SBL algorithm for the recovery the sparse vector $\mathbf{x}$ from (1.1). SBL uses a parametrized prior to obtain sparse solutions, given by

$$p(\mathbf{x}; \boldsymbol{\gamma}) = \prod_{i=1}^{L} (\pi \gamma(i))^{-1} \exp\left(-\frac{|x(i)|^2}{\gamma(i)}\right). \tag{5.13}$$

Typically, the hyperparameters $\boldsymbol{\gamma}$ can be estimated using the type-II ML procedure [34], i.e., by maximizing the marginalized pdf $p(\mathbf{y}; \boldsymbol{\gamma})$ as

$$\hat{\boldsymbol{\gamma}}_{ML} = \arg\max_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} p(\mathbf{y}; \boldsymbol{\gamma}). \tag{5.14}$$

Since the above problem cannot be solved in closed form, iterative estimators such as the EM based SBL algorithm [49] have to be employed. The sparse channel $\mathbf{h}$ is considered as the hidden variable and the ML estimate of $\boldsymbol{\gamma}$ is obtained in the M-step.

The steps of the algorithm can be given as

$$\text{E-step} : Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})] \tag{5.15}$$

$$\text{M-step} : \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \tag{5.16}$$

The E-step above requires the posterior density of the sparse vector with the hyperparameter $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(r)}$, which can be expressed as

$$p\left(\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}\right) = \mathcal{CN}(\boldsymbol{\mu},\boldsymbol{\Sigma}), \tag{5.17}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^{(r)} - \boldsymbol{\Gamma}^{(r)}\boldsymbol{\Phi}^H\left(\sigma^2\mathbf{I}_N + \boldsymbol{\Phi}\boldsymbol{\Gamma}^{(r)}\boldsymbol{\Phi}^H\right)^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}^{(r)}$, and $\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^H\mathbf{y}$. The M-step in (5.16) can be simplified, to obtain

$$\gamma^{(r+1)}(i) = \underset{\gamma(i)\in\mathbb{R}_+}{\arg\max}\, \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}\left[\log p(\mathbf{x};\boldsymbol{\gamma})\right] \tag{5.18}$$

$$= \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}\left[|x(i)|^2\right] = \Sigma(i,i) + |\mu(i)|^2. \tag{5.19}$$

In (5.18), the term $\mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}|\mathbf{x};\boldsymbol{\gamma})]$ has been discarded, as it is not a function of $\gamma(i)$. Note that the EM framework enjoys the monotonicity property, i.e., the likelihood is guaranteed to increase at each iteration [126, 127]. We have found, empirically, that the straightforward initialization such as $\boldsymbol{\Gamma}^{(0)} = \mathbf{I}_L$ leads to accurate solutions. Note that this algorithm recovers the sparse vector with no additional constraints on $\mathbf{x}$ such as its sparsity level etc.

In the following section, we discuss the generalizations of the SBL framework for estimation of group-sparse and cluster-sparse signals.

5.3 SBL: Group and Cluster Sparsity

In this section, we discuss various extensions of SBL that exploit the group and cluster sparsity in sparse vectors. First, we consider recovering M vectors in (5.6)/(5.7), where the columns of $\mathbf{X}$ are group-sparse, i.e., locations of non-zero entries of the vectors coincide, leading to several zero rows in $\mathbf{X}$. Subsequently, we consider the problem of recovery of cluster-sparse vectors from (5.9), where clusters are equi-sized and consist of either all zero or all non-zero entries. We consider cluster-sparse signals where the entries of non-zero clusters are correlated within the cluster and independent between clusters.

In this chapter, we design algorithms which exploit the intra-cluster correlation when the correlation is known. Further, when the correlation is unknown, we provide EM based update equations for learning the correlation.

5.3.1 Existing Algorithms for Group Sparsity: MSBL and TSBL

In this subsection, we first describe the MSBL algorithm [104] for recovery of uncorrelated group-sparse vectors in $\mathbf{X}$ from the observation matrix $\mathbf{Y}$ in (2.2). Next, we describe the TSBL algorithm [79] for recovering the group-sparse correlated vectors.

MSBL

MSBL incorporates a parameterized prior to obtain sparse solutions in regression, which can be written as

$$p(\mathbf{X}; \mathbf{\Gamma}) = \prod_{k=1}^{K} p(\mathbf{x}_k; \mathbf{\Gamma}), \tag{5.20}$$

where $\mathbf{x}_k$ represents the kth column of $\mathbf{X}$, with prior density given by $\mathbf{x}_k \sim \mathcal{CN}(0, \boldsymbol{\Gamma})$. The hyperparameters in $\boldsymbol{\Gamma}$ are estimated using the type-II ML procedure [34], i.e., by maximizing the marginalized pdf $p(\mathbf{y}_k; \boldsymbol{\gamma})$ as follows:

$$\gamma_{ML}(i) = \arg\max_{\gamma(i) \in \mathbb{R}_+} p(\mathbf{y}_k; \boldsymbol{\gamma}), \quad 1 \leq i \leq L. \tag{5.21}$$

As in the case of SBL, the above problem cannot be solved in closed form. Unlike the SBL algorithm, the above problem involves solving for $\mathbf{x}_1, \ldots, \mathbf{x}_K$, for which iterative estimators such as the EM based MSBL algorithm [104] are employed. In order to use the MSBL algorithm, $\mathbf{X}$ is treated as the hidden variable and the posterior distribution of $\mathbf{X}$ is obtained in the E-step, and the updates of $\boldsymbol{\gamma}$ is obtained in the M-step. The steps of the algorithm are given as

$$\text{E-step}: Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \mathbb{E}_{\mathbf{X}|\mathbf{Y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}, \mathbf{X}; \boldsymbol{\gamma})] \tag{5.22}$$

$$\text{M-step}: \gamma^{(r+1)}(i) = \arg\max_{\gamma(i) \in \mathbb{R}_+} Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}), \tag{5.23}$$

for $1 \leq i \leq L$, and these steps are iterated until convergence. The E-step requires the posterior distribution $p(\mathbf{X}|\mathbf{Y}; \boldsymbol{\gamma}^{(r)})$, which can be obtained from the likelihood given by

$$p(\mathbf{y}_k|\mathbf{x}_k) = \frac{1}{(\pi\sigma^2)^K} \exp\left(-\frac{\|\mathbf{y}_k - \boldsymbol{\Phi}\mathbf{x}_k\|_2^2}{\sigma^2}\right). \tag{5.24}$$

Combining the likelihood and the prior distribution, the posterior distribution of $\mathbf{x}_k$ is given by $p\left(\mathbf{x}_k|\mathbf{y}_k; \boldsymbol{\gamma}^{(r)}\right) \sim \mathcal{CN}(\boldsymbol{\mu}_{\text{msbl},k}, \boldsymbol{\Sigma}_{\text{msbl}})$, with mean and covariance given by

$$\boldsymbol{\mu}_{\text{msbl},k} = \sigma^{-2}\boldsymbol{\Sigma}_{\text{msbl}}\boldsymbol{\Phi}^H\mathbf{y}_k \quad \boldsymbol{\Sigma}_{\text{msbl}} = \left(\frac{\boldsymbol{\Phi}^H\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Gamma}^{(r)^{-1}}\right)^{-1}. \tag{5.25}$$

Here, $\mathbf{\Gamma}^{(r)}$ is the hyperparameter update in the rth iteration.

The M-step given by (5.23), can be simplified to obtain the update equation for γ as

$$\gamma^{(r+1)}(i) = \frac{1}{K} \sum_{k=1}^{K} \left(\sum_{k=1}^{K} \mu_{\text{msbl,k}}(i)^2 + \Sigma_{\text{msbl}}(i,i) \right). \tag{5.26}$$

Note that, in the above equation, the group-sparse nature of the channel results in the update of γ which is averaged over the K instantiations of the sparse vector.

The MSBL algorithm consists of executing the E and the M steps iteratively, until the algorithm reaches convergence, i.e., the difference $\left\| \gamma^{(r)} - \gamma^{(r-1)} \right\|_2^2 \le \epsilon$, where ϵ is a small value, for e.g. $\epsilon = 10^{-6}$. The E-step involves computing the posterior mean and variance of the sparse vector as given in (5.25), incurring a computational complexity given by $\mathcal{O}(N^2 L)$ [104], while M-step computes the hyperparameter update as given in (5.26). In practice, it is found that an initial estimate for $\mathbf{\Gamma}$ given by

$$\mathbf{\Gamma}^{(0)} = \mathbf{I}_{L \times L}, \tag{5.27}$$

is sufficient for the MSBL algorithm.

In Fig. 5.1 and Fig. 5.2, we compare the the MSE performance and the support recovery of the MSBL algorithm with the CS based Simultaneous OMP (SOMP) technique [83], the conventional SBL and the OMP algorithms [12], which are unaware of the group-sparse nature of the vectors. First, we observe that SBL based techniques have a better MSE performance compared to CS based greedy techniques such as the OMP and the SOMP algorithms. MSBL also outperforms the conventional SBL algorithm, i.e., utilizing the group-sparse nature of the channel leads to superior support recovery and MSE performance.
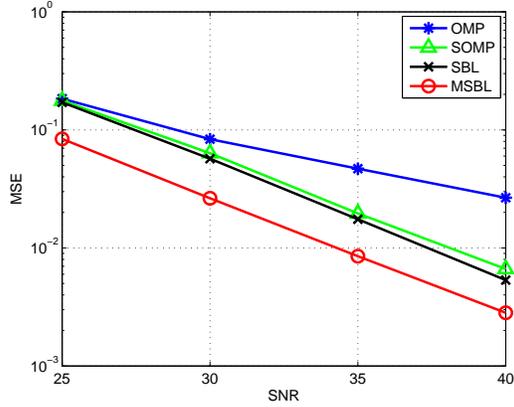
Figure 5.1: MSE performance of the algorithms MSBL and SBL compared to SOMP and OMP algorithms for $N = 30$, $L = 64$ and $S = 3$.
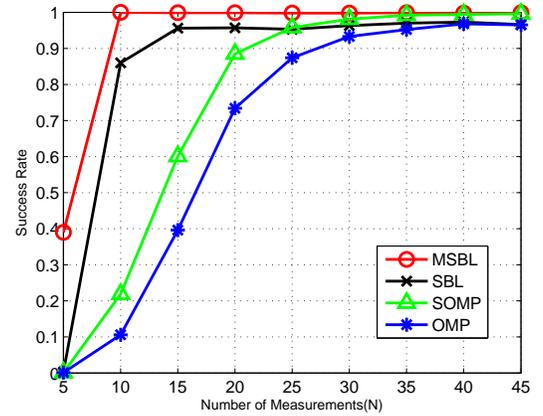
Figure 5.2: Success rate of the the algorithms MSBL and SBL compared to SOMP and OMP algorithms for SNR $= 20$, $L = 64$ and $S = 3$.

TSBL

Here, we describe the TSBL algorithm proposed in [79]. As mentioned earlier, the TSBL algorithm recovers the group-sparse vectors by transforming the MMV model in (2.2) to an SMV model.

As in the case of MSBL, each column of $\mathbf{X}$ is governed by a prior density given by $\mathbf{x}_k \sim \mathcal{CN}(0, \boldsymbol{\Gamma})$. However, since the group-sparse columns are correlated, the prior density associated with every row of $\mathbf{X}$ (the ith row of $\mathbf{X}$ is denoted as $\mathbf{x}(i,:)$), is as follows:

$$p(\mathbf{x}(i,:); \boldsymbol{\gamma}(i), \mathbf{B}_i) = \mathcal{CN}(0, \boldsymbol{\gamma}(i)\mathbf{B}_i), \quad 1 \leq i \leq N \tag{5.28}$$

where $\boldsymbol{\gamma}(i)$ is the hyperparameter associated with each row of $\mathbf{X}$, i.e., $\boldsymbol{\gamma}(i) = 0$ corresponds to a zero row of the matrix, and $\mathbf{B}_i$ is a non-diagonal positive definite matrix that captures the correlation structure of $\mathbf{x}(i,:)$.

Let $\mathbf{y} = \text{vec}(\mathbf{Y}^H) \in \mathbb{C}^{NM \times 1}$, $\mathbf{D} = \boldsymbol{\Phi} \otimes \mathbf{I}_L$ and $\mathbf{x} = \text{vec}(\mathbf{X}^H)$ and $\mathbf{v} = \text{vec}(\mathbf{V}^H)$. Then the

system model given in (2.2) can be rewritten as

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{v}. \tag{5.29}$$

Hence, in the process of converting a MMV model to the SMV model, a row-sparse matrix $\mathbf{X}$ is transformed into a cluster-sparse vector $\mathbf{x}$. The row correlation of $\mathbf{X}$ governs the correlation of the ith cluster of $\mathbf{x}$, i.e., the ith cluster is correlated according to $\mathbf{B}_i$.

Using the likelihood and the prior distribution given in (5.28), we obtain the posterior distribution using Bayes' rule, as follows:

$$p(\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}) = \mathcal{N}(\boldsymbol{\mu}_{\text{tsbl}}, \boldsymbol{\Sigma}_{\text{tsbl}}), \tag{5.30}$$

where the posterior mean and covariance are given by

$$\boldsymbol{\mu}_{\text{tsbl}} = \sigma^{-2}\boldsymbol{\Sigma}_{\text{tsbl}}\mathbf{D}^H\mathbf{y},$$

$$\boldsymbol{\Sigma}_{\text{tsbl}} = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0\mathbf{D}^H(\sigma^2\mathbf{I}_{NK} + \mathbf{D}\boldsymbol{\Sigma}_0\mathbf{D}^H)^{-1}\mathbf{D}\boldsymbol{\Sigma}_0, \tag{5.31}$$

where $\boldsymbol{\Sigma}_0$ is a block diagonal matrix with the ith block diagonal entry given by $\gamma(i)\mathbf{B}_i$. Here, $\boldsymbol{\Sigma}_0$ represents the overall covariance matrix of the cluster-sparse vector $\mathbf{x}$. In order to avoid the over-fitting, a single positive definite matrix $\mathbf{B} = \mathbf{B}_1 = \ldots = \mathbf{B}_N$ is used to model all the covariance matrices.

In the M-step, the update equations for the unknown parameters can be obtained as

follows:

$$\gamma(i) = \text{Tr}\left[\mathbf{B}^{-1}(\mathbf{\Sigma}^i_{\text{tsbl}} + \boldsymbol{\mu}^i_{\text{tsbl}}\boldsymbol{\mu}^i_{\text{tsbl}}{}^H)\right] \tag{5.32}$$

$$\mathbf{B} = \frac{1}{M}\sum_{m=1}^{M}\frac{\mathbf{\Sigma}^i_{\text{tsbl}} + \boldsymbol{\mu}^i_{\text{tsbl}}\boldsymbol{\mu}^i_{\text{tsbl}}{}^H}{\gamma(i)} \tag{5.33}$$

$$\sigma^2 = \frac{\|\mathbf{y} - \mathbf{D}\boldsymbol{\mu}_{\text{tsbl}}\|^2 + \sigma^2[ML - \text{Tr}(\mathbf{\Sigma}_{\text{tsbl}}\mathbf{\Sigma}_0)]}{NM}, \tag{5.34}$$

where $\mathbf{\Sigma}^i_{\text{tsbl}} = \mathbf{\Sigma}_{\text{tsbl}}((i-1)L+1 : iL, (i-1)L+1 : iL)$ and $\boldsymbol{\mu}^i_{tsbl} = \boldsymbol{\mu}_{\text{tsbl}}((i-1)L+1 : iL)$.

The posterior mean, covariance and the update equations together constitute the TSBL algorithm.

A special case of TSBL algorithm is considered in [125], where the correlation among the entries of the ith row ($1 \le k \le K$) of $\mathbf{X}$ is modeled using a first order AR model as follows:

$$x_{k+1}(i) = \rho x_k(i) + \sqrt{(1-\rho^2)}z_{k+1}(i), \tag{5.35}$$

where ρ is the AR coefficient, $z_k(i) \sim \mathcal{CN}(0, \gamma(i))$ and $x_k(i) \sim \mathcal{CN}(0, \gamma(i))$.

TSBL and the ARSBL algorithms are computationally complex since the measurement matrix is tranformed from a $N \times L$ matrix to a $NK \times KL$ matrix. As a low-complexity alternative to TSBL/ARSBL, we propose the KSBL algorithm [90] in the following subsection. In contrast to the ARSBL algorithm, we show that the update equations of the KSBL algorithm involve measurement matrices of dimension $N \times L$ and not $NK \times KL$. Moreover, the KSBL algorithm is capable of recovering correlated simultaneously sparse vectors for the general system model given in (5.7).

5.3.2 Proposed Technique: Kalman SBL

In this section, we describe the KSBL algorithm [90] which is a recursive, low-complexity recovery technique based on the Kalman Filter and Smoother (KFS). The algorithm works by fitting a first order AR model for the temporal evolution of the sparse vectors.

As explained in the previous section, $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are considered to be simultaneously sparse and correlated. We model the correlation using a first order AR process given by (5.35). The correlation in columns of $\mathbf{X}$ is given by

$$\mathbf{x}_k = \rho \mathbf{x}_{k-1} + \mathbf{z}_k, \quad k = 1, \ldots, K, \tag{5.36}$$

where by definition, $\mathbf{x}_0 = \mathbf{0}_L$. The state space model comprises of the measurement equation given by in (5.7), and the state equation given by the first order AR channel model in (5.36). The joint pdf of the observations and the K sparse vectors is given by

$$p(\mathbf{Y}, \mathbf{x}_1, \ldots, \mathbf{x}_K; \boldsymbol{\gamma}) = \prod_{k=1}^{K} p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_{k-1}; \boldsymbol{\gamma}). \tag{5.37}$$

We propose the KSBL algorithm using the EM updates, as follows:

$$\text{E-step} : Q\left(\boldsymbol{\gamma} | \boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{x}_1, \ldots, \mathbf{x}_K | \mathbf{Y}; \boldsymbol{\gamma}^{(r)}} [\log p(\mathbf{Y}, \mathbf{x}_1, \ldots, \mathbf{x}_K; \boldsymbol{\gamma})] \tag{5.38}$$

$$\text{M-step} : \boldsymbol{\gamma}^{(r+1)} = \arg\max_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} Q\left(\boldsymbol{\gamma} | \boldsymbol{\gamma}^{(r)}\right). \tag{5.39}$$

To compute the E-step given above, we require the posterior distribution of the unknown a-sparse channel, which is obtained using the KFS recursive update equations

given as follows [124, 128]:

> **for** $k = 1, \ldots, K$ **do**
>
> Prediction: $\hat{\mathbf{x}}_{k|k-1} = \rho \hat{\mathbf{x}}_{k-1|k-1}$ (5.40)
>
> $\mathbf{P}_{k|k-1} = \rho^2 \mathbf{P}_{k-1|k-1} + (1 - \rho^2)\mathbf{\Gamma}$ (5.41)
>
> Filtering:
>
> $\mathbf{G}_k = \mathbf{P}_{k|k-1}\mathbf{\Phi}_k^T \left(\sigma^2 \mathbf{I}_N + \mathbf{\Phi}_k \mathbf{P}_{k|k-1}\mathbf{\Phi}_k^T\right)^{-1}$ (5.42)
>
> $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}_k - \mathbf{\Phi}_k \hat{\mathbf{x}}_{k|k-1})$ (5.43)
>
> $\mathbf{P}_{k|k} = (\mathbf{I}_L - \mathbf{G}_k \mathbf{\Phi}_k)\mathbf{P}_{k|k-1}$ (5.44)
>
> **end**
>
> **for** $j = K, K-1, \ldots, 2$ **do**
>
> Smoothing: $\hat{\mathbf{x}}_{j-1|K} = \hat{\mathbf{x}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{x}}_{j|K} - \hat{\mathbf{x}}_{j|j-1})$ (5.45)
>
> $\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^T$ (5.46)
>
> **end**

where the mean and the covariance matrix of sparse vector in the kth instant is represented by $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$, respectively, for $1 \leq k \leq K$, $\mathbf{J}_{j-1} \triangleq \rho \mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{G}_k$ is the Kalman gain. In the above, the symbols $\hat{\mathbf{x}}_{k|k-1}$, $\mathbf{P}_{k|k-1}$, etc. have their usual meanings as in the KF literature [124]. For example, $\hat{\mathbf{x}}_{k|k-1}$ is the estimate of the kth sparse vector given the observations $\mathbf{Y}_{k-1} = [\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}]$ and $\mathbf{P}_{k|k-1}$ is the covariance of the kth channel estimate given $\mathbf{Y}_{k-1}$. The above KFS equations are initialized by setting $\hat{\mathbf{x}}_{0|0} = \mathbf{0}_L$ and $\mathbf{P}_{0|0} = \mathbf{\Gamma}$.

Typically, in a Kalman Filtering (KF) approach, the goal is to recursively estimate the

channel state and its covariance matrix using forward and backward recursions, given the observations $\mathbf{Y}$. In the forward recursion, the KF operates on the observations to obtain the estimates of the sparse vector as a weighted average of the previous estimate and the current received symbol using prediction and the filtering equations in (5.40)-(5.44). These weights are given by the the Kalman gain matrix, and are updated for each sparse vector. In the backward recursion, the Kalman *smoother* ensures that the observations of the K instantiations are included in the estimation of the sparse vector corresponding to the kth instant for $1 \leq k < K$ using (5.45)-(5.46). Hence, it improves the accuracy of the estimates of the sparse vectors in every recursion.

In order to obtain an ML estimate of $\boldsymbol{\gamma}$, KSBL incorporates an M-step, which, in turn, utilizes the mean and covariance of the posterior distribution from the E-step. From (5.37), the M-step results in the following optimization problem:

$$\boldsymbol{\gamma}^{(r+1)} = \arg\min_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} \mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_K | \mathbf{Y}; \boldsymbol{\gamma}^{(r)}} \left[K \log |\boldsymbol{\Gamma}| + \sum_{j=2}^{K} \frac{(\mathbf{x}_j - \rho \mathbf{x}_{j-1})^H \boldsymbol{\Gamma}^{-1} (\mathbf{x}_j - \rho \mathbf{x}_{j-1})}{(1 - \rho^2)} + \mathbf{x}_1^H \boldsymbol{\Gamma}^{-1} \mathbf{x}_1 \right].$$

$$(5.47)$$

As mentioned earlier, we see that the M-step requires the computation of $\hat{\mathbf{x}}_{j|K} \triangleq \mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_K | \mathbf{Y}; \boldsymbol{\gamma}^{(r)}} [\mathbf{x}_j]$, and covariance $\mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_K | \mathbf{Y}; \boldsymbol{\gamma}^{(r)}} [\mathbf{x}_j \mathbf{x}_j^T] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{x}}_{j|K} \hat{\mathbf{x}}_{j|K}^H$ for $j = 1,\ldots,K$, which is obtained from (5.40)-(5.46). The M-step also requires the computation of $\mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_K | \mathbf{Y}; \boldsymbol{\gamma}^{(r)}} [\mathbf{x}_j \mathbf{x}_{j-1}^T] \triangleq \mathbf{P}_{j,j-1|K} + \hat{\mathbf{x}}_{j|K} \hat{\mathbf{x}}_{j-1|K}^T$ for $j = K, K-1,\ldots,2$, which we obtain from [124] as follows:

$$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1} \mathbf{J}_{j-2}^T + \mathbf{J}_{j-1}^T (\mathbf{P}_{j,j-1|K} - \rho \mathbf{P}_{j-1|j-1}) \mathbf{J}_{j-2}. \tag{5.48}$$

The above recursion is initialized using $\mathbf{P}_{K,K-1|K} = \rho(\mathbf{I}_L - \mathbf{G}_K \boldsymbol{\Phi}_K) \mathbf{P}_{K-1|K-1}$. Using the

above expressions, (5.47) simplifies as

$$\boldsymbol{\gamma}^{(r+1)} = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}} \left\{ K\log|\boldsymbol{\Gamma}| + \text{Trace}(\boldsymbol{\Gamma}^{-1}\mathbf{M}_{1|K}) + \frac{1}{(1-\rho^2)}\sum_{j=2}^{K}\text{Trace}(\boldsymbol{\Gamma}^{-1}\mathbf{M}_{j|K}) \right\}, \quad (5.49)$$

where $\mathbf{M}_{j|K} \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{x}}_{j|K}\hat{\mathbf{x}}_{j|K}^T + \rho^2(\mathbf{P}_{j-1|K} + \hat{\mathbf{x}}_{j-1|K}\hat{\mathbf{x}}_{j-1|K}^T) - 2\rho\text{Re}(\mathbf{P}_{j,j-1|K} + \hat{\mathbf{x}}_{j|K}\hat{\mathbf{x}}_{j-1|K}^T)$

and $\mathbf{M}_{1|K} \triangleq \mathbf{P}_{1|K} + \hat{\mathbf{x}}_{1|K}\hat{\mathbf{x}}_{1|K}^T$. Differentiating (5.49) w.r.t. $\gamma(i)$ and setting the resulting

equation to zero gives the update for the ith hyperparameter as follows:

$$\gamma^{(r+1)}(i) = \frac{1}{K}\left(\sum_{j=2}^{K}\frac{M_{j|K}(i,i)}{(1-\rho^2)} + M_{1|K}(i,i)\right), \quad (5.50)$$

for $i = 1,\ldots,L$. Thus the KSBL algorithm learns $\boldsymbol{\gamma}$ in the M-step and provides low-complexity and recursive estimates of the correlated simultaneously sparse vectors in the E-step.

 Using a flop-count analysis [129], for K ($K > 1$) group-sparse vectors, the computations of the KSBL algorithm is dominated by the computation of the $\mathbf{J}_{K-1}$ term in the smoothing step, which has a complexity of $\mathcal{O}(KL^3)$ per iteration. In ARSBL/TSBL, the computation of the covariance matrix $\boldsymbol{\Sigma}$ incurs a complexity of $\mathcal{O}(K^3N^2L)$ per iteration. Hence, we see that if the number of OFDM symbols to be tracked are such that $KN > L$, the complexity of the ARSBL/TSBL algorithm is larger than the KSBL algorithm. In other words, the KSBL algorithm is a good choice among SBL based techniques when the number of group-sparse vectors is large.

 In Fig. 5.4 and Fig. 5.3, we demonstrate the support recovery and the MSE performance of KSBL algorithm as compared to conventional SBL and the OMP algorithms, which are unaware of the correlated group-sparse nature of sparse vectors. Note that KSBL uses distinct measurement matrices for every k, and hence, we cannot implement
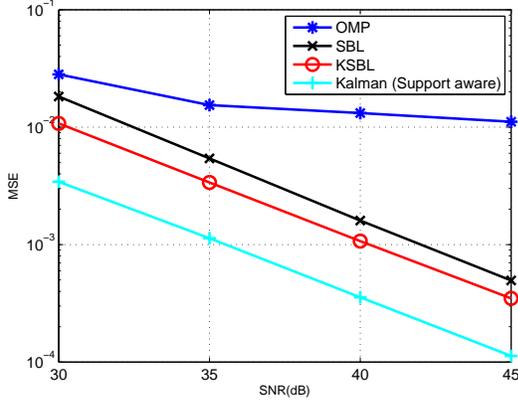
Figure 5.3: MSE performance of the KSBL algorithm compared to SBL and OMP algorithms and support-aware Kalman filter for $N = 30$, $L = 64$ and $S = 3$.
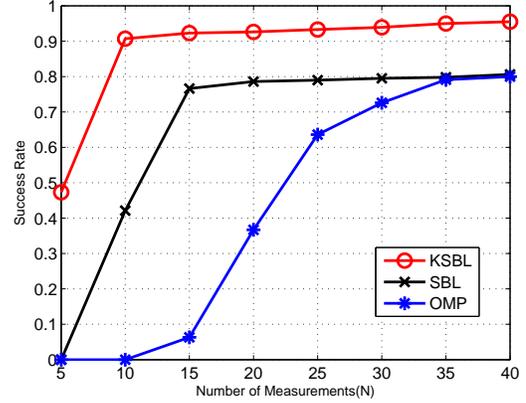
Figure 5.4: Success rate of the KSBL algorithm compared to SBL and OMP algorithms for SNR $= 20$, $L = 64$ and $S = 3$.

TSBL and ARSBL algorithms in such a scenario. We see that KSBL which benefits from group sparsity and temporal correlation among sparse vectors has a superior performance compared to SBL and OMP algorithms.

Furthermore, the KSBL algorithm can be generalized to learn ρ as well. The $Q(\cdot)$ function in the E-step w.r.t. ρ is given by

$$Q\left(\rho|\rho^{(r)}\right) = c'' - (K-1)\log(1-\rho^2) - \mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_K|\mathbf{Y}_K;\rho^{(r)}} \left[\sum_{j=2}^{K} \frac{(\mathbf{x}_j - \rho\mathbf{x}_{j-1})^H \Gamma^{-1}(\mathbf{x}_j - \rho\mathbf{x}_{j-1})}{(1-\rho^2)} \right]$$

$$= c'' - (K-1)\log(1-\rho^2) - \frac{1}{(1-\rho^2)}\text{Tr}\left\{\mathbf{T}_1 - \rho(\mathbf{T}_2 + \mathbf{T}_3) + \rho^2\mathbf{T}_4\right\}, \qquad (5.51)$$

where,

$$\mathbf{T}_1 = \Gamma^{-1} \sum_{j=2}^{K} [\mathbf{P}_{j|k} + \hat{\mathbf{x}}_{j|k} \hat{\mathbf{x}}_{j|k}^H]$$

$$\mathbf{T}_2 = \Gamma^{-1} \sum_{j=2}^{K} [\mathbf{P}_{j,j-1|k} + \hat{\mathbf{x}}_{j|k} \hat{\mathbf{x}}_{j-1|k}^H]$$

$$\mathbf{T}_3 = \Gamma^{-1} \sum_{j=2}^{K} [\mathbf{P}_{j,j-1|k} + \hat{\mathbf{x}}_{j-1|k} \hat{\mathbf{x}}_{j|k}^H]$$

$$\mathbf{T}_4 = \Gamma^{-1} \sum_{j=2}^{K} [\mathbf{P}_{j-1|k} + \hat{\mathbf{x}}_{j-1|k} \hat{\mathbf{x}}_{j-1|k}^H]. \tag{5.52}$$

Differentiating the above expression w.r.t. ρ, we get

$$\frac{\partial(Q\left(\rho|\rho^{(r)}\right))}{\partial \rho} = \frac{2(K-1)\rho}{(1-\rho^2)} - \frac{2\rho}{(1-\rho^2)^2} \mathrm{Tr}\left\{\mathbf{T}_1 - \rho(\mathbf{T}_2 + \mathbf{T}_3) + \rho^2 \mathbf{T}_4\right\}$$

$$- \frac{1}{(1-\rho^2)} \mathrm{Tr}\left\{-\mathbf{T}_2 - \mathbf{T}_3 + 2\rho \mathbf{T}_4\right\}. \tag{5.53}$$

Hence, ρ can be obtained as a solution to the cubic equation

$$2(K-1)\rho^3 - \mathrm{Tr}\left\{\mathbf{T}_2 + \mathbf{T}_3\right\}\rho^2 - [2(K-1) - 2\mathrm{Tr}\left\{\mathbf{T}_1 + \mathbf{T}_4\right\}]\rho - \mathrm{Tr}\left\{\mathbf{T}_2 + \mathbf{T}_3\right\} = 0, \tag{5.54}$$

subject to the constraint $0 < \rho \leq 1$.

In the following subsection, we discuss the cluster-sparse recovery problem based on the system model given in (5.9). Further, we discuss existing SBL based techniques for the recovery of cluster-sparse vectors from (5.9) and propose novel techniques such as the NSBL approach that overcomes the in the existing SBL based techniques.

5.3.3   Existing Algorithms for Cluster Sparsity: BSBL

In this section, we present the BSBL algorithm [134] recovering cluster-sparse vectors from an underdetermined set of linear equations as given in (5.9). Unlike the existing CS based techniques, BSBL is capable of handling intra-block correlation in block-sparse vectors.

BSBL

In this subsection, we present the BSBL algorithm proposed in [134] for recovery of cluster-sparse vectors in the presence of unknown intra-block correlation.

We consider the system model given in (5.9), where the unknown $L = BM$ length cluster-sparse vector $\mathbf{x}$ consists of B blocks denoted by $\mathbf{b}_1, \ldots, \mathbf{b}_B$, as follows:

$$\mathbf{x}_c = \big[\underbrace{x_{11}, x_{12}, \ldots, x_{1M}}_{\mathbf{b}_1^H \in \mathbb{C}^{1 \times M}}; \ldots; \underbrace{x_{B1}, x_{B2}, \ldots, x_{BM}}_{\mathbf{b}_B^H \in \mathbb{C}^{1 \times M}}\big]. \tag{5.55}$$

The M entries of each cluster $\mathbf{b}_i$ are constrained to be either all-zero or all-nonzero. In the cluster-sparse framework, the underlying structure is exploited by modeling $\mathbf{b}_i \sim \mathcal{CN}(0, \gamma_i \mathbf{B}_i)$, where γ_i is an unknown hyperparameter such that when $\gamma_i = 0$, the ith cluster of $\mathbf{x}_c$ is zero [134]. Here, $\mathbf{B}_i \in \mathbb{C}^{M \times M}$ is a positive-definite covariance matrix that captures the intra-block correlation of the ith block, which is also unknown. Moreover, different clusters are mutually uncorrelated, and hence, the cluster-sparse vector $\mathbf{x}_c \sim \mathcal{CN}(0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0$ is a block-diagonal matrix with principal blocks given by $\gamma_i \mathbf{B}_i$, $1 \leq i \leq B$.

The posterior mean and covariance are given by

$$\boldsymbol{\mu}_{\text{bsbl}} = \sigma^{-2}\boldsymbol{\Sigma}_{\text{bsbl}}\boldsymbol{\Phi}_c^H\mathbf{y}_c,$$

$$\boldsymbol{\Sigma}_{\text{bsbl}} = \boldsymbol{\Sigma}_0 - \quad \boldsymbol{\Sigma}_0\boldsymbol{\Phi}_c^H(\sigma^2\mathbf{I}_N + \boldsymbol{\Phi}_c\boldsymbol{\Sigma}_0\boldsymbol{\Phi}_c^H)^{-1}\boldsymbol{\Phi}_c\boldsymbol{\Sigma}_0, \tag{5.56}$$

Note that the subtle different between the BSBL framework and the TSBL framework arises in the construction of the measurement matrix - the BSBL framework employs a general measurement matrix $\boldsymbol{\Phi}_c$, while the TSBL assumes a block-diagonal matrix. As a consequence, the learning rules for the unknown parameters $\boldsymbol{\gamma}_c, \mathbf{B}$ ($\mathbf{B} = \mathbf{B}_i,\ 1 \leq i \leq B$), σ^2 remain the same as in (5.34).

Since learning the entire matrix $\mathbf{B}$ could lead to over fitting, $\mathbf{B}$ is further constrained such that it depends on a single parameter [134]. A first-order AR process with parameter ρ is suitable in such scenarios. The corresponding correlation matrix of each block is given by

$$\mathbf{B} = \begin{bmatrix} 1 & \rho & \dots & \rho^{M-1} \\ \rho & \rho^2 & \dots & \rho^{M-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{M-1} & \rho^{M-1} & \dots & 1 \end{bmatrix} \tag{5.57}$$

When ρ is unknown, an EM based closed form update for ρ is not available in the BSBL framework. The approach adopted in [134] is to heuristically look for a correlation matrix which is close to $\mathbf{B}$ especially along the main diagonal and the main sub-diagonal. Such a non-EM based, heuristic update procedure offers satisfactory performance and reasonably fast convergence in practice. However, it disrupts the EM based monotonicity properties of the BSBL algorithm.

In order to overcome the above drawback of a heuristic update method, we propose

the Nested SBL approach. In the following subsection, we show that, using the NSBL approach, we are able to provide a convergent, low-complexity solution to the problem of cluster-sparse recovery with EM based updates for all the unknowns.

5.3.4 Proposed Technique: NSBL

In this subsection, we describe the NSBL algorithm [135], which is based on a divide and conquer approach: the problem of estimating a high-dimensional cluster-sparse vector is reformulated as a set of smaller problems, each involving the estimation of low-dimensional correlated group-sparse vectors. We show that the NSBL framework enjoys the unique distinction of being a low-complexity recovery technique comprising of *EM based updates* for the unknown intra-block correlation, unlike the BSBL algorithm which uses a heuristic method to estimate the unknown intra-block correlation.

NSBL

The NSBL algorithm is an EM-based convergent technique which is capable of recovering cluster-sparse vectors in the presence of unknown intra-block correlation ($\mathbf{B} \neq \mathbf{I}_M$). As a special case, the NSBL approach can be simplified to a Parallel Cluster SBL (PCSBL) approach for the case when the entries within a block are not correlated ($\mathbf{B} = \mathbf{I}_M$).

As depicted in Fig. 5.6, restructuring the cluster-sparse vector $\mathbf{x}_c$, the problem of recovering $\mathbf{x}_c$ from $\mathbf{y}_c$ is equivalent to finding the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$, where $\mathbf{x}_i = [x_{1i}, \ldots, x_{Bi}]$. Since $\mathbf{b}_i \sim \mathcal{CN}(0, \gamma_i \mathbf{B}_i)$ for $1 \leq i \leq B$, $\mathbf{x}_i \sim \mathcal{CN}(0, \mathbf{\Gamma}_c)$ where $\mathbf{\Gamma}_c = \mathrm{diag}(\gamma_c(1), \ldots, \gamma_c(B))$, i.e., $\mathbf{x}_1, \ldots, \mathbf{x}_M$ represent *group-sparse vectors*.
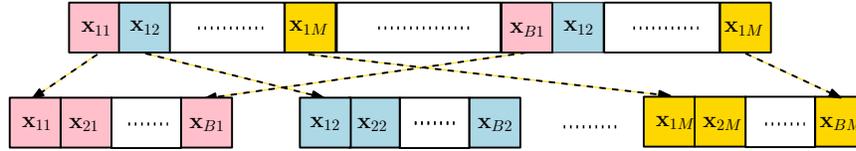
Figure 5.5: Restructuring the block-sparse recovery problem such that the B length vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are *group-sparse vectors*.

By rearranging the columns of $\boldsymbol{\Phi}_c$, the system model in (5.9) can be equivalently written as

$$\mathbf{y}_c = \sum_{m=1}^{M} \mathbf{t}_m, \quad \text{where} \quad \mathbf{t}_m \triangleq \boldsymbol{\Phi}_m \mathbf{x}_m + \mathbf{n}_m, \quad 1 \leq m \leq M. \tag{5.58}$$

In (5.58), $\boldsymbol{\Phi}_m \in \mathbb{C}^{N \times B}$ consists of the columns of $\boldsymbol{\Phi}_c$ such that the coefficients corresponding to its columns are given by $\mathbf{x}_m$. Although $\mathbf{n}_m$ cannot be explicitly obtained, we note that its covariance can be written as $\mathbf{n}_m \sim \mathcal{CN}(0, \beta_m \sigma^2 \mathbf{I}_N)$ where, $0 \leq \beta_m \leq 1$ and $\sum_{m=1}^{M} \beta_m = 1$. If $\mathbf{t}_m$ is known, recovering $\mathbf{x}_m$ from $\mathbf{t}_m$ is a group-sparse recovery problem [104] in a lower dimensional space (B), as compared to the dimension of the original problem (MB). Here, we focus on recovering the cluster-sparse vector by recovering its group-sparse components $\mathbf{x}_1, \ldots, \mathbf{x}_M$, using the restructured problem given by (5.58).

The conventional SBL framework in Sec. 5.2 treats $(\mathbf{y}_c, \mathbf{x}_c)$ in (5.9) as the complete data, and $\mathbf{x}_c$ as the hidden variable. However, for the reformulated system model in (5.58), it is necessary to augment the set of hidden variables $\mathbf{x}_c$ with $\mathbf{t} = [\mathbf{t}_1^T, \ldots, \mathbf{t}_M^T]^T$ since $\mathbf{t}$ is also hidden [143]. Accordingly, the complete information is given by $(\mathbf{y}_c, \mathbf{t}, \mathbf{x}_c)$, and $(\mathbf{t}, \mathbf{x}_c)$ constitute the hidden variables. Since closed-form expressions for the maximum likelihood estimates of the unknown parameter $\boldsymbol{\gamma}_c$ cannot be obtained, we adopt the

iterative EM algorithm for estimating γ_c, as follows:

$$\text{E-step}: Q\left(\gamma_c|\gamma_c^{(r)}\right) = \mathbb{E}_{\mathbf{t}, \mathbf{x}_c|\mathbf{y}_c; \gamma_c^{(r)}}[\log p(\mathbf{y}_c, \mathbf{t}, \mathbf{x}_c; \gamma_c)]$$

$$\text{M-step}: \gamma_c^{(r+1)} = \underset{\gamma_c \in \mathbb{R}_+^{B \times 1}}{\arg\max} Q\left(\gamma_c|\gamma_c^{(r)}\right). \tag{5.59}$$

The E-step in (5.59) requires the computation of $p(\mathbf{t}, \mathbf{x}_c|\mathbf{y}_c; \gamma_c^{(r)})$, which is given by

$$p(\mathbf{t}, \mathbf{x}_c|\mathbf{y}_c; \gamma_c^{(r)}) = p(\mathbf{x}_c|\mathbf{t}, \mathbf{y}_c; \gamma_c^{(r)})p(\mathbf{t}|\mathbf{y}_c; \gamma_c^{(r)}) = p(\mathbf{x}_c|\mathbf{t}; \gamma_c^{(r)})p(\mathbf{t}|\mathbf{y}_c; \gamma_c^{(r)}). \tag{5.60}$$

Hence, the E-step can be rewritten as

$$\text{E-step}: Q\left(\gamma_c|\gamma_c^{(r)}\right) = \underbrace{\mathbb{E}_{\mathbf{t}|\mathbf{y}_c; \gamma_c^{(r)}}}_{\mathbb{E}_{\mathbf{t}}} \underbrace{\mathbb{E}_{\mathbf{x}_c|\mathbf{t}; \gamma_c^{(r)}}}_{\mathbb{E}_{\mathbf{x}_c}}[\log p(\mathbf{y}_c, \mathbf{t}, \mathbf{x}_c; \gamma_c)]. \tag{5.61}$$

To compute $Q(\gamma_c|\gamma_c^{(r)})$, we first compute the posterior distribution $p(\mathbf{t}|\mathbf{y}_c; \gamma_c^{(r)})$ using the likelihood $p(\mathbf{t}_m|\mathbf{x}_m) = \mathcal{CN}(\boldsymbol{\Phi}_m \mathbf{x}_m, \beta_m \sigma^2 \mathbf{I}_N)$ for $1 \leq m \leq M$, and the prior $p(\mathbf{x}_c; \gamma_c) = \mathcal{CN}(0, \boldsymbol{\Gamma}_B)$. Given $\mathbf{H} = \mathbf{1}_M \otimes \mathbf{I}_N$, where $\mathbf{1}_M$ is a M length vector of ones, and $\mathbf{y}_c = \mathbf{Ht}$, we have $p(\mathbf{t}|\mathbf{y}_c; \gamma_c^{(r)}) = \mathcal{CN}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, where

$$\boldsymbol{\mu}_t = (\mathbf{R} + \boldsymbol{\Phi}_B \boldsymbol{\Gamma}_B \boldsymbol{\Phi}_B^H)\mathbf{H}^H(\mathbf{H}(\mathbf{R} + \boldsymbol{\Phi}_B \boldsymbol{\Gamma}_B \boldsymbol{\Phi}_B^H)\mathbf{H}^H)^{-1}\mathbf{y}_c$$

$$\boldsymbol{\Sigma}_t = (\mathbf{R} + \boldsymbol{\Phi}_B \boldsymbol{\Gamma}_B \boldsymbol{\Phi}_B^H) - (\mathbf{R} + \boldsymbol{\Phi}_B \boldsymbol{\Gamma}_B \boldsymbol{\Phi}_B^H)\mathbf{H}^H(\mathbf{H}(\mathbf{R} + \boldsymbol{\Phi}_B \boldsymbol{\Gamma}_B \boldsymbol{\Phi}_B^H)\mathbf{H}^H)^{-1}\mathbf{H}(\mathbf{R} + \boldsymbol{\Phi}_B \boldsymbol{\Gamma}_B \boldsymbol{\Phi}_B^H). \tag{5.62}$$

Here, $\boldsymbol{\Phi}_B \in \mathbb{C}^{NM \times BM}$ is a block diagonal matrix with $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_M$ along the diagonal, and $\boldsymbol{\Gamma}_B = \mathbf{B} \otimes \boldsymbol{\Gamma}_c$, where $\boldsymbol{\Gamma}_c = \text{diag}(\gamma_c)$. The block diagonal matrix $\mathbf{R}$ has m^{th} diagonal entry $\mathbf{R}_m = \beta_m \sigma^2 \mathbf{I}_N$. Note that the posterior mean $\boldsymbol{\mu}_t \in \mathbb{C}^{MN \times 1}$ consists of M vectors, $\boldsymbol{\mu}_{\mathbf{t}_1}, \ldots, \boldsymbol{\mu}_{\mathbf{t}_M}$ such that $\mathbf{H}\boldsymbol{\mu}_t = \mathbf{y}_c$, i.e., $\mathbf{y}_c = \sum_{m=1}^{M} \boldsymbol{\mu}_{\mathbf{t}_m}$.

The posterior distribution $p(\mathbf{x}_c|\mathbf{t}; \boldsymbol{\gamma}_c^{(r)})$ depends on the correlation between the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$. We model the intra-block correlation (i.e., correlation among $\mathbf{x}_1, \ldots, \mathbf{x}_M$) using a first-order AR model. The first order AR model is a widely accepted model, and is used in a variety of applications [144–146]. It also has the advantage that it avoids over-fitting [134] and allows for a Kalman filtering based learning framework. The evolution of the mth group-sparse vector is modeled as

$$\mathbf{x}_m = \rho\mathbf{x}_{m-1} + \mathbf{u}_m, \quad m = 1, \ldots, M, \tag{5.63}$$

where the driving noise $\mathbf{u}_m$ is distributed as $\mathbf{u}_m(i) \sim \mathcal{CN}(0, (1-\rho^2)\gamma(i))$, $\rho \in \mathbb{R}$ is the AR coefficient and $0 \leq \rho \leq 1$. Overall, this leads to a common correlation matrix given by $\mathbf{B}_1 = \ldots = \mathbf{B}_B = \mathbf{B} = \text{Toep}([1, \rho, \ldots, \rho^{M-1}])$, where $\text{Toep}(\mathbf{a})$ represents the symmetric Toeplitz matrix defined by its first row $\mathbf{a}$ [134]. The state space model for $\mathbf{t}_m$ and $\mathbf{x}_m$ is given as

$$\mathbf{t}_m = \boldsymbol{\Phi}_m\mathbf{x}_m + \mathbf{n}_m, \tag{5.64}$$

$$\mathbf{x}_m = \rho\mathbf{x}_{m-1} + \mathbf{u}_m, \quad m = 1, \ldots, M. \tag{5.65}$$

Since $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are group-sparse, from the above model, we have

$$p(\mathbf{t}, \mathbf{x}_1, \ldots, \mathbf{x}_M; \boldsymbol{\gamma}_c) = \prod_{m=1}^{M} p(\mathbf{t}_m|\mathbf{x}_m)p(\mathbf{x}_m|\mathbf{x}_{m-1}; \boldsymbol{\gamma}_c), \tag{5.66}$$

where $p(\mathbf{x}_1|\mathbf{x}_0; \boldsymbol{\gamma}_c) \triangleq p(\mathbf{x}_1; \boldsymbol{\gamma}_c)$. Using (5.66), the posterior distribution of the sparse vectors $p(\mathbf{x}_1, \ldots, \mathbf{x}_M|\mathbf{t}; \boldsymbol{\gamma}_c^{(r)})$ is computed using the recursive Kalman Filter and Smoother

(KFS) equations for $1 \leq m \leq M$ as follows [124, 128]:

for $m = 1, \ldots, M$ **do**

Prediction: $\hat{\mathbf{x}}_{m|m-1} = \rho \hat{\mathbf{x}}_{m-1|m-1}$ (5.67)

$\mathbf{P}_{m|m-1} = \rho^2 \mathbf{P}_{m-1|m-1} + (1 - \rho^2)\mathbf{\Gamma}_c$ (5.68)

Filtering: $\mathbf{G}_m = \mathbf{P}_{m|m-1}\mathbf{\Phi}_m^T \left(\sigma^2 \mathbf{I}_N + \mathbf{\Phi}_m \mathbf{P}_{m|m-1}\mathbf{\Phi}_m^T\right)^{-1}$ (5.69)

$\hat{\mathbf{x}}_{m|m} = \hat{\mathbf{x}}_{m|m-1} + \mathbf{G}_m(\mathbf{t}_m - \mathbf{\Phi}_m \hat{\mathbf{x}}_{m|m-1})$ (5.70)

$\mathbf{P}_{m|m} = (\mathbf{I}_B - \mathbf{G}_m \mathbf{\Phi}_m)\mathbf{P}_{m|m-1}$ (5.71)

end

for $j = M, M - 1, \ldots, 2$ **do**

Smoothing: $\hat{\mathbf{x}}_{j-1|m} = \hat{\mathbf{x}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{x}}_{j|m} - \hat{\mathbf{x}}_{j|j-1})$ (5.72)

$\mathbf{P}_{j-1|m} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|m} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^T$ (5.73)

end,

where $\mathbf{J}_{j-1} = \rho \mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{G}_m$ is the Kalman gain matrix. The above mentioned KFS equations are initialized by setting $\hat{\mathbf{x}}_{0|0} = \mathbf{0}_B$, i.e., a B length zero vector, and $\mathbf{P}_{0|0} = \mathbf{\Gamma}_c$. The E-step requires the computation of $\mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_M|\mathbf{t};\boldsymbol{\gamma}_c^{(r)}}[\mathbf{x}_j \mathbf{x}_{j-1}^T] \triangleq \mathbf{P}_{j,j-1|m} + \hat{\mathbf{x}}_{j|m}\hat{\mathbf{x}}_{j-1|m}^T$ for $m = M, M - 1, \ldots, 2$, which we obtain from [124] as follows:

$$\mathbf{P}_{j-1,j-2|m} = \mathbf{P}_{j-1|j-1}\mathbf{J}_{j-2}^T + \mathbf{J}_{j-1}^T(\mathbf{P}_{j,j-1|m} - \rho \mathbf{P}_{j-1|j-1})\mathbf{J}_{j-2}.$$ (5.74)

The above recursion is initialized using $\mathbf{P}_{m,m-1|M} = \rho(\mathbf{I}_B - \mathbf{G}_m \mathbf{\Phi}_m)\mathbf{P}_{m-1|m-1}$. Note that $\mathbf{x}_{i|M}$ and $\mathbf{P}_{i|M}, 1 \leq i \leq M$ represent the posterior mean and covariance of $\mathbf{x}$ given $\mathbf{t}$, respectively.

The expectation $\mathbb{E}_{\mathbf{t}}$ involves computing $p(\mathbf{t}|\mathbf{y}_c; \boldsymbol{\gamma}_c^{(r)})$ using (5.62). As mentioned earlier, $\boldsymbol{\Gamma}_B$ is given by $\boldsymbol{\Gamma}_B = \mathbf{B} \otimes \boldsymbol{\Gamma}_c$. The KFS equations in (5.67)-(5.75) constitute the $\mathbb{E}_{\mathbf{x}}$ step, after which we compute $\mathbb{E}_{\mathbf{t}}$. However, due to the recursive nature of the inner E-step, $\mathbb{E}_{\mathbf{x}}$, the expectation of $\boldsymbol{\mu}_{\mathbf{x}_m}$ w.r.t. the posterior density of $\mathbf{t}$ is a recursive function of $\mathbf{t}_m, \dots, \mathbf{t}_1$. As M increases, the complexity of computing such a recursive expectation becomes prohibitive. In order to circumvent this problem, we employ an alternate technique, known as the Nested EM approach [147]. This monotonically convergent approach allows us to simplify the overall algorithm into an inner and outer EM loop, while the unknown parameter $\boldsymbol{\gamma}_c$ is the common factor between the two loops. We call this algorithm as the NSBL algorithm, where the nested E- and M-steps are given as

$$
\begin{aligned}
\text{E-step}: Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r+\frac{k}{K})}, \boldsymbol{\gamma}_c^{(r)}\right) &= \mathbb{E}_{\mathbf{t}|\mathbf{y}_c; \boldsymbol{\gamma}_c^{(r)}}\left[\mathbb{E}_{\mathbf{x}_1,\dots,\mathbf{x}_M|\mathbf{t}; \boldsymbol{\gamma}_c^{(r+\frac{k}{K})}}[\log p(\mathbf{y}_c, \mathbf{t}, \mathbf{x}_1, \dots, \mathbf{x}_M; \boldsymbol{\gamma}_c)]\right] \\
\text{M-step}: \boldsymbol{\gamma}_c^{(r+\frac{k+1}{K})} &= \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B\times 1}}{\arg\max} \, Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r+\frac{k}{K})}, \boldsymbol{\gamma}_c^{(r)}\right).
\end{aligned}
\tag{5.75}
$$

The inner EM loop is initialized by $\boldsymbol{\gamma}_c^{(r+\frac{0}{K})} = \boldsymbol{\gamma}_c^{(r)}$. Note that, when $\boldsymbol{\gamma}_c$ is updated in every iteration, only the inner E-step ($\mathbb{E}_{\mathbf{x}} = \mathbb{E}_{\mathbf{x}_1,\dots,\mathbf{x}_M|\mathbf{t}; \boldsymbol{\gamma}_c^{(r+\frac{k}{K})}}[\cdot]$) is updated. The overall NSBL algorithm is executed by nesting one EM loop within the other, as depicted in Fig. 5.6. The inner EM loop consists of $\mathbb{E}_{\mathbf{x}}$ and the corresponding posterior distribution is given by (5.67)-(5.71). Further, the M-step for the inner EM loop is given by

$$
\gamma^{(r+\frac{(k+1)}{K})}(i) = \frac{1}{M}\left(\sum_{j=2}^{M} \frac{\mathbf{M}_{j|M}(i,i)}{(1-\rho^2)} + \mathbf{M}_{1|M}(i,i)\right)
\tag{5.76}
$$

for $1 \leq i \leq B$, where $\mathbf{M}_{j|M} \triangleq \mathbf{P}_{j|M} + \hat{\mathbf{x}}_{j|M}\hat{\mathbf{x}}_{j|M}^T + \rho^2(\mathbf{P}_{j-1|M} + \hat{\mathbf{x}}_{j-1|M}\hat{\mathbf{x}}_{j-1|M}^T) - 2\rho(\mathbf{P}_{j,j-1|M} + \hat{\mathbf{x}}_{j|M}\hat{\mathbf{x}}_{j-1|M}^T)$ and $\mathbf{M}_{1|M} \triangleq \mathbf{P}_{1|M} + \hat{\mathbf{x}}_{1|M}\hat{\mathbf{x}}_{1|M}^T$. After K iterations of the inner EM loop, we obtain $\boldsymbol{\gamma}_c^{r+\frac{K}{K}} = \boldsymbol{\gamma}_c^{r+1}$, which affects the posterior distribution of $\mathbf{t}$. The outer EM loop
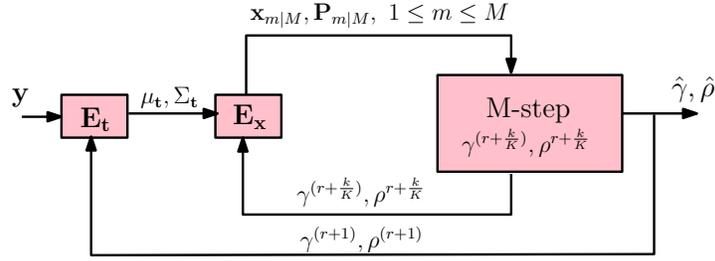
Figure 5.6: Pictorial representation of the Nested SBL implementation.

consists of updating the posterior distribution of $\mathbf{t}$ given in (5.62).

An update step for the unknown correlation coefficient ρ can also be incorporated into the M-step of the NSBL algorithm. The correlation coefficient $\rho^{(r+\frac{(k+1)}{K})}$ in the $(r+\frac{(k+1)}{K})^{\text{th}}$ iteration is obtained as a solution to the cubic equation

$$(2B(M-1))\rho^3 + \text{Tr}\{\mathbf{T}_2 + \mathbf{T}_3\}\rho^2 - \text{Tr}\{\mathbf{T}_2 + \mathbf{T}_3\} - [2B(M-1) - 2\text{Tr}\{\mathbf{T}_1 + \mathbf{T}_4)\}]\rho = 0,$$

$$(5.77)$$

where the matrices $\mathbf{T}_1$ through $\mathbf{T}_4$ are defined as

$$\mathbf{T}_1 = \mathbf{\Gamma}_c^{-1} \sum_{j=2}^M [\mathbf{P}_{j|M} + \hat{\mathbf{x}}_{j|M}\hat{\mathbf{x}}_{j|M}^T], \tag{5.78}$$

$$\mathbf{T}_2 = \mathbf{\Gamma}_c^{-1} \sum_{j=2}^M [\mathbf{P}_{j,j-1|M} + \hat{\mathbf{x}}_{j|M}\hat{\mathbf{x}}_{j-1|M}^T], \tag{5.79}$$

$$\mathbf{T}_3 = \mathbf{\Gamma}_c^{-1} \sum_{j=2}^M [\mathbf{P}_{j,j-1|M} + \hat{\mathbf{x}}_{j-1|M}\hat{\mathbf{x}}_{j|M}^T], \tag{5.80}$$

$$\mathbf{T}_4 = \mathbf{\Gamma}_c^{-1} \sum_{j=2}^M [\mathbf{P}_{j-1|M} + \hat{\mathbf{x}}_{j-1|M}\hat{\mathbf{x}}_{j-1|M}^T]. \tag{5.81}$$

Among the possible solutions of the above cubic equation, we pick the $\rho \in \mathbb{R}$ that satisfies $0 \leq \rho \leq 1$.

Using a flop-count analysis [129], we note that the computations in BSBL are dominated by the E-step, which incurs a computational complexity of $\mathcal{O}(N^2MB)$. On the

other hand, NSBL consists of two EM loops, where the maximum complexity of the outer and inner EM loop are $\mathcal{O}(N^2MB)$ and $\mathcal{O}(MB^3)$, respectively. Typically, in the nested EM approach, the number of inner EM iterations are fixed, so that the outer EM loop is guaranteed to converge, and the inner EM loop ensures likelihood increase [147]. Consequently, the complexity of the NSBL algorithm is dominated by $\mathcal{O}(N^2MB)$. However, since the number of outer EM iterations are far lower than that of BSBL, the NSBL entails a lower computational complexity than the BSBL approach.
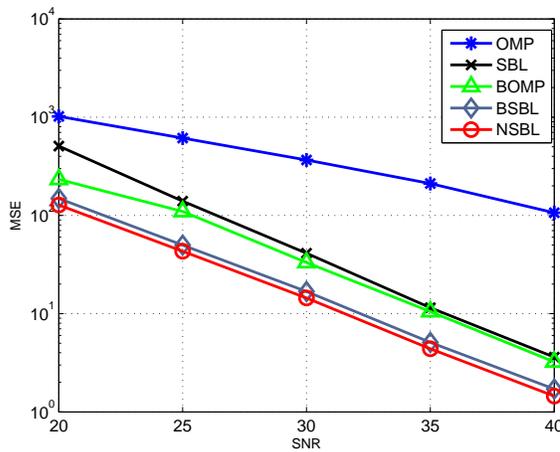
Figure 5.7: MSE performance of the NSBL as compared to BSBL, BOMP, SBL and OMP algorithms for $N = 40$, $L = 100$, $B = 5$ and $S = 1$.
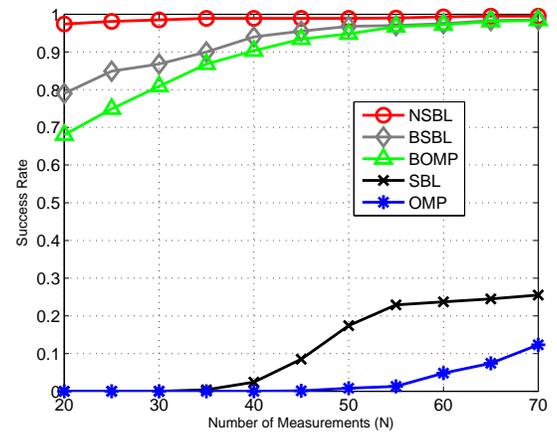
Figure 5.8: Success rate of the NSBL as compared to BSBL, BOMP, SBL and OMP algorithms for SNR $= 40$, $L = 100$, $B = 5$ and $S = 1$.

In Fig. 5.7 and Fig. 5.8, we demonstrate the MSE performance and the support recovery performance of the NSBL algorithm as compared to BSBL [134] and the conventional SBL and the OMP algorithms, which are unaware of the correlated cluster-sparse nature of sparse vectors. Since the NSBL algorithm employs EM based updates for updating the unknown correlation co-efficient ρ unlike the heuristic updates based BSBL algorithm, NSBL has a superior performance compared to the BSBL algorithm. Further,

we see that the cluster-sparse structure aware NSBL and the BSBL algorithms have a superior performance compared to SBL and OMP algorithms.

Remarks: In the literature on cluster-sparse vector recovery, past work has mainly focused on sparse vectors that consist of clusters where intra-cluster correlation is absent, i.e., $\mathbf{B} = \mathbf{I}_B$ [138,139]. The NSBL proposed earlier has a very simple form when $\mathbf{B} = \mathbf{I}_B$, which we refer to as the PCSBL algorithm.

When $\mathbf{B} = \mathbf{I}_B$, the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are uncorrelated, and hence, we have $p(\mathbf{x}_c|\mathbf{t}; \boldsymbol{\gamma}_c) = \prod_{m=1}^{M} p(\mathbf{x}_m|\mathbf{t}_m; \boldsymbol{\gamma}_c)$. This decomposes the cluster-sparse recovery problem in (5.9) into a MMV problem [104], where the goal is recovering group-sparse vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ from multiple measurements $\boldsymbol{\mu}_{\mathbf{t}_1}, \ldots, \boldsymbol{\mu}_{\mathbf{t}_M}$. The posterior distribution of $\mathbf{x}_m$ is given by $p(\mathbf{x}_m|\mathbf{t}_m; \boldsymbol{\gamma}_c^{(r)}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_m}, \boldsymbol{\Sigma}_{\mathbf{x}_m})$, where $\boldsymbol{\mu}_{\mathbf{x}_m} = \beta_m^{-1}\sigma^{-2}\boldsymbol{\Sigma}_{\mathbf{x}_m}\boldsymbol{\Phi}_m^T\mathbf{t}_m$ and $\boldsymbol{\Sigma}_{\mathbf{x}_m} = \left(\frac{\boldsymbol{\Phi}_m^T\boldsymbol{\Phi}_m}{\beta_m\sigma^2} + \boldsymbol{\Gamma}_c^{(r)^{-1}}\right)^{-1}$. Using the posterior distribution computed above, the update for $\boldsymbol{\gamma}_c$ is obtained as follows:

$$
\begin{aligned}
\boldsymbol{\gamma}_c^{(r+1)} &= \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg\max} \; \mathbb{E}_{\mathbf{t}, \mathbf{x}_c|\mathbf{y}_c; \boldsymbol{\gamma}_c^{(r)}}[\log p(\mathbf{t}, \mathbf{x}_c; \boldsymbol{\gamma}_c)] \\
&= \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg\max}(c' - \mathbb{E}_{\mathbf{t}|\mathbf{y}_c; \boldsymbol{\gamma}_c^{(r)}}\mathbb{E}_{\mathbf{x}_c|\mathbf{t}; \boldsymbol{\gamma}_c^{(r)}}[\tfrac{\mathbf{x}_c^T\boldsymbol{\Gamma}_B^{-1}\mathbf{x}_c}{2} + \tfrac{1}{2}\log|\boldsymbol{\Gamma}_B|])
\end{aligned}
\tag{5.82}
$$

In the above expression, $\log|\boldsymbol{\Gamma}_B|$ simplifies as $M\log|\boldsymbol{\Gamma}_c|$ and $\mathbf{x}_c^T\boldsymbol{\Gamma}_B^{-1}\mathbf{x}_c = \sum_{m=1}^{M}\mathbf{x}_m^T\boldsymbol{\Gamma}_c^{-1}\mathbf{x}_m$. Further, $\mathbb{E}_{\mathbf{x}_c|\mathbf{t}; \boldsymbol{\gamma}_c^{(r)}}[\mathbf{x}_m^T\boldsymbol{\Gamma}_c^{-1}\mathbf{x}_m] = \text{Tr}(\boldsymbol{\Gamma}_c^{-1}(\boldsymbol{\Sigma}_{\mathbf{x}_m} + \boldsymbol{\mu}_{\mathbf{x}_m}\boldsymbol{\mu}_{\mathbf{x}_m}^T))$. Substituting for $\boldsymbol{\mu}_{\mathbf{x}_m}$, we obtain the overall optimization problem in the M-step as

$$
\boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B \times 1}}{\arg\min}(c' + \tfrac{M}{2}\log|\boldsymbol{\Gamma}_c| + \tfrac{1}{2}\sum_{m=1}^{M}\text{Tr}(\boldsymbol{\Gamma}_c^{-1}\boldsymbol{\Sigma}_{\mathbf{x}_m})\text{Tr}\left(\boldsymbol{\Gamma}_c^{-1}\tfrac{\boldsymbol{\Sigma}_{\mathbf{x}_m}\boldsymbol{\Phi}_m^T\mathbf{R}_m\boldsymbol{\Phi}_m\boldsymbol{\Sigma}_{\mathbf{x}_m}}{\beta_m^2\sigma^4}\right)), \tag{5.83}
$$

where $\mathbf{R}_m = \boldsymbol{\Sigma}_{\mathbf{t}_m} + \boldsymbol{\mu}_{\mathbf{t}_m}\boldsymbol{\mu}_{\mathbf{t}_m}^T$, $\boldsymbol{\Sigma}_{\mathbf{t}_m} \in \mathbb{R}^{N \times N}$ is the mth entry of blkdiag($\boldsymbol{\Sigma}_{\mathbf{t}}$) and blkdiag($\mathbf{A}$)

returns the block diagonal matrices of $\mathbf{A}$. Maximizing (5.83) w.r.t. $\boldsymbol{\gamma}_c$, we get

$$\boldsymbol{\gamma}_c^{(r+1)} = \frac{1}{M} \sum_{m=1}^{M} \text{diag} \left(\boldsymbol{\Sigma}_{\mathbf{x}_m} + \frac{\boldsymbol{\Sigma}_{\mathbf{x}_m} \boldsymbol{\Phi}_m^T \mathbf{R}_m \boldsymbol{\Phi}_m \boldsymbol{\Sigma}_{\mathbf{x}_m}}{\beta_m^2 \sigma^4} \right). \tag{5.84}$$

The proposed PCSBL and the BSBL algorithm [142] are mathematically equivalent for the case when $\mathbf{B} = \mathbf{I}_B$. However, the PCSBL approach allows for parallel implementation of the algorithm, since the cluster-sparse vector is recovered by solving M parallel problems. Further, from (5.84), we see that the overall M-step is simply the average of the hyperparameter updates obtained from the M parallel problems.

Note that several cluster-sparse extensions of the algorithms proposed in the context of group-sparse signals are possible. For e.g., a simple modification of the prior density leads to the cluster-sparse variant of the MSBL algorithm, which we refer to as the BMSBL algorithm, for the recovery of multiple uncorrelated cluster-sparse vectors in an MMV framework. Similarly, a modification in the prior density following a modification in the AR model leads to the KBMSBL algorithm as the block-sparse MMV variant of the KSBL algorithm in order to track multiple correlated cluster-sparse vectors in the MMV framework. The novelty in the proposed algorithms lies in effectively incorporating the block-sparse structure of the vectors in the prior pdf, leading to effective utilization of the information regarding the structure of the vectors. In chapter 4, we consider such cluster-sparse variants of the algorithms proposed in this chapter for OFDM channel estimation and tracking.

Until now, we have considered scenarios where the measurement matrices are completely known. In the following section, we consider a case when $\boldsymbol{\Phi}$ is not known completely; some entries of $\boldsymbol{\Phi}$ are missing. In particular, we propose algorithms that

facilitate learning the missing entries along with sparse vector estimation in a general

SBL framework.

5.3.5 Partially Unknown Measurement Matrix and SBL

In the algorithms proposed thus far, we assumed that the sparse vector $\mathbf{x}$ is unknown

and $\boldsymbol{\Phi}$ is completely known. In this section, we consider a partially known measure-

ment matrix, i.e., a measurement matrix in which a few entries are missing. Since SBL

uses an EM based framework, we show that SBL can be elegantly extended not only

to estimate such missing variables, but also incorporate their estimates into the update

equations of unknown parameters.

When $\boldsymbol{\Phi}$ is partially known, the unknown parameter set is given by $\Theta = [\boldsymbol{\Phi}, \boldsymbol{\gamma}]$. The

EM algorithm in such a case is given by

$$\text{E-step: } Q(\Theta|\Theta^{(r)}) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\Theta^{(r)}}\left[\log(p(\mathbf{y}, \mathbf{x}; \Theta))\right]$$

$$\text{M-step: } \Theta^{(r+1)} = \arg\max_{\Theta} Q(\Theta|\Theta^{(r)}) \tag{5.85}$$

The E-step given above can be further simplified as

$$\log(p(\mathbf{y}, \mathbf{x}; \Theta)) = \log(p(\mathbf{y}|\mathbf{x}; \boldsymbol{\Phi})) + \log(p(\mathbf{x}; \boldsymbol{\gamma})) \tag{5.86}$$

The $Q(\Theta|\Theta^{(r)})$ function in the E-step given above splits into two independent functions,

as follows:

$$Q(\Theta|\Theta^{(r)}) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\Theta^{(r)}}\left[\log(p(\mathbf{y}|\mathbf{x}; \boldsymbol{\Phi}))\right] + \mathbb{E}_{\mathbf{x}|\mathbf{y};\Theta^{(r)}}\left[\log(p(\mathbf{x}; \boldsymbol{\gamma}))\right] \tag{5.87}$$

The optimization problem w.r.t. the missing variables in $\boldsymbol{\Phi}$ can be given by

$$Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(r)}) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\Theta^{(r)}}\left[\log(p(\mathbf{y}|\mathbf{x};\boldsymbol{\Phi}))\right], \tag{5.88}$$

since other terms in (5.87) are independent of the missing variables in $\boldsymbol{\Phi}$. Further, $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(r)})$ is independent of $\mathbf{x}$ and parameters of the pdf of $\mathbf{x}$. Specifically, the missing entries of $\boldsymbol{\Phi}$ are independent of the sparse vector estimation process, and in turn independent of the inherent structure of the sparse vector (the sparse vector being correlated, group-sparse or cluster-sparse). Hence, the expressions given above hold for all the algorithms proposed in this chapter.

Since estimating all the entries in $\boldsymbol{\Phi}$ leads to overfitting, we restrain ourselves to measurement matrices of the form $\boldsymbol{\Phi} = \mathbf{A}\psi$, where $\mathbf{A}$ is a diagonal matrix with a few missing entries and ψ is a known basis matrix (such as wavelet or Fourier). Such matrices are naturally found in applications such as OFDM channel estimation and data detection [89]. The function $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(r)})$ given above simplifies as

$$Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(r)}) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\mathbf{A}^{(r)}}\left[\log(p(\mathbf{y}|\mathbf{x};\mathbf{A}^{(r)}))\right] \tag{5.89}$$

As discussed earlier, the update for γ depends on the optimization problem with objective function given by the second term in (5.87). The optimization problem w.r.t $\mathbf{A}$ is given by

$$\begin{aligned}
\mathbf{A}^{(r+1)} &= \arg\max_{\mathbf{A}} Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(r)}) \\
&= \arg\max_{\mathbf{A}} \mathbb{E}_{\mathbf{x}|\mathbf{y};\mathbf{A}^{(r)}}\left[-\frac{(\mathbf{y}-\mathbf{A}\psi\mathbf{x})^T(\mathbf{y}-\mathbf{A}\psi\mathbf{x})}{\sigma^2}\right]
\end{aligned} \tag{5.90}$$

The optimization problem given above can be simplified as

$$\mathbf{A}^{(r+1)} = \arg\min_{\mathbf{A}} \left[\mathbf{y}^T \mathbf{A} \boldsymbol{\psi} \boldsymbol{\mu}_{\text{sbl}} + \boldsymbol{\mu}_{\text{sbl}}^T \boldsymbol{\psi}^T \mathbf{A}^T \mathbf{y} - \mathbf{y}^T \mathbf{A} \boldsymbol{\psi} [\boldsymbol{\mu}_{\text{sbl}} \boldsymbol{\mu}_{\text{sbl}}^T + \boldsymbol{\Sigma}_{\text{sbl}}] \boldsymbol{\psi}^T \mathbf{A}^T \mathbf{y} \right] \quad (5.91)$$

The above expression can be solved in closed form to obtain the update equation for elements of $\mathbf{A}$ as follows:

$$\mathbf{a} = (\mathbf{Y} \boldsymbol{\psi} \boldsymbol{\mu}_{\text{sbl}} + \boldsymbol{\mu}_{\text{sbl}}^T \boldsymbol{\psi}^T \mathbf{Y}^T)^{-1} \mathbf{Y} \boldsymbol{\psi} \boldsymbol{\mu}_{\text{sbl}}, \quad (5.92)$$

where $\mathbf{Y} = \text{diag}(\mathbf{y})$ and $\mathbf{a} = \text{diag}(\mathbf{A})$. The theory of EM algorithm shows that $Q(\Theta | \Theta^{(r)})$ increases monotonically for $1 \leq r \leq r_{\text{max}}$, where r_{max} is the maximum number of iterations. That is,

$$Q(\Theta^{(r+1)} | \Theta^{(r)}) \geq Q(\Theta^{(r)} | \Theta^{(r-1)}), \quad \text{for } 1 \leq r \leq r_{\text{max}}. \quad (5.93)$$

Note that the above function $Q(.)$ monotonically (in Θ) approaches the likelihood function, which in turn is bounded. This guarantees the convergence of the proposed approach. Further, the convergence guarantee holds for correlated/uncorrelated or sparse/block-sparse type of sparse vectors as well.

In Fig. 5.9 and Fig. 5.10, we demonstrate the MSE performance and the support recovery performance of the SBL algorithm as compared to the novel J-SBL algorithm. In order to run the SBL algorithm, we assume that the diagonal values of $\mathbf{A}$ are known at a few locations, and some of the entries are missing. Since the J-SBL incorporates estimates of the missing entries as well as the observations corresponding to the missing entries of $\mathbf{A}$, it outperforms the SBL algorithm, both in MSE and support recovery.
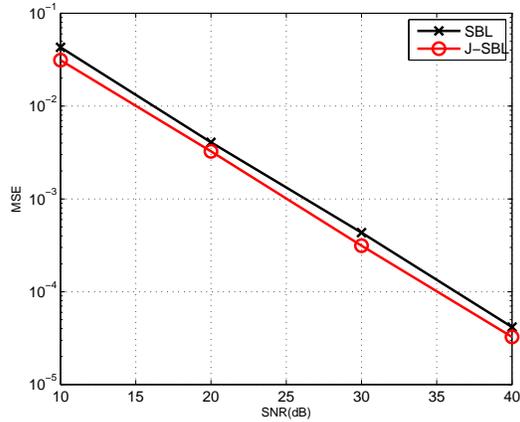
Figure 5.9: MSE performance of the SBL compared to J-SBL for $N = 30$, $L = 64$ and $S = 3$.
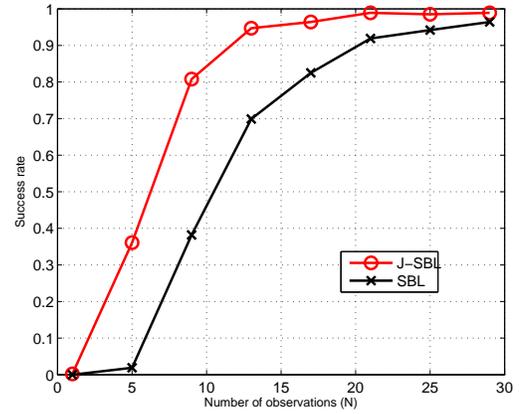
Figure 5.10: Success rate of the SBL compared to J-SBL for SNR $= 20$, $L = 64$ and $S = 3$.

5.4 Conclusions

In this chapter, we summarized the novel algorithms proposed in this thesis. We first stated the SBL and the MSBL algorithm. In order to handle inter-vector correlation in the MMV framework, we derived the KSBL algorithm as a low-complexity alternative to the known TSBL and ARSBL algorithm. The KSBL algorithm exploited the KFS framework, while the correlated sparse vectors were modeled using a first order AR model. In the context of recovering cluster-sparse vectors, we proposed the NSBL algorithm in order to exploit the intra-cluster correlation. First, we showed that in contrast to the heuristic updates of the correlation coefficient in the BSBL algorithm, the NSBL framework consists of EM updates with global convergence guarantee to a local optimum, i.e., it converges to a local minimum from any initial point. The NSBL and the PCSBL algorithms entail a complexity much lower than the BSBL algorithm. We also extended the SBL framework to jointly update a partially known matrix along with

recovering the sparse vector. Using Monte Carlo simulations, we presented a performance comparison of the proposed algorithms and their SBL and CS based counterparts.

Chapter 6

Conclusions and Future Work

In this thesis, we investigated the Bayesian learning technique known as sparse Bayesian learning for the problem of channel estimation and data detection in approximately-sparse (a-sparse) SISO and MIMO-OFDM systems. We also derived Cramér Type lower bounds to benchmark the performance of the SBL based estimators in the presence of compressible vectors. The concluding remarks of this thesis are summarized below.

6.1   Conclusions

In chapter 2, we derived Cramér Rao type lower Bounds (CRB) on the MSE for the single measurement and the multiple measurement SBL problem of estimating compressible signals. In particular, we derived the hybrid, Bayesian and marginalized CRBs under various assumptions on the unknown parameters, using a hierarchical model for the compressible priors. The derived bounds provided key insights into the MSE performance of SBL and the values of the parameters that govern these hyperpriors. We showed that the MCRB is tighter than the BCRB. We compared the lower bounds with the MSE performance of the ARD-SBL and the EM algorithm using Monte Carlo

simulations. The numerical results illustrated the near-optimality of EM based updates for SBL, which makes it interesting for practical implementations.

Chapter 3 considered joint approximately sparse channel estimation and data detection for block-fading and time-varying channels in SISO-OFDM systems, using SBL-based techniques. To estimate the a-sparse block-fading channel, we proposed the pilot-based SBL algorithm and the J-SBL algorithm for joint a-sparse channel estimation and data detection, and a mathematically equivalent low-complexity RJ-SBL algorithm. For time-varying channels, we used a first order AR model to capture the temporal correlation of the a-sparse channel and proposed a novel pilot-based KSBL algorithm. We generalized the KSBL algorithm to obtain the J-KSBL algorithm for joint channel estimation and data detection. We discussed the implementation issues of the recursive algorithms and showed that the proposed algorithms entail a significantly lower computational complexity compared to the previously known SBL techniques. Simulation results showed that the proposed recursive techniques are able to exploit the temporal correlation of the channel, resulting in an enhanced channel estimation and data detection capability compared to the per-symbol SBL and J-SBL algorithms, and also learn the hyperparameters within a few OFDM symbols.

In chapter 4, we considered pilot-only channel estimation and joint group a-sparse (ga-sparse) and group a-cluster sparse (ga-cpsarse) channel estimation and data detection for block-fading and time-varying channels in MIMO-OFDM systems, using the SBL framework. To estimate the ga-sparse and ga-csparse block-fading channels, we adapted the existing MSBL and BMSBL algorithms and generalized it to obtain the J-MSBL and J-BMSBL algorithms, respectively, for joint ga-sparse and ga-csparse channel

estimation and data detection. We used a first order AR model to capture the temporal correlation of the ga-sparse and ga-csparse channels and proposed a novel pilot-only KMSBL and KBMSL algorithms, respectively. We generalized these algorithms to obtain the J-KMSBL and J-KBMSBL algorithms, respectively, for joint channel estimation and data detection. We discussed the computational aspects of the proposed algorithms and showed that the proposed recursive algorithms entail a significantly lower computational complexity compared to the previously known SBL based techniques. Further, we also discussed efficient implementation structures for ga-csparse channels in block-fading and time-varying scenarios. Simulation results showed that (i) joint algorithms outperformed their pilot-only counterparts, (ii) recursive techniques outperformed the per-symbol algorithms, and (iii) algorithms proposed in the context of ga-csparse channels outperformed their ga-csparse counterparts.

In chapter 5 of this thesis, we summarized the algorithms proposed in this thesis in a generic SBL framework. We showed that many of the proposed algorithms are capable of handling correlated sparse vectors along with being efficient and low-complexity. We demonstrated the efficacy of the proposed techniques (MSE and support recovery performance) as compared to the existing CS and SBL based algorithms using Monte Carlo simulations. We also showed that these performance improvements come at a minimal computational cost. In some cases, exploiting the structure in the sparse signal enables us to improve both the performance and computational complexity.

6.2   Future Work

Some interesting directions for future work are as follows.

- In the Kalman based KSBL framework for correlated sparse MMVs, the EM framework involves a fixed-interval type smoothing. Further, this necessitates batch processing and not *online* processing of the observations. It would be interesting to derive fixed-lag type smoother for estimating correlated sparse MMVs.

- In this work, we have assumed the noise to be white across the measurements. It would be interesting to study the effect of colored noise along with exploring the learning rules for the same.

- Throughout this work, we propose SBL based iterative receivers where we perform data detection and employ the detected symbols as pilots in the channel estimation procedure. The channel estimates are used to compute the log likelihood ratios and decode the data bits. It would be interesting to explore SBL based iterative receivers that perform joint channel estimation and data decoding.

Appendix A

Appendix for Chapter 2

A.1 Proof of Proposition 1

Using the graphical model of Fig. 1.1 in (2.5), $\mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x})$ is computed as

$$
\begin{aligned}
\mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x}) &\triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla_{\mathbf{x}}^2 \log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})\right] \\
&= -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla_{\mathbf{x}}\left(\frac{\boldsymbol{\Phi}^T(\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})}{\sigma^2} - \boldsymbol{\Upsilon}^{-1}\mathbf{x}\right)\right] \\
&= \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}.
\end{aligned}
\tag{A.1}
$$

Similarly, it is straightforward to show that $\nabla_{\mathbf{x}}\nabla_{\boldsymbol{\gamma}} \log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma}) = \operatorname{diag}\left(\frac{x_1}{\gamma_1^2}, \frac{x_2}{\gamma_2^2}, \ldots, \frac{x_L}{\gamma_L^2}\right)$.
Since x_i are zero mean random variables,

$$
\mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\mathbf{x}) = -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla_{\boldsymbol{\gamma}}\nabla_{\mathbf{x}} \log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})\right] = \mathbf{0}_{L\times L}.
\tag{A.2}
$$

Further,

$$
\mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) \triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla_{\boldsymbol{\gamma}}^2(\log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) + \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma}))\right].
\tag{A.3}
$$

Now, since $\log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma}) = \sum_{i=1}^{L} \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x}_i;\gamma_i)$, we get,

$$
\frac{\partial^2 \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma})}{\partial \gamma_i \gamma_j} = \begin{cases} \frac{1}{2\gamma_i^2} - \frac{x_i^2}{\gamma_i^3} & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{A.4}
$$

Taking $-\mathbb{E}_{\mathbf{X};\boldsymbol{\gamma}}(\cdot)$ on both the sides of the equation above and noting that $\mathbb{E}_{\mathbf{X};\boldsymbol{\gamma}}(x_i^2) = \gamma_i$,

we obtain

$$
\begin{aligned}
\mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) &= \text{diag}\left(-\mathbb{E}_{\mathbf{X};\boldsymbol{\gamma}}\left[\frac{\partial^2 \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma})}{\partial \gamma_i^2}\right]\right) \\
&= \text{diag}\left(\left[\frac{1}{2\gamma_1^2}, \cdots, \frac{1}{2\gamma_L^2}\right]\right).
\end{aligned} \tag{A.5}
$$

This completes the proof.

A.2   Proof of Proposition 2

Using the graphical model of Fig. 1.1 in (2.5), $\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x})$ is computed as

$$
\begin{aligned}
\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) &\triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X},\boldsymbol{\Gamma}}\left[\nabla_{\mathbf{x}}^2 \log p_{\mathbf{Y},\mathbf{X},\boldsymbol{\Gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})\right] \\
&= -\mathbb{E}_{\mathbf{Y},\mathbf{X},\boldsymbol{\Gamma}}\left[\nabla_{\mathbf{x}}\left(\frac{\boldsymbol{\Phi}^T(\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})}{\sigma^2} - \boldsymbol{\Upsilon}^{-1}\mathbf{x}\right)\right] \\
&= \mathbb{E}_{\boldsymbol{\Gamma}}\left[\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right] \\
&= \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{\Upsilon}^{-1}\right].
\end{aligned} \tag{A.6} \tag{A.7}
$$

The expression for $\mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{\Upsilon}^{-1}\right]$ w.r.t. γ_i is given by,

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\Gamma}}\left[\frac{1}{\gamma_i}\right] &= K_{\gamma}\int_{\gamma_i=0}^{\infty}\gamma_i^{\left(-\frac{\nu}{2}-2\right)}\exp\left\{-\frac{\nu}{2\lambda\gamma_i}\right\}\mathrm{d}\gamma_i \\
&= K_{\gamma}\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}+1}}\underbrace{\int_{\gamma_i=0}^{\infty}\mathcal{IG}\left(\frac{\nu}{2}+1,\frac{\nu}{2\lambda}\right)\mathrm{d}\gamma_i}_{=1} \\
&= \lambda,
\end{aligned} \tag{A.8} \tag{A.9}
$$

since $K_\gamma = \left(\frac{\nu}{2\lambda}\right)^{\nu/2} \left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}$. Hence, the overall bound is given by

$$\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2} + \lambda \mathbf{I}_{L \times L}. \tag{A.10}$$

Using the graphical model of Fig. 1.1 in (2.5), for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ is defined as

$$\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) \triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X},\Gamma}\left[\nabla^2_{\boldsymbol{\gamma}}\left(\log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\right.\right.$$

$$\left.\left. + \log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma}) + \log p_\Gamma(\boldsymbol{\gamma})\right)\right]. \tag{A.11}$$

Since the expressions for $\log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma})$ and $\log p_\Gamma(\boldsymbol{\gamma})$ are separable and symmetric w.r.t. γ_i, the off-diagonal terms of $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ are zero, and it is sufficient to evaluate the diagonal terms

$-\mathbb{E}_{\mathbf{Y},\mathbf{X},\Gamma}\left(\frac{\partial^2 (\log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma}) + \log p_\Gamma(\boldsymbol{\gamma}))}{\partial \gamma_i^2}\right)$. Differentiating the expression w.r.t. γ_i twice,

$$\frac{\partial^2 \left(\log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma}) + \log p_\Gamma(\boldsymbol{\gamma})\right)}{\partial \gamma_i^2} = -\frac{(\nu+1)}{2\gamma_i^2} + \frac{\nu}{\lambda \gamma_i^3}. \tag{A.12}$$

The expression for $-\mathbb{E}_\Gamma\left[-\frac{(\nu+1)}{2\gamma_i^2} + \frac{\nu}{\lambda\gamma_i^3}\right]$ is given by

$$\mathbb{E}_\Gamma\left[\frac{(\nu+1)}{2\gamma_i^2} - \frac{\nu}{\lambda\gamma_i^3}\right] =$$

$$K_\gamma \int_{\gamma_i=0}^{\infty} \left[\frac{(\nu+1)\gamma_i^{-2}}{2} - \frac{\nu\gamma_i^{-3}}{\lambda}\right] \gamma_i^{\left(-\frac{\nu}{2}-1\right)} \exp\{-\frac{\nu}{2\lambda\gamma_i}\}d\gamma_i, \tag{A.13}$$

where $K_\gamma = \left(\frac{\nu}{2\lambda}\right)^{\nu/2} \left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}$. After some manipulation, it can be shown that the above integral reduces to

$$-\mathbb{E}_\Gamma\left[-\frac{(\nu+1)}{2\gamma_i^2} + \frac{\nu}{\lambda\gamma_i^3}\right] = \frac{\lambda^2(\nu+2)(\nu+7)}{2\nu}. \tag{A.14}$$

Using (2.5), the $(ij)^{th}$ component of the matrix $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \mathbf{x})$ is obtained as

$$(\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \mathbf{x}))_{ij} = \frac{\partial^2 \log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma})}{\partial \gamma_i \partial x_i} = -\frac{x_i}{\gamma_i^2}, \tag{A.15}$$

and $\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}) = (\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \mathbf{x}))^T$. Since $\mathbb{E}_{\mathbf{X}|\Gamma}(x_i) = 0$, $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \mathbf{x}) = \mathbf{0}_{L \times L}$. This completes the proof.

A.3 Proof of Theorem 1

To establish the regularity condition, the first order derivative of the log likelihood $\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y}; \boldsymbol{\gamma})$ is required. This, in turn, requires the evaluation of $\frac{\partial \log |\boldsymbol{\Sigma}_y|}{\partial \gamma_j}$ and $\frac{\partial \mathbf{y}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}}{\partial \gamma_j}$. The derivative of the log likelihood w.r.t. γ_j is obtained using the chain rule [148] as follows:

$$\begin{aligned} \frac{\partial \log |\boldsymbol{\Sigma}_y|}{\partial \gamma_j} &= \mathrm{Tr}\left\{ \left(\frac{\partial \log |\boldsymbol{\Sigma}_y|}{\partial \boldsymbol{\Sigma}_y} \right)^T \frac{\partial \boldsymbol{\Sigma}_y}{\partial \gamma_j} \right\} \\ &= \mathrm{Tr}\left\{ (\boldsymbol{\Sigma}_y^{-1})^T \Phi_j \Phi_j^T \right\} = \Phi_j^T \boldsymbol{\Sigma}_y^{-1} \Phi_j. \end{aligned} \tag{A.16}$$

Here, we have used the identity $\nabla_X \log |X| = X^{-1}$ [148], and results from vector calculus [148] to obtain $\frac{\partial \boldsymbol{\Sigma}_y}{\partial \gamma_j} = \Phi_j \Phi_j^T$, where Φ_j is the jth column of Φ. Similarly, the derivative of $\mathbf{y}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}$ can be obtained as

$$\begin{aligned} \frac{\partial \mathbf{y}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}}{\partial \gamma_j} &= \mathrm{Tr}\left\{ \left(\frac{\partial \mathbf{y}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}}{\partial \boldsymbol{\Sigma}_y^{-1}} \right)^T \frac{\partial \boldsymbol{\Sigma}_y^{-1}}{\partial \gamma_j} \right\} \\ &= -\Phi_j^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y} \mathbf{y}^T \boldsymbol{\Sigma}_y^{-1} \Phi_j, \end{aligned} \tag{A.17}$$

and hence,

$$\frac{\partial}{\partial \gamma_j} \log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma}) = -\left(\frac{\Phi_j^T \Sigma_y^{-1} \Phi_j - \Phi_j^T \Sigma_y^{-1} \mathbf{y}\mathbf{y}^T \Sigma_y^{-1} \Phi_j}{2}\right). \tag{A.18}$$

Taking $\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}(\cdot)$ on both the sides of the above equation, it can be seen that

$$\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}\left[\frac{\partial}{\partial \gamma_j} \log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})\right]$$

$$= -\left(\frac{\Phi_j^T \Sigma_y^{-1} \Phi_j - \Phi_j^T \Sigma_y^{-1}\left\{\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y}\mathbf{y}^T)\right\} \Sigma_y^{-1} \Phi_j}{2}\right) = 0, \tag{A.19}$$

since $\mathbb{E}_{\mathbf{Y}}(\mathbf{y}\mathbf{y}^T) = \Sigma_y$. Hence, the pdf satisfies the required regularity constraint.

Using the regularity condition above, the MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$ is obtained by computing the second derivative of the log likelihood, as follows:

$$-\frac{\partial^2}{\partial \gamma_i \gamma_j} \log p_{\mathbf{Y},\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma}) = \frac{1}{2}\frac{\partial}{\partial \gamma_i}(\Phi_j^T \Sigma_y^{-1} \Phi_j - (\Phi_j^T \Sigma_y^{-1}\mathbf{y})^2)$$

$$= \frac{1}{2}\mathrm{Tr}\left\{\Phi_j \Phi_j^T(-\Sigma_y^{-1}\Phi_i \Phi_i^T \Sigma_y^{-1})\right\}$$

$$- (\Phi_j^T \Sigma_y^{-1}\mathbf{y})\mathrm{Tr}\left\{\left(\frac{\partial(\Phi_j^T \Sigma_y^{-1}\mathbf{y})}{\partial \Sigma_y^{-1}}\right)^T \frac{\partial \Sigma_y^{-1}}{\partial \gamma_i}\right\}$$

$$= \frac{1}{2}\Phi_j^T(-\Sigma_y^{-1}\Phi_i \Phi_i^T \Sigma_y^{-1})\Phi_j -$$

$$(\Phi_j^T \Sigma_y^{-1}\mathbf{y})\mathbf{y}^T \left(-\Sigma_y^{-1}\Phi_i \Phi_i^T \Sigma_y^{-1}\right)\Phi_j. \tag{A.20}$$

Taking $-\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}(\cdot)$ on both the sides of the above expression, we obtain

$$(\mathbf{M}^{\boldsymbol{\gamma}})_{ij} \triangleq -\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}\left[\frac{\partial^2 \log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})}{\partial \gamma_i \gamma_j}\right] = \frac{(\Phi_j^T \Sigma_y^{-1}\Phi_i)^2}{2}, \tag{A.21}$$

as stated in (2.11). This completes the proof.

A.3.1   Proof of Theorem 2

The proof follows from the proof for Theorem 3 in Appendix H by substituting $\tau = 2$.

A.3.2   Proof of Theorem 3

The MCRB for estimation of the compressible random vector with $\boldsymbol{\theta} = [\mathbf{x}]$ is given by

$$\mathbf{M}^{\mathbf{x}} = -\mathbb{E}_{\mathbf{Y},\mathbf{X}}[\nabla_{\mathbf{x}}^2 \log p_{\mathbf{Y},\mathbf{X}}(\mathbf{y},\mathbf{x})]$$

$$= -\mathbb{E}_{\mathbf{Y},\mathbf{X}}[\nabla_{\mathbf{x}}^2 \log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) + \nabla_{\mathbf{x}}^2 \log p_{\mathbf{X}}(\mathbf{x})]. \tag{A.22}$$

The first term above is given by

$$-\mathbb{E}_{\mathbf{Y},\mathbf{X}}\left[\nabla_{\mathbf{x}}^2 \log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\right] = -\mathbb{E}_{\mathbf{Y},\mathbf{X}}\left[\nabla_{\mathbf{x}}\frac{\boldsymbol{\Phi}^T(\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})}{\sigma^2}\right]$$

$$= -\mathbb{E}_{\mathbf{Y},\mathbf{X}}\left[\frac{-\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2}\right] = \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2}. \tag{A.23}$$

Note that $p_{\mathbf{X}}(\mathbf{x})$ is not differentiable if any of its components $x_i = 0$. However, the measure of $x_i = 0$ is zero since the distribution is continuous, and hence, this condition can be safely ignored. Note that

$$\frac{\partial \log p_{\mathbf{X}}(\mathbf{x})}{\partial x_i} = \begin{cases} -\frac{(\nu+1)\lambda}{\nu}\frac{x_i^{\tau-1}}{\left(1+\frac{\lambda x_i^{\tau}}{\nu}\right)} & \text{if } x_i > 0 \\ (-1)^{\tau}\frac{(\nu+1)\lambda}{\nu}\frac{x_i^{\tau-1}}{\left(1+(-1)^{\tau}\frac{\lambda x_i^{\tau}}{\nu}\right)} & \text{if } x_i < 0. \end{cases} \tag{A.24}$$

First, we consider the case of $x_i > 0$. Differentiating the above w.r.t. x_i again, we obtain

$$\frac{\partial}{\partial x_i^2} \log p_{\mathbf{X}}(\mathbf{x}) = \frac{-(\nu+1)\lambda(\tau-1)}{\nu}\frac{x_i^{\tau-2}}{\left(1+\frac{\lambda x_i^{\tau}}{\nu}\right)} + \frac{\lambda^2\tau(\nu+1)}{\nu^2}\frac{x_i^{2\tau-2}}{\left(1+\frac{\lambda x_i^{\tau}}{\nu}\right)^2}. \tag{A.25}$$

Taking $-\mathbb{E}_{\mathbf{X}}(\cdot)$ on both sides of the above equation, we get

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{K(\nu+1)\lambda}{\nu}$$

$$\int_0^\infty \left\{\frac{(\tau-1)x_i^{\tau-2}}{\left(1+\frac{\lambda x_i^\tau}{\nu}\right)^{(\nu+\tau+1)/\tau}} - \frac{\lambda\tau x_i^{2\tau-2}}{\nu\left(1+\frac{\lambda x_i^\tau}{\nu}\right)^{(\nu+2\tau+1)/\tau}}\right\}\mathrm{d}x_i. \tag{A.26}$$

The above can be simplified using the transformation $t_i = \frac{\lambda x_i^\tau}{\nu}$ and using $\int_0^\infty \frac{t^{u-1}}{(1+t)^{u+v}}\mathrm{d}t = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$, we get

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{K(\nu+1)(\tau-1)}{\tau}\left(\frac{\lambda}{\nu}\right)^{1/\tau}$$

$$\Gamma\left(1-\frac{1}{\tau}\right)\left\{\frac{\Gamma\left(\frac{\nu+\tau+2}{\tau}\right)-\frac{1}{\tau}\Gamma\left(\frac{\nu+2}{\tau}\right)}{\Gamma\left(\frac{\nu+2\tau+1}{\tau}\right)}\right\} \quad \text{for} \quad x_i > 0. \tag{A.27}$$

For the case of $x_i < 0$, we see that the integral simplifies to the integral given in (A.26).

Hence, the overall expression is given by

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{K(\nu+1)^2(\tau-1)}{\tau(\nu+\tau+1)}\left(\frac{\lambda}{\nu}\right)^{1/\tau}$$

$$\Gamma\left(1-\frac{1}{\tau}\right)\left\{\frac{\Gamma\left(\frac{\nu+2}{\tau}\right)}{\Gamma\left(\frac{\nu+\tau+1}{\tau}\right)}\right\}. \tag{A.28}$$

Substituting the expression for K in the above, we get

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{\tau^2(\nu+1)}{(\nu+\tau+1)}\left(\frac{\lambda}{\nu}\right)^{2/\tau}\frac{\Gamma\left(\frac{\nu+2}{\tau}\right)\Gamma\left(2-\frac{1}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)\Gamma\left(\frac{\nu}{\tau}\right)}. \tag{A.29}$$

Combining the expression above and (A.23), we obtain the MCRB in (2.17).

A.4   Proof of Proposition 3

In this case, we define $\boldsymbol{\theta}' = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$ and hence, $\boldsymbol{\theta} = [\boldsymbol{\theta}'^T, \xi]^T$. In order to compute the HCRB, we need to find $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi)$, $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}')$ and $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}', \xi)$. We have $\log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma},\xi}(\mathbf{y}, \mathbf{x}; \boldsymbol{\gamma}, \xi) =$

$\log p_{\mathbf{Y}|\mathbf{X};\xi}(\mathbf{y}|\mathbf{x};\xi) + \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma})$, where $\xi = \sigma^2$. Using (2.5), the submatrix $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\boldsymbol{\theta}') = \mathbf{H}^{\boldsymbol{\theta}'}$,

i.e., the same as computed earlier in (2.8). Hence, we focus on the block matrices that

occur due to the additional parameter ξ. First, $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\xi)$ is computed as in Sec. 3.6 in [105],

from which, $-\mathbb{E}_{\mathbf{Y},\mathbf{X};\xi}\left[-\frac{N}{2\xi^2}\right] = \frac{N}{2\xi^2}$.

From Lemma 2, it directly follows that $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\xi) = \mathbf{0}_{L\times 1}$. Using (2.5), we compute

$\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\mathbf{x},\xi)$ as follows:

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\mathbf{x},\xi) = \mathbb{E}_{\mathbf{X}}(\mathbb{E}_{\mathbf{Y}|\mathbf{X};\xi}(\boldsymbol{\Phi}^T\mathbf{y} - \boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{x})). \tag{A.30}$$

Since $\mathbb{E}_{\mathbf{Y}|\mathbf{X};\xi}(\mathbf{y}) = \boldsymbol{\Phi}\mathbf{x}$, $\mathbb{E}_{\mathbf{X}}(\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\mathbf{x}) - \boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{x}) = \mathbf{0}_{L\times 1}$. This completes the proof.

A.5   Proof of Proposition 4

In this case, we define $\boldsymbol{\theta} \triangleq [\boldsymbol{\theta}'^T, \xi]$ and $\boldsymbol{\theta}' \triangleq [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$. In order to compute the HCRB,

we need to find $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\xi)$, $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\boldsymbol{\theta}')$ and $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\boldsymbol{\theta}',\xi)$. Using (2.5), the expression for $\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\boldsymbol{\theta}')$ is the

same as computed earlier in (2.8). Since ξ is random, the expectation operator includes

an expectation over ξ, and hence,

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\xi) = -\mathbb{E}_{\mathbf{Y},\mathbf{X},\Xi}\left[\frac{\partial^2}{\partial\xi^2}(\log p_{\mathbf{Y}|\mathbf{X},\Xi}(\mathbf{y}|\mathbf{x},\xi) + \log p_{\Xi}(\xi))\right]$$
$$= \mathbb{E}_{\Xi}\left(\frac{N/2-c-1}{\xi^2} + \frac{2d}{\xi^3}\right). \tag{A.31}$$

The above expectation is evaluated as

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}}(\xi) = \frac{(N/2-c-1)d^c}{\Gamma(c)}\int_{\xi=0}^{\infty}\xi^{-2}\xi^{(-c-1)}\exp\left\{-\frac{d}{\xi}\right\}d\xi +$$
$$\frac{2d^{(c+1)}}{\Gamma(c)}\int_{\xi=0}^{\infty}\xi^{-3}\xi^{(-c-1)}\exp\left\{-\frac{d}{\xi}\right\}d\xi = \frac{c(c+1)\left(\frac{N}{2}+c+3\right)}{d^2}. \tag{A.32}$$

To find the other components of the matrix, we compute $\mathbf{H}_\xi^\theta(\theta', \xi) = (\mathbf{H}_\xi^\theta(\xi, \theta'))^T$, which consists of $\mathbf{H}_\xi^\theta(\gamma, \xi)$ and $\mathbf{H}_\xi^\theta(\mathbf{x}, \xi)$. From Lemma 2, $\mathbf{H}_\xi^\theta(\gamma, \xi) = \mathbf{0}_{L \times 1}$. Using the definition of $\mathbf{H}_\xi^\theta(\mathbf{x}, \xi)$, we see that $\mathbf{H}_\xi^\theta(\mathbf{x}, \xi) = (\mathbf{H}_\xi^\theta(\xi, \mathbf{x}))^T = \mathbf{0}_{L \times 1}$, from (A.30) and since $p_\Xi(\xi)$ is not a function of x_i. Thus, we obtain the FIM given by (2.21).

A.6   Proof of Theorem 4

The proof of Theorem 3 requires the regularity condition to be satisfied when the noise variance is unknown. In (A.19), we showed that the log likelihood function satisfies the regularity condition when $\theta = [\gamma]$. In this subsection, we show that the log likelihood $\log(p_{\mathbf{Y}, \gamma, \xi}(\mathbf{y}, \gamma, \xi))$ in (2.3) satisfies the regularity condition w.r.t. ξ. Differentiating the log likelihood w.r.t. ξ and taking $-\mathbb{E}_{\mathbf{Y}; \gamma, \xi}(\cdot)$ on both the sides of the equation,

$$
\begin{aligned}
\frac{\partial}{\partial \xi} \log(p_{\mathbf{Y}; \gamma, \xi}(\mathbf{y}, \gamma, \xi)) &= \frac{1}{2} \frac{\partial}{\partial \xi} (-\log|\mathbf{\Sigma}_y| - \mathbf{y}^T \mathbf{\Sigma}_y^{-1} \mathbf{y}) \\
&= -\frac{1}{2} \left[\mathrm{Tr}(\mathbf{\Sigma}_y^{-1}) - \mathrm{Tr}(\mathbf{y}\mathbf{y}^T(\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_y^{-1})) \right],
\end{aligned}
\tag{A.33}
$$

$$
\begin{aligned}
\mathbb{E}_{\mathbf{Y}; \gamma, \xi} &\left[\mathrm{Tr}(-\tfrac{1}{2}\mathbf{\Sigma}_y^{-1}) + \tfrac{1}{2}\mathrm{Tr}(\mathbf{y}\mathbf{y}^T(\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_y^{-1})) \right] \\
&= \frac{1}{2} \left[\mathrm{Tr}(\mathbf{\Sigma}_y^{-1}) - \mathrm{Tr}(\mathbf{\Sigma}_y^{-1}) \right] = 0.
\end{aligned}
\tag{A.34}
$$

Hence, the regularity condition is satisfied. From (A.21), we have $(\mathbf{M}_\xi^\theta(\gamma))_{ij} = -\frac{(\Phi_j^T \mathbf{\Sigma}_y^{-1} \Phi_i)^2}{2}$. To obtain $\mathbf{M}_\xi^\theta(\xi)$, we differentiate (A.33) w.r.t. ξ to obtain

$$
\frac{\partial^2}{\partial \xi^2}(\log p_{\mathbf{Y}; \gamma, \xi}(\mathbf{y}; \gamma, \xi)) = \frac{1}{2}\mathrm{Tr}(\mathbf{\Sigma}_y^{-2}) - \mathrm{Tr}(\mathbf{y}\mathbf{y}^T(\mathbf{\Sigma}_y^{-3})).
\tag{A.35}
$$

Taking $-\mathbb{E}_{\mathbf{Y};\gamma,\xi}(\cdot)$ on both sides of the above equation,

$$\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi) = -\mathbb{E}_{\mathbf{Y};\gamma,\xi}\left[\tfrac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) - \mathrm{Tr}(\mathbf{y}\mathbf{y}^T\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-3}))\right]$$

$$= \mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) - \tfrac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) = \tfrac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}). \qquad (A.36)$$

The vector $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\gamma,\xi)$ is found by differentiating (A.18) w.r.t. ξ and taking the negative expectation:

$$(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\gamma,\xi))_i = \mathbb{E}_{\mathbf{Y};\gamma,\xi}\left[\tfrac{\partial}{\partial\xi}\left(\tfrac{\Phi_i^T\boldsymbol{\Sigma}_y^{-1}\Phi_i - \Phi_i^T\boldsymbol{\Sigma}_y^{-1}\mathbf{y}\mathbf{y}^T\boldsymbol{\Sigma}_y^{-1}\Phi_i}{2}\right)\right]$$

$$= \frac{1}{2}\Phi_i^T\boldsymbol{\Sigma}_y^{-2}\Phi_i. \qquad (A.37)$$

Since $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi,\gamma) = (\mathbf{M}_\xi^{\boldsymbol{\theta}}(\gamma,\xi))^T$, the ith term of $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi,\gamma))_i = \tfrac{1}{2}\Phi_i^T\boldsymbol{\Sigma}_y^{-2}\Phi_i$. The MCRB $\mathbf{M}_\xi^{\boldsymbol{\theta}}$ can now be obtained by combining the expressions in (A.21), (A.36) and (A.37); this completes the proof.

Bibliography

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[2] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," in *Picture Coding Symposium*, vol. 1, 2006.

[3] C. Deng, W. Lin, B.-s. Lee, and C. T. Lau, "Robust image compression based on compressive sensing," in *Proc. of ICME*, 2010, pp. 462–467.

[4] S. Aviyente, "Compressed sensing framework for eeg compression," in *Proc. of SSP*, 2007, pp. 181–184.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] D. Venkatraman and A. Makur, "A compressive sensing approach to object-based surveillance video coding," in *Proc. ICASSP*, 2009, pp. 3513–3516.

[7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[8] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. ICASSP*. IEEE, 2008, pp. 3469–3472.

[9] R. Prasad and C. Murthy, "Bayesian learning for joint sparse OFDM channel estimation and data detection," in *Proc. Globecom*, 2010, pp. 1–6.

[10] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[11] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[12] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[13] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[14] M. Fornasier and H. Rauhut, "Iterative thresholding algorithms," *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.

[15] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[16] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[17] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.

[18] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[19] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, 2007.

[20] I. Daubechies, M. Fornasier, and I. Loris, "Accelerated projected gradient method for linear inverse problems with sparsity constraints," *Journal of Fourier Analysis and Appl.*, vol. 14, no. 5-6, pp. 764–792, 2008.

[21] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, 1997.

[22] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Let.*, vol. 14, no. 10, pp. 707–710, 2007.

[23] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. ICASSP*, 2008, pp. 3869–3872.

[24] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, 2010.

[25] H. V. Pham, W. Dai, and O. Milenkovic, "Sublinear compressive sensing reconstruction via belief propagation decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 674–678.

[26] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of compu. mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[27] P. Garrigues and L. E. Ghaoui, "An homotopy algorithm for the LASSO with online observations," in *Advances in NIPS*, 2009, pp. 489–496.

[28] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted balls," *Signal Processing, IEEE Transactions on*, vol. 59, no. 3, pp. 936–952, 2011.

[29] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. ICIP*, 2008, pp. 893–896.

[30] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1-norm," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3436–3447, 2010.

[31] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4013–4025, 2010.

[32] P. M. Williams, "Bayesian regularisation and pruning using a Laplace prior," *Neural Computation*, vol. 7, pp. 117–143, 1994.

[33] M. E. Tipping, "The relevance vector machine," in *Advances in NIPS*, vol. 12, 2000.

[34] ——, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–214, 2001.

[35] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[36] ——, "Comparison of approximate methods for handling hyperparameters," *Neural computation*, vol. 11, no. 5, pp. 1035–1068, 1999.

[37] Y. Grandvalet, "Least absolute shrinkage is equivalent to quadratic penalization," in *Proc. of ICANN*. Springer, 1998, pp. 201–206.

[38] D. Wipf, J. Palmer, B. Rao, and K. Kreutz-Delgado, "Performance evaluation of latent variable models with sparse priors," in *Proc. ICASSP*, vol. 2, 2007, pp. 448–453.

[39] T. Park and G. Casella, "The Bayesian LASSO," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

[40] M. W. Seeger, "Bayesian inference and optimal design for the sparse linear model," *The Journal of Machine Learning Research*, vol. 9, pp. 759–813, 2008.

[41] S. Mohamed, K. Heller, and Z. Ghahramani, "Bayesian and ℓ_1 approaches to sparse unsupervised learning," in *Proc. of ICML*, 2012.

[42] T. J. Mitchell and J. J. Beauchamp, "Bayesian variable selection in linear regression," *Journal of the American Statistical Association*, vol. 83, no. 404, pp. 1023–1032, 1988.

[43] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 2168–2172.

[44] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE J. Sel. Topics Signal Process.*, vol. 61, no. 2, pp. 340–354, 2013.

[45] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, 2010.

[46] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Info. Th. and Appli. Workshop*, 2008, pp. 326–333.

[47] E. G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, no. 2, p. 451, 2007.

[48] M. Masood and T. Al-Naffouri, "Sparse reconstruction using distribution agnostic Bayesian matching pursuit," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5298–5309, Nov 2013.

[49] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[50] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. ICIP*, 2009, pp. 1393–1396.

[51] H. Jiang, W. Deng, and Z. Shen, "Surveillance video processing using compressive sensing," *Inverse Problems & Imaging*, vol. 6, no. 2, 2012.

[52] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. of CVPR*, 2008, pp. 1–8.

[53] J. Provost and F. Lesage, "The application of compressed sensing for photo-acoustic tomography," *IEEE Trans. on Med. Imag.*, vol. 28, no. 4, pp. 585–594, 2009.

[54] D. J. Wilkinson, "Bayesian methods in bioinformatics and computational systems biology," *Briefings in bioinformatics*, vol. 8, no. 2, pp. 109–116, 2007.

[55] R. Pique-Regi, J. Monso-Varona, A. Ortega, R. C. Seeger, T. J. Triche, and S. Asgharzadeh, "Sparse representation and bayesian detection of genome copy number alterations from microarray data," *Bioinformatics*, vol. 24, no. 3, pp. 309–318, 2008.

[56] O. Williams, A. Blake, and R. Cipolla, "Sparse Bayesian learning for efficient visual tracking," *Trans. on Pat. Analy. and Machine Intelligence*, pp. 1292–1304, 2005.

[57] ——, "A sparse probabilistic learning algorithm for real-time tracking," *Proc. of Intl. Conf. on Computer Vision*, pp. 353–360, 2003.

[58] D. Wipf, J. Owen, H. Attias, K. Sekihara, and S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *NeuroImage*, vol. 49, no. 1, pp. 641–655, 2010.

[59] D. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," *Proc. Intl. Conf. on Machine Learning*, pp. 1023–1030, 2007.

[60] J. Zyren and W. McCoy, "Overview of the 3GPP long term evolution physical layer," *Freescale Semiconductor, Inc., white paper*, 2007.

[61] "Universal Mobile Telecommunications System (UMTS), Selection procedures for the choice of radio transmission technologies of the UMTS," ETSI technical report, UMTS 21.01 version 3.0.1, Nov., 1997.

[62] D. Tse, and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[63] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, 2011.

[64] L. Hanzo, M. Munster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-user communications, WLANs and Broadcasting*. John Wiley, 2003.

[65] S. S. Ghassemzadeh, L. J. Greenstein, T. Sveinsson, A. Kavcic, and V. Tarokh, "UWB delay profile models for residential and commercial indoor environments," *IEEE Trans. Veh. Tech.*, vol. 54, no. 4, pp. 1235–1244, 2005.

[66] J. Van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Borjesson, "On channel estimation in OFDM systems," in *Proc. VTC*, vol. 2, 1995, pp. 815–819.

[67] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, 2002.

[68] D. Meng, "Approximate message passing for multi-carrier transmission over doubly selective channels," Ph.D. dissertation, Ohio State University, 2012.

[69] N. L. Pedersen, C. N. Manchón, D. Shutin, and B. H. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *Proc. ICC*, 2012, pp. 3487–3492.

[70] A. P. Kannu and P. Schniter, "On communication over unknown sparse frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6619–6632, 2011.

[71] ITU-R Rec. M.1225, "Guidelines for evaluation of radio transmission technologies (RTTs) for IMT-2000," 1997.

[72] X. Ma, H. Kobayashi, and S. Schwartz, "An EM-based estimation of OFDM signals," in *Proc. WCNC*, 2002, pp. 228–232.

[73] T.Y. Al Naffouri, A. Bahai, and A. Paulraj, "An EM-based OFDM receiver for time-variant channels," in *Proc. Globecom*, vol. 1, 2002.

[74] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 315–325, 2004.

[75] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, 2010.

[76] C. Wu and D. Lin, "A group matching pursuit algorithm for sparse channel estimation for OFDM transmission," in *Proc. ICASSP*, vol. 4, 2006.

[77] S. Barembruch, E. Moulines, and A. Scaglione, "A sparse EM algorithm for blind and semi-blind identification of doubly selective OFDM channels," in *Proc. SPAWC*, 2010, pp. 1–5.

[78] J. Ziniel, L. C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, 2010, pp. 808–812.

[79] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[80] F. Wan, W.-P. Zhu, and M. Swamy, "Semiblind sparse channel estimation for mimo-ofdm systems," *IEEE Trans. Veh. Tech.*, vol. 60, no. 6, pp. 2569–2582, 2011.

[81] A. Gilbert and J. Tropp, "Applications of sparse approximation in communications," in *Proc. IEEE Int. Symp. Inf. Theory*, 2005, pp. 1000–1004.

[82] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *Arxiv preprint arXiv:1107.1339*, 2011.

[83] Y. Peng, X. Yang, X. Zhang, W. Wang, and B. Wu, "Compressed mimo-ofdm channel estimation," in *Proc. of ICCT*, 2010, pp. 1291–1294.

[84] M. Khojastepour, K. Gomadam, and X. Wang, "Pilot-assisted channel estimation

for mimo ofdm systems using theory of sparse signal recovery," in *Proc. ICASSP*, 2009, pp. 2693–2696.

[85] C. Qi and L. Wu, "A hybrid compressed sensing algorithm for sparse channel estimation in mimo ofdm systems," in *Proc. ICASSP*, 2011, pp. 3488–3491.

[86] D. Eiwen, G. Taubock, F. Hlawatsch, and H. Feichtinger, "Group sparsity methods for compressive channel estimation in doubly dispersive multicarrier systems," in *Proc. SPAWC*, 2010, pp. 1–5.

[87] G. Gui, N. Zheng, N. Wang, A. Mehbodniya, and F. Adachi, "Compressive estimation of cluster-sparse channels." *Progress In Electromagnetics Research C*, vol. 24, 2011.

[88] R. Prasad and C. Murthy, "Bayesian learning for joint sparse OFDM channel estimation and data detection," in *Proc. Globecom*, 2010, pp. 1–6.

[89] P. Ranjitha and C. R. Murthy, "Cramér-Rao-type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, 2013.

[90] R. Prasad, C. R. Murthy, B. D. Rao, "Joint approximately sparse channel estimation and data detection in ofdm systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, July 2014.

[91] ——, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *Submitted to IEEE Trans. Signal Process.*, 2014.

[92] R. Prasad and C. R. Murthy, "Joint channel estimation and data detection in

mimo-ofdm systems using sparse Bayesian learning," in *Proc. Nat. Conf. Commun. (NCC)*, Feb 2014, pp. 1–6.

[93] V. Cevher, "Learning with compressible priors," *Advances in NIPS*, pp. 7–12, 2008.

[94] R. Gribonval, V. Cevher, and M. Davies, "Compressible priors for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.

[95] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. of Artificial Intelligence and Statistics*, vol. 1, no. 3, 2003.

[96] D. Wipf and S. Nagarajan, "A new view of automatic relevance determination," *Advances in NIPS*, vol. 20, pp. 1625–1632, 2008.

[97] ——, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *Proc. of SPARS*, 2009.

[98] R. Prasad and C. R. Murthy, "Bayesian learning for joint sparse OFDM channel estimation and data detection," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2010, pp. 1–6.

[99] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, 2009.

[100] Z. Ben-Haim and Y. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, 2010.

[101] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér-Rao bound in noisy compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 518 –526, 2012.

[102] J. Helferty and D. Mmiscudgett, "Optimal observer trajectories for bearings only tracking by minimizing the trace of the Cramér-Rao lower bound," in *Proc. of Decision and Control*, 1993, pp. 936–939.

[103] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[104] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[105] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*.   Prentice Hall, 1993.

[106] H. V. Trees, *Detection, Estimation, and Modulation Theory - Part I*.   New York: Wiley, 1968, vol. 2.

[107] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations-part I: far-field sources," *Proc. ICASSP*, vol. 35, no. 3, pp. 286–299, 1987.

[108] J. Dauwels, "On graphical models for communications and machine learning: Algorithms, bounds, and analog implementation," Ph.D. dissertation, Swiss Federal Institute of Technology, Zürich, 2005. [Online]. Available: http://www.dauwels.com/Justin/PhD.htm

[109] D. Wipf, J. Palmer, and B. Rao, "Perspectives on sparse Bayesian learning," in *Advances in NIPS*, vol. 16, 2004.

[110] A. Armagan, D. Dunson, and J. Lee, "Generalized double Pareto shrinkage," *Arxiv preprint arXiv:1104.0861*, Aug. 2011.

[111] S. Balakrishnan and D. Madigan, "Priors on the variance in sparse Bayesian learning; the demi-Bayesian LASSO," *Online: http://www.stat.columbia.edu/ madigan/PAPERS/recentPapers.html*, 2009.

[112] S. J. Hwang and P. Schniter, "Efficient multicarrier communication for highly spread underwater acoustic channels," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 9, pp. 1674–1683, 2008.

[113] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[114] P. Schniter, "Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels," *Physical Commun.*, vol. 5, no. 2, pp. 91–101, 2012.

[115] G. Taubock and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. ICASSP*, 2008, pp. 2885 –2888.

[116] C. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing

to sparse channel estimation," *IEEE Comm. Magazine*, vol. 48, no. 11, pp. 164–174, 2010.

[117] K. Murphy, "Dynamic Bayesian networks: representation, inference and learning," Ph.D. dissertation, University of California, 2002.

[118] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, University of London, 2003.

[119] B. J. Frey and N. Jojic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *IEEE Trans. on Patt. Analysis and Machine Intelli.*, vol. 27, no. 9, pp. 1392–1416, 2005.

[120] T. Feng, T. Field, and S. Haykin, "Stochastic differential equation theory applied to wireless channels," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1478–1483, 2007.

[121] W. Chen and R. Zhang, "Kalman-filter channel estimator for OFDM systems in time and frequency-selective fading environment," in *Proc. ICASSP*, vol. 4, 2004, p. 377.

[122] Z. Yuanjin, "A novel channel estimation and tracking method for wireless OFDM systems based on pilots and Kalman filtering," *IEEE Trans. Consum. Electron.*, vol. 49, no. 2, pp. 275–283, 2003.

[123] T. Y. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 7, p. 3924, 2007.

[124] Z. Ghahramani and G. E. Hinton, "*Parameter estimation for linear dynamical systems*," Tech. Rep., 1996.

[125] Z. Zhang and B. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *Proc. ICASSP*, 2010, pp. 3986–3989.

[126] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[127] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*.   Wiley New York, 1997, vol. 274.

[128] B. Anderson and J. Moore, *Optimal filtering*.   Courier Dover Publications, 2005.

[129] R. Hunger, "Floating point operations in matrix-vector calculus," *Munich University of Technology, TUM-LNS-TR-05-05*, 2005.

[130] C. Studer, C. Benkeser, S. Belfanti, and Q. Huang, "Design and implementation of a parallel turbo-decoder ASIC for 3GPP-LTE," *IEEE J. Solid State Circuits.*, vol. 46, no. 1, pp. 8–17, 2011.

[131] Y. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, 2003.

[132] M. Wang, D. Quevedo, and G. Goodwin, "Joint data detection and channel estimation for MIMO-OFDM systems via EM algorithm and sphere decoding," in *Proc. Globecom*, Dec. 2006, pp. 1–5.

[133] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sens-
ing to sparse channel estimation," *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 164–
174, 2010.

[134] Z. Zhang and B. Rao, "Extension of SBL algorithms for the recovery of block
sparse signals with intra-block correlation," *IEEE J. Sel. Topics Signal Process.*,
vol. 61, no. 8, pp. 2009–2015, 2013.

[135] R. Prasad, C. Murthy, and B. D. Rao, "Nested sparse Bayesian learning for block-
sparse signals with intra-block correlation," in *Proc. ICASSP*, 2010, pp. 1–6.

[136] A. Juditsky, F. K. Karzan, A. Nemirovski, and B. Polyak, "On the accuracy of l1-
filtering of signals with block-sparse structure," *Advances in NIPS*, vol. 24, 2011.

[137] Z. Zhang, T. P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for
energy-efficient wireless telemonitoring of noninvasive fetal ecg via block sparse
bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 300–309, 2013.

[138] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse
signals with an optimal number of measurements," *IEEE J. Sel. Topics Signal Pro-
cess.*, vol. 57, no. 8, pp. 3075–3085, 2009.

[139] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union
of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[140] S. Negahban and M. J. Wainwright, "Joint support recovery under high-
dimensional scaling: Benefits and perils of $\ell_1 - \ell_\infty$-regularization," *Advances in
NIPS*, vol. 21, pp. 1161–1168, 2008.

[141] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[142] Z. Zhang and B. D. Rao, "Recovery of block sparse signals using the framework of block sparse Bayesian learning," in *Proc. ICASSP*, 2012, pp. 3345–3348.

[143] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 477–489, 1988.

[144] H. Akaike, "Fitting autoregressive models for prediction," *Annals of the institute of Statistical Mathematics*, vol. 21, no. 1, pp. 243–247, 1969.

[145] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, 1999.

[146] W. Wei, *Time series analysis*. Addison-Wesley Redwood City, California, 1994.

[147] D. A. Van Dyk, "Nesting EM algorithms for computational efficiency," *Statistica Sinica*, vol. 10, no. 1, pp. 203–226, 2000.

[148] K. Petersen and M. Pedersen, "The matrix cookbook," *Online: http://matrixcookbook.com*, 2008.