Structured sparse signal recovery algorithms and their applications

Wang, Lu

2014

Wang, L. (2014). Structured sparse signal recovery algorithms and their applications.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/61762>

<https://doi.org/10.32657/10356/61762>

STRUCTURED SPARSE SIGNAL RECOVERY ALGORITHMS AND THEIR APPLICATIONS

WANG LU

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2014

Acknowledgments

I would like to express my sincere thanks and appreciation to my supervisor Dr. Guoan Bi and co-supervisor, Dr. Chunru Wan (DSO) for their invaluable guidance, support and suggestions. Their knowledge, suggestions, and discussions help me to become a capable researcher. Their encouragement also helps me to overcome the difficulties encountered in my research.

I also want to thank my senior graduate fellow Dr. Xiaolei Lv for his generous help, technical suggestions. I am very grateful to the members of our team in NTU. It is a pleasure to collaborate with my team mates, Dr. Yong Xiao, Dr. Jiang Qian, Dr. Haijian Zhang, Dr. Junyang Pan, Dr. Lei Yang, Dr. Shan Luo, Mr. Lifan Zhao, Mr. Pu Yuan and Mr. Zhe Wang.

I thank all the members of Information System Research Lab for being great colleagues to work with, and in particular thanks to Ms. Ye Luo, Ms. Tzu-Yi Hung, Dr. Fuxi Wen, Dr. Maodong Li, Ms. Likun Huang, Dr. Tao Wang, Dr. Seong Ping Chuah, Dr. Wenwen Wang, Ms. Lili Yi, Mr. Yaotao Wang, Mr. Li Wang, Mr. Bing Wang and Mr. Xingxing Wang, for their kind support.

Last but not least, I want to thank my parents, my husband and other family members in China for their constant love and encouragement.

Wang Lu

Contents

Acknowledgments	I
Summary	vii
List of Figures	x
List of Tables	xiv
List of Abbreviations	xv
List of Notations	xvi
1 Introduction	1
1.1 Background and Motivations	1
1.1.1 Background	1
1.1.2 Motivations	3
1.2 Contributions of the Thesis	6
1.3 Organization of the Thesis	8
2 Literature Review	9
2.1 Methods for Standard Sparse Recovery	10
2.1.1 Convex Optimization based Methods	10

2.1.2	Iterative Greedy Algorithms	11
2.1.3	Sparse Bayesian Learning Algorithms	12
2.1.4	Summary	13
2.2	Methods for Structured Sparse Recovery	14
2.2.1	Group Lasso	15
2.2.2	BOGA	15
2.2.3	BSBL	16
2.2.4	CluSS	17
2.2.5	Summary	18
3	Improved Stability Conditions of BOGA for Noisy Block-Sparse Signals	21
3.1	Introduction	21
3.2	Preliminary Definitions and Results	24
3.3	Stability and Support Recovery of BOGA	25
3.4	Numerical Experiments	33
3.5	Conclusion	37
4	Harmonic Tonal Detection Architectures based on BOGA	39
4.1	Introduction	39
4.2	Signal Model	42
4.2.1	Harmonic Signal Model	42
4.2.2	Dictionary Construction	44

4.2.3	Block-sparse signal Model	46
4.3	Harmonic Signal Detection	47
4.3.1	BOGA Description	48
4.3.2	Architecture 1	49
4.3.3	Architecture 2	51
4.4	Discussion	53
4.4.1	The effect of β on the distribution of test statistic	53
4.4.2	Difference between the two proposed architectures	53
4.4.3	Model mismatch	53
4.5	Simulation Results	54
4.5.1	PMF of K' and Verification of Equations in (B.23)	54
4.5.2	Detection performance evaluation	56
4.6	Conclusion	63
5	Hierarchical Sparse Signal Recovery by Variational Bayesian Inference	65
5.1	Introduction	65
5.2	Signal Model and Recovery	67
5.2.1	Probabilistic Signal Model	67
5.2.2	Hierarchical Sparse Signal Recovery	69
5.3	Relation to Prior Work	71
5.4	Simulations	72

5.5	Conclusion	78
6	Enhanced ISAR Imaging by Exploiting the Continuity of the Target Scene	79
6.1	Introduction	79
6.2	ISAR Imaging	81
6.2.1	Signal model for ISAR imaging	81
6.2.2	Sparse probing frequencies	83
6.2.3	Maximal mutual coherence of $\mathbf{D}$	85
6.3	Continuity Patterns and Bayesian model	85
6.3.1	Continuity Patterns	86
6.3.2	Bayesian Model	87
6.4	Proposed Imagery Algorithm	91
6.4.1	Sparse Solution and the Gibbs Sampler	91
6.4.2	Discussion on the Proposed Imaging Algorithm	92
6.5	Experiments and Performance Comparisons	95
6.5.1	Dictionary Coherence Evaluation	95
6.5.2	Synthetic Experiments	96
6.5.3	Performance evaluation with respect to the target sparsity . .	103
6.5.4	Experiment with data set “Mig-25”	103
6.5.5	Practical Data Experiments with data set “Yak-42”	107
6.5.6	Discussion on the Results	111

6.6	Conclusion	111
7	Conclusions and Future Work	115
7.1	Conclusions	115
7.2	Future work	117
A	Proof of Lemma 1 in Chapter 3	119
B	Detection Performance Analysis in Chapter 4	121
B.1	Analysis on architecture 1	121
B.1.1	PDF of T_{arch1} under H_0 for fixed K'	121
B.1.2	PDF of T_{arch1} under H_1 for fixed K'	124
B.1.3	PDF of T_{arch1} under H_0 for Random K'	125
B.2	Analysis on architecture 2	127
B.2.1	K' is fixed	127
B.2.2	K' is random	128
B.3	Detection Probability of GLRT	130
B.4	Detection Probability of Energy Detector (ED)	131
C	Maximal Coherence of the Dictionary D in Chapter 6	133
	Author's Publications	135
	References	137

Summary

Getting more sophisticated, the theory of sparse representation (SR) and the successful SR based applications in various aspects of signal processing have been extensively investigated and discussed over the past several decades. Besides sparsity, underlying structures of the signal have been considered and exploited recently to enhance the performance of the standard SR recovery methods. While a large number of recovery algorithms have been proposed to incorporate different signal structures in the literature and have achieved reasonably good performance, there is still a need for further improvement and extending them for more real world applications. This thesis focuses on various issues of structured sparse representations, including theoretical performance analysis for certain existing structured sparse recovery methods, sparse recovery algorithm developments for signal with specific structure and their applications.

One of the most commonly exploited structures in the literature is the block sparsity. For signals with block sparsity, a greedy recovery method, known as block orthogonal greedy algorithm (BOGA), is attractive due to its computational efficiency. BOGA has been proven to successfully recover block sparse signals in noiseless environments and outperforms its standard counterpart, i.e., orthogonal greedy algorithm (OGA) in terms of achieving smaller mean square error (MSE) by using less measurements. The associated stability problem dealing with noisy signals is re-studied in this thesis. It is demonstrated that the recovery conditions of the BOGA previously reported can be relaxed by using a different definition of the block-coherence. The theoretical results obtained provide a generalization of those reported by Tseng for block-sparse signal and serve as a complement of the BOGA

reported by Eldar for noisy signals. Motivated by the success of the BOGA, new detection architectures for harmonic tonals are proposed for tonal detection in low signal to noise ratio (SNR) environments. The distributions of the test statistics of detection architectures under the orthogonal dictionary are analyzed theoretically and comprehensively based on the theory of order statistics for the first time. Detection performances are also analyzed and compared theoretically and experimentally. Significant improvements on detection performance in low SNR environments are shown over the conventional detectors that do not consider the block structure.

As a natural generation of block sparse signal, hierarchical sparse signal, exhibiting two levels of sparsity, i.e., block sparsity among different blocks and internal sparsity within individual block, is of interest. It is mainly recovered by solving a convex optimization problem with two user-selected regularization parameters. To avoid the subjective parameter selection, a recovery method is developed under a Bayesian framework. To enforce the two-level hierarchical sparsity, the variance of the sparse coefficient is modeled by the product of two kinds of random variables controlling the block sparsity and the internal sparsity, respectively. Under the proposed probabilistic model, variational Bayesian inference is used to recover the coefficient vector from the noise corrupted data. Numerical simulation and experimental results show that the proposed method outperforms those state-of-the-art recovery methods for hierarchical sparse signals.

Both the block sparse and the hierarchical sparse models require a known block partition which is often unrealistic. For real world applications, nonzero elements of the sparse coefficient vector may cluster randomly with unknown locations and sizes. We propose to impose this spatial cluster structure statistically in a Bayesian framework and successfully apply it to Inverse Synthetic Aperture Radar (ISAR) imaging by using the concept of spatial continuity. Following the Bayesian compressive sensing theory, a hierarchical Bayesian prior is imposed on the scatterers on the range-Doppler plane to encourage the sparsity. Then a correlated prior is used to statistically encourage a two dimensional cluster structure of the scatterers in the target region. However, the posterior distribution is intractable under

the formulated probabilistic model. To overcome this difficulty, the Gibbs sampling strategy is used for Bayesian inference and the parameters of the signal model are inferred efficiently from samples obtained by the Gibbs sampler. Both the synthetic and the experimental results demonstrate that the proposed algorithm exploits efficiently the continuity of the target scene and can achieve enhanced ISAR images in the scenarios of limited measurements and low SNR compared with other reported algorithms for ISAR imaging problems.

List of Figures

3.1	Representation errors obtained by using the BOGA and the OGA for $K=1, 2, 3$ and 4. Dotted lines represent the bounds of Theorem 5.1 in [1]; solid lines depict the bounds of Theorem 1 in [2]; Cross lines depict the bounds of Theorem 1 in this Chapter.	34
3.2	Signal errors achieved by the BOGA and the OGA for $K =1, 2, 3$ and 4. Dotted lines represent the input noise power.	35
3.3	Support recovery success achieved by the BOGA and OGA for $K=1, 2, 3$ and 4.	36
4.1	Density function comparisons. The histograms are obtained by the Monte Carlo method and the dash-dot curves are the theoretical density functions.	57
4.2	Detection performances of architecture 1 for different values of β . . .	58
4.3	Detection performance of proposed architectures.	61
4.4	Detection performance comparisons when $K = 1$. “arch1 _f ” refers to detection architecture 1 with K' fixed to 1; “arch1 _r ” refers to detection architecture 1 with K' randomly determined by the stopping criterion $\ \mathbf{r}_t \ _2^2 \leq \beta N \sigma^2$ with $\beta = 1.1$; similarly, “arch2 _f ” and “arch2 _r ” refer to detection architecture 2 with fixed and randomly K' , respectively.	62

4.5	Detection performance comparisons when $K = 2$	62
4.6	Detection performance comparisons when $K = 1$	63
4.7	False alarm probability of architecture 2	63
4.8	Performance loss evaluation due to model mismatch	64
5.1	Recovered signals by various methods.	73
5.2	The MSEs of reconstructed signal versus the number of measurements.	74
5.3	The MSEs of reconstructed signal versus internal sparsity.	75
5.4	The MSEs of reconstructed signal versus noise power.	75
5.5	Performance of variable selection.	77
6.1	The geometry of the radar imaging of a target when $t = 0$	82
6.2	Cross-range and range domain with a designed resolution.	84
6.3	Continuity structure illustration by Yak-42.	87
6.4	Three kinds of continuity patterns	87
6.5	Maximal mutual coherence versus the number of measurements (a) $B/F_c = 0.15$, (b) $B/F_c = 0.2$	96
6.6	The scatterer locations	97
6.7	ISAR images achieved by different methods: (a) The target scene, (b) The RDA with full measurements; (c) The RDA using SPF; (d) The BP using SPF; (e) The BCS using SPF; (f) The proposed method using SPF.	98

6.8	ISAR image comparison in terms of the number of SPF. (a) The RDA using 1/4 of measurements; (b) The RDA using 1/8 of measurements; (c) The BP using 1/4 of measurements; (d) The BP using of 1/8 measurements; (e) The proposed method using of 1/4 measurements; (f) The proposed method using of 1/8 measurements.	100
6.9	The converge curve.	101
6.10	The NMSEs: (a) against the number of measurements (noise power =0.2), (b) against the noise power with 10 measurements in the range domain.	102
6.11	Performance of the proposed method versus the degree of the target sparsity. The first row presents the target scenes with increased degree of target sparsity; the second row shows the images obtained by the RDA; the third row presents the images obtained by the proposed method; and the last row indicates the degree of the target sparsity and the NMSE of the results obtained by the proposed method. . . .	104
6.12	The NMSE versus the degree of the target sparsity.	105
6.13	Comparison of ISAR images against number of SPFs achieved by the (a) RDA with full measurements; (b) RDA using 1/2 of measurements; (c) BP using 1/2 of measurements; (d) proposed method using 1/2 of measurements; (e) RDA with full measurements ($SNR = 7.5\text{dB}$); (f) RDA using 1/4 of measurements; (g) BP using 1/4 of measurements; (h) proposed method using 1/4 of measurements. . . .	106
6.14	The ISAR image of Yak-42 airplane obtained by the RDA with full measurements.	108
6.15	ISAR images versus the number of SPFs obtained by the (a) RDA using 1/2 of measurements; (b) RDA using 1/4 of measurements; (c) BP using 1/2 of measurements; (d) BP using 1/4 of measurements; (e) proposed method using 1/2 of measurements; (f) proposed method using 1/4 of measurements.	109

6.16 ISAR images comparison with different noise powers. (a) Image achieved by RDA under noise ($SNR = 11.3\text{dB}$); (b) Image achieved by RDA ($SNR = 5.3\text{dB}$); (c) Image achieved by RDA by using 1/2 of measurements ($SNR = 11.3\text{dB}$); (d) Image achieved by RDA using 1/2 of measurements ($SNR = 5.3\text{dB}$); (e) Image achieved by BP by using 1/2 of measurements ($SNR = 11.3\text{dB}$); (f) Image achieved by BP using 1/2 of measurements ($SNR = 5.3\text{dB}$); (g) Image achieved by proposed method using 1/2 of measurements ($SNR = 11.3\text{dB}$); (h) Image achieved by proposed method using 1/2 of measurements ($SNR = 5.3\text{dB}$).	113
--	-----

List of Tables

2.1	Summary of the representative methods for standard sparse recovery	14
3.1	Bounds on minimal block energy versus noise power	33
4.1	Estimated PMF of K'	55
5.1	Performance evaluation for variable selection	77
5.2	The MSEs of the recovered digits	78
6.1	System parameters for data “Yak-42”	86
6.2	Comparison of the computational complexity	94
6.3	System parameters for the simulated data	97
6.4	The required CPU time (s)	102
6.5	System model parameters for data “Mig-25”	105
6.6	Performance evaluation by Corr and Entropy	111
C.1	The maximal coherence	133

List of Abbreviations

BCS	Bayesian Compressive Sensing
BOGA	Block Orthonormal Greedy Algorithm
BP	Basic Pursuit
BSBL	Block Sparse Bayesian Learning
Cluss	Clustered Sparse Solver
CoSaMP	Compressive Sampling Matching Pursuit
CS	Compressive Sensing
DFT	Discrete Fourier Transform
EM	Expectation Maximization
FOCUSS	FOCal Under-determined System Solver
GMF	Generalized Matched Filter
GLRT	Generalized Likelihood Ratio Test
HiLasso	Hierarchical Lasso
ISAR	Inverse Synthetic Aperture Radar
Lasso	Least Absolute Shrinkage and Selection Operator
MCMC	Monte Carlo Markov Chain
ML(E)	Maximum Likelihood (Estimator)
MP	Matching Pursuit
MRF	Markov random field
MSE	Mean Square Error
OGA	Orthonormal Greedy Algorithm
OMP	Orthonormal Matching Pursuit
RDA	Range-Doppler Algorithm
SAR	Synthetic Aperture Radar
SBL	Sparse Bayesian Learning
SNR	Signal to Noise Ratio
SPF	Sparse Probing Frequency
SR	Sparse Representation
StOMP	Stage-wise Orthogonal Matching Pursuit
RIP	Restricted Isometry Property
VB(I)	Variational Bayes Inference

List of Notations

$(\cdot)^T$	Matrix or vector transpose
$(\cdot)^H$	Matrix or vector conjugate transpose
$\mathbf{X}^{-1}$	Matrix inverse and pseudo-inverse
$\mathbf{X}^\dagger$	Matrix pseudo-inverse
$\circ$	Elementary multiplication
$ x $	The absolute value of scalar x
$ \mathbf{x} $	The modulus of vector $\mathbf{x}$
$\ \mathbf{X}\ _p$	Matrix p -norm defined as $\max_{\ \mathbf{u}\ _p=1} \ \mathbf{X}\mathbf{u}\ _p$
$\ \mathbf{x}\ _0$	l_0 -norm of $\mathbf{x}$ which counts the number of nonzero components in $\mathbf{x}$
$\ \mathbf{x}\ _1$	l_1 -norm of $\mathbf{x}$ defined as $\sum_n x_n $
$\ \mathbf{x}\ _2$	l_2 -norm of $\mathbf{x}$ defined as $\sqrt{\sum_n x_n^2}$
$\ \mathbf{x}\ _\infty$	l_∞ -norm of $\mathbf{x}$ defined as $\max(x_1 , \dots, x_N)$
$\ \mathbf{x}\ _{2,1}$	The mixed l_2/l_1 -norm defined as $\sum_{l=1}^L \ \mathbf{b}[l]\ _2$ for $\mathbf{x}$ with L blocks
$\ \mathbf{x}\ _{2,\infty}$	The mixed l_2/l_∞ -norm defined as $\max_{1 \leq l \leq L} \{\ \mathbf{b}[l]\ _2\}$ for $\mathbf{x}$ with L blocks
$\lambda_{\max}(\mathbf{X}^H \mathbf{X})$	The largest eigenvalue of matrix $\mathbf{X}^H \mathbf{X}$
$\rho(\mathbf{X})$	The spectral norm of matrix $\mathbf{X}$ defined as $\rho(\mathbf{X}) = \sqrt{\lambda_{\max}(\mathbf{D}^H \mathbf{D})}$
$\text{diag}(\mathbf{x})$	Diagonal matrix whose diagonal elements are the elements of $\mathbf{x}$
$\ \mathbf{x} - \mathbf{u}\ ^2$	Euclidian distance between vectors $\mathbf{x}$ and $\mathbf{u}$
$\langle \mathbf{x}, \mathbf{y} \rangle$	Inner product of $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{x} \in \mathcal{R}^{N \times 1}$	A vector $\mathbf{x}$ of length N with real elements
$\mathbf{x} \in \mathcal{C}^{N \times 1}$	A vector $\mathbf{x}$ of length N with complex elements
$\bigcup_{i=1}^L \mathbf{D}[i]$	A concatenation of sub-matrix $\mathbf{D}[i]$ for $1 \leq i \leq L$
$\Gamma(\mathbf{x})$	The support of vector $\mathbf{x}$
$\Gamma_l \subseteq \Gamma(\mathbf{x})$	The support Γ_l is contained in the support of $\mathbf{x}$
$\Gamma \cup i$	New support by including the index i
$ \Gamma(\mathbf{x}) $	The cardinality of the support of $\mathbf{x}$
$\mathbf{x} \setminus x_i$	The sub-vector of $\mathbf{x}$ excluding the element x_i
$x \sim q(a, b)$	Random variable x follows distribution q with parameter a and b
$\langle x \rangle_q$	The expectation of variable x with respect to distribution q
$\exp(\cdot)$	The exponential function

Chapter 1

Introduction

1.1 Background and Motivations

1.1.1 Background

Sparse signal recovery from the noisy observation is attracting growing attention in many applications, such as audio and music retrieval, image denoising and compression, reconstruction of Radar target image, event detection in Radar and Sonar, etc.. The standard sparse signal model assumes that signals can be efficiently represented by a linear combination of a small number of atoms from an over-complete dictionary. In other word, the number of the nonzero components of the coefficient vector in the representation is significantly smaller than the number of atoms in the dictionary. The recovery problem is under-determined due to the over-completeness of the dictionary. Finding a sparse representation (SR) of the signal with a small number of active atoms from the over-complete dictionary has been shown to be a NP-hard problem [3].

This has led to considerable effort being put into the development of many sub-optimal schemes. In general, the signal can be reconstructed by using three kinds of strategies to find its parsimonious representation. One strategy is to solve a minimization problem that recovers all the coefficients simultaneously, known as convex optimization techniques, such as l_2 -norm based FOCal Under-determined System

Solver (FOCUSS) [4] and l_1 -norm based Basis Pursuit (BP) [5] or Least Absolute Shrinkage and Selection Operator (Lasso) [6]. Another kind of methods to achieve sparse solutions is iterative greedy algorithms that iteratively build up the signal approximation one coefficient at a time, such as matching pursuit (MP) [7], and its variations [8, 9], including orthonormal greedy algorithm (OGA, i.e., orthonormal MP). These greedy algorithms find the support set of the sparse signals based on the correlation values between the atoms in the over-complete dictionary and the observation. Analysis of the recovery conditions in noiseless environments [10–12] and the stability property in noisy environments [1, 13, 14] have been reported. Last but not the least, the third kind of methods statistically solve the problem by using the Bayesian inference and is collectively known as sparse Bayesian Learning (SBL) or Bayesian Compressive Sensing (BCS) in applications of compressive sensing (CS). Comparing to the two aforementioned formulations, SBL based methods provide several advantages including giving probabilistic predictions, flexible incorporation and estimation of model parameters, and the uncertainty estimation of reconstruction. The SBL is firstly proposed in the field of machine learning [15], further analyzed in [16] and adapted for the compressive sensing in [17]. Those methods differ from each other on the aspects of priors imposing the sparsity [18–23] and the statistical tools [15, 16] for unknown inference. The efficient inferences of the model parameters are usually performed by the iterative learning methods such as Expectation Maximization (EM) algorithm [17], variational Bayesian (VB) inference [24, 25] and stochastic sampling methods [26, 27] such as Monte Carlo Markov chain (MCMC) sampling.

Those recovery methods are designed for the standard SR, where no dependency has been assumed or considered in finding the sparse representation. However, elements of the sparse coefficient may behave according to some deterministic or probabilistic inter-rules, i.e., they depend on each other in most real applications, since natural signals always show some structures and physical properties which in turn exert links among the sparse coefficients [28]. Most recently, structural constraints on the signal support have been added to enhance the sparse recovery performance. In contrast to the standard sparse signal whose components are treated

independently, the structured sparse representation problem takes the underlying signal structure into consideration.

Different algorithms for exploiting different types of signal structure have been developed from perspective of the three strategies mentioned above, respectively. A commonly exploited structure, the block or group sparsity, is considered in [29] where the coefficients in one of the pre-partitioned blocks tend to be zero or nonzero simultaneously. The convex optimization strategy based group Lasso is proposed in [30] and further analyzed in [31] to recover the signal with block sparsity. Parallely, block orthonormal greedy algorithm (BOGA), a direct extension of OGA, is proposed in [32] and further analyzed in [2, 33]. Block sparse signal recovery has also been considered in the Bayesian framework in [34] where block sparse Bayesian learning (BSBL) method is proposed. The above mentioned block sparse recovery methods have a common assumption that the block partition is known a priori. Work in [35–37] employs the neighborhood system to enforce the nonzero components to appear in clusters with their locations and sizes inferred automatically from data. Another widely exploited structure in wavelet decomposition is tree structured sparsity [25, 26] where the structure is imposed in a probabilistic way and the unknowns are given by either the Monte Carlo Markov chain (MCMC) or mean-field approximation method. More general structural sparsity is considered in [28, 38–40] where Boltzmann machine is used to model the structure which allows dependencies between distant atoms.

1.1.2 Motivations

Even though a range of empirical evidence is corroborated for the performance of methods built on sparse and structured sparse representation, there are still some fundamental theoretical and application questions remaining to be addressed:

- 1) As mentioned above, the block sparsity is put forth in [29] and the BOGA algorithm is proposed to deal with the signal with block sparsity [32]. Stable recovery conditions of BOGA under noiseless case are analyzed in [32] and further extended

to noisy case in [2]. Both of the derivations in [2] and [32] are based on the maximal coherence and block-coherence of dictionary. It is found both in [2] and [32] that the performance of the BOGA deteriorates as the maximal block-coherence of the dictionary increases. However, it is noted that basic definitions of the maximal block-coherence of the dictionary are different in [2] and [32] and the definition in [32] always gives a smaller maximal block-coherence value than that in [2] for the same dictionary. Therefore, it is expected that more positive results on the stability of the BOGA in noisy environment may be obtained with the definition of [32]. How can we incorporate that definition in our derivations? How much improvement can we achieve when using that definition?

2) In sonar application, the radiated noise from ship or submarine inherently has a harmonic line spectrum, i.e., a series of discrete lines or tonals that are harmonically related, due to the rotating and periodic moving machinery parts, such as the shaft and blades, gear box and firing engine. Transforming the signal into its discrete Fourier spectrum, it is found that tonal signal is of a block sparsity, where nonzero blocks correspond to the fundamental frequencies and their higher order harmonics. It has been shown in [41, 42] that improved harmonic tonal recovery performance can be achieved by exploiting the block-sparsity. Motivated by the success of BOGA in recovery of the signal with block sparsity, we are interested in how to incorporate the BOGA in harmonic tonal detection. Detectors for harmonic tonals by using the BOGA are to be designed and their theoretical performances are to be analyzed. The performance comparisons with that of the other existing conventional detectors are to be made.

3) As a direct generation of block sparse model, hierarchical sparse model is first formally put forth in [43] where signals typically exhibit two levels of sparsity, i.e., block-sparsity among different blocks and internal sparsity inside an individual block. The recovery algorithms for this kind of signal are successively proposed in [42–44] which mainly focus on solving a convex optimization problem with two user-selected regularization parameters. The selection of the parameters are subjective and some parameters always lead to a poor recovery. To make a good selection of

regularization parameters, cross-validation is usually performed which is extremely time-consuming. Since the two regularization parameters control the block sparsity and the internal sparsity, respectively, the balance of their values is of great importance in accurate signal recovery, but is hard to be adjusted. How to simultaneously avoid the subjective selections of those parameters and correctly regularize the block and internal sparsity to recover the hierarchical sparse signal is to be investigated.

4) SR methods have been successfully used in inverse synthetic aperture radar (ISAR) imaging due to their capability of achieving high-resolution image with a limited number of pulses [45–48]. Apart from the sparsity, we further notice that the nonzero scatterers of the target are not isolated, but located continuously in the target scene. This target scene continuity is naturally captured by a two dimensional cluster structure of the scatterers. To the best of the author’s knowledge, this cluster structure of the target scene has not been taken into consideration in ISAR imaging. However, in ISAR problem, unlike the block sparse signal which requires a known block(cluster) partition, nonzero scatterers cluster randomly with unknown location and size in two dimensional target scene. Motivated by the success of the algorithms for recovering one dimensional sparse signal with cluster [35, 36] and tree structures [25, 26], where the structure of the coefficients is imposed in a probabilistic way and the unknowns can be further inferred by either the Monte Carlo Markov chain (MCMC) or mean-field approximation method, we realize that those ideas can be devised and generalized to statistically incorporate the two dimensional cluster structure. It will be of great potential to apply that structure to the ISAR imagery. However, what kind of prior should we choose to encourage the structure of two dimensional cluster? How to design the over-complete dictionary for the high-resolution ISAR imagery? How the recovery algorithm performs under the designed dictionary? How much improvement on the ISAR imaging could be achieved by exploiting the continuity structure of the target scene? Attempts to answer these questions will be made in this thesis.

1.2 Contributions of the Thesis

Aiming at addressing the aforementioned problems, we re-study the stability conditions of BOGA, develop two BOGA based harmonic tonal detectors and analyze their theoretical performances, propose algorithm for recovering hierarchical sparse signal and apply the structured sparse representation in ISAR imaging to obtain enhanced target images. The main contributions of the thesis are summarized below.

1) Extensive stability condition analysis of the BOGA using the maximum block-coherence definition in [32] is presented. By derivations, we demonstrate that recovery conditions of the BOGA previously reported can be relaxed by using the definition of the maximum block-coherence defined in [32]. The theoretical results obtained are more accurate than those presented in [2], provide a generalization of those reported by Tseng [13] for block-sparse signal and serve as a complement of the BOGA reported by Eldar [32] for noisy signals. The obtained theoretical results are validated by a number of designed simulations. Comparisons with OGA, which does not take the block sparsity into consideration, are also made to show the improvements of recovery performance of BOGA in terms of mean square error (MSE) and support recovery probability under different settings.

2) We formulate the harmonic tonals by the block-sparse signal model and suggest using the block sparse recovery techniques in harmonic detection. Two harmonic tonal detectors based on the BOGA are proposed. The first one is a subspace harmonic detector in which the BOGA is applied to find the signal space before an energy detector used to detect the existence of the signal. The second one is a sequential harmonic detector by inserting a sequential detection procedure into each step of BOGA. Both proposed detectors accumulate the signal energy from the fundamental frequencies and their associated higher order harmonics to achieve improved detection performance in low SNR environments. The theoretical analysis of the test statistics and the detection performances using the theory of the order statistics are given comprehensively under an orthogonal dictionary. Significant improvements on detection performance over the conventional detectors in low SNR environments are shown by extensive simulations.

3) We propose a novel hierarchical sparse signal recovery algorithm under a Bayesian framework. Following the sparse Bayesian learning, each component of the sparse coefficient vector is firstly modeled as a Gaussian distributed variable with zero mean. To enforce the two-level hierarchical sparsity, the variance is further modeled by the product of two classes of hidden variables controlling the block-sparsity and the internal sparsity, respectively. Variational Bayesian inference is then used to recover the coefficient vector and other unknown hidden variables from the noise corrupted data. This method is found to avoid the tedious regularization parameter selection procedure required by the convex optimization based methods [42–44] and can automatically learn the block-sparsity and the internal sparsity of the signal from the input data. We also conduct numerical simulations and experiments with real data to demonstrate advantages of the proposed algorithm for hierarchical sparse signal recovery over other reported methods.

4) We propose a novel ISAR imaging method by exploiting the inherent continuity (two dimensional cluster structure) of the scatterers on the target scene to obtain enhanced target images within a Bayesian framework. A simplified radar system is utilized by transmitting the sparse probing frequency signal, where the ISAR imaging problem is firstly converted into one dealing with under-determined linear inverse scattering. Following the Bayesian compressive sensing theory, a hierarchical Bayesian prior is employed to model the scatterers on the range-Doppler plane. In contrast to the independent prior on each scatterer in conventional Bayesian compressive sensing, a correlated prior is proposed to statistically encourage the continuity (cluster) structure of the scatterers in the target region. To overcome the intractability of the posterior distribution under our formulated probabilistic model, the Gibbs sampling strategy is used for Bayesian inference. The parameters of the signal model are inferred efficiently from samples obtained by the Gibbs sampler. Because the proposed method is a data-driven learning process, the tedious parameter tuning process required by the convex optimization based approaches can be avoided. Both the synthetic and the experimental results demonstrate that the proposed algorithm can achieve substantial improvements in the scenarios of limited measurements and low SNR compared with other reported algorithms for ISAR imaging problems.

1.3 Organization of the Thesis

The rest of this thesis is organized as follows: Chapter 2 provides a general overview of existing sparse recovery and structured sparse recovery techniques. Chapter 3 analyzes the stable recovery conditions of the BOGA using a different definition of block-coherence. In Chapter 4, two harmonic tonal detectors based on the BOGA are proposed and their detection performances are analyzed theoretically and experimentally. Chapter 5 proposes a novel hierarchical sparse signal recovery algorithm under a Bayesian framework. Chapter 6 develops a novel ISAR imaging method based on the structured sparse representation to obtain enhanced target images by exploiting the inherent continuity of the target scene. Finally, conclusions and future work are presented in Chapter 7.

Chapter 2

Literature Review

Noise corrupted sparse signal recovery is typically regarded as a problem of estimating a set of unknown sparse coefficients through a linear function, i.e.,

$$\mathbf{x} = \mathbf{D}\mathbf{b} + \mathbf{w} \quad (2.1)$$

where the noisy observation , $\mathbf{x} \in \mathcal{C}^N$ (or $\mathbf{x} \in \mathcal{R}^N$), is a column vector, $\mathbf{b}$ is the coefficient vector of size $M \times 1$, $\mathbf{D}$ is an $N \times M$ dictionary, i.e., a matrix whose columns are the atoms expanding an over-complete signal space, and $\mathbf{w}$ is an additive noise vector. A signal $\mathbf{x}$ is k -sparse when it has at most k nonzero entries, i.e., $\|\mathbf{b}\|_0 \leq k$, where $k \ll M$. A standard sparse signal recovery is formulated as [49]

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \|\mathbf{x} - \mathbf{D}\mathbf{b}\|_2 \text{ subject to } \|\mathbf{b}\|_0 \leq k \quad (2.2)$$

or

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \|\mathbf{x} - \mathbf{D}\mathbf{b}\|_2 + \lambda \|\mathbf{b}\|_0 \quad (2.3)$$

where $\|\mathbf{x}\|_p$ with $p = 0, 2$ represents the l_p -norm of the vector $\mathbf{x}$, l_0 -norm counts the number of the nonzero components in $\mathbf{x}$, λ is the user-selected regularization parameter which controls the tradeoff between signal sparsity and reconstructed error and $\hat{\mathbf{b}}$ is the estimate of $\mathbf{b}$. Solving (2.2) and (2.3) is NP-hard and requires a time-consuming combinatorial search [3].

Efficient algorithms have been proposed in recent years to avoid solving (2.2) and (2.3) and been successfully used to recover the sparse coefficient vector $\mathbf{b}$ as

well as the denoised signal $\mathbf{x}$. In this chapter, we briefly review some representative algorithms dealing with standard and structured sparse recovery problems.

2.1 Methods for Standard Sparse Recovery

Methods for standard sparse recovery in this part are presented according to the three solving strategies aforementioned in section 1.1.1.

2.1.1 Convex Optimization based Methods

Convex optimization [50] is the most commonly used strategy for solving sparse signal presentation problem due to the availability of sophisticated algorithms and toolbox. Two basic convex optimization based sparse recovery methods are summarized herein.

One is the Basis Pursuit (BP) using l_1 -norm minimization [5]

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \|\mathbf{b}\|_1 \text{ subject to } \|\mathbf{x} - \mathbf{D}\mathbf{b}\|_2 \leq \lambda\sigma, \quad (2.4)$$

where σ is the standard derivation of the noise.

The other is the Least Absolute Shrinkage and Selection Operator (Lasso) [6], given as

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \|\mathbf{x} - \mathbf{D}\mathbf{b}\|_2 + \lambda\|\mathbf{b}\|_1, \quad (2.5)$$

where λ is the user-selected regularization parameter. To obtain a correct estimation, one needs to select λ carefully. Usually, it is selected by cross validation which is always time consuming.

Eq. (2.4) and Eq. (2.5) are found to be equivalent [1]. They are the most common and fundamental solvers used for standard sparse representation problems. Some improved versions, such as the reweighted l_1 and l_2 minimizations [51–53], are proposed to achieve better recovery performance in the sense that substantially fewer measurements are needed.

2.1.2 Iterative Greedy Algorithms

The iterative greedy algorithms are attractive for their computational efficiency. The most commonly used methods include MP, OMP [7–9, 54, 55] and their variations, such as, stage-wise orthogonal matching pursuit (StOMP) [56] and compressive sampling matching pursuit (CoSaMP) [57]. For conciseness, only the MP and OMP algorithms are presented in Algorithm 1 in this thesis, where Γ is the support of the coefficient vector $\mathbf{b}$ and $\mathbf{r}_l$ is the residual at the l -th iteration. At each step, one of the support is selected based on the correlation of atoms with the current residual.

Algorithm 1 MP and OMP

Initialization:

The residual $\mathbf{r}_0 = \mathbf{x}$, step counter $l = 0$, $\Gamma_0 = \emptyset$.

Repeat:

$i_l = \arg \max_i \| \mathbf{d}_i^H \mathbf{r}_{l-1} \|_2^2$ for all $i \notin \Gamma_{l-1}$;

$\Gamma_l = \Gamma_{l-1} \cup i_l$;

Estimate the coefficient for the selected blocks:

For MP:

$$\hat{b}_{i_l} = \langle \mathbf{r}_{l-1}, \mathbf{d}_{i_l} \rangle;$$

For OMP:

$$\hat{\mathbf{b}} = \mathbf{D}_{\Gamma_l}^\dagger \mathbf{x};$$

Update the residual:

For MP:

$$\mathbf{r}_l = \mathbf{r}_{l-1} - \mathbf{d}_{i_l} \hat{b}_{i_l};$$

For OMP:

$$\mathbf{r}_l = \mathbf{x} - \mathbf{D}_{\Gamma_l} \hat{\mathbf{b}};$$

Update $l = l + 1$;

Until stopping condition.

The stopping conditions are set according to different prior knowledge of the signal. There are mainly three kinds of stopping conditions [58] adopted in the literature. One is $l \leq k$ with a known k which is the maximum number of the nonzero components. Another commonly used stopping condition is to halt when the residual

has a small magnitude, i.e., $\|\mathbf{r}_l\|_2^2 \leq \delta^{stop}$ where δ^{stop} is a predefined threshold which usually relates to the noise level. The third is to halt when no column explains a significant amount of energy in the residual, i.e., $\|\mathbf{D}^H \mathbf{r}_{l-1}\|_\infty \leq \delta^{stop}$.

2.1.3 Sparse Bayesian Learning Algorithms

Sparse recovery methods from the point view of Bayesian inference are known collectively as sparse Bayesian learning (SBL) algorithms. They are originally proposed in the field of machine learning where Bayesian model with sparse promoting priors are used and the estimation of the sparse coefficient vector is given by an Expectation-Maximization (EM), type II Maximum Likelihood (ML) [15–17] or variational approximation algorithm [18]. Those sparse Bayesian learning algorithms mainly distinguish each other by the prior model imposed to promote the sparsity. The sparse coefficient $\mathbf{b}$ is either modeled as variable whose distribution has a sharp peak at zero and heavy tails, such as, Laplace [18, 19, 59, 60], Student- t [15, 17, 20], Jeffrey [21] distributions, or as multiplication of Gaussian and Bernoulli variables [22, 23, 61–63]. The SBL based methods are also known as Bayesian compressive sensing (BCS) in the field of CS. The performances of SBL based methods differ each other slightly. We summarize below one of the popular methods given in [17] using type-II ML for conciseness. The details of the model using Gaussian-Bernoulli prior will be left latter in Section 2.2.4.

Rather than modeling the sparse coefficient as a determined variable in the convex optimization and iterative greedy based methods, SBL treats the elements of sparse coefficient $\mathbf{b}$ as random variables and further imposes a hierarchical prior on them to encourage the sparsity. Firstly, a zero-mean Gaussian prior is imposed on each element of $\mathbf{b}$:

$$p(\mathbf{b} | \boldsymbol{\alpha}) = \prod_{i=1}^M \mathcal{N}(b_i | 0, \alpha_i^{-1})$$

with α_i , for $i = 1, \dots, M$, being the precision (inverse-variance) of a Gaussian density function, $\mathcal{N}(\cdot, \cdot)$. Furthermore, a Gamma prior is considered over $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_M]^T$:

$$p(\boldsymbol{\alpha} | a, c) = \prod_{i=1}^M \text{Gamma}(\alpha_i | a, c).$$

By marginalizing over the hyperparameters $\boldsymbol{\alpha}$, it is found that $p(\mathbf{b} | a, c)$ follows the Student- t distribution which is sharply peaked at zero like the Laplace distribution and can be utilized to encourage the sparsity. Assuming that the noise follows a zero mean Gaussian distribution with precision α_0 , it is found that the posterior of $\mathbf{b}$ can be expressed analytically as a multivariate Gaussian distribution with mean and covariance:

$$\begin{aligned} \boldsymbol{\mu} &= \alpha_0 \boldsymbol{\Sigma} \mathbf{D}^T \mathbf{x} \\ \boldsymbol{\Sigma} &= (\alpha_0 \mathbf{D}^T \mathbf{D} + \boldsymbol{\Lambda})^{-1} \end{aligned}$$

where $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\alpha})$ presents a diagonal matrix whose diagonal entries are from vector $\boldsymbol{\alpha}$. The hyperparameters $\boldsymbol{\alpha}$ and α_0 are estimated from the data $\mathbf{x}$ by performing a type-II ML (or evidence maximization) procedure and are given by:

$$\alpha_i^{\text{new}} = \frac{1 - \alpha_i \Sigma_{ii}}{\mu_i^2}$$

and

$$1/\alpha_0^{\text{new}} = \frac{\|\mathbf{x} - \mathbf{D}\boldsymbol{\mu}\|_2^2}{N - \sum_{i=1}^M (1 - \alpha_i \Sigma_{ii})}.$$

The estimation of $\mathbf{b}$ is finally given by $\hat{\mathbf{b}} = \boldsymbol{\mu}$.

2.1.4 Summary

The aforementioned standard sparse signal recovery methods have their own advantages and constraints. They are widely used in different applications where one or more specific methods are chosen based on their characteristics. For example, greedy methods are always used in real-time applications with a known signal sparsity where computational efficiency is required. When there is no prior knowledge to

determine the regularization parameter in the convex optimization based methods, the sparse Bayesian learning algorithms are good choices. The computational complexity and the prior knowledge required for the representative methods of different strategies are summarized in Table. 2.1, where $N_{Maxiter}$ is the maximum number of iterations. For OMP, the maximum number of iterations is approximately set to be the maximal sparsity of the signal. The prior distribution imposed on the

Table 2.1: Summary of the representative methods for standard sparse recovery

Algorithms	Prior knowledge	Computational complexity
BP	a proper regularization parameter	$O(N^3)$ [64]
OMP	the sparsity or a proper stopping criterion	$O(MN \times N_{Maxiter})$ [58]
SBL (BCS)	a proper prior distribution	$O(M^3 \times N_{Maxiter})$ [16]

coefficient vector encourages statistically the sparse property of the signal in SBL or BCS. However, it is worthy to emphasize that more general structures other than the sparsity can be encouraged by imposing different priors on the coefficient vector, as presented in the next section.

2.2 Methods for Structured Sparse Recovery

We mainly focus on the structure of block sparsity, since it is the most commonly exploited structure in the literature and is related to our work presented in the latter chapters of this thesis.

Signals with coefficients being either zero or nonzero simultaneously within pre-specified blocks are referred to as block-sparse [29, 32, 65]. Under signal model (2.1), the coefficient vector $\mathbf{b}$ of block sparse signal is a concatenation of sub-blocks $\mathbf{b}[l]$, denoting the l th block of size d , that is:

$$\mathbf{b} = [\underbrace{b_1 \cdots b_d}_{\mathbf{b}^T[1]} \underbrace{b_{d+1} \cdots b_{2d}}_{\mathbf{b}^T[2]} \cdots \underbrace{b_{M-d+1} \cdots b_M}_{\mathbf{b}^T[L]}]^T. \quad (2.6)$$

Accordingly, the dictionary

$$\mathbf{D} = [\underbrace{\mathbf{d}_1 \cdots \mathbf{d}_d}_{\mathbf{D}[1]} \underbrace{\mathbf{d}_{d+1} \cdots \mathbf{d}_{2d}}_{\mathbf{D}[2]} \cdots \underbrace{\mathbf{d}_{M-d+1} \cdots \mathbf{d}_M}_{\mathbf{D}[L]}] \quad (2.7)$$

is a matrix of size $N \times M$ containing a concatenation of column-block $\mathbf{D}[l]$ of size $N \times d$. Unlike the conventional sparse signals, the block-sparse signals with K sparse degree have K out of L nonzero blocks in their coefficient vectors.

Following the three strategies aforementioned in section 1.1.1 for standard sparse signal recovery, considerable effort has been put into the development of reconstruction algorithms for block sparse signal. Those algorithms are presented below according to the strategies used.

2.2.1 Group Lasso

The convex optimization strategy based group Lasso is proposed in [30] and further analyzed in [31] to recover the signal with block sparsity under an over-complete dictionary. The block sparse signal is recovered by solving a mixed l_2/l_1 -norm optimization:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_{2,1}$$

where $\|\mathbf{b}\|_{2,1} = \sum_{l=1}^L \|\mathbf{b}[l]\|_2$ and λ is the user-selected regularization parameter controlling the block sparsity. It is shown in [31] that the group Lasso has a better reconstruction performance than Lasso if the underlying structure of the observation complies with the block sparsity pattern. However, similar to that of the Lasso, a proper regularization parameter λ is to be carefully selected.

2.2.2 BOGA

Block orthonormal greedy algorithm (BOGA), an direct extension of OGA to recover the signal with block-sparsity, is proposed in [29, 32]. BOGA starts from an initial residual $\mathbf{r}_0 = \mathbf{x}$ and finds a low-dimensional block support at each iteration step. At the l th step, a column block is selected based on its correlation with the current residual

$$i_l = \arg \max \|\mathbf{D}^H[i]\mathbf{r}_{l-1}\|_2. \quad (2.8)$$

The following decomposition consisting of the blocks selected through the current and previous steps is built by

$$\hat{\mathbf{x}}_l = \sum_{j=1}^l \mathbf{D}[i_j] \mathbf{b}[i_j] \quad (2.9)$$

where the block coefficients $\mathbf{b}[i_j]$ fitted by least squares to minimize $\|\mathbf{x} - \hat{\mathbf{x}}_l\|_2$ are given by

$$\hat{\mathbf{b}} = \mathbf{D}_l^\dagger \mathbf{x} \quad (2.10)$$

with $\mathbf{D}_l^\dagger = (\mathbf{D}_l^H \mathbf{D}_l)^{-1} \mathbf{D}_l^H$ and $\mathbf{D}_l = [\mathbf{D}[i_1] \cdots \mathbf{D}[i_l]]$. A new residual is then obtained by

$$\mathbf{r}_l = \hat{\mathbf{x}}_l - \mathbf{x} \quad (2.11)$$

as the input to the next step of the algorithm. The algorithm is stopped when the residual norm is below some predetermined threshold or $l \leq K$.

2.2.3 BSBL

Block sparse signal recovery has also been considered in the Bayesian framework in [34] where block sparse Bayesian learning (BSBL) method is proposed. It is assumed in [34] that each block of $\mathbf{b}[l]$ obeys a parameterized multivariate Gaussian distribution:

$$p(\mathbf{b}[l], \gamma_l, \mathbf{B}_l) \sim \mathcal{N}(\mathbf{0}, \gamma_l \mathbf{B}_l)$$

where the unknown parameter γ_l is nonnegative which controls the block-sparsity of $\mathbf{b}$; $\mathbf{B}_l$ is a positive definite matrix which captures the correlation structure of the l -th block. Under the assumption that blocks are mutually uncorrelated, the prior of $\mathbf{b}$ is $p(\mathbf{b}; \{\gamma_l, \mathbf{B}_l\}_{l=1}^L) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\Sigma}_0 = \text{diag}\{\gamma_1 \mathbf{B}_1, \cdots, \gamma_L \mathbf{B}_L\}$. Assuming that the noise follows a zero mean Gaussian distribution with variance γ_0 , it is found that the posterior of $\mathbf{b}$ can be expressed analytically as

$$p(\mathbf{b} | \mathbf{x}; \gamma_0, \{\gamma_l, \mathbf{B}_l\}_{l=1}^L) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}_0 \mathbf{D}^T (\gamma_0 \mathbf{I} + \mathbf{D} \boldsymbol{\Sigma}_0 \mathbf{D}^T)^{-1} \mathbf{x}$$

$$\mathbf{\Sigma} = \left(\mathbf{\Sigma}_0^{-1} + \frac{1}{\gamma_0} \mathbf{D} \mathbf{D}^T \right)^{-1}.$$

The unknown parameters $\{\gamma_l, \mathbf{B}_l\}_{l=1}^L$ and γ_0 are estimated by EM approach, given respectively by

$$\gamma_l = \frac{1}{d} \text{Tr} \left[\mathbf{B}_l^{-1} (\mathbf{\Sigma}_l + \boldsymbol{\mu}[l] \boldsymbol{\mu}^H[l]) \right],$$

and

$$\gamma_0 = \frac{\|\mathbf{x} - \mathbf{D} \boldsymbol{\mu}\|_2^2 + \text{Tr}(\mathbf{\Sigma} \mathbf{D}^T \mathbf{D})}{N}.$$

To avoid the over-fitting, it is assumed in [34] that $\mathbf{B}_l = \mathbf{B}$ for all l and the estimate of $\mathbf{B}$ is given by

$$\mathbf{B} = \frac{1}{L} \sum_{l=1}^L \frac{\boldsymbol{\mu}[l] \boldsymbol{\mu}^H[l] + \mathbf{\Sigma}_l}{\gamma_l}$$

where $\mathbf{\Sigma}_l \in \mathcal{R}^{d \times d}$ is the l -th principal diagonal block in $\mathbf{\Sigma}$.

Intra-block correlation is captured in the matrix $\mathbf{B}$. It is find in [34] that modeling elements of a block as a first-order Auto-Regressive (AR) process is sufficient to model the intra-block correlation of the signal for many applications. The specific choice of $\mathbf{B}$ being identity matrix indicates the belief that components inside the blocks are independent to each other.

2.2.4 CluSS

The above mentioned block sparse recovery methods have a common assumption that the block partition is known a priori. Work in [35, 36] employs the neighborhood system to enforce the nonzero components to appear in clusters with their locations and sizes inferred automatically from the input data, where the structure is imposed in a probabilistic way and the unknowns are given by either the Monte Carlo Markov chain (MCMC) [35] or variational Bayes (VB) approximation method [36]. The algorithms developed in [35] and [36] are referred to as CluSSMCMC and CluSSVB, respectively, where CluSS is the abbreviation for Clustered Sparse Solver. They differ from each other in the statistical solving tools. Since the VB is guaranteed to

converge and its convergence is easy to diagnose, we summarize herein the CluSSVB method.

To facilitate the constraint on the signal support, the Gaussian-Bernoulli prior is used over $\mathbf{b}$ in [36], i.e.,

$$\mathbf{b} = \mathbf{z} \circ \boldsymbol{\alpha}$$

where $\circ$ denoting the point-wise multiplication. Following the SBL [17], hierarchical priors are imposed on $\boldsymbol{\alpha}$, i.e., $\alpha_i \sim \mathcal{N}(0, \gamma_i^{-1})$ and $\gamma_i \sim \text{Gamma}(a, c)$ to encourage the sparsity of $\mathbf{b}$, where γ_i is the precision. It is assumed that z_i follows the Bernoulli distribution with probability π_i , i.e., $z_i \sim \text{Bernoulli}(\pi_i)$, which is a binary variable with state $\{1, 0\}$ indicating whether b_i is nonzero or not.

The cluster structure over the coefficients is encouraged by assuming that neighboring coefficients tend to be nonzero or zero simultaneously. Three cluster patterns are defined in [36]. They are patterns for strong elimination, weak elimination, and strong supporting. Supposing z_i is under investigation, the value of two adjacent neighbors z_{i-1} and z_{i+1} are considered for pattern selection. If both of z_{i+1} and z_{i-1} are nonzero, pattern for strong supporting is chosen and hyper-prior $\pi_i \sim \text{Beta}(e_2, f_2)$ with $e_2 > f_2$ is used to encourage $z_i = 1$ with higher probability. If both of z_{i+1} and z_{i-1} are zero, pattern for strong elimination is chosen and hyper-prior $\pi_i \sim \text{Beta}(e_0, f_0)$ with $e_0 < f_0$ is used to encourage $z_i = 0$ with higher probability. Otherwise, pattern for weak elimination is chosen and hyper-prior $\pi_i \sim \text{Beta}(e_1, f_1)$ with $e_1 = f_1$ is used to exert non-informative prior. Under the above described Bayesian model, the VB method is used to infer all the hidden variables. The inference procedures are standard and the detailed updating rules for the hidden variables can be found in [36].

2.2.5 Summary

Algorithms of group Lasso, BOGA, and BSBL explicitly make a hard imposition of the block sparse structure on the coefficient vector, therefore a prior knowledge of the block partition is required. The block structural information is statistically imposed

on the Bayesian prior by Cluss, therefore it implicitly encourages a clustering effect on the coefficient vector. In other words, merely the belief that nonzero components are likely to be present in the neighborhood of each other, is utilized by Cluss and the nonzero block sizes and locations are flexibly inferred from the input data. Therefore, compared to those methods with hard imposition of the block structure, the Cluss algorithm requires much less prior knowledge.

Our work in this thesis is more or less related to the aforementioned algorithms. Chapter 3 and Chapter 4 are based on the BOGA method. Most of the algorithms, such as MP, BP, SBL(BCS), BSBL, and ClussVB, are served as baselines and compared with our proposed methods for certain purposes and applications in Chapters 3, 5 and 6. In Chapter 6, the basic idea of Cluss in [35, 36] is devised and generalized to statistically incorporate the two dimensional cluster structure of the scatterers in ISAR imagery to obtain enhanced target images.

Chapter 3

Improved Stability Conditions of BOGA for Noisy Block-Sparse Signals

3.1 Introduction

The block sparse signal model is briefly reviewed below for completeness and self-consistency. This chapter will focus on the stability conditions of BOGA under bounded noise. The statistical behavior of BOGA can be found in [66] under random noise. A signal with block-sparsity corrupted by bounded noise is modeled as [2]

$$\mathbf{x} = \mathbf{D}\mathbf{b} + \mathbf{w} \quad (3.1)$$

where $\mathbf{x}$ is the observed signal vector of size $N \times 1$, $\mathbf{w}$ is a bounded noise,

$$\|\mathbf{w}\|_2 \leq \varepsilon, \quad (3.2)$$

with $\varepsilon > 0$, and $\mathbf{b}$ represents the coefficient vector which is a concatenation of sub-blocks $\mathbf{b}[i]$, denoting the i th block of size d , that is:

$$\mathbf{b} = \underbrace{[b_1 \cdots b_d]}_{\mathbf{b}^T[1]} \underbrace{[b_{d+1} \cdots b_{2d}]}_{\mathbf{b}^T[2]} \cdots \underbrace{[b_{M-d+1} \cdots b_M]}_{\mathbf{b}^T[L]}^T \quad (3.3)$$

and the dictionary

$$\mathbf{D} = [\underbrace{\mathbf{d}_1 \cdots \mathbf{d}_d}_{\mathbf{D}[1]} \underbrace{\mathbf{d}_{d+1} \cdots \mathbf{d}_{2d}}_{\mathbf{D}[2]} \cdots \underbrace{\mathbf{d}_{M-d+1} \cdots \mathbf{d}_M}_{\mathbf{D}[L]}] \quad (3.4)$$

is a matrix of size $N \times M$ containing a concatenation of column-block $\mathbf{D}[i]$ of size $N \times d$. Unlike the conventional sparse signals, the block-sparse signals with K sparse degree have K nonzero blocks in their coefficient vectors.

The corresponding block-version of orthonormal MP, also known as block orthonormal greedy algorithm (BOGA) in [2, 29, 32]), and the block version of Lasso (group Lasso) [31], have been proved to successfully recover block-sparse signals under their own recovery conditions on the block-coherence of the over-complete dictionary in both noiseless and noisy environments, respectively.

It was shown in Theorem 3 [32] that if

$$K < \frac{1}{2d} \left(\mu_B^{-1} + d - (d-1) \frac{\nu}{\mu_B} \right) \quad (3.5)$$

where K is the sparse degree, ν is the sub-coherence and d is the size of each block, the BOGA picks up a correct block in each step and consequently recovers the support of the signal completely with a certain convergence rate. In (3.5), the block-coherence is defined as

$$\mu_B = \max_{l \neq r} \frac{1}{d} \rho(\mathbf{D}^H[l] \mathbf{D}[r]) \quad (3.6)$$

with $\mathbf{d}_i \in \mathbf{D}[l], \mathbf{d}_j \in \mathbf{D}[r]$, which measures the similarity between blocks of the dictionary. Similar results, with a definition

$$\mu_B = \max_{l \neq r} \max_{i,j} |\mathbf{d}_i^H \mathbf{d}_j| \quad (3.7)$$

different from that given by Eldar [32], have also been reported for signals in noiseless environments [2]. The stability and support recovery performances, when the BOGA is applied for noisy signals, are further studied [2]. It is shown in [2] that for $\delta^{\text{BOMP}} \geq \varepsilon$, if K satisfies (3.5), and

$$\|\mathbf{b}\|_{2,\min} > \max \left\{ \frac{2\sqrt{d}\varepsilon}{1-(d-1)\nu-d\mu_B(2K-1)}, \frac{(\delta^{\text{BOGA}}+\varepsilon)\sqrt{1-[(K-2)d\mu_B+(d-1)\nu]}}{\sqrt{1-[(K-2)d\mu_B+(d-1)\nu]} \cdot \sqrt{1-(d-1)\nu} \cdot \sqrt{(K-1)d\mu_B}} \right\} \quad (3.8)$$

where $\|\mathbf{b}\|_{2,\min} = \min_{i \in \Gamma(\mathbf{b})} (\|\mathbf{b}[i]\|_2)$, $\Gamma(\mathbf{b})$ is the support of the coefficient vector $\mathbf{b}$, δ^{BOMP} is stopping threshold, and ε is the power level of the bounded noise, the BOGA can completely recover the support of ideal block-sparse signal from noise-corrupted data. Both the basic conclusions of the recovery condition in (3.5) obtained in [2] and [32] show that the block-sparse degree allowed to fully recover the ideal block-sparse signal is constrained by a small value of block-coherence μ_B . A small μ_B of the dictionary with block structure leads to a large degree of signal sparsity, which is often the case since signals in many practical applications are not always of low sparsity. Comparing both definitions of μ_B in [2] and [32], it is easy to conclude that the definition of block-coherence in [32] is less restrictive than that in [2] since

$$\begin{aligned} \mu_{B,L} &= \max_{l \neq r} \max_{i,j} |\mathbf{d}_i^H \mathbf{d}_j| = \max_{l \neq r} \frac{1}{d} \max_i \sum_{j=1}^d \left| (\mathbf{D}^H[l] \mathbf{D}[r])_{i,j} \right| \\ &\geq \max_{l \neq r} \frac{1}{d} \rho(\mathbf{D}^H[l] \mathbf{D}[r]) = \mu_{B,E} \end{aligned} \quad (3.9)$$

where $\mu_{B,L}$ and $\mu_{B,E}$ represent the definitions given by Lv and Eldar, respectively. With the definition of block-coherence $\mu_{B,E}$, which is also adopted in our thesis, a smaller value of the block-coherence can be computed for a given dictionary. Using such a block-coherence, the BOGA can correctly recover the ideal signal with a larger block-sparse degree in noiseless case than that obtained based on the definition by Lv. To the best of the authors' knowledge, the recovery conditions and the stability of BOGA with $\mu_{B,E}$ for recovering the noisy block-sparse signals have not been reported in the literature.

This chapter investigates the stable recovery problems, i.e., recovery conditions and representation error bound of the BOGA, using the block-coherence defined in [32], i.e.,

$$\mu_B = \max_{l \neq r} \frac{1}{d} \rho(\mathbf{D}^H[l] \mathbf{D}[r]) \quad (3.10)$$

for noisy signals. Although the recovery conditions and approximate representation error have already been studied in [2], and the results obtained agree well with those reported in [13], the definition of the block-coherence is too restrictive. Under the definition of block-coherence in (3.10), the theoretical recovery conditions to

be reported in this chapter are more relaxed than those presented in [2], which also provide a generalization of those for the orthogonal greedy algorithm (OGA) reported in [13] for conventional sparse signals.

The remainder of this chapter is organized as follows: Section 3.2 gives some preliminary definitions used in [2] and [32] for recovery condition derivations from the perspective of dictionary coherence. The detailed derivations and the main results obtained in this chapter are presented in Section 3.3. Section 3.4 provides the numerical simulations to validate the theoretical results obtained. Conclusion is given finally in Section 3.5.

3.2 Preliminary Definitions and Results

In this section, we will present some important results and the definitions of the dictionary that will be used in the next section. The definition of the coherence used in this Chapter is the same as that in [32]. It is assumed that the columns of the dictionary $\mathbf{D}$ are normalized to unit l_2 -norm, i.e., $\|\mathbf{d}_i\|_2 = 1$.

Definition1: Given a block decomposition dictionary $\mathbf{D} = \bigcup_{i=1}^L \mathbf{D}[i]$, define

$$\mu = \max_{i,j \neq i} |\mathbf{d}_i^H \mathbf{d}_j|$$

where $\mathbf{d}_i, \mathbf{d}_j \in \mathbf{D}$,

$$\mu_B = \max_{l \neq r} \frac{1}{d} \rho(\mathbf{D}^H[l] \mathbf{D}[r])$$

where $\mathbf{d}_i \in \mathbf{D}[l], \mathbf{d}_j \in \mathbf{D}[r]$,

$$\nu = \max_l \max_{i,j \neq i} |\mathbf{d}_i^H \mathbf{d}_j|$$

where $\mathbf{d}_i, \mathbf{d}_j \in \mathbf{D}[l]$. μ is the coherence of the dictionary $\mathbf{D}$, μ_B is the coherence between two different blocks and ν is the sub-coherence of $\mathbf{D}$.

Definition2 [2]: The support of the coefficient vector $\mathbf{b} = (\mathbf{b}[i]) \in \mathcal{C}^M$, where $\mathbf{b}[i] = (b_j) \in \mathcal{C}^{d_i}$, is $\Gamma(\mathbf{b}) := \{i, \|\mathbf{b}[i]\|_2 \neq 0\}$ and $\Gamma(\mathbf{b}[i]) := \{j, b_j \neq 0\}$ and the

cardinality of set $\Gamma(\mathbf{b})$, denoted by $|\Gamma(\mathbf{b})|$, measures the “number of elements of the set”.

Lemma1: Suppose $\mathbf{D}_K$ is an $N \times Kd$ submatrix formed by concatenating K column-blocks of $\mathbf{D}$. Then any singular value of $\mathbf{D}_K$ denoted as σ_D satisfies

$$\sqrt{1 - [(K-1)d\mu_B + (d-1)\nu]} \leq \sigma_D \leq \sqrt{1 + [(K-1)d\mu_B + (d-1)\nu]}. \quad (3.11)$$

Proof: See the Appendix A.

Another form of (3.11) to be used in the following sections is

$$\sqrt{1 - [(K-1)d\mu_B - (d-1)\nu]} \|\mathbf{u}\|_2 \leq \|\mathbf{D}_K \mathbf{u}\|_2 \leq \sqrt{1 + [(K-1)d\mu_B + (d-1)\nu]} \|\mathbf{u}\|_2 \quad (3.12)$$

where $\mathbf{u}$ is an arbitrary vector. Supposing $\mathbf{D} \in \mathcal{C}^{d \times d}$ and $\mu = \max_{i,j \neq i} |\mathbf{d}_i^H \mathbf{d}_j|$ in the conventional sparse case ($d = 1, \nu = 0, \mu_B = \mu$), (3.12) degenerates into

$$\sqrt{1 - (d-1)\mu} \|\mathbf{u}\|_2 \leq \|\mathbf{D} \mathbf{u}\|_2 \leq \sqrt{1 + (d-1)\mu} \|\mathbf{u}\|_2 \quad (3.13)$$

which is also proven by Donoho in Lemma 2.2 [1].

Lemma2: Defining $\mathbf{E} = (\mathbf{D}_K^H \mathbf{D}_K)^{-1}$, then $\mathbf{E}$ satisfies

$$\|\mathbf{E}\|_2 \leq \frac{1}{1 - (K-1)d\mu_B - (d-1)\nu} \quad (3.14)$$

$$\|\mathbf{E} \mathbf{D}_K^H\|_2 \leq \frac{1}{\sqrt{1 - (K-1)d\mu_B - (d-1)\nu}} \quad (3.15)$$

Proof: According to **Lemma1** and the property of singular value, (3.14) and (3.15) can be easily obtained.

3.3 Stability and Support Recovery of BOGA

We study the recovery conditions and the stability of the BOGA under a relaxed definition of block-coherence μ_B originally defined in [32] in the presence of noise. The following theorem presents the derived stability results of the BOGA in this

chapter. The corollary derived from the theorem gives the recovery conditions of the BOGA, which is similar to, but better than those results provided in [2]. The derivation presented bellow partially follows the ideas in [13] and [32] with necessary changes.

Theorem1: If $K = |\Gamma(\mathbf{b})|$ satisfies

$$K < \frac{1}{2d} \left(\mu_B^{-1} + d - (d-1) \frac{\nu}{\mu_B} \right), \quad (3.16)$$

the following results hold.

(a) If the stopping threshold $\delta^{\text{BOGA}} \geq \varepsilon$ and

$$\|\mathbf{b}[\Gamma_l] + 1\|_2 \geq \frac{2\sqrt{1 + (d-1)\nu}\varepsilon}{1 - (d-1)\nu + d\mu_B - \frac{2d\mu_B(1-(d-1)\nu+d\mu_B)(K-l)}{1-(l-1)d\mu_B-(d-1)\nu}}, \quad (3.17)$$

then the output of the BOGA at the $(l+1)$ th step follows $\Gamma_{l+1}^{\text{BOGA}} \subseteq \Gamma(\mathbf{b})$. In addition, if

$$\sqrt{1 - (d-1)\nu} \|\mathbf{b}\|_{2,\min} - \frac{\sqrt{(K-1)d\mu_B} \|\mathbf{b}\|_{2,\min}}{\sqrt{1 - [(K-2)d\mu_B + (d-1)\nu]}} > \delta^{\text{BOGA}} + \varepsilon, \quad (3.18)$$

then $\Gamma_K^{\text{BOGA}} = \Gamma(\mathbf{b})$.

(b) The representation error by the BOGA can be bounded by

$$\left\| \hat{\mathbf{b}}_{\text{BOGA}} - \mathbf{b} \right\|_2^2 \leq \frac{\varepsilon^2}{1 - [(K-1)d\mu_B + (d-1)\nu]} \quad (3.19)$$

where $\hat{\mathbf{b}}_{\text{BOGA}}$ is the block sparse solution.

Proof: (a)

Supposing $\Gamma_l \subseteq \Gamma(\mathbf{b})$, let us derive the proof by induction. Initially, this is true when $\Gamma_0 = \emptyset$. Assuming it is true at the l th step of the BOGA, then it requires that

$$\|\mathbf{D}^{*H} \mathbf{r}_l\|_{2,\infty} > \|\bar{\mathbf{D}}^{*H} \mathbf{r}_l\|_{2,\infty} \quad (3.20)$$

be also true at the $(l+1)$ th step, where sub-matrix $\mathbf{D}^*$ contains all the blocks whose indices belong to $\Gamma(\mathbf{b})$, $\mathbf{r}_l = \sum_{j \in \Gamma_l} \mathbf{D}[j] \hat{\mathbf{b}}_l[j] - \mathbf{x}$ is the residual signal at the l th step,

and $\hat{\mathbf{b}}_l = \mathbf{D}_l^\dagger \mathbf{x} = \mathbf{E} \mathbf{D}_l^H \mathbf{x}$ is the solution obtained at step l with $\mathbf{E} = (\mathbf{D}_l^H \mathbf{D}_l)^{-1}$. For simplicity, we assume that the first $|\Gamma(\mathbf{b})|$ blocks in the corresponding coefficient vector $\mathbf{b}$ are ordered by

$$\|\mathbf{b}[1]\|_2 \geq \|\mathbf{b}[2]\|_2 \geq \cdots \geq \|\mathbf{b}[|\Gamma(\mathbf{b})|]\|_2 \quad (3.21)$$

and the corresponding blocks are ordered to be $\mathbf{D}[1], \mathbf{D}[2], \dots$ in the sub-matrix $\mathbf{D}^*$ containing the first $|\Gamma(\mathbf{b})|$ blocks of the dictionary $\mathbf{D}$. The remainder of $\mathbf{D}$ is denoted by $\bar{\mathbf{D}}^*$.

For r_l , BOGA picks up block $\mathbf{D}[|\Gamma_l| + 1]$ at the $l + 1$ th step. Then (3.20) is equivalent to

$$\left\| \mathbf{D}[|\Gamma_l| + 1]^H \mathbf{r}_l \right\|_2 > \left\| \mathbf{D}[i]^H \mathbf{r}_l \right\|_2 \quad (3.22)$$

with $i \in \Gamma(\mathbf{b})^C$, where $\Gamma(\mathbf{b})^C$ denotes the complement of the set $\Gamma(\mathbf{b})$.

Let us develop a lower bound on the left-hand side and an upper bound on the right-hand side of (3.22), respectively. Both sides of (3.22) can be written into

$$\begin{aligned} & \left\| \mathbf{D}[i]^H (\mathbf{D}_l \hat{\mathbf{b}}_l - \mathbf{x}) \right\|_2 \\ &= \left\| \mathbf{D}[i]^H \left(\mathbf{D}_l (\hat{\mathbf{b}}_l - \mathbf{b}_l) + \mathbf{D}_l \mathbf{b}_l - \mathbf{s}_0 + \mathbf{s}_0 - \mathbf{x} \right) \right\|_2 \\ &= \left\| \mathbf{D}[i]^H \left(\mathbf{D}_l (\hat{\mathbf{b}}_l - \mathbf{b}_l) - \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} - (\mathbf{x} - \mathbf{s}_0) \right) \right\|_2 \\ &= \left\| \mathbf{D}[i]^H \left((\mathbf{D}_l \mathbf{E} \mathbf{D}_l^H - \mathbf{I}) (\mathbf{x} - \mathbf{s}_0) - \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} - \mathbf{D}_l \mathbf{E} \mathbf{D}_l^H \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right) \right\|_2 \end{aligned} \quad (3.23)$$

where $\Gamma(\mathbf{b}) \setminus \Gamma_l$ represents those indices belong to $\Gamma(\mathbf{b})$ instead of Γ_l , for $i \in \Gamma(\mathbf{b})^C$ or $i = |\Gamma_l| + 1$, and $\mathbf{s}_0 = \sum_{i \in \Gamma(\mathbf{b})} \mathbf{D}[i] \mathbf{b}[i]$. Then we have $\mathbf{x} - \mathbf{s}_0 = \mathbf{w}$.

The first item in (3.23) becomes

$$\left\| \mathbf{D}[i]^H (\mathbf{D}_l \mathbf{E} \mathbf{D}_l^H - \mathbf{I}) (\mathbf{x} - \mathbf{s}_0) \right\|_2 \leq \|\mathbf{D}[i]\|_2 \left\| (\mathbf{D}_l \mathbf{E} \mathbf{D}_l^H - \mathbf{I}) (\mathbf{x} - \mathbf{s}_0) \right\|_2 \leq \sqrt{1 + (d-1)\nu} \varepsilon. \quad (3.24)$$

Equation (3.24) holds because of $\|\mathbf{D}[i]\|_2 \leq \sqrt{1 + (d-1)\nu}$, which can be easily verified by Lemma 2.2 in [1], the fact that the term $(\mathbf{D}_l \mathbf{E} \mathbf{D}_l^H - \mathbf{I})$ in (3.24) is a projection matrix which projects a vector onto the space orthogonal to that spanned by the columns of $\mathbf{D}_l$, as well as $\|\mathbf{x} - \mathbf{s}_0\|_2 = \|\mathbf{w}\|_2 = \varepsilon$.

By Lemma 2, we have

$$\begin{aligned}
 & \left\| \mathbf{D}[i]^H \mathbf{D}_l \mathbf{E} \mathbf{D}_l^H \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_2 \\
 & \leq \left\| \mathbf{D}[i]^H \mathbf{D}_l \right\|_2 \left\| \mathbf{E} \right\|_2 \left\| \mathbf{D}_l^H \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_2 \\
 & \leq \left\| \mathbf{D}_l^H \mathbf{D}[i] \right\|_2 \frac{1}{1-(l-1)d\mu_B-(d-1)\nu} \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} \left\| \mathbf{D}_l^H \mathbf{D}[j] \right\|_2 \left\| \mathbf{b}[j] \right\|_2 \\
 & = \frac{1}{1-(l-1)d\mu_B-(d-1)\nu} \sqrt{\sum_{k=1}^l \left\| \mathbf{D}^H[k] \mathbf{D}[i] \right\|_2^2} \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} \sqrt{\sum_{k=1}^l \left\| \mathbf{D}^H[k] \mathbf{D}[j] \right\|_2^2} \left\| \mathbf{b}[j] \right\|_2 \\
 & \leq \frac{ld^2\mu_B^2}{1-(l-1)d\mu_B-(d-1)\nu} \left\| \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_{2,1}
 \end{aligned} \tag{3.25}$$

where the last inequality holds because $\left\| \mathbf{D}^H[i] \mathbf{D}[j] \right\|_2^2 \leq d^2\mu_B^2$ and $\sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} \left\| \mathbf{b}[j] \right\|_2 = \left\| \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_{2,1}$.

When $i \in \Gamma(\mathbf{b})^C$, we have

$$\begin{aligned}
 & \left\| \mathbf{D}^H[i] \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_2 = \left\| \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{D}^H[i] \mathbf{D}[j] \mathbf{b}[j] \right\|_2 \\
 & \leq \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} \left\| \mathbf{D}^H[i] \mathbf{D}[j] \right\|_2 \left\| \mathbf{b}[j] \right\|_2 \\
 & \leq d\mu_B \left\| \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_{2,1}
 \end{aligned} \tag{3.26}$$

and when $i = |\Gamma_l| + 1$, we obtain

$$\begin{aligned}
 & \left\| \mathbf{D}^H[|\Gamma_l| + 1] \mathbf{D}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_2 \\
 & = \left\| \mathbf{D}^H[|\Gamma_l| + 1] \mathbf{D}[|\Gamma_l| + 1] \mathbf{b}[|\Gamma_l| + 1] + \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l \setminus |\Gamma_l| + 1} \mathbf{D}^H[|\Gamma_l| + 1] \mathbf{D}[j] \mathbf{b}[j] \right\|_2 \\
 & \geq \left\| \mathbf{D}^H[|\Gamma_l| + 1] \mathbf{D}[|\Gamma_l| + 1] \right\|_2 \left\| \mathbf{b}[|\Gamma_l| + 1] \right\|_2 - \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l \setminus |\Gamma_l| + 1} \left\| \mathbf{D}^H[i] \mathbf{D}[j] \right\|_2 \left\| \mathbf{b}[j] \right\|_2 \\
 & \geq (1 - (d-1)\nu) \left\| \mathbf{b}[|\Gamma_l| + 1] \right\|_2 - d\mu_B \left\| \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_{2,1}.
 \end{aligned} \tag{3.27}$$

According to (3.24), (3.25), (3.26) and (3.27), it is easy to know that (3.22) requires

$$\begin{aligned}
 & (1 - (d-1)\nu) \left\| \mathbf{b}[|\Gamma_l| + 1] \right\|_2 \\
 & \stackrel{(1)}{\geq} 2\sqrt{1 + (d-1)\nu\varepsilon} + 2d\mu_B \left(1 + \frac{ld\mu_B}{1-(l-1)d\mu_B-(d-1)\nu} \right) \left\| \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_{2,1} \\
 & \stackrel{(2)}{\geq} 2\sqrt{1 + (d-1)\nu\varepsilon} + 2d\mu_B \left(\frac{(1+d\mu_B-(d-1)\nu)(K-l)}{1-(l-1)d\mu_B-(d-1)\nu} \right) \left\| \mathbf{b}[|\Gamma_l| + 1] \right\|_2
 \end{aligned} \tag{3.28}$$

where (3.28.(1)) is implied by (3.28.(2)) because $(K-l) \left\| \mathbf{b}[|\Gamma_l| + 1] \right\|_2 \geq \sum_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} \left\| \mathbf{b}[j] \right\|_2 = \left\| \mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l} \right\|_{2,1}$.

Inequality (3.17) is then obtained by rearranging the terms in (3.28.(2)). Hence, the BOGA chooses the index $|\Gamma_l| + 1$ from $\Gamma(\mathbf{b})$, which means that the output of the BOGA $\Gamma_{l+1}^{BOGA} \subseteq \Gamma(\mathbf{b})$ for $1 \leq l \leq K - 1$.

Because every step of the BOGA selects correctly one block from $\Gamma(\mathbf{b})$, then the residual $\|r_K\|_2 = \|\mathbf{x} - \mathbf{s}_0\|_2 = \varepsilon \leq \delta^{\text{BOGA}}$. It means that the stopping criterion must be reached at step K . Then $\Gamma_K^{BOGA} = \Gamma(\mathbf{b})$ requires that the BOGA does not stop at the step $K - 1$, that is

$$\left\| \mathbf{D}_{K-1} \hat{\mathbf{b}}_{K-1} - \mathbf{x} \right\|_2 > \delta^{\text{BOGA}} \quad (3.29)$$

where $\hat{\mathbf{b}}_{K-1}$ comprises the $K - 1$ largest blocks of $\hat{\mathbf{b}}_{\text{BOGA}}$, whose corresponding dictionary is denoted by $\mathbf{D}_{K-1}$. The lower bound on the left-hand side of (3.29) can be derived in the same way as that used in [2]. Due to the definition difference of block-coherence, some necessary manipulations are needed during the derivation. From [2], the left-hand side of (3.29) is bounded by

$$\begin{aligned} & \left\| \mathbf{D}_{K-1} \hat{\mathbf{b}}_{K-1} - \mathbf{x} \right\|_2 \\ &= \left\| \mathbf{D}_{K-1} \left(\mathbf{D}_{K-1}^\dagger \mathbf{x} - \mathbf{b}_{K-1} \right) + \mathbf{D}_{K-1} \mathbf{b}_{K-1} - \mathbf{s}_0 + \mathbf{s}_0 - \mathbf{x} \right\|_2 \\ &\geq \left\| \mathbf{D}_{K-1} \mathbf{b}_{K-1} - \mathbf{s}_0 \right\|_2 - \left\| \mathbf{D}_{K-1} \left(\mathbf{D}_{K-1}^\dagger \mathbf{x} - \mathbf{b}_{K-1} \right) + \mathbf{s}_0 - \mathbf{x} \right\|_2 \\ &\geq \sqrt{1 - (d-1)\nu} \|\mathbf{b}[K]\|_2 - \left\| \mathbf{D}_{K-1} \left(\mathbf{D}_{K-1}^\dagger \mathbf{x} - \mathbf{b}_{K-1} \right) + \mathbf{s}_0 - \mathbf{x} \right\|_2 \end{aligned} \quad (3.30)$$

where $\mathbf{b}_{K-1}$ comprises the $K - 1$ largest blocks of $\mathbf{b}$. The second term in (3.30) is bounded by

$$\begin{aligned} & \left\| \mathbf{D}_{K-1} \left(\mathbf{D}_{K-1}^\dagger \mathbf{x} - \mathbf{b}_{K-1} \right) + \mathbf{s}_0 - \mathbf{x} \right\|_2 \\ &= \left\| \mathbf{D}_{K-1} \mathbf{D}_{K-1}^\dagger (\mathbf{D}[K] \mathbf{b}[K] - \mathbf{s}_0 + \mathbf{x}) + \mathbf{s}_0 - \mathbf{x} \right\|_2 \\ &\leq \left\| \mathbf{D}_{K-1} (\mathbf{D}_{K-1}^H \mathbf{D}_{K-1})^{-1} \right\|_2 \left\| \mathbf{D}_{K-1}^H \mathbf{D}[K] \mathbf{b}[K] \right\|_2 + \left\| \mathbf{I} - \mathbf{D}_{K-1} \mathbf{D}_{K-1}^\dagger \right\|_2 \|\mathbf{s}_0 - \mathbf{x}\|_2 \\ &\leq \frac{\sqrt{(K-1)d\mu_B \|\mathbf{b}[K]\|_2}}{\sqrt{1 - [(K-2)d\mu_B + (d-1)\mu_L]}} + \varepsilon. \end{aligned} \quad (3.31)$$

The last inequality holds due to Lemma 2, the fact that $\mathbf{I} - \mathbf{D}_{K-1} \mathbf{D}_{K-1}^\dagger$ is a projection matrix, and the inequality

$$\left\| \mathbf{D}_{K-1}^H \mathbf{D}[K] \mathbf{b}[K] \right\|_2 \leq \sqrt{\sum_{i=1}^{K-1} \|\mathbf{D}^H[i] \mathbf{D}[K] \mathbf{b}[K]\|_2^2} \leq \sqrt{(K-1)d\mu_B} \|\mathbf{b}[K]\|_2. \quad (3.32)$$

Let us define $\|\mathbf{b}\|_{2,\min} = \min_{i \in \Gamma(\mathbf{b})} \|\mathbf{b}[i]\|_2$, then $\|\mathbf{b}[K]\|_2 = \|\mathbf{b}[\Gamma(\mathbf{b})]\|_2 = \|\mathbf{b}\|_{2,\min}$. By substituting (3.30) into (3.29) and replacing $\|\mathbf{b}[K]\|_2$ with $\|\mathbf{b}\|_{2,\min}$, we obtain

$$\sqrt{1 - (d-1)\nu} \|\mathbf{b}\|_{2,\min} - \frac{\sqrt{(K-1)d\mu_B} \|\mathbf{b}\|_{2,\min}}{\sqrt{1 - [(K-2)d\mu_B + (d-1)\nu]}} > \delta^{\text{BOGA}} + \varepsilon \quad (3.33)$$

Proof: (b)

By Lemma 2, the representation error can be bounded by

$$\|\hat{\mathbf{b}}_{\text{BOGA}} - \mathbf{b}\|_2^2 = \|(\mathbf{D}^*)^\dagger \mathbf{x} - \mathbf{b}\|_2^2 = \|(\mathbf{D}^*)^\dagger \mathbf{w}\|_2^2 \leq \frac{\varepsilon^2}{1 - [(K-1)d\mu_B + (d-1)\nu]} \quad (3.34)$$

Corollary1. Suppose $\delta^{\text{BOMP}} \geq \varepsilon$ and $K = |\Gamma(\mathbf{b})|$ satisfies (3.16). If

$$\|\mathbf{b}\|_{2,\min} > \max \left\{ \frac{2\sqrt{1+(d-1)\nu}\varepsilon}{1-(d-1)\nu-2d\mu_B(1+d\mu_B/(1-(d-1)\nu))(K-1)}, \frac{(\delta^{\text{BOGA}}+\varepsilon)\sqrt{1-[(K-2)d\mu_B+(d-1)\nu]}}{\sqrt{1-[(K-2)d\mu_B+(d-1)\nu]}\cdot\sqrt{1-(d-1)\nu}-\sqrt{(K-1)d\mu_B}} \right\}, \quad (3.35)$$

the BOGA can completely recover the support of ideal block-sparse signals from noise-corrupted data.

Proof:

For $1 \leq l \leq K-1$, (3.28.(2)) is implied by

$$\begin{aligned} & (1 - (d-1)\nu) \|\mathbf{b}[\Gamma_l] + 1\|_2 \\ & \geq 2\sqrt{1 + (d-1)\nu}\varepsilon + 2d\mu_B \max_{1 \leq l \leq K-1} \left(\frac{(1+d\mu_B-(d-1)\nu)(K-l)}{1-(l-1)d\mu_B-(d-1)\nu} \right) \|\mathbf{b}[\Gamma_l] + 1\|_2 \\ & \geq 2\sqrt{1 + (d-1)\nu}\varepsilon + 2d\mu_B \left(\frac{(1+d\mu_B-(d-1)\nu)(K-1)}{1-(d-1)\nu} \right) \|\mathbf{b}[\Gamma_l] + 1\|_2. \end{aligned} \quad (3.36)$$

Rearrange the terms in (3.36), we get

$$\|\mathbf{b}[\Gamma_l] + 1\|_2 \geq \frac{2\sqrt{1 + (d-1)\nu}\varepsilon}{1 - (d-1)\nu - 2d\mu_B(1 + d\mu_B/(1 - (d-1)\nu))(K-1)}. \quad (3.37)$$

For $\|\mathbf{b}\|_{2,\min} = \min_{i \in \Gamma(\mathbf{b})} \|\mathbf{b}[i]\|_2$, (3.37) can be implied by

$$\|\mathbf{b}\|_{2,\min} \geq \frac{2\sqrt{1 + (d-1)\nu}\varepsilon}{1 - (d-1)\nu - 2d\mu_B(1 + d\mu_B/(1 - (d-1)\nu))(K-1)}. \quad (3.38)$$

Rearranging (3.18) and combining with (3.38), (3.35) is obtained.

Remark: a)

The above results are similar to those obtained in [2] with a completely different definition of μ_B . Comparing (3.35) with the recovery conditions in (12) obtained in Corollary 1 [2], we conclude that:

i. The block-coherence in this Chapter is smaller than that in [2], that is, $\mu_B \leq \mu_{B,L}$. Then the recovery condition in (3.16) allows a larger block-sparse degree K since

$$0.5d^{-1} (\mu_B^{-1} + d - (d-1)\nu\mu_B^{-1}) \geq 0.5d^{-1} (\mu_{B,L}^{-1} + d - (d-1)\nu\mu_{B,L}^{-1}).$$

ii. The representation error in (3.19) is smaller than that in [2] simply due to $\mu_B \leq \mu_{B,L}$, and

$$\|\hat{\mathbf{b}}_{\text{BOGA}} - \mathbf{b}\|_2^2 \leq \frac{\varepsilon^2}{1 - [(K-1)d\mu_B + (d-1)\nu]} \leq \frac{\varepsilon^2}{1 - [(K-1)d\mu_{B,L} + (d-1)\nu]}.$$

iii. Under the condition in (3.16), which is equivalent to $d\mu_B(2K-1) < 1 - (d-1)\nu$, we have

$$\begin{aligned} & \frac{2\sqrt{1+(d-1)\nu\varepsilon}}{1-(d-1)\nu-2d\mu_B(1+d\mu_B/(1-(d-1)\nu))(K-1)} \\ & \leq \frac{2\sqrt{1+(d-1)\nu\varepsilon}}{1-(d-1)\nu-d\mu_B(2K-1)} \leq \frac{2\sqrt{d\varepsilon}}{1-(d-1)\nu-d\mu_B(2K-1)} \\ & \leq \frac{2\sqrt{d\varepsilon}}{1-(d-1)\nu-d\mu_{B,L}(2K-1)}. \end{aligned}$$

Therefore, the first item on the right-hand side of (3.35) is smaller than the corresponding part of (12) in Corollary 1 [2].

iv. The second item on the right-hand side of (3.35) can be rewritten as:

$$\frac{(\delta^{\text{BOGA}} + \varepsilon)}{\sqrt{1 - (d-1)\nu} - \sqrt{(K-1)df(\mu_B)}}$$

where

$$f(\mu_B) = \frac{\mu_B}{\sqrt{1 - [(K-2)d\mu_B + (d-1)\nu]}}$$

is a function of μ_B . Under the recovery condition in (3.16), it is ready to prove that $f(\mu_B)$ is a monotonic increasing function of μ_B . Then the second term on the right-hand side of (3.35) is monotonic decreasing. We have:

$$\frac{(\delta^{\text{BOGA}} + \varepsilon)}{\sqrt{1 - (d-1)\nu} - \sqrt{(K-1)df(\mu_B)}} \leq \frac{(\delta^{\text{BOGA}} + \varepsilon)}{\sqrt{1 - (d-1)\nu} - \sqrt{(K-1)df(\mu_{B\perp})}}.$$

Thus, the second item on the right-hand side of (3.35) is also smaller than the corresponding part of (12) in Corollary 1 [2].

Based on our derivation, it is concluded that the recovery condition in (3.16) is relaxed since it allows a larger block-sparse degree. The recovery condition in **Corollary1** in (3.35) is also relaxed since it allows a smaller threshold on the minimum block energy $\|\mathbf{b}\|_{2,\min}$ of the coefficient vector. The representation error is also smaller.

Remark: b)

In the conventional sparse case, that is, $d = 1$, $\nu = 0$, and $\mu_B = \mu$, inequality (3.28.(1)) degenerates into

$$\|b_{|\Gamma_l|+1}\|_2 \geq 2 \left(\varepsilon + \left(\frac{\mu + \mu^2}{1 - \mu(l-1)} \right) \|\mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l}\|_1 \right) \quad (3.39)$$

where $\|b_{|\Gamma_l|+1}\|_2$ is the l_2 -norm of coefficient $b_{|\Gamma_l|+1}$, which can also be represented as $\max_{j \in \Gamma(\mathbf{b}) \setminus \Gamma_l} |b_j|$ or $\|\mathbf{b}_{\Gamma(\mathbf{b}) \setminus \Gamma_l}\|_\infty$. The result in (3.39) has also been reported in Theorem 3 [13].

Inequality (3.35) degenerates into

$$b_{\min} > \max \left\{ \frac{2\varepsilon}{1 - 2\mu(1+\mu)(K-1)}, \frac{1 - \mu(K-2) + \mu\sqrt{K-1}\sqrt{1 - \mu(K-2)}}{(1+\mu)(1+\mu-\mu K)} (\delta^{\text{BOGA}} + \varepsilon) \right\}$$

which is the recovery condition reported in Lemma 2 [13], where $b_{\min}$ is the minimum value of the coefficient vector $\mathbf{b}$. As shown in *Remark b)*, stability and the recovery conditions in **Theorem1** and **Corollary1** reported in this Chapter are a generation of those given in [13] for signals with block-sparsity.

Table 3.1: Bounds on minimal block energy versus noise power

Minimum Energy \ Noise Power	0	0.1	0.3	0.5	0.7	0.9	1
$K=1$ (ours)	0	0.2063	0.6189	1.0315	1.4441	1.8567	2.0630
$K'=1$ ([2])	0	0.3762	1.1287	1.8811	2.6335	3.3860	3.7622
$K=2$ (ours)	0	0.2999	0.8997	1.4996	2.0994	2.6992	2.9992
$K'=2$ ([2])	0	0.9536	2.8609	4.7681	6.6753	8.5826	9.5362

3.4 Numerical Experiments

In this section, simulations are conducted to show that under the recovery conditions given in this Chapter, the estimator based on the BOGA gives the correct signal support and the associated representation errors are within the bound in (3.19). The simulation parameters used for our experiments are: signal length $N = 100$, the size of each block $d = 2$, the number of blocks $L = 60$. The block dictionary is generated in the same way as described in [2] with $\mu_B = 0.0675$ and $\nu = 0.0206$, and value of the block-coherence computed by the definition in [2] is $\mu_{B,L} = 0.1138$ for the same dictionary. The condition in (3.16) requires that the block-sparse degree should satisfy $K < 0.5d^{-1}(\mu_B^{-1} + d - (d-1)\nu\mu_B^{-1}) = 4.1296$. The block-sparse degree K is relatively relaxed comparing to the block-sparse degree computed by $\mu_{B,L}$ i.e., $K_L < 0.5d^{-1}(\mu_{B,L}^{-1} + d - (d-1)\nu\mu_{B,L}^{-1}) = 2.6524$. Apparently, $K = 3$ and 4 violates the recovery condition in (3.16) provided by [2]. However the BOGA can still work efficiently as to be shown in our simulations. The coefficient vector $\mathbf{b}$ is randomly generated and satisfies the condition in (3.16), that is $K = 1, 2, 3, 4$. The positions of the nonzero blocks are randomly distributed and the entry values in each block are drawn from a complex normal distribution. The noise added follows zero-mean complex white Gaussian distribution and the power value varies from 0 to 1.0. For block-sparse degree $K = 1, 2$ and $\delta^{\text{BOGA}} = \varepsilon$, a comparison, as shown in Table 1, is made between our derived bound and that given in [2] on the minimum block energy $\|\mathbf{b}\|_{2,\min}$ of the coefficient vector to show the relaxation on the recovery condition. The ideal noiseless signals are also normalized to have a unit l_2 -norm. For each given block-sparse degree and noise power, 500 trials are performed in our simulations.

Figure 3.1: Representation errors obtained by using the BOGA and the OGA for $K=1, 2, 3$ and 4 . Dotted lines represent the bounds of Theorem 5.1 in [1]; solid lines depict the bounds of Theorem 1 in [2]; Cross lines depict the bounds of Theorem 1 in this Chapter.

Table 3.1 shows that, for $K = 1, 2$, the minimum block energy required in this Chapter for recovering the block-sparse signals is much lower than that reported in [2]. Clearly, minimal block energy of coefficient vector is not required when the BOGA is applied to recover the ideal noiseless block-sparse signals.

Let us now consider the performance in terms of representation errors, recovery signal errors and percentage of support recovery. For comparison, the performance of the OGA for signals with block-sparsity is also given.

Figure 3.1 shows the representation errors of the BOGA and OGA for different values of K . The theoretical representation error upper bounds of the BOGA derived in this Chapter and those reported in [2] and the bound of the OGA [1] are presented

Figure 3.2: Signal errors achieved by the BOGA and the OGA for $K = 1, 2, 3$ and 4. Dotted lines represent the input noise power.

Figure 3.3: Support recovery success achieved by the BOGA and OGA for $K=1, 2, 3$ and 4 .

by the cross lines, solid lines, and dotted lines, respectively. Although there is still a large gap between the worst experiment results and the three theoretical bounds, it is seen that both the representation error bounds for block-sparse model have a smaller gap than the conventional bound given in [1] and our error bound has a smaller gap than that given in [2]. Figure 3.2 presents the reconstructed signal errors $\|\mathbf{D}\hat{\mathbf{b}}_{BOGA} - \mathbf{D}\mathbf{b}\|_2$ to show the denoising effect of both methods. It is observed that even with the block-sparse degree $K=3$ and 4, the denoising effects of the BOGA decrease slightly comparing to those when $K=1$ and 2. Figure 3.3 shows the support recoveries of both methods for different values of K . The support recovery percentage drops as the noise power increases because the high noise power requires a large minimal block energy to fully recover the support. As expected, the support recovery percentage decreases as the value of K increases for the same noise power. From all the figures, it is seen that the BOGA has a better performance than the OGA when it is applied to signals with block-sparsity. Furthermore, the BOGA is still able to work efficiently for $K=3$ and 4, which violates the recovery condition given in [2].

3.5 Conclusion

This chapter studies the stable recovery conditions and the approximation error bound of the BOGA for noisy signals. Although the recovery conditions have already been studied in [2] for noisy block-sparse signals, the definition of block-coherence used in [2] is found to result in tight recovery conditions. The stable recovery conditions and the representation error bound of the BOGA with a different block-coherence defined in [32], which is always smaller than that used in [2], are studied. For signals with block-sparsity, it is shown that the theoretical stable recovery conditions presented in this chapter are more relaxed than those in [2] and the representation error bound is also tighter than those given in [2]. For conventional sparse signals, stable recovery conditions of the BOGA obtained are degenerated to those of the OGA reported in [13].

Chapter 4

Harmonic Tonal Detection Architectures based on BOGA

4.1 Introduction

The detection of acoustic signals produced by the motorized vessels or vehicles is needed for various applications. Acoustic signals in complex underwater environments are usually contaminated by ambient sounds due to seismic disturbances, oceanic turbulence, and various anthropogenic sources. For simplicity, the ambient noise is generally assumed to be a white Gaussian stationary process and the underwater sources, produced by rotating machineries such as shafts, blades and propellers, are typically modeled as a periodic envelope multiplying broadband noise. This audible phenomenon has been widely exploited to extract useful information, such as the number of blades and rotational rate [67]. Demodulated-noise (DEMON) processing is commonly used to obtain the required information based on the Fourier analysis, in which the received signals are presumed to have a harmonic line spectrum, i.e., a series of discrete lines or tonals that are harmonically related [68]. In [69], the received signal, $\mathbf{x}(t)$, is defined as cavitation noise, $\mathbf{s}(t)$, plus white Gaussian stationary ambient noise $\mathbf{v}(t)$:

$$\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{v}(t) = \mathbf{m}_\omega(t)\mathbf{c}(t) + \mathbf{v}(t)$$

where $\mathbf{m}_\omega(t)$ is the periodic modulating signal with fundamental frequency ω determined by the motion of the rotating machineries, and $\mathbf{c}(t)$ is assumed to be stationary and changes slowly with time t . Within a short time period, we can reasonably assume that $\mathbf{c}(t)$ is a constant. Since $\mathbf{m}_\omega$ is presumed to be periodic, the spectrum of the acoustic signal contains several harmonically related narrow-band peaks (or tonals) which can be modeled as sinusoids. Because engines typically have more than one rotating or vibrating elements, these harmonics appear in different blocks related to their fundamental frequencies determined by the period of the rotating elements.

Several types of tonal detectors exist in the literature. The most common technique, low frequency analysis and recording (LOFAR) [70], averages the spectra of data segments after noise-mean removal and normalization. Wagstaff's integration silence processor (WISPR) [71] achieves better performance by integrating the spectrum of each data segment nonlinearly based on the fluctuations between signal and noise. Unlike the LOFAR and WISPR, which perform simple integration, a generalized likelihood ratio test based on power spectrum (PGLRT) [72] optimally integrates the power spectra of data segments according to their statistical properties. Although the PGLRT is optimal from the perspective of statistical properties, it loses the phase information of the signal by squaring the amplitude of the discrete Fourier transform (DFT). The generalized likelihood ratio test based on coherent integration (CGLRT) [73] was then proposed to make full use of the phase information to achieve coherent signal integration for improvement on the detection performance.

Previously reported detectors only focus on the detection of a single spectrum line without sufficiently considering the harmonic structure. It is generally assumed that observation time is long enough such that incoherent and coherent integration can be applied to increase the SNR. This chapter considers more critical application environments with very low SNR and short observation duration, in which $\mathbf{c}(t)$ is considered to be constant. In such cases, most detectors based on long time integration cannot provide satisfactory performance. Conventionally, one may try

to make use of the concept of a generalized matched filter (GMF) to achieve the optimal detection performance under the assumption that the signal space is known [74]. However, the only prior information is that these tonals are harmonically related to their own fundamental frequencies. Since the number of the fundamental frequencies and the harmonic order associated with each fundamental frequency are unknown, it becomes impossible to implement the GMF technique.

This chapter shows that the harmonic tonals associated with different fundamental frequencies are block-sparse in nature. That is, only a few harmonic blocks of the tonals are nonzero. Our basic idea is to take the block-sparsity into consideration to find the support of the signal and then design the detector based on the estimation results. Block-sparsity has been recently exploited for many applications such as: face recognition [75], motion segmentation [76], multiband signal reconstruction [77], and harmonic decomposition of audio signals [41]. Algorithms, such as $l_{p,q}$ -norm based convex optimization, group lasso [31], block orthogonal greedy algorithm (BOGA) [2, 32, 33, 66, 78], as well as the block-sparse Bayesian learning (BSBL) algorithm [34], have been reported in the literature. In this chapter, the BOGA method is used due to the following two reasons. Firstly, since group lasso algorithm requires convex optimization and BSBL requires an iterative expectation maximization (EM) procedure, both methods are much more expensive in computation complexity compared to the BOGA. Secondly, since the BOGA method finds the signal support in a successive manner, it is easy to insert a sequential detection procedure into each step of the BOGA. Two BOGA based architectures are proposed. The first one is a subspace harmonic detector in which the BOGA is firstly applied to find the signal space and then the energy detector is used to detect the existence of the signal. By inserting a sequential detection procedure into each step of BOGA, the second one is a sequential harmonic detector based on order statistics. Both proposed detectors accumulate the energy from the fundamental frequencies and their associated harmonics to achieve improved tonal detection performance in low SNR environments. There are three main contributions of this chapter. The first one is to formulate the harmonic tonals by the block-sparse signal model and suggest using the block sparse recovery techniques in harmonic detection and

estimation. The second contribution is that two new detection architectures are proposed by modifying the BOGA which theoretically uses the maximal signal energy in the tonal detection with no prior information of the number of harmonic sets. The theoretical analysis of the test statistics and the detection performances using the theory of the order statistics are given comprehensively, which is the third contribution of this chapter.

This chapter is organized as follows. In the next section, the harmonic signal model, dictionary construction and a signal model with block-sparsity are presented. Based on the constructed dictionary, two detection architectures are proposed in Section 4.3. By studying the test statistics and thresholds under the null hypothesis, the performances of the proposed architectures are also analyzed in this section. Some issues on the proposed architectures are discussed in Section 4.4. The results of numerical simulations are given in Section 4.5 to show the effectiveness of the proposed detection architectures. The comparisons with the performances of the conventional detectors are also made in this section. Conclusion is given in Section 4.6

4.2 Signal Model

In this section, the model of the harmonic signals and its relation with the block-sparsity signal model are introduced.

4.2.1 Harmonic Signal Model

It is assumed that the received signals simultaneously contain K fundamental frequencies. During a short observation period, the tonal with a harmonic structure [69] can be expressed into:

$$\mathbf{x}(n) = \sum_{k=1}^K \sum_{m=1}^M a_{m,k} \exp(j(m\omega_k n + \theta_{m,k})) + \mathbf{v}(n) \quad (4.1)$$

where $n = 0, 1, \dots, N - 1$, ω_k is the k th fundamental frequency, K is the total number of fundamental frequencies, $a_{m,k}$ and $\theta_{m,k}$ are, respectively, the amplitude and the phase of the m -th harmonic of the k -th fundamental frequency, M is the maximum order of the harmonics and $\mathbf{v}(n)$ is the additive complex white Gaussian noise with a variance σ^2 . Because of the short duration of the signal, the amplitude, $a_{m,k}$, is assumed to be constant.

The model defined in (4.1) is expressed in matrix notation as

$$\mathbf{x} = \mathbf{s} + \mathbf{v} = \mathbf{D}_0 \mathbf{b}_0 + \mathbf{v}, \quad (4.2)$$

where

$$\begin{aligned} \mathbf{D}_0 &= [\mathbf{D}[1] \ \mathbf{D}[2] \ \dots \ \mathbf{D}[K]], \\ \mathbf{D}[k] &= \begin{bmatrix} 1 & \dots & 1 \\ e^{j\omega_k} & \dots & e^{jM\omega_k} \\ \vdots & \ddots & \vdots \\ e^{j\omega_k(N-1)} & \dots & e^{jM\omega_k(N-1)} \end{bmatrix}, \\ \mathbf{b}_0 &= \underbrace{[b_1 \ \dots \ b_M]}_{\mathbf{b}^T[1]} \underbrace{[b_{M+1} \ \dots \ b_{2M}]}_{\mathbf{b}^T[2]} \dots \underbrace{[b_{(K-1)M+1} \ \dots \ b_{KM}]}_{\mathbf{b}^T[K]}^T, \\ b_{km} &= a_{m,k} e^{j\theta_{m,k}}. \end{aligned}$$

In the above equations, harmonic structure is included in the observation matrix $\mathbf{D}_0$, which contains K column-blocks, i.e., sub-matrix $\mathbf{D}[k]$ for $k = 1, \dots, K$, corresponding to the k th fundamental frequency. This signal model allows us to develop detection algorithms utilizing the inherent harmonic structure.

To reformulate the model into a standard block-sparse signal model, a proper dictionary with a block structure is constructed. Similar to the time-frequency analysis, signals are represented as a product between the corresponding coefficient vector and a dictionary whose atoms are represented by Gabor atoms, wavelets, or Fourier basis. As shown in (4.2), the atoms used to represent the harmonics will be the Fourier atoms. Without any prior information about the fundamental frequencies, an over-complete dictionary is constructed to include all the possible

fundamental frequencies, which unavoidably results in a large dictionary and more computation complexity. In this section, we attempt to simplify the dictionary construction for practical applications by coarsely estimating possible fundamental frequencies before other processing steps are performed.

4.2.2 Dictionary Construction

The size of the sub-dictionary is determined by the maximum harmonic order which is assumed to be known since it can be determined from the properties of the sound sources in a particular application. The exact harmonic order of each fundamental frequency is not necessarily required because our detection methods are not sensitive to the inaccuracies of the maximum harmonic order. Compared with the Fourier transform, the sparse representation can achieve high resolution by setting narrow frequency bins in the fundamental frequency range given that the signal has sparse representation. However, narrow frequency bins lead to high dictionary coherence, defined by $\mu = \max_{r \neq l} (\mathbf{d}_r^H \mathbf{d}_l)$, which results in a strict constraint of low sparsity, as shown in [1, 79, 80]. For example, the normalized Fourier basis has atoms

$$\begin{aligned} \mathbf{d}_l(k) &= \exp\{-j2\pi l(k-1)/(hN)\}/\sqrt{N}, \\ k &= 1, \dots, N, \quad l = 1, \dots, hN. \end{aligned}$$

When the width of the frequency bin is $2\pi/N$, i.e., $h = 1$, the dictionary, $\mathbf{D}$, becomes the same as that of the conventional Fourier transform, which is ideal for the sparse recovery due to its orthogonality. When $h = 2$, however, the narrower frequency bin results in an over-complete dictionary with a much higher coherence $\mu = 1/(N \sin(N\pi))$ [12]. This issue has been dealt with in [80] to achieve super-resolution by using over-sampled DFT frame and model based sparse signal recovery algorithm. Since this chapter mainly focuses on the detection of the harmonics rather than recovering exact signal frequencies, the dictionary is chosen to be orthogonal, i.e., $h = 1$.

So far, the possible fundamental frequencies are still unknown for dictionary construction. In some applications, the range of the fundamental frequencies can

be approximately determined by the velocity of the signal sources. Assuming a known fundamental frequency range, the dictionary can be constructed by a set of fundamental frequencies evenly distributed in the range. If such information is unknown, a general solution is to estimate the fundamental frequencies from the data.

4.2.2.1 Fundamental Frequency Estimator

The maximum likelihood (ML) estimator, finding the spectrum peak of the input signal, is the optimum estimator in the presence of white Gaussian noise. However, this estimator fails in extremely low SNR environments since it relies on the energy from only one tone. Here we present a method that will accumulate the energy from all the harmonics of a fundamental frequency. To coarsely estimate the fundamental frequencies, a maximum likelihood estimator making use of the harmonic structure is introduced.

Let us assume that there is only one fundamental frequency ω in the signal. Then $\mathbf{D}_0$ in (4.2) degenerates to $\mathbf{D}_\omega$. Based on this inherent harmonic structure, a variable, $E = (\mathbf{D}_\omega \mathbf{b})^H (\mathbf{D}_\omega \mathbf{b})$, is defined to represent the coherent integration of the energy from each harmonic component. Since both ω and $\mathbf{b}$ are unknown, it is necessary to estimate them beforehand. Assuming the noise follows a complex Gaussian distribution, we have

$$p(\mathbf{x}; \mathbf{b}, \sigma^2, \omega) = \frac{1}{\pi^N \sigma^{2N}} \exp \left[-\frac{1}{\sigma^2} (\mathbf{x} - \mathbf{D}_\omega \mathbf{b})^H (\mathbf{x} - \mathbf{D}_\omega \mathbf{b}) \right]. \quad (4.3)$$

If ω is known, the MLE of $\mathbf{b}$ can be found by minimizing the following cost function:

$$J(\mathbf{b}) = (\mathbf{x} - \mathbf{D}_\omega \mathbf{b})^H (\mathbf{x} - \mathbf{D}_\omega \mathbf{b}). \quad (4.4)$$

Therefore the MLE of $\mathbf{b}$ can be represented by

$$\hat{\mathbf{b}} = (\mathbf{D}_\omega^H \mathbf{D}_\omega)^{-1} \mathbf{D}_\omega^H \mathbf{x}. \quad (4.5)$$

Putting (4.5) into (4.4), the cost function becomes

$$J(\hat{\mathbf{b}}) = \mathbf{x}^H (I - \mathbf{D}_\omega (\mathbf{D}_\omega^H \mathbf{D}_\omega)^{-1} \mathbf{D}_\omega^H) \mathbf{x}. \quad (4.6)$$

Then, ω is found by minimizing (4.6), or equivalently maximizing

$$\mathbf{x}^H (\mathbf{D}_\omega (\mathbf{D}_\omega^H \mathbf{D}_\omega)^{-1} \mathbf{D}_\omega^H) \mathbf{x}. \quad (4.7)$$

Submitting (4.5) and (4.7) into $E = (\mathbf{D}_\omega \mathbf{b})^H (\mathbf{D}_\omega \mathbf{b})$, we obtain the total estimated harmonic energy for a fundamental frequency ω :

$$\begin{aligned} \hat{E} &= \max_{\omega} (\mathbf{x}^H (\mathbf{D}_\omega (\mathbf{D}_\omega^H \mathbf{D}_\omega)^{-1} \mathbf{D}_\omega^H) \mathbf{x}) \\ &= \max_{\omega} (\mathbf{x}^H \mathbf{P}_{\mathbf{D}_\omega} \mathbf{x}) \end{aligned} \quad (4.8)$$

where $\mathbf{P}_{\mathbf{D}_\omega}$ represents the projection matrix onto the signal space spanned by the columns of $\mathbf{D}_\omega$, and $\mathbf{x}^H \mathbf{P}_{\mathbf{D}_\omega} \mathbf{x}$ represents the projection energy of the observations upon the space. From (4.8), it is concluded that the summation of the energy from all harmonic components of a fundamental frequency can be obtained by searching for the maximum value of $\hat{E}$, and the corresponding frequency bin is the MLE of the fundamental frequency ω . A similar idea of combining the harmonic structure into pitch tracking has been addressed in [81][82]. It is shown in [81] that the variance of the pitch estimation by considering the harmonic structure is significantly lower than that obtained by considering the fundamental frequency alone. However, the existence of multiple fundamental frequencies may lead to false peak estimation and sometimes the true peaks are submerged by noise in the low SNR environment. In general, a number of peaks are estimated simultaneously to ensure that the estimates cover all the possible fundamental frequencies.

4.2.3 Block-sparse signal Model

Without loss of generality, let us assume that L peaks are estimated in a set $[\omega_1 \ \omega_2 \ \dots \ \omega_L]$. Based on the estimated fundamental frequencies, a dictionary $\mathbf{D}$ composed of L blocks is constructed to represent the original signal:

$$\mathbf{D} = [\mathbf{D}[1] \ \dots \ \mathbf{D}[L]], \quad (4.9)$$

where

$$\mathbf{D}[l] = [\mathbf{d}_{1l} \ \dots \ \mathbf{d}_{Ml}],$$

$$\mathbf{d}_{ml} = \exp(j2\pi m\omega_l n), \quad 1 \leq m \leq M.$$

The coefficient vector is represented by $\mathbf{b}$, which is a concatenation of sub-blocks $\mathbf{b}[l]$, denoting the l th block of size M , that is:

$$\mathbf{b} = [\mathbf{b}[1] \ \dots \ \mathbf{b}[L]]^T \quad (4.10)$$

where

$$\mathbf{b}[l] = [b_{1l} \ \dots \ b_{Ml}]^T.$$

Then (4.2) becomes

$$\mathbf{x} = \mathbf{D}\mathbf{b} + \mathbf{v}. \quad (4.11)$$

Comparing (4.2) with (4.11), the K nonzero blocks of coefficient vector $\mathbf{b}$ correspond to the K fundamental frequencies in the signal. It is suggested that the original time-domain signal $\mathbf{s}$ defined in (4.1) is a block-sparse signal with block-sparse degree of K when it is represented on dictionary $\mathbf{D}$ since $K \ll L$. Therefore, amplitude vector $\mathbf{b}_0$ is the support of the coefficient vector $\mathbf{b}$ and the observation matrix $\mathbf{D}_0$ is the corresponding part in $\mathbf{D}$. This complex block-sparse signal model is quite a good fit to the harmonic tonal, which is one of the contributions of this chapter. By estimating the complex coefficient vector $\mathbf{b}$ based on the block sparse signal recovery techniques, the information of harmonic tonal can be totally retrieved. In the remaining parts, the block-sparse structure is exploited to show its potentials in weak signal detection.

4.3 Harmonic Signal Detection

Two architectures are developed for detecting the harmonic signals with multiple fundamental frequencies based on the BOGA and their detection performances are analyzed theoretically in this section. The proposed architectures create test statistics gradually in steps of BOGA to accumulate signal energy including that of the

higher order harmonics of each fundamental frequency. These architectures require to know the noise variance and are suitable for the scenario that additional data are available to estimate the noise power σ^2 .

In the proposed architectures, complex coefficient vector $\mathbf{b}$ is assumed to be unknown but determinative. Generally, it is a binary hypothesis testing problem:

$$H_0 : \mathbf{b} = 0; \quad H_1 : \mathbf{b} \neq 0.$$

where H_0 is the noise-only hypothesis and H_1 is the signal-present hypothesis.

4.3.1 BOGA Description

BOGA is an extension of orthogonal matching pursuit (OMP) and has been reported to successfully recover the block-sparse signal in a recursive manner in both noisy and noiseless environments [32]. The BOGA starts with an initial residual $\mathbf{r}_0 = \mathbf{x}$ and finds a low-dimensional block support at each iteration step. The BOGA is terminated once the residual signal energy is equal or below a threshold, i.e., δ^{BOGA} . If the signal is corrupted by bounded noise, it is stated in [2, 78] that for signal with K fundamental frequencies, the BOGA selects one correct block from the dictionary at each step and stops exactly at K steps when δ^{BOGA} is equal or larger than the energy of the bounded noise. If the noise is unbounded, such as the white Gaussian noise, it is shown in [66] that the BOGA selects the first K correct blocks with a high probability and stops after the K th step under the assumption that K is known. When K is unknown, the stopping criterion that BOGA is terminated once the residual signal energy is equal or below a threshold, i.e., $\delta^{BOGA} = \beta N \sigma^2$, is often used in practical applications, where β is a user selected parameter. In this chapter, we assume that the BOGA stops at the K' th step, where the value of K' can be determined in two ways. One way is to select K' as a fixed number based on the prior information on the number of fundamental frequencies. When there is no such a prior information, K' will be given randomly by the criterion that BOGA stops when residual signal energy is equal or below δ^{BOGA} , i.e., $\|\mathbf{r}_l\|_2^2 \leq \beta N \sigma^2$. To the best of the authors' knowledge, there has been no theoretical analysis on the performance of BOGA when this criterion is used under Gaussian noise.

4.3.2 Architecture 1

The details of the architecture are given below in Algorithm 2.

Algorithm 2 Detection architecture 1

Initialize the residual $\mathbf{r}_0 = \mathbf{x}$, and step counter $l = 0$.

while $l \leq K'$ **do**

$$i_l = \arg \max_i \|\mathbf{D}[i]^H \mathbf{r}_{l-1}\|_2^2$$

$$T_l = \|\mathbf{D}[i_l]^H \mathbf{r}_{l-1}\|_2^2$$

Estimate the coefficient for the selected blocks:

$$\hat{\mathbf{b}} = \mathbf{D}_l^\dagger \mathbf{x}$$

Update the residual:

$$\mathbf{r}_l = \mathbf{x} - \hat{\mathbf{x}}_l$$

end while

Create test statistic T_{arch1}

$$T_{arch1} = T_1 + T_2 + \cdots + T_{K'}$$

Make decision

$$T_{arch1} \underset{H_0}{\overset{H_1}{>}} \gamma_{arch1}$$

where γ_{arch1} is the threshold of architecture 1.

A similar idea has been used for incoherent detection and estimation algorithm (IDEA) [83] in which matching pursuit (MP) is used to estimate the coefficient vector and the decision is simply made by $\|\hat{\mathbf{b}}\|_\infty > \gamma$, where γ is the threshold. Architecture 1 in Algorithm 2 can be thought as an extension of IDEA to block-sparse signal with a different form of test statistic. In each step of the BOGA, the partial signal energy is recorded after selecting one correct block. Since the BOGA algorithm never chooses the same block more than once, summation of signal energy in K' steps, which is the test statistic used in architecture 1, i.e., $T_{arch1} = \sum_{l=1}^{K'} T_l$, yields the estimation of the total signal energy.

The theoretical performance of architecture 1 is studied under different selections of K' in section B.1. When K' is fixed and known a priori, the distribution T_{arch1} is studied in sections B.1.1 and B.1.2. The theoretical detection probability (PD) under the assumption that BOGA selects the correct blocks in first K steps and the probability of false alarm (PFA) for fixed K' is given in (B.8), (B.10) and (B.13) to approximately evaluate the performance of architecture 1. They are summarized below:

Under H_0 , $T'_{arch1} = \frac{T_{arch1}}{N\sigma^2/2}$ is the sum of K' largest order statistics in a sample of L observations from χ^2_{2M} distribution. The PFA is given by (B.8) presented below

$$P_{FA} = \int_{\frac{2\gamma_{arch1}}{N\sigma^2}}^{\infty} f_{T'_{arch1}}(x) dx \quad (4.12)$$

where $f_{T'_{arch1}}(x)$ is given by

$$f_{T'_{arch1}}(x) = 1/2 f_{K'}(2x) \quad (4.13)$$

and $f_{K'}(2x)$ can be found in (B.6).

Under H_1 , assume that BOGA selects the correct signal blocks in the first K steps and that $K' \geq K$. When $K' = K$, the test statistic follows noncentral chi-squared distribution with density function $f_{\chi^2_{2KM}}(\lambda)$, where $\lambda = \frac{N\|\mathbf{b}_0\|_2^2}{\sigma^2/2}$. The PD of architecture 1 is given by (B.10) presented below

$$P_D = \int_{\gamma}^{\infty} f_{\chi^2_{2KM}(\lambda)}(x) dx \quad (4.14)$$

where $\gamma = \frac{2\gamma_{arch1}}{N\sigma^2}$. When $K' > K$, the density function $f_{T'_{arch1}}$ is the convolution of density functions of two independent distributions, given by

$$f_{T'_{arch1}}(y) = \int f_{\chi^2_{2KM}(\lambda)}(x) f_{T^2_{arch1}}(y-x) dx \quad (4.15)$$

where $f_{T^2_{arch1}} = 1/2 f_{K'-K}(2x)$ with $f_{K'-K}$ given in (B.6). The PD of architecture 1 when $K' > K$ is given by (B.13) presented below

$$P_D = \int_{\gamma}^{\infty} f_{T'_{arch1}}(y) dy. \quad (4.16)$$

When K' is randomly determined by the stopping criterion, $\|\mathbf{r}_l\|_2^2 \leq \beta N\sigma^2$, the theoretical distribution of test statistic, T_{arch1} , is hard to find. To determine the threshold, this distribution under H_0 is found by Monte Carlo method. It is seen in section B.1.3 that the density function of the test statistic under the null hypothesis is a linear combination of order statistics conditioned on K' with coefficients given by the probability of mass function of K' . Assume the conditional distribution of

T_{arch1} on $K' = l$ and $\Pr\{K' = l\}$ are known and are represented by $f_{T_{arch1}|K'=l}$ and $C_l(\beta)$. The PFA of architecture 1 for random K' is given by (B.18) presented below

$$P_{FA} = \int_{\frac{2\gamma_{arch1}}{N\sigma^2}}^{\infty} f'_{T_{arch1}}(x) dx \quad (4.17)$$

with

$$f'_{T_{arch1}}(x) = \sum_{K'} f'_{T_{arch1}|K'=k}(x) \Pr(K' = k) . \quad (4.18)$$

4.3.3 Architecture 2

Detection architecture 1 tries to estimate the support of the signal and uses the estimated signal energy as the test statistic. The detection performance depends on the selection of K' . To alleviate this dependency, a detector, known as architecture 2, is proposed to sequentially integrate the energy of the signal in the successive steps of BOGA. As shown in Algorithm 3, a decision is made in each step of BOGA. Once

Algorithm 3 Detection architecture 2

Initialize the residual $\mathbf{r}_0 = \mathbf{x}$ and step counter $l = 0$.

while $l \leq K'$ **do**

 Selected a block and construct:

$$i_l = \arg \max_i \| \mathbf{D}[i]^H \mathbf{r}_{l-1} \|_2^2$$

$$T_l = \| \mathbf{D}[i_l]^H \mathbf{r}_{l-1} \|_2^2$$

$$\hat{\mathbf{b}} = \mathbf{D}_l^\dagger \mathbf{x}$$

 Update the residual:

$$\mathbf{r}_l = \mathbf{x} - \hat{\mathbf{x}}_l$$

if $T_{arch2} = \sum_{i=1}^l T_i > \gamma_l$ **then**

 Choose H_1 and exit the loop

else

$$l = l + 1$$

end if

end while

 Choose H_0 , where γ_l is the threshold at step l .

the decision H_1 is made, the detection process is terminated directly. Otherwise, if decision H_0 is made at l -th step, estimated block energy T_{l+1} is added to T_{arch2} and

a new decision is made for step $l + 1$. Decision on H_0 is made if $l \leq K'$. Similarly, K' can be set as a fixed value based on the prior on the number of fundamental frequencies or be determined randomly by stopping criterion $\|\mathbf{r}_l\|_2^2 \leq \beta N \sigma^2$.

The thresholds for the steps of architecture 2 can be determined theoretically by finding the distribution of test statistics under null hypothesis if K' is fixed. The derivations are given in section B.2.1. The theoretical probability bound of the false alarm for this case is given by (B.21) presented below

$$\begin{aligned} P_{\text{FA}} &= \sum_{l=1}^{K'} P_{\text{FA},l} \\ &\leq Q_{f_1} + (1 - Q_{f_1}) Q_{f_2} + \cdots + (1 - Q_{f_1}) \cdots (1 - Q_{f_{K'-1}}) Q_{f_{K'}} \end{aligned} \quad (4.19)$$

where $Q_{f_l}(\frac{\gamma_l}{N\sigma^2}) = \int_{\frac{\gamma_l}{N\sigma^2}}^{\infty} f_l(x) dx$ for $1 \leq l \leq K'$ is the right-tail probability of a random variable with the PDF f_l in (B.6). When $K' = 1$, the equality in (4.19) holds strictly.

When K' is randomly determined by the stopping criterion as it is used in architecture 1, the distribution of the test statistic T_{arch2} under H_0 can be found by Monte Carlo method to determine the threshold at each step. The final probability bound of false alarm for this case under the assumption that the conditional distribution of $T_{\text{arch2}}|K' = l$ and $\Pr\{K' = l\}$ are known is given in section B.2.2:

$$\begin{aligned} P_{\text{FA}} &= \sum_k P_{\text{FA},k} |_{K'=k} \Pr(K' = k) \\ &\leq c_1(\beta) Q_{f'_1} + c_2(\beta) (Q_{f'_1} + (1 - Q_{f'_1}) Q_{f'_2}) + \cdots \\ &\quad + c_l(\beta) \{Q_{f'_1} + (1 - Q_{f'_1}) Q_{f'_2} + \cdots + (1 - Q_{f'_1}) \cdots (1 - Q_{f'_{k-1}}) Q_{f'_k}\} + \cdots \end{aligned} \quad (4.20)$$

where $Q_{f'_l}$, given in (B.25), is the right tail function of the density function of T_{arch2} at the l -th step.

The theoretical detection probability of architecture 2 can not be determined theoretically. We evaluate its detection probabilities under both ways of selecting K' by Monte Carlo method in the simulations.

4.4 Discussion

4.4.1 The effect of β on the distribution of test statistic

The test statistic distributions of both architectures under the stopping criterion $\|\mathbf{r}_l\|_2^2 \leq \beta N \sigma^2$ depend on β . As discussed in section B.1.3, β controls the PMF of K' . A small β will result in a high possibility of a relatively large K' . It leads to a heavy tail distribution of test statistic under H_0 hypothesis, thus resulting in a high threshold γ in order to keep a low probability of false alarm. On the contrary, a large β results in a relatively small K' , which gives a small threshold, but a high possibility that the BOGA stops before the K -th step when there are still signal components present in low SNR environments. Therefore, there must be an optimal value of β which yields the best detection performance. Unfortunately, neither the PMF of K' nor the distribution of test statistic under H_1 hypothesis can be found theoretically. We resort to the Monte Carlo method to select the optimal β .

4.4.2 Difference between the two proposed architectures

The detection architecture 1 tries to give an estimated number of fundamental frequencies and detect the signal based on the estimated signal subspace. In contrast, the detection architecture 2 detects the existence of the signal in each step of the BOGA. Once the decision on H_1 is made, the BOGA stops. Therefore architecture 2 is more efficient and saves computational time. However, the optimal thresholds in architecture 2 are difficult to find. We determine the threshold at each step of the BOGA by setting a constant false alarm probability in the current step to provide a theoretical false alarm probability bound of architecture 2 as a reference.

4.4.3 Model mismatch

A model mismatch occurs when the value of M used in the signal model is not equal to the true harmonic order of the received signal. When M is smaller than

the true harmonic order, the created test statistic contains only part of the signal energy. When M is larger, the test statistic contains the partial noise energy. Both situations result in a gracefully degraded detection performance as shown in the next section.

4.5 Simulation Results

In this section, two sets of simulations are conducted. The first set is conducted to estimate the PMF of K' and verify some theoretical derivations of the distributions of test statistics. The second set aims at evaluating the detection performances of the proposed architectures by Monte Carlo method.

4.5.1 PMF of K' and Verification of Equations in (B.23)

4.5.1.1 Estimate PMF of K' under H_0

As demonstrated in section B.1.3, only the probability of $K' = 0$, i.e., $C_0(\beta)$ can be found exactly. The Monte Carlo method is used to estimate the PMF of K' when $K' \neq 0$. As shown in section B.1.3, C_l depends on β and N only. With fixed β and N , any σ^2 results in the same PMF of K' . Therefore in our simulation, it is assumed the noise follows a complex white normal distribution with a length $N = 256$ and the sampling frequency is 256Hz. The block length is $M = 4$. We perform 30 experiments, each having 10000 independent realizations. In each experiment, the relative frequency of $K' = l$ is recorded and the mean of the relative frequencies of $K' = l$ in 30 experiments are computed as $\hat{C}_l$. The estimated PMF of K' with different values of β is given in Table 4.1, where C_0 is theoretically computed by (B.14).

As seen from Table 4.1, the estimated values of $\hat{C}_0$ s are quite similar to the theoretical ones. Since the theoretical C_0 is given, the estimated C_l can be further scaled to satisfy $\hat{C}_1 + \hat{C}_2 + \dots = 1 - C_0$. These scaled estimated $\hat{C}_l$ s are used to give the PDF of test statistics for both architectures.

Table 4.1: Estimated PMF of K'

	C_0	$\hat{C}_0$	$\hat{C}_1$	$\hat{C}_2$	$\hat{C}_3$	$\hat{C}_4$	$\hat{C}_5$	$\hat{C}_6$	$\hat{C}_7$	$\hat{C}_8$	$\hat{C}_9$
$\beta=0.95$	0.2926	0.29256	0.31013	0.23035	0.11329	0.04033	0.01040	0.00247	0.00041	0.000045	0.000003
$\beta=1.00$	0.5118	0.51209	0.28769	0.13863	0.04724	0.01169	0.00229	0.00030	0.00005	0.000013	0
$\beta=1.05$	0.7210	0.72076	0.19814	0.06365	0.01459	0.00254	0.00028	0.00003	0.000003	0	0
$\beta=1.10$	0.8697	0.86937	0.10455	0.02206	0.00352	0.00045	0.00005	0	0	0	0
$\beta=1.15$	0.9503	0.95025	0.04312	0.00599	0.00055	0.00008	0.000006	0	0	0	0
$\beta=1.20$	0.9845	0.98440	0.01408	0.00140	0.00013	0.000003	0	0	0	0	0
$\beta=1.25$	0.9960	0.99607	0.00366	0.00025	0.00002	0	0	0	0	0	0

4.5.1.2 Verification

To verify the equations in (B.23) by Monte Carlo method, we take T_1 as an example. As shown in Fig. 4.1, the estimated histogram of T_1 for fixed K' , denoted by $\hat{f}_{T_1}$ and the estimated histograms of T_1 for random K' under the conditions that $K' = 1, 2$, and 3 , denoted by $\hat{f}'_{T_1|K'=1}$, $\hat{f}'_{T_1|K'=2}$ and $\hat{f}'_{T_1|K'=3}$, respectively, are given when $\beta = 1.1$. Those estimated histograms $\hat{f}'_{T_1|K'=3}$, $\hat{f}'_{T_1|K'=2}$ and $\hat{f}'_{T_1|K'=1}$ are quite similar to each other, which shows the correctness of our derivations in (B.23). The curves marked by dash-dot give the theoretical density function of the largest order statistics without any constraint, which fits the estimated histogram of T_1 quite well for fixed K' . It is seen that the density function of T_1 for fixed K' does differ from $f'_{T_1|K'=1}$ due to the dependence of T_1 on random integer variable K' . As long as density functions $f'_{T_1+T_2+\dots+T_k|K'=k}$ and C_k are estimated, thresholds for architecture 1 and architecture 2 can be computed according to (B.18) and (B.32) for random K' with different values of β .

4.5.2 Detection performance evaluation

In this subsection, we first conduct a serial of simulations with different values of β to find the best one producing the highest detection probability. Then the performances of the proposed architectures, the single tone detector based on the Fourier power spectrum given in (B.36) and the energy detector given in (B.39) are compared to show the improvement obtained by exploring the sparse harmonic structure. The performances of the proposed architectures are also compared with that of the GLRT detector given in (B.33) to measure the performance loss. Thirdly, the comparisons of PFA of architecture 2 under different selection of K' , given in (B.29) and (B.31), with those obtained by Monte Carlo method are made to verify the theoretical derivations in Appendix B. Finally, we measure the performance loss due to model mismatch, i.e., incorrect information of harmonic order.

In our simulations, the false alarm rate is $P_{FA} = 0.03$ for architecture 1. To determine the threshold in each detection of architecture 2 for fixed K' and random

Figure 4.1: Density function comparisons. The histograms are obtained by the Monte Carlo method and the dash-dot curves are the theoretical density functions.

K' , respectively, $P_{\text{FA}}^g = 0.03$ and $P_{\text{FA}}^g = 0.2$ are used. The input signal for all the proposed architectures is sampled at a frequency $f_s = 256\text{Hz}$ and has a length of $N = 256$. Simulations are conducted using different signals which contain one ($K = 1$) and two ($K = 2$) harmonic sets. Fundamental frequencies are $\omega_1 = 11\text{Hz}$ and $\omega_2 = 37\text{Hz}$, respectively, with harmonic order $M = 4$. The SNR is defined as the ratio of signal power and noise power in a harmonic set, i.e., $\|\mathbf{D}[i]\mathbf{b}[i]\|_2^2/(N\sigma^2)$. It is also assumed that each harmonic set has the same SNR in our simulations.

4.5.2.1 Selection of β

Figures 4.2 (a) and (b) give the simulated detection probabilities obtained by architecture 1 with different values of β for $K = 1$ and 2, respectively. As seen from both figures, the detection performance does depend on β and the best detection performance is achieved when $\beta = 1.1$. Therefore, we chose $\beta = 1.1$ for all the following

simulations. It is also seen in Fig. 4.2 (a) and Fig. 4.2 (b), the detection probability of architecture 1 improves when the number of harmonic sets, K , increases.

Figure 4.2: Detection performances of architecture 1 for different values of β .

4.5.2.2 Detection comparisons

Figures 4.3(a) and (b) give the detection probabilities obtained by architecture 1 with different SNR values for $K = 1$ and 2, respectively. The dashed curves give the detection probabilities of architecture 1 for fixed K' obtained by the Monte Carlo method with the thresholds computed according to (B.8) when $K' = 1, 2$ and 3, respectively. The solid curves marked by the symbol, $+$, represents the detection probability of the architecture 1 for random K' when $\beta = 1.1$. As seen from both figures, the detection performance decreases when K' is larger than the true number of harmonic sets K . The detection probability of architecture 1 for random K' is a slightly lower than that for the fixed K' . However, it requires no prior information of the number of the harmonic sets.

Figures 4.3(c) and (d) give the detection probabilities of architecture 2 with different SNR values for $K = 1$ and 2, respectively. The dashed curves give the

detection probabilities of architecture 2 for fixed K' obtained by the Monte Carlo method with the thresholds computed according to (B.30) when $K' = 1, 2$ and 3 , respectively. The solid curves marked by the symbol, $*$, represents the detection probability of architecture 2 for random K' when $\beta = 1.1$. Similarly, the detection probability of architecture 2 for random K' is slightly lower than that for the fixed K' without requiring the prior information on the number of the harmonic sets. The detection performance of architecture 2 when K' is slightly larger than the true number of harmonic sets is similar to that when $K' = K$ with increased probability of false alarm, as seen in Fig. 4.7 on which more details on false alarm rates are to be presented later.

Figures 4.4 and 4.5 present the comparisons on the detection probabilities obtained by both architectures with various values of SNR for $K = 1$ and 2 , respectively. In Fig. 4.4, the solid and dashed curves marked by $*$ and $\circ$ show the detection probabilities of architectures 1 and 2 for fixed $K' = 1$ and random K' , respectively. The curves marked by $\bullet$, $\diamond$ and $\triangle$ represent the performances obtained by the GLRT given in (B.34), energy detector in (B.40) and single tone detector in (B.37). The detection performances achieved by architecture 1 and 2 for $K' = 1$ and the GLRT are quite similar. The detection performance of architecture 2 for random K' is better than that of architecture 1 and with a lower probability of false alarm, as shown in Fig. 4.7. The detection performances obtained by both architectures significantly outperform the energy detector and single tone detector. Similar conclusions can be made from Fig. 4.5 for $K = 2$ and $K' = 2$, where the dotted curve marked with $\square$ gives the detection probability obtained by the GMF for harmonics with known fundamental frequency. The detection curves of single-tone detector and the GMF are meant to give some sense of performance bounds: the single tone detector does not exploit the harmonic structure and therefore one would expect the proposed methods to do better, and the GMF gives the “best” when the harmonic structure is exploited and all the parameters are known. Figure 4.6 gives the approximated theoretical detection probabilities of architecture 1 in (B.13) for fixed K' when $K = 1$. The solid curves marked by $\circ$, $*$ and $\triangle$ are the theoretical approximations to the detection probabilities of architecture 1

for $K' = 1, 2$ and 3 , respectively. The dashed curves marked by $\circ, *$ and $\triangle$ give the corresponding detection probabilities of architecture 1 obtained by Monte Carlo simulations for $K' = 1, 2$ and 3 , respectively. The theoretical approximations of the detection probabilities are similar to those obtained by simulations when $K' = K$. The larger the difference between K' and K , the worse the detection probability in (B.13) approximates the detection performance.

It is concluded that although there are gaps between the performances obtained by the proposed architectures and the GMF, the proposed ones significantly outperform those detectors without making use of the harmonic (block) structure of the signals.

4.5.2.3 False alarm rates

Since there is no theoretical expression for the final false alarm probability of architecture 2, we evaluate the false alarm by Monte Carlo methods and make comparisons with the theoretical bounds of the false alarm rate derived in Appendix B. According to (B.29), the theoretical false alarm bounds for fixed K' with $P_{FA}^g = 0.03$ and $K' = 1, 2$ and 3 are $0.03, 0.0591$ and 0.0873 , respectively. Similarly, the theoretical false alarm bound for random K' with $P_{FA}^g = 0.2$ can be computed, according to (B.31), to be 0.031 . Figure 4.7 shows the false alarm probability of architecture 2. In Fig. 4.7(a), solid and dashed curves marked by $*, \circ$ and $\triangle$ represent the theoretical probability bounds of the false alarm rate and the simulated ones for $K' = 1, 2$ and 3 , respectively. Although the simulated probabilities of false alarm are below the theoretical ones, there are still large gaps. In Fig. 4.7(b), the theoretical bound and the simulated probability of false alarm are given by the solid and dashed lines, respectively. The false alarm probability of architecture 2 for random K' is smaller than 0.03 which is the false alarm rate in architecture 1. Therefore, for random K' , architecture 2 outperforms architecture 1 in terms of higher detection probability, lower probability of false alarm, and lower computation complexity.

Figure 4.3: Detection performance of proposed architectures.

4.5.2.4 Model mismatch

Figure 4.8 presents the detection probabilities of the proposed architectures when there is a model mismatch. In this simulation, the true harmonic order is 4, and the assumed harmonic order is $M = 3, 4$, and 5, respectively. There is only one harmonic set in the signal, i.e., $K = 1$, and $K' = 1$ is used when K' is fixed for both detection architectures. Dashed and solid curves represent the detection probabilities with fixed and random K' , respectively. The curves marked by $*$, $\triangle$ and $\circ$ give the detection probabilities for $M = 3, 4$ and 5. Comparing to the detection probability for $M = 4$, as discussed in section 4.4.3, there are slight performance losses when

Figure 4.4: Detection performance comparisons when $K = 1$. “arch1_f” refers to detection architecture 1 with K' fixed to 1; “arch1_r” refers to detection architecture 1 with K' randomly determined by the stopping criterion $\|\mathbf{r}_l\|_2^2 \leq \beta N \sigma^2$ with $\beta = 1.1$; similarly, “arch2_f” and “arch2_r” refer to detection architecture 2 with fixed and randomly K' , respectively.

Figure 4.5: Detection performance comparisons when $K = 2$.

 Figure 4.6: Detection performance comparisons when $K = 1$.

$M = 3$ and 5 for both architectures.

4.6 Conclusion

Target detection of mechanical harmonics using a block-sparsity model with an orthogonal dictionary was shown. Based on this new model, two detection architec-

Figure 4.7: False alarm probability of architecture 2

Figure 4.8: Performance loss evaluation due to model mismatch

tures are proposed for improved detection. Both of them presume the knowledge of the maximum harmonic order and the noise power. Theoretically, those architectures use the maximum signal energy with no prior knowledge of the number of the harmonic sets and achieve much better performance than the conventional single tone detector especially in low SNR environments. In addition, the proposed architectures try to estimate the signal subspace based on the prior that signal is block-sparse. Comparing to the energy detector in which the signal subspace is assumed unknown, the proposed architectures achieve better detection performances. The threshold used in the detection architectures are given theoretically and the corresponding theoretical detection performances are also provided based on the theory of order statistics under an orthogonal dictionary.

Chapter 5

Hierarchical Sparse Signal Recovery by Variational Bayesian Inference

5.1 Introduction

Hierarchical sparsity is a natural extension of common sparsity and block-sparsity, which can be used to model more generalized class of practical signals. These signals typically exhibit two levels of sparsity, i.e., block-sparsity among different blocks and internal sparsity inside an individual block. With a dictionary $\mathbf{D} \in \mathcal{R}^{N \times N}$ and a sensing matrix $\mathbf{A} \in \mathcal{R}^{M \times N}$, the hierarchical sparse signal can be represented by

$$\mathbf{y} = \mathbf{A}\mathbf{D}\mathbf{b} + \mathbf{w} = \mathbf{\Phi}\mathbf{b} + \mathbf{w}, \quad (5.1)$$

where $\mathbf{y} \in \mathcal{R}^M$ is the measurement vector, $\mathbf{b} \in \mathcal{R}^N$ is the sparse coefficient vector, and $\mathbf{w} \in \mathcal{R}^M$ is the additive noise vector. Let us denote $\mathbf{\Phi} = \mathbf{A}\mathbf{D}$ for simplicity. In block sparse modeling, coefficient vector $\mathbf{b}$ is partitioned into L blocks of length d_l for $l = 1, \dots, L$:

$$\mathbf{b} = [\underbrace{b_{1,1} \dots b_{d_1,1}}_{\mathbf{b}_1^T} \underbrace{b_{1,2} \dots b_{d_2,2}}_{\mathbf{b}_2^T} \dots \underbrace{b_{1,L} \dots b_{d_L,L}}_{\mathbf{b}_L^T}]^T,$$

where $\sum_{l=1}^L d_l = N$. The dictionary is partitioned accordingly into

$$\mathbf{D} = [\mathbf{D}_1 \ \mathbf{D}_2 \ \dots \ \mathbf{D}_L].$$

It is assumed that only K out of L blocks of $\mathbf{b}$ have nonzero norms. To further exert the internal sparsity in each block, it is assumed that only a few components in each block are nonzero, that is, only a few atoms in sub-dictionary $\mathbf{D}_l$ are active to represent the signals. Hierarchical sparse signals widely exist in audio and image signal modeling [42–44]. In [43], this hierarchical sparsity structure is used to separate textures from sample mixture.

Recovery methods for sparse and block-sparse signals have been extensively studied in [5, 8] and [31–33, 78], respectively. It has been shown that exploiting the block-sparsity leads to relaxed recovery conditions and smaller representation errors. However, they are not accurate enough to deal with the hierarchical sparse signals which inherently have both block and internal sparsity. The recovery algorithms of this kind of signals are recently considered in [42–44] via a convex optimization framework to solve the objective function:

$$\min_{\mathbf{b}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{b}\|_2^2 + \lambda_1 \|\mathbf{b}\|_{2,1} + \lambda_2 \|\mathbf{b}\|_1, \quad (5.2)$$

where $\|\mathbf{b}\|_{2,1} = \sum_{l=1}^L \|\mathbf{b}_l\|_2$, $\|\mathbf{b}_l\|_2 = \sqrt{\mathbf{b}_l^T \mathbf{b}_l}$, λ_1 and λ_2 are the regularization parameters. This method, denoted as HiLasso [43], works well with carefully selected regularization parameters. The main disadvantage of this method is the subjective tuning of the parameters. Hierarchical sparse signal recovery is also considered in a Bayesian framework [35]. The block-sparsity is encouraged by three different cluster patterns describing the neighborhood information of each component in the coefficient vector. However, this neighborhood information is too restrictive to be true since the components do not necessarily appear in each other’s neighborhood within one block in our considered problem.

To motivate a more general framework, the recovery of hierarchical sparse signals is re-addressed in a Bayesian framework. Firstly, each component of the coefficient vector is modeled as a Gaussian distributed variable with zero mean as in sparse Bayesian learning (SBL) [15, 19]. To impose the aforementioned two-level sparsity prior, we introduce two classes of hidden variables, i.e., block-wise variance and element-wise variance, which can be used to encourage sparsity between and within

blocks simultaneously. Then the covariance matrix of coefficient vector $\mathbf{b}$ is given by the multiplication of the block- and element-wise variances. Under this new model, the variational Bayesian inference (VBI) [84] is used to estimate the hidden variables. Finally, the coefficient vector is estimated by the mean of the posterior distribution.

The rest of this chapter is organized as follows. In Section 5.2, hierarchical sparsity is reformulated within the Bayesian framework and the variational Bayesian inference is used to estimate the hidden variables. Relationship with prior work is addressed in Section 5.3. Illustrative simulations and real-data experiments are presented in Section 5.4. Section 5.5 concludes this chapter.

5.2 Signal Model and Recovery

In Bayesian framework, we first introduce a new probabilistic model which enforces a hierarchical sparsity to the coefficient vector $\mathbf{b}$. Then under this new probabilistic model, rules for updating sparse signal and hidden variables are derived based on the VBI technique.

5.2.1 Probabilistic Signal Model

It is assumed that the noise is independent Gaussian distributed with zero mean and covariance matrix $a_0 \mathbf{I}_M$, where $\mathbf{I}_M$ is an identity matrix of M dimensions. Thus, measurement $\mathbf{y}$ follows

$$p(\mathbf{y} | \mathbf{b}; a_0) = \mathcal{N}(\mathbf{y} | \Phi \mathbf{b}, a_0 \mathbf{I}_M). \quad (5.3)$$

Let us introduce two kinds of variables of the covariance matrix of $\mathbf{b}$ to enforce the block and internal sparsity simultaneously. To exert the block sparsity, firstly, we assume all blocks $\mathbf{b}_l \in \mathcal{R}^{d_l}$ for $l = 1, \dots, L$ are mutually independent, and the probability density of each $\mathbf{b}_l$ is Gaussian given Λ_l and γ_l , i.e., $p(\mathbf{b}_l | \Lambda_l, \gamma_l) \sim \mathcal{N}(\mathbf{0}, \gamma_l \Lambda_l)$, where γ_l is a nonnegative hyperparameter controlling the block-sparsity

used in BSBL [34]. When $\gamma_l = 0$, the variance of the associated block $\mathbf{b}_l$ is zero, and thus the block is inactive. The number of nonzero, γ_l , for all l , determines the block-sparsity. We further assume that $\bar{\gamma} = [\gamma_1 \dots \gamma_L]^T$ follows a Gamma distribution

$$p(\bar{\gamma}) = \prod_{l=1}^L \Gamma(\gamma_l | 1, \lambda) = (\lambda)^L \exp \left(-\lambda \sum_{l=1}^L \gamma_l \right), \quad (5.4)$$

and $\lambda \sim \Gamma(v, v)$ with v being a very small value. Note that the prior of $\mathbf{b}$ is

$$p(\mathbf{b} | \{\Lambda_l, \gamma_l\}_{l=1:L}) \sim \mathcal{N}(\mathbf{0}, \Sigma_0)$$

where Σ_0 is a block-diagonal matrix with the l -th principal block given by $\gamma_l \Lambda_l$. As shown in [60], block sparsity is encouraged because this hierarchical modeling results in a multivariate generalized Laplace distribution on blocks given Λ_l :

$$\begin{aligned} p(\mathbf{b}_l | \Lambda_l, \lambda) &= \int_0^\infty p(\mathbf{b}_l | \Lambda_l, \gamma_l) p(\gamma_l | \lambda) d\gamma_l \\ &= \int_0^\infty \frac{\lambda}{\sqrt{|\Lambda_l|} (2\pi\gamma_l)^{\frac{d_l}{2}}} \exp \left(-\frac{\mathbf{b}_l^T \Lambda_l^{-1} \mathbf{b}_l}{2\gamma_l} - \lambda\gamma_l \right) d\gamma_l \\ &= \frac{2^{(3/2-d_l/4)} \lambda^{d_l/2+1} \kappa_{1-d_l/2} \left(\sqrt{2\lambda} \sqrt{\mathbf{b}_l^T \Lambda_l^{-1} \mathbf{b}_l} \right)}{\sqrt{(2\pi)^{d_l} |\Lambda_l|} \left(\sqrt{\mathbf{b}_l^T \Lambda_l^{-1} \mathbf{b}_l} \right)^{d_l/2-1}}, \end{aligned} \quad (5.5)$$

where κ is the modified Bessel function of the second kind.

To exert the internal sparsity, Λ_l is further assumed to be a diagonal matrix with diagonal components $\bar{\alpha}_l = [\alpha_{1,l} \dots \alpha_{d_l,l}]^T$ following a Gamma distribution with a density function

$$p(\bar{\alpha}_l) = \prod_{j=1}^{d_l} \Gamma(\alpha_{j,l} | 1, \beta_l) = (\beta_l)^{d_l} \exp \left(-\beta_l \sum_{j=1}^{d_l} \alpha_{j,l} \right), \quad (5.6)$$

where $\beta_l \sim \Gamma(u, u)$ with u being a very small value. Similar to block sparsity, we model $\mathbf{b}_l$ hierarchically to impose a Laplace prior on $\mathbf{b}_l$ given γ_l to provide a sparser estimation in individual block [19].

5.2.2 Hierarchical Sparse Signal Recovery

5.2.2.1 Updating Sparse Signal

Based on the probabilistic model proposed, given $\bar{\gamma}$ and $\bar{\alpha}_l$, the posterior for $\mathbf{b}$ is a Gaussian distribution

$$p(\mathbf{b} | \mathbf{y}, \bar{\gamma}, \{\bar{\alpha}_l\}_{l=1}^L) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (5.7)$$

with the mean and covariance given by

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}^T (\alpha_0 \mathbf{I}_M + \boldsymbol{\Phi} \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}^T)^{-1} \mathbf{y} \quad (5.8)$$

$$\begin{aligned} \boldsymbol{\Sigma} &= \left(\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\alpha_0} \boldsymbol{\Phi}^T \boldsymbol{\Phi} \right)^{-1} \\ &= \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}^T (\alpha_0 \mathbf{I}_M + \boldsymbol{\Phi} \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}^T)^{-1} \boldsymbol{\Phi} \boldsymbol{\Sigma}_0 \end{aligned} \quad (5.9)$$

where $\boldsymbol{\mu}$ gives the estimated $\mathbf{b}$.

To calculate $\boldsymbol{\mu}$, we need to estimate $\bar{\gamma}$ and $\{\bar{\alpha}_l\}_{l=1}^L$, and update their parameters λ and $\{\beta_l\}_{l=1}^L$. However, the exact Bayesian inference of these parameters are intractable due to the intractability of the posterior $p(\boldsymbol{\Theta} | \mathbf{y})$ where $\boldsymbol{\Theta} = \{\bar{\gamma}, \lambda, \{\bar{\alpha}_l\}_{l=1}^L, \{\beta_l\}_{l=1}^L\}$. In this chapter, the VBI techniques [24, 84] is employed to approximate the posterior and to estimate $\boldsymbol{\Theta}$.

An distribution $q(\boldsymbol{\Theta})$ with a factorized form is assumed to approximate the true posterior, which minimizes the Kullback-Leibler (KL) divergence between $q(\boldsymbol{\Theta})$ and the true posterior $p(\boldsymbol{\Theta} | \mathbf{y})$:

$$\begin{aligned} q^*(\boldsymbol{\Theta}) &= \arg \min_{q(\boldsymbol{\Theta})} \text{KL}(q(\boldsymbol{\Theta}) | p(\boldsymbol{\Theta} | \mathbf{y})) \\ &= \arg \min_{q(\boldsymbol{\Theta})} \left(\int q(\boldsymbol{\Theta}) \log \frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta} | \mathbf{y})} d\boldsymbol{\Theta} \right). \end{aligned} \quad (5.10)$$

According to [24, 84], when $q(\boldsymbol{\Theta})$ is of the factorization form:

$$\begin{aligned} q(\boldsymbol{\Theta}) &= q\left(\bar{\gamma}, \lambda, \{\bar{\alpha}_l\}_{l=1}^L, \{\beta_l\}_{l=1}^L\right) \\ &= q(\lambda) \prod_l \gamma_l \prod_l \beta_l \prod_{l,j} \alpha_{j,l}. \end{aligned} \quad (5.11)$$

the optimal distribution that minimizes (5.10) can be expressed as:

$$\ln q^*(\Theta_k) = \langle \ln p(\mathbf{y}, \Theta) \rangle_{q(\Theta \setminus \Theta_k)} \quad (5.12)$$

where $\langle \cdot \rangle_{q(\Theta \setminus \Theta_k)}$ denotes the expectation with respect to $q(\Theta \setminus \Theta_k)$ and $\Theta \setminus \Theta_k$ represents the set Θ without Θ_k .

5.2.2.2 Updating Hidden Variables

Based on the assumption in (5.11) and (5.12), the updating rules are derived as follows.

a). According to (5.12), for $q(\gamma_l)$, we have

$$\begin{aligned} \ln q^*(\gamma_l) &= \langle \ln p(\mathbf{b}_l | \Lambda_l, \gamma_l) p(\gamma_l | \lambda) \rangle_{q(\mathbf{b}_l)q(\bar{\alpha}_l)q(\lambda)} \\ &\propto -\frac{d_l}{2} \ln \gamma_l - \frac{1}{2} \left(\sum_{j=1}^{d_l} \langle \alpha_{j,l}^{-1} \rangle \langle b_{j,l}^2 \rangle \gamma_l^{-1} + \langle \lambda \rangle \gamma_l \right), \end{aligned} \quad (5.13)$$

where the posterior of γ_l is a generalized inverse Gaussian distribution. Therefore, the i -th moment of γ_l is given by [85]

$$\langle \gamma_l^i \rangle = \left(e/f \right)^{\frac{i}{2}} \times \frac{\kappa_{1-d_l/2+i}(\sqrt{ef})}{\kappa_{1-d_l/2}(\sqrt{ef})}, \quad (5.14)$$

where $e = \sum_{j=1}^{d_l} \langle \alpha_{j,l}^{-1} \rangle \langle b_{j,l}^2 \rangle$, $f = \langle \lambda \rangle$ and

$$\langle b_{j,l}^2 \rangle = \mu_{j,l}^2 + \Sigma_{(j,l)(j,l)}. \quad (5.15)$$

Therefore, γ_l and γ_l^{-1} are updated by $\langle \gamma_l^i \rangle$ with $i = 1$ and $i = -1$, respectively.

b). Similarly, for $q(\alpha_{j,l})$, we have

$$\begin{aligned} \ln q^*(\alpha_{j,l}) &= \langle \ln p(b_{j,l} | \alpha_{j,l}, \gamma_l) p(\alpha_{j,l} | \beta_l) \rangle_{q(b_{j,l})q(\gamma_l)q(\beta_l)} \\ &\propto -\frac{\ln \alpha_{j,l}}{2} - \frac{\langle \gamma_l^{-1} \rangle \langle b_{j,l}^2 \rangle \alpha_{j,l}^{-1} + \langle \beta_l \rangle \alpha_{j,l}}{2}. \end{aligned} \quad (5.16)$$

It is seen from (5.16) that $\alpha_{j,l}$ also obeys a generalized inverse Gaussian distribution with its i -th moment given by

$$\langle \alpha_{j,l}^i \rangle = \left(\frac{\langle \gamma_l^{-1} \rangle \langle b_{j,l}^2 \rangle}{\langle \beta_l \rangle} \right)^{\frac{i}{2}} \frac{\kappa_{0.5+i} \left(\sqrt{\langle \beta_l \rangle \langle \gamma_l^{-1} \rangle \langle b_{j,l}^2 \rangle} \right)}{\kappa_{0.5} \left(\sqrt{\langle \beta_l \rangle \langle \gamma_l^{-1} \rangle \langle b_{j,l}^2 \rangle} \right)}, \quad (5.17)$$

where $\langle \gamma_l^{-1} \rangle$ is given in (5.14) with $i = -1$ and $\langle b_{j,l}^2 \rangle$ is given in (5.15). Hence, $\alpha_{j,l}$ and $\alpha_{j,l}^{-1}$ can be updated by $\langle \alpha_{j,l} \rangle$ and $\langle \alpha_{j,l}^{-1} \rangle$, respectively. Subsequently, $\boldsymbol{\mu}$ in (5.8) and $\boldsymbol{\Sigma}$ in (5.9) can be updated accordingly with given $\langle \alpha_{j,l}^i \rangle$ and $\langle \gamma_l \rangle$.

c). For $q(\lambda)$, we have

$$\begin{aligned} \ln q^*(\lambda) &= \langle \ln p(\bar{\gamma} | \lambda) p(\lambda; v) \rangle_{q(\bar{\gamma})} \\ &\propto (L + v - 1) \ln(\lambda) - \left(v + \sum_{l=1}^L \langle \gamma_l \rangle \right) \lambda. \end{aligned} \quad (5.18)$$

Submitting (5.14) into (5.18), it is shown that the posterior of λ is a Gamma distribution and $\langle \lambda \rangle$ is given by

$$\langle \lambda \rangle = \frac{L + v}{v + \sum_{l=1}^L \langle \gamma_l \rangle}. \quad (5.19)$$

d). Similarly, β_l obeys a Gamma distribution and $\langle \beta_l \rangle$ is computed by

$$\langle \beta_l \rangle = \frac{d_l + u}{u + \sum_{j=1}^{d_l} \langle \alpha_{j,l} \rangle}. \quad (5.20)$$

Therefore, the proposed method, referred as HiSBL in the following, can be implemented by iteratively calculating (5.8), (5.9), (5.14), (5.17), (5.19) and (5.20).

5.3 Relation to Prior Work

This chapter focuses on the recovery of the hierarchical sparse signals in a Bayesian framework. Prior work in [42–44] has tried to recover hierarchical sparse signal in (5.1) by convex optimization techniques. However, optimal regularization parameter

selection under different scenarios remains unsolved. How to balance λ_1 and λ_2 to produce a satisfactory result is still a problem. We avoid this problem by formulating the recovery of the signal in a Bayesian framework. Following SBL [15, 19, 86] and BSBL [34], the hierarchical sparsity is encouraged via hierarchically modeling of two kinds of hidden variables to automatically control the block-sparsity and the internal sparsity, respectively.

The recovery of hierarchical sparse signal from Bayesian framework is also considered in [35], where the block-sparsity is implicitly exerted on the coefficient vector $\mathbf{b}$. It is assumed that nonzero components are in the neighborhood of each other to follow the predefined patterns. By this assumption, the algorithm can automatically estimate the partition of the blocks. However, it is also possible that the nonzero components in an active block are randomly located so that ClussVB in [35] gives a noisy estimation. Our proposed method does not require the neighborhood assumption and nonzero components in each block can be allocated randomly. However, like BSBL and HiLasso, HiSBL requires the prior knowledge of the block partition.

5.4 Simulations

The recovery performance of the proposed method is evaluated and compared with existing ones in terms of the mean square error (MSE), defined as:

$$\left\| \mathbf{b} - \hat{\mathbf{b}} \right\|_2^2 / \left\| \mathbf{b} \right\|_2^2$$

where $\hat{\mathbf{b}}$ is the recovered signal. It is assumed that Φ is a random Gaussian matrix, $\mathbf{b}$ is the signal with $L = 10$, $N = 150$, and each $\mathbf{b}_l$ has the same length $d = 15$. Direct performance comparisons of the methods with BSBL, SBL, ClussVB and HiLasso are conducted, as shown in Fig. 5.1. The signal has $K = 4$ nonzero blocks and each block has 4 nonzero components located randomly with a noise power 8×10^{-3} . The number of measurements is $M = 100$.

It is shown in Fig. 5.1(c) that the SBL method almost fails to recover the signal. Other methods work much better than the SBL, because they take the advantage

Figure 5.1: Recovered signals by various methods.

of block sparsity. Since BSBL does not exploit the internal sparsity, the recovered blocks remain noisy. The result obtained by the ClussVB is better than that by the

Figure 5.2: The MSEs of reconstructed signal versus the number of measurements.

BSBL; however, noise components are introduced into the active blocks due to the neighborhood assumption. In Fig. 5.1(b), the amplitude of the recovered signal is relatively low since the solution of HiLasso is a biased estimation for the amplitude and the good result can be obtained only by a suitable selection of regularization parameters. Compared to all those methods, our method gives the best recovered signal.

The proposed method is tested with different numbers of measurements, M , degrees of internal sparsity and noise power. In each test, 50 independent trails are carried out and the values of MSE are calculated by averaging the results of 50 trails. The signal parameters used in the simulations remain the same as those used for Fig. 5.1. Note that all the MSEs are plotted in log-scale.

Figure 5.2 gives the MSEs obtained by various methods versus different values of M with a fixed noise power 8×10^{-3} . The MSEs of all methods tend to be small as the number of the measurements used increases. Figure 5.3 gives the MSEs of different methods versus the degree of internal sparsity when $M = 100$ and noise power 8×10^{-3} . It is shown in Fig. 5.3 that a large internal sparsity results in increased MSEs for most methods, except the BSBL. It is noted in Fig. 5.3, the MSEs obtained by HiSBL and ClussVB tend to that of the BSBL as the internal

Figure 5.3: The MSEs of reconstructed signal versus internal sparsity.

Figure 5.4: The MSEs of reconstructed signal versus noise power.

sparsity increases. However, compared to that of the SBL and HiLasso, the MSE of HiSBL increases slowly as the internal sparsity changes. This behavior indicates that, in contrast to that of HiLasso, the HiSBL has the ability to automatically learn the sparsity of the dataset and adjusts itself to achieve a relatively low MSE. Figure 5.4 gives the MSEs of different methods versus noise power α_0 when $M = 100$. As the noise power increases, the MSEs obtained by all methods tend to be larger and all of them fail to give a relatively lower MSE under heavy noise case as shown in Fig. 5.4. It is shown in Figs. 5.2-5.4, HiLasso, ClussVB and HiSBL achieve lower MSEs than BSBL and SBL due to their utility of both the block sparsity and the internal sparsity.

The performances of different algorithms for variable selection in terms of detection rate (DR) and false alarm rate (FAR) are also compared. The DR is defined as

$$\left| \left\{ b_j \neq 0, \hat{b}_j \neq 0 \right\}_{j=1:N} \right| / \left| \{ b_j \neq 0 \}_{j=1:N} \right|$$

and FAR is defined as

$$\left| \left\{ b_j = 0, \hat{b}_j \neq 0 \right\}_{j=1:N} \right| / \left| \{ b_j = 0 \}_{j=1:N} \right|,$$

where $|\{\cdot\}|$ represents the cardinality of a set. We also define the block-wise FAR as

$$\left| \left\{ \mathbf{b}_l = 0, \hat{\mathbf{b}}_l \neq 0 \right\}_{l=1:L} \right| / \left| \{ \mathbf{b}_l = 0 \}_{l=1:L} \right|$$

to evaluate the algorithms' performance of selecting inactive blocks. The results of different methods are summarized in Table. 5.1, where 50 independent trails are used, the noise power is 0.01 and the other parameters remain the same as those in Fig. 5.1. Also $\hat{b}_j \neq 0$ and $\hat{\mathbf{b}}_l \neq 0$ are, respectively, implemented by $|\hat{b}_j| > 0.01$ and $\|\hat{\mathbf{b}}_l\|_2 > 0.01$ in our simulations. As shown in Table. 5.1, HiLasso, ClussVB and HiSBL achieve a relatively lower FAR compared to BSBL and SBL. The fact that BSBL obtains the low BFAR with a relatively high FAR indicates that it is robust in block recovery but is ineffective in suppressing noise inside the active blocks.

In addition to the simulations, a simple real-world application of image recovery and separation is conducted by using the USPS digits dataset, which has been

Table 5.1: Performance evaluation for variable selection

Algorithm	DR	FAR	block-wise FAR
BSBL	1	0.3152	0.1300
SBL	1	0.3018	0.9953
HiLasso	0.9989	0.1188	0.6100
ClussVB	1	0.1357	0.2000
HiSBL	0.9978	0.1765	0.0433

Figure 5.5: Performance of variable selection.

shown to be well represented in the sparse modeling framework. Following [43], a sub-dictionary is firstly learned by the K-SVD algorithm [87] for each digit and the complete dictionary is created by concatenating them. Each sub-dictionary of size 256×40 is learned by 200 samples and a dictionary $\Phi \in R^{256 \times 400}$ is created for the ten digits, i.e. $0, 1, \dots, 9$. The testing data is set to be a mixture of digits 3 and 5 randomly selected from the samples. The noisy testing data is generated by adding the Gaussian noise with standard deviation of 5% of the peak value of the digits.

The results evaluating the variable selection performances of different algorithms are presented in Fig. 5.5, where the line segments indicate the selection of the atoms, where a atom is selected if the absolute value of its corresponding coefficient is large than 0.05. As shown in Fig. 5.5, algorithms BSBL, HiLasso and HiSBL can detect the active blocks (digits 3 and 5) while others cannot. The MSEs of the recovered digits using different algorithms are given in Table. 5.2, where 50 independent trails are conducted. It is shown in Table. 5.2, algorithms HiLasso, BSBL and HiSBL achieve relatively low MSEs with HiSBL performing slightly better.

Table 5.2: The MSEs of the recovered digits

Algorithm	HiLasso	SBL	ClussVB	BSBL	HiSBL
MSE	0.0351	0.0473	0.0539	0.0361	0.0292

5.5 Conclusion

In this chapter, the hierarchical sparse signal recovery is discussed under a Bayesian framework. The hierarchical sparsity is enforced by controlling the covariance matrix of the coefficient vector using two classes of hidden variables to regulate the block-sparsity and the internal sparsity, respectively. Under the formulated probabilistic model, variational Bayesian inference is used to estimate the hidden variables and the estimate of the signal is given by the mean of the posterior of the coefficient vector. Our method uses both the block- and internal-sparse property simultaneously in recovering the signal compared to the SBL or BSBL, and is free of regularization parameter selection compared to the l_1 -norm minimization based methods. Simulation and experimental results show that the proposed method outperforms other existing methods in terms of the recovery performance.

Chapter 6

Enhanced ISAR Imaging by Exploiting the Continuity of the Target Scene

6.1 Introduction

High resolution radar imaging based on sparse representation has recently received much attention. It has been shown in [45–48, 88–90] that high-resolution target image can be achieved with a limited number of pulses by reformulating the ISAR or SAR imaging problem into an under-determined linear inverse scattering problem.

Taking the uncertainties of the under-determined inverse problem into consideration, all the three solving strategies mentioned in Chapter 1 have already been used in radar imagery applications via sparse representation. By exploiting the sparsity of ISAR/SAR images, sparsity based imagery methods are shown to outperform the conventional range-Doppler algorithm (RDA) [91] in terms of achieving high resolution with fewer measurements. The l_1 -norm based BP is used to obtain a high-resolution image with fewer pulses in ISAR [45, 48] and in SAR [92]. Apart from the l_1 -norm regularization, additional constraints and procedures are added for other purposes. For example, the Total Variation (TV) constraint is added along with the l_1 -norm constraint [88, 93] to deal with the speckle noise, and an

iterative procedure for estimating the phase error is incorporated [94] to obtain a focused image. The MP is used for imagery of the target with uniform or even maneuvering rotation [46] due to its computational efficiency and the BCS strategy has been found to obtain an auto-focused ISAR image [47, 86] and high-resolution SAR image [89]. To impose the sparsity, the Laplace prior is employed for each scatterer on the target scene where the posterior distribution is intractable [47] and an estimate of the sparse signal is obtained by maximizing its posterior probability via a quasi-Newton technique. In [89], a hierarchical prior which facilitates a tractable posterior distribution is used and the sparse signal is inferred by the EM algorithm. In the recent literature of ISAR imagery based on sparse representation, undersampling is performed for each range cell in the cross-range domain after the pulse compression. Instead of transmitting a wide-band signal and performing the conventional pulse compression in the range domain, a different and simplified radar system which transmits the sparse probing frequencies (SPF) is proposed to obtain a high-resolution ISAR image by using the convex optimization technique with l_1 -norm constraint [48].

In general, ISAR images of the practical targets always exhibit strong continuity. More concretely, nonzero scatterers on the target region are continuously located in either the range or the cross-range domain. Considering that radar transmits sparse probing frequencies [48], we reformulate the ISAR imaging problem into the Bayesian framework to exploit the inherent continuity of the scatterers in the target scene for enhanced ISAR images. In contrast to the work reported in [47] and [89], where independent prior for each scatterer in target scene is simply used to impose the sparsity of the target image, we propose a correlated prior to enforce the continuity of the target image. A similar idea has been explored for recovering a tree-structured sparse signal by compressive sensing [26], cluster-structured sparse signals [35] and the ultra sound image by using the Markov random field (MRF) to exploit the spatial correlation of the image pixels in the Fourier domain [95]. These methods exploiting the inherent structure of the signal can be generalized as model-based compressive sensing [96]. Some research focusing on different structures of the signal has been reported, such as block-sparsity [29, 32, 97] where a rigid signal

structure is directly imposed on the signal vector itself, tree-structure [25, 26, 98, 99] and other spatial correlation [35, 95] where structures are statistically imposed on the Bayesian prior to offer a flexible modeling of the signal structure. Although the prior knowledge about the structure of the signal and its specific imposing strategy can be quite different, the fundamental principle to exploit the structure of the signal to achieve the guarantees of better performance remains the same. Inspired by the similar idea, this chapter introduces the continuity structure of the target images into the ISAR imagery problem within a Bayesian framework. By exploiting the continuity structure of the target image, the proposed imagery algorithm offers the advantages of high recovery accuracy and resolution with fewer measurements in low signal-to-noise ratio (SNR) environments.

This chapter is organized as follows. Section 6.2 briefly reviews the principles of the ISAR imaging and the mathematical signal model based on sparse probing frequencies. Section 6.3 introduces continuity patterns of the scatterers on the target scene and constitutes the Bayesian model to enforce such continuities. The efficient inferences of the parameters and hidden variables via the Markov chain Monte Carlo (MCMC) sampling are given in Section 6.4. Results of experiments with synthetic and practical data are presented in Section 6.5 to show the effectiveness of the proposed algorithm. Conclusions are drawn in Section 6.6.

6.2 ISAR Imaging

6.2.1 Signal model for ISAR imaging

The ISAR imaging geometry for a plane target is shown in Fig. 6.1 [100], where three coordinate systems T_z , T_x and T_y are used. T_z represents the reference system embedded in the radar with the axis z_2 oriented along the line of sight (LOS). T_x is a time varying coordinate system chosen to keep the x_2 axis oriented along the LOS, the x_3 axis along the effective rotation vector $\boldsymbol{\Omega}_{eff}$ and the origin fixed at the target center O . T_y is a reference system embedded in the target which is coincident with

Figure 6.1: The geometry of the radar imaging of a target when $t = 0$.

T_x when and only when $t = 0$. The distance between target and radar is R_0 . Ω_T is the total angular rotation vector caused by the translation and angular motions of the target, and Ω_{eff} is the effective rotation vector which is the projection of Ω_T on the plane orthogonal to the LOS. Both Ω_T and Ω_{eff} are applied to the target center O . P is the k th scatterer with position vector $\mathbf{y}_k = [y_{1,k} \ y_{2,k} \ y_{3,k}]^T$ in T_y coordinate system. The Doppler induced by the rotation of scatterer P is generated by [100]

$$x_2(\mathbf{y}_k, t) = a_2(\mathbf{y}_k) + b_2(\mathbf{y}_k) \cos(|\Omega_T| t) + \frac{c_2(\mathbf{y}_k)}{|\Omega_T|} \sin(|\Omega_T| t) \quad (6.1)$$

where

$$\begin{aligned} a_2(\mathbf{y}_k) &= \frac{(\Omega_T \cdot \mathbf{y}_k)}{|\Omega_T|^2} \Omega_{T_2}, \\ b_2(\mathbf{y}_k) &= y_{2,k} - \frac{(\Omega_T \cdot \mathbf{y}_k)}{|\Omega_T|^2} \Omega_{T_2}, \\ c_2(\mathbf{y}_k) &= |\Omega_{eff}| y_{1,k}, \end{aligned}$$

$|\Omega|$ is the modulus of vector Ω , and Ω_{T_2} is the coordinate of Ω_T along the x_2 axis. For target with a small constant angular velocity, (6.1) can be approximated by its

first-order polynomial:

$$x_2(\mathbf{y}_k, t) = y_{2,k} + |\boldsymbol{\Omega}_{eff}| y_{1,k} t \quad (6.2)$$

and the cross-range profile in a particular range cell can be rewritten as

$$s(\tau, t) = \sum_{k=1}^K a_k B \cdot \text{rect}\left(\frac{t}{T_{ob}}\right) \text{sinc}\left[B\left(\tau - \frac{2}{c}(x_2(\mathbf{y}_k, t) + R_0)\right)\right] e^{j2\pi(\gamma_k t)} \quad (6.3)$$

where

$$\gamma_k = \frac{2F_c}{c}(x_2(\mathbf{y}_k, t) + R_0),$$

a_k is the complex amplitude of the scatterer k , B is the transmitted signal bandwidth, F_c is the carrier frequency, c is the speed of light, T_{ob} is the observation time and K is the number of point scatterers in a particular range cell, **rect** and **sinc** denote the rectangular and the sinc functions, respectively. Expression $\frac{2|\boldsymbol{\Omega}_{eff}|y_{1,k}}{\lambda}$ in (6.3) denotes the Doppler frequency of scatterer k , where $\lambda = c/F_c$. The conventional RDA has been used to obtain the target image based on the signal model described in (6.3).

6.2.2 Sparse probing frequencies

With the compressive sensing theory, it is unnecessary for Radar to transmit a wide-band signal because a few probing frequencies are sufficient to reconstruct the ISAR image by solving a convex optimization problem with l_1 -norm constraint as demonstrated in [48]. Therefore, a simplified Radar system which transmits the sparse probing frequency (SPF) signal is proposed to obtain a high resolution ISAR image. Let us denote l as the counter of the measurements, f_l and t_l as the frequency and the time of measurement l , respectively. It is shown in [48] that if a probing frequency signal $\exp(j2\pi f_l \tau)$ with frequency f_l at time t_l is transmitted, the received signal at the range frequency $f_l \in [F_c - B/2, F_c + B/2]$ is given by

$$\mathbf{r}(l) = \sum_{k=1}^K a_k \exp\left(-j4\pi f_l \frac{R_0 + y_{2,k} + |\boldsymbol{\Omega}_{eff}| y_{1,k} t_l}{c}\right), \quad (6.4)$$

where $l \in [1, L]$ and L is the number of the total measurements. Discretizing the target scene of interest into an $M \times N$ grid in the cross-range and range domains as

illustrated in Fig. 6.2, each scatterer located at grid (m, n) is denoted by $(y_{2,m}, y_{1,n})$ with amplitude $a_{m,n}$, where $y_{1,n}$ in the cross-range domain belongs to

$$\left\{ -\frac{ND_r}{2}, -\frac{ND_r}{2} + D_r, \dots, \frac{ND_r}{2} - D_r \right\} \quad (6.5)$$

with $D_r = \frac{F_D}{N} \frac{\lambda}{2|\mathbf{\Omega}_{eff}|}$, F_D is the maximal unambiguous Doppler shift which is determined by the pulse repetition frequency (PRF), and $y_{2,m}$ in the range domain belongs to $\left\{ -\frac{MR_r}{2}, -\frac{MR_r}{2} + R_r, \dots, \frac{MR_r}{2} - R_r \right\}$ with $R_r = c/2B$. The resolution is defined by (R_r, D_r) . An over-completed dictionary $\mathbf{D}$ of size $L \times MN$ can be

Figure 6.2: Cross-range and range domain with a designed resolution.

constructed according to the target scene in Fig. 6.2 with atom $\mathbf{d}_{m,n}$ whose element is defined as

$$\mathbf{d}_{m,n}(l) = \exp \left\{ -j4\pi \frac{f_l}{c} (R_0 + y_{2,m} + |\mathbf{\Omega}_{eff}| y_{1,n} t_l) \right\}. \quad (6.6)$$

Denoting the coefficient vector

$$\mathbf{a} = [a_{1,1} \dots a_{m,1} a_{1,2} \dots a_{1,n} \dots a_{m,n}]^T,$$

we have

$$\mathbf{r} = \mathbf{D}\mathbf{a}. \quad (6.7)$$

Because $L < MN$, (6.7) is under-determined. However, since vector $\mathbf{a}$ is sparse due to the fact $K \ll MN$, there exists a unique solution to (6.7) that can be obtained by sparse recovery methods provided that the dictionary $\mathbf{D}$ satisfies the restricted isometry property (RIP)[48, 101].

6.2.3 Maximal mutual coherence of $\mathbf{D}$

Since the RIP property of a dictionary is generally hard to analyze, instead we analyze theoretically the maximal mutual coherence of the dictionary in Appendix C when full measurements are used. As shown in the Appendix, the maximal coherence of the dictionary varies with ratio B/F_c for a given N . For X-band signal which is commonly used in ISAR, we generally have $0.04 \leq \frac{B}{F_c} \leq 0.2$. By designing the radar system with a proper B/F_c , a dictionary with relatively small maximal coherence can be obtained.

However, random sampling in the range domain will result in a dictionary with increased maximal mutual coherence. The maximal coherence of $\mathbf{D}$ depends on both the number of samples used and their positions. Therefore, it is hard to theoretically analyze the maximal mutual coherence under random sampling. We evaluate the maximal mutual coherence of the dictionary $\mathbf{D}$ versus the number of random samples by Monte Carlo method for fixed B/F_c and N in section 6.5.1. As shown in latter sections, the maximal mutual coherence of $\mathbf{D}$ is either similar to or smaller than that of the random Gaussian dictionary of the same size which has been found to satisfy the RIP with a high probability [101].

6.3 Continuity Patterns and Bayesian model

The conventional CS-based methods have used only the sparse property of ISAR and SAR images to solve the radar imagery problems [45–48, 88–90]. In addition to the sparsity, we further exploit the inherent continuity of the target scatterers in either the range or the cross-range domain to obtain an enhanced ISAR image under the Bayesian framework. Following the work reported in [48], let us consider a simple radar system which transmits the sparse probing frequencies instead of the wide-band linear frequency modulated (LFM) signal. Before reformulating the ISAR imagery problem in the Bayesian framework, let us first introduce three continuity patterns of the scatterers in range and cross-range domains that are to be exploited in this chapter.

6.3.1 Continuity Patterns

In real world scenarios, the target region is continuous in the range and cross-range domains. It is therefore assumed in this chapter that the scatterers of the target are not isolated, but located continuously in the target scene. This assumption is reasonable for ISAR image since the target of interest in ISAR is usually small and can be regarded as a mass of small scatterers. This phenomenon can also be observed in the practical ISAR data.

The motivation for considering the continuity structures of the range and cross-range of the target scene into ISAR imagery is illustrated by the real data image of Yak-42 airplane with the Radar system parameters given in Table 6.1. The image obtained by the RDA [91] with full measurements is given in Fig. 6.3 (a) where the scatterers obviously exhibit a sparse property.

Table 6.1: System parameters for data “Yak-42”

Carrier frequency F_c	10 GHz
Bandwidth B	400 MHz
Range resolution $R_r = C/(2B)$	0.375 m
Pulse repetition frequency (PRF)	25 Hz

In addition to sparsity, continuities in cross-range and range domains can be observed. Figure 6.3 (b) shows the profiles of range cells 45, 46 and 47. An example of the continuity in cross-range is clearly shown in each range cell where the scatterers with nonzero magnitude are grouped in two clusters corresponding to the head and the wing of the plane, respectively. The continuity in range domain is shown by the fact that the scatterers of the neighboring range cells 45, 46 and 47 tend to be nonzero or zero almost simultaneously. A closer inspection of the scatter points in cross-range cell 63 and range cell 46 shows that both the continuities in range and cross-range are demonstrated simultaneously. Motivated by this phenomenon, three continuity patterns for a particular scatterer indexed by (m, n) in the target scene are summarized in Fig. 6.4, where the shaded squares denote the scatterers with nonzero amplitudes and the white ones are the scatterers which are irrelevant to the

Figure 6.3: Continuity structure illustration by Yak-42.

Figure 6.4: Three kinds of continuity patterns

scatterer (m, n) . The amplitudes of the white ones can be nonzero or zero. If any of the patterns in Fig. 6.4 (a), (b) and (c) is observed, it is reasonable to assume that $a_{m,n}$ for scatterer (m, n) is nonzero with a high probability. Therefore the continuity either in range domain, cross-range domain, or in both domains can be encouraged accordingly.

6.3.2 Bayesian Model

Under the framework of Bayesian compressive sensing, a hierarchical prior is used to enforce the sparsity in the signal coefficient vector $\mathbf{a}$, where elements in $\mathbf{a}$ are

independent and each element $a_{m,n}$ follows a Bernoulli-Gaussian distribution, which is also known as a “spike-and-slab” prior [25, 102, 103]:

$$p(a_{m,n} | q_{m,n}, \beta_{m,n}) = \sum_{\varepsilon \in \{0,1\}} p(a_{m,n} | w_{m,n} = \varepsilon, \beta_{m,n}) Pr(w_{m,n} = \varepsilon | q_{m,n}), \quad (6.8)$$

where the hyperparameter $q_{m,n}$ is the prior probability of having a non-zero $w_{m,n}$ which is a hidden binary variable indicating if $a_{m,n}$ is nonzero or not and follows the Bernoulli distributions with parameter $q_{m,n}$

$$Pr(w_{m,n} | q_{m,n} = 1) = q_{m,n},$$

and

$$Pr(w_{m,n} | q_{m,n} = 0) = 1 - q_{m,n}.$$

Also $p(a_{m,n} | w_{m,n} = 1, \beta_{m,n})$ follows $\mathcal{CN}(a_{m,n} | 0, \beta_{m,n}^{-1})$ when $w_{m,n} = 1$, where $\mathcal{CN}(0, \beta_{m,n}^{-1})$ denotes the complex Gaussian distribution with zero mean and precision $\beta_{m,n}$. The conditional prior of $a_{m,n}$ can be directly obtained by

$$p(a_{m,n} | q_{m,n}, \beta_{m,n}) = (1 - q_{m,n}) \delta(|a_{m,n}|) + \frac{q_{m,n} \beta_{m,n}}{\pi} \exp(-\beta_{m,n} |a_{m,n}|^2). \quad (6.9)$$

With the prior in (6.9) on each element of the signal vector $\mathbf{a}$, the distribution of $\mathbf{a}$ given $\mathbf{q}^T = [q_{1,1} \cdots q_{M,N}]$ and $\boldsymbol{\beta}^T = [\beta_{1,1} \cdots \beta_{M,N}]$ is

$$p(\mathbf{a} | \mathbf{q}, \boldsymbol{\beta}) = \prod_{i=1}^{MN} [(1 - q_i) \delta(|a_i|) + \frac{q_i \beta_i}{\pi} \exp(-\beta_i |a_i|^2)] \quad (6.10)$$

where $1 \leq i \leq MN$ ¹.

Despite the sparsity, the scatterer distribution of a practical target image also exhibits a strong continuity which has not been exploited in most CS-based ISAR and SAR imagery such as the method proposed in [89]. In our framework, the correlated priors for $\beta_{m,n}$ and $q_{m,n}$ are proposed to encourage the continuity, which is similar to the model used in [25] and [35], where the structural information is statistically imposed on the Bayesian prior rather than explicitly making a hard imposition of the structure on the signal vector [32, 96]. In this way, signals are not restricted by the imposed structure and the corresponding algorithms would be more flexible and robust for signals with weak structural information.

¹We abuse the subscripts i and m, n for easy representation, where $i = (n - 1)M + m$.

6.3.2.1 Correlated prior for β

Let us assume $a_{m-1,n}$, $a_{m+1,n}$, $a_{m,n-1}$ and $a_{m,n+1}$ are all nonzero. To enforce continuity pattern in both range and cross-range domains in Fig. 6.4 (c), it is proper to choose a correlated prior for the precision of the scatterers in sets $\{j = n \mid m - 1 \leq i \leq m + 1\}$ and $\{i = m \mid n - 1 \leq j \leq n + 1\}$, i.e., they share the same precision:

$$\beta_{m,n} = \beta_{m-1,n} = \beta_{m+1,n} = \beta_{m,n-1} = \beta_{m,n+1}.$$

To obtain a tractable posterior distribution for $\beta_{m,n}$, we further impose a Gamma prior, $\text{Gamma}(b, d)$, on $\beta_{m,n}$ with parameters b and d . Since the Gamma prior is conjugate to Gaussian distribution with an unknown precision, the posterior distribution of $\beta_{m,n}$ given $\mathbf{a}$ also follows a Gamma distribution²:

$$\begin{aligned} p(\beta_{m,n} \mid \mathbf{a}, b, d) &= p(\beta_{m,n} \mid \bar{\mathbf{a}}_{m,n}, b, d) \\ &= \text{Gamma}\left(b + 5, d + \|\bar{\mathbf{a}}_{m,n}\|_2^2\right) \end{aligned} \quad (6.11)$$

with $\bar{\mathbf{a}}_{m,n}$ defined as the neighbor set of $a_{m,n}$:

$$\bar{\mathbf{a}}_{m,n}^T = \begin{bmatrix} a_{m,n} & a_{m-1,n} & a_{m+1,n} & a_{m,n-1} & a_{m,n+1} \end{bmatrix}.$$

Similarly, the posterior distribution of $\beta_{m,n}$ for continuity pattern (a) in range domain and continuity pattern (b) in cross-range domain can be derived as follows:

$$p(\beta_{m,n} \mid \mathbf{a}, b, d) = \text{Gamma}\left(b + 3, d + \|\bar{\mathbf{a}}_{m,n}^k\|_2^2\right) \quad (6.12)$$

with $k = 1, 2$ and $\bar{\mathbf{a}}_{m,n}^1 = \begin{bmatrix} a_{m,n} & a_{m-1,n} & a_{m+1,n} \end{bmatrix}^T$ for continuity pattern (a) and $\bar{\mathbf{a}}_{m,n}^2 = \begin{bmatrix} a_{m,n} & a_{m,n-1} & a_{m,n+1} \end{bmatrix}^T$ for continuity pattern (b).

If none of the continuity patterns occurs, then $\beta_{m,n}$ is assumed to be independent to others. By exerting a Gamma prior to it, the posterior distribution of $\beta_{m,n}$ can be easily calculated by

$$p(\beta_{m,n} \mid a_{m,n}, b, d) = \text{Gamma}\left(b + 1, d + |a_{m,n}|^2\right). \quad (6.13)$$

²Gamma distribution: $f(x; b, d) = d^b x^{b-1} e^{-dx} / \Gamma(b)$, where $\Gamma(b)$ is the Gamma function evaluated at b .

6.3.2.2 Prior for hidden variable $\mathbf{q}$

To incorporate the continuity information in the inference of the hidden variable $\mathbf{q}$, three kinds of sparsity patterns, similar to those in [26] and [35], are introduced. They are the patterns for strong acceptance, strong rejection and weak rejection, respectively. To facilitate the inference, a Beta prior, which is conjugate to Bernoulli distribution with three different sets of parameters $\{e_k, f_k\}_{k=0,1,2}$, is assigned for $q_{m,n}$ according to the introduced sparsity patterns. The sparsity patterns are defined below based on the idea of continuity:

Strong rejection (referred as Sparsity Pattern 0 (SP0)): If $\bar{\mathbf{a}}_{m,n} \setminus a_{m,n}$, i.e., the sub-vector of $\bar{\mathbf{a}}_{m,n}$ excluding the element of $a_{m,n}$, are all zeros, it is highly possible that $a_{m,n}$ is also zero due to the continuity of the target scene. The prior $Beta(e_0, f_0)$ with $e_0 < f_0$ is used to reject scatterer (m, n) and to guarantee the probability, $q_{m,n}$, for $w_{m,n} = 1$, being small.

Strong acceptance (referred as Sparsity Pattern 2 (SP2)): If any of the continuity patterns in Fig. (6.4) is observed, then it is highly possible that $a_{m,n}$ is nonzero to encourage the continuity of the target image. The prior $Beta(e_2, f_2)$ with $e_2 > f_2$ is used to accept scatterer (m, n) and to guarantee the probability, $q_{m,n}$, for $w_{m,n} = 1$, being large.

Weak rejection (referred as Sparsity Pattern 1 (SP1)): Apart from those situations mentioned in SP0 and SP2, no apparent conclusion can be drawn on the value of $q_{m,n}$. The prior $Beta(e_1, f_1)$ with $e_1 = f_1$ is used to exert non-informative prior on the $q_{m,n}$.

Since Beta distribution is conjugate to Bernoulli for its unknown parameter $q_{m,n}$, the posterior distribution for $q_{m,n}$ also follows a Beta distribution given by:

$$\begin{aligned} p(q_{m,n}^k | \mathbf{a}, e_k, f_k) &= p(q_{m,n}^k | \bar{\mathbf{a}}_{m,n}, e_k, f_k) \\ &= \mathbf{Beta}(e_k + \|\bar{\mathbf{a}}_{m,n}\|_0, f_k + 5 - \|\bar{\mathbf{a}}_{m,n}\|_0) \end{aligned} \quad (6.14)$$

where $k \in \{0, 1, 2\}$ and $\|\mathbf{a}\|_0$ gives the number of the nonzero components in $\mathbf{a}$.

6.4 Proposed Imagery Algorithm

6.4.1 Sparse Solution and the Gibbs Sampler

Supposing that the signal is corrupted by a complex white Gaussian noise $\mathbf{n}$ with zero mean and unknown precision β_0 , according to (6.7), the measurement vector $\mathbf{r}$ is

$$\mathbf{r} = \mathbf{D}\mathbf{a} + \mathbf{n} \quad (6.15)$$

where $\mathbf{n} \in \mathcal{C}^{L \times 1}$ and follows $\mathcal{CN}(0, \beta_0^{-1}\mathbf{I})$. The posterior distribution of all the variables given $\mathbf{r}$,

$$p(\mathbf{a}, \mathbf{q}, \boldsymbol{\beta}, \beta_0 | \mathbf{r}) \propto p(\mathbf{r} | \mathbf{D}\mathbf{a}, \beta_0) p(\beta_0) p(\mathbf{a} | \mathbf{q}, \boldsymbol{\beta}) p(\mathbf{q}) p(\boldsymbol{\beta}), \quad (6.16)$$

is intractable due to the implicit priors for $\boldsymbol{\beta}$ and $\mathbf{q}$ and no analytical estimation can be found for $\{\mathbf{a}, \mathbf{q}, \boldsymbol{\beta}, \beta_0\}$. The Gibbs sampler [27], the standard MCMC sampling, is used to generate samples asymptotically distributed according to (6.16) and to find the estimation of $\{\mathbf{a}, \mathbf{q}, \boldsymbol{\beta}, \beta_0\}$. The sequential sampling procedures of the Gibbs sampler are given in the following four steps:

- (a) Sampling a_i according to its posterior distribution;
- (b) Sampling β_i according to its posterior distribution given in (6.11), (6.12) and (6.13) for different continuity patterns;
- (c) Sampling q_i according to its posterior distribution given in (6.14) for different sparsity patterns;
- (d) Sampling noise precision β_0 according to its posterior distribution.

In step (a), the posterior distribution of a_i is required to be derived. Based on (6.10) and (6.15), the posterior distribution of $\mathbf{a}$ is proportional to

$$p(\mathbf{a} | \mathbf{r}, \mathbf{q}, \boldsymbol{\beta}, \beta_0) \propto p(\mathbf{r} | \mathbf{D}\mathbf{a}, \beta_0) \times \left\{ \prod_i^{MN} \left[(1 - q_i) \delta(|a_i|) + \frac{q_i \beta_i}{\pi} \exp(-\beta_i |a_i|^2) \right] \right\} \quad (6.17)$$

By (6.17), one can easily compute the posterior distribution $p(a_i)$ of a_i given $\{\mathbf{a} \setminus a_i, \mathbf{r}, \mathbf{q}, \boldsymbol{\beta}, \beta_0\}$ for $1 \leq i \leq MN$, which follows

$$p(a_i | \mathbf{a} \setminus a_i, \mathbf{r}, \mathbf{q}, \boldsymbol{\beta}, \beta_0) \propto (1 - \tilde{q}_i) \delta(|a_i|) + \tilde{q}_i \mathcal{CN}(a_i | \tilde{\mu}_i, \tilde{\beta}_i^{-1}) \quad (6.18)$$

where

$$\begin{aligned} \tilde{\beta}_i &= \beta_i + \beta_0 \mathbf{d}_i^H \mathbf{d}_i, \\ \tilde{\mu}_i &= \tilde{\beta}_i^{-1} \beta_0 \mathbf{d}_i^H \left(\mathbf{r} - \sum_{j=1, j \neq i}^{MN} a_j \mathbf{d}_j \right), \\ \frac{\tilde{q}_i}{1 - \tilde{q}_i} &= \frac{q_i}{1 - q_i} \times \frac{\mathcal{CN}(0, \beta_i^{-1})}{\mathcal{CN}(\tilde{\mu}_i, \tilde{\beta}_i^{-1})}. \end{aligned}$$

To ensure a tractable posterior distribution $p(\beta_0 | \mathbf{a}, \mathbf{r}, b_n, d_n)$ of β_0 , it is also assumed that β_0 follows a Gamma distribution $\text{Gamma}(b_n, d_n)$. Therefore $p(\beta_0 | \mathbf{a}, \mathbf{r}, b_n, d_n)$ is easily given by

$$p(\beta_0 | \mathbf{a}, \mathbf{r}, b_n, d_n) = \text{Gamma}(b_n + L, d_n + \|\mathbf{r} - \mathbf{D}\mathbf{a}\|_2^2). \quad (6.19)$$

The estimate of the coefficient vector $\mathbf{a}$ is given empirically by averaging the last few samples of the Gibbs sampler. The successive steps of Gibbs sampler for our problem are summarized in Algorithm 4. The convergence of the MCMC sampling is generally hard to diagnose. However, empirical experiments show that the algorithm will converge after several dozens of iterations, which will be shown later.

6.4.2 Discussion on the Proposed Imaging Algorithm

Comparing to the previous work on ISAR and SAR imaging based on compressive sensing [45–48, 88–90], the main difference and contribution of this chapter is to exploit the continuity structure of the target scene in the sparse recovery method. Since the continuity structure is imposed statistically on the Bayesian prior instead of directly on the signal vector, the proposed method is flexible enough to perform

Algorithm 4 Gibbs sampler

Initialize the maximum iteration number $N_{Maxiter}$, iteration counter $l = 0$ and initial value of parameters $\{\boldsymbol{\beta}, \mathbf{q}, \beta_0, b, d, b_n, d_n, e_k, f_k\}$.

while $l \leq N_{Maxiter}$ **do**

for $i = 1 : MN$ **do**

 Sample a_i according to $p(a_i | \mathbf{a} \setminus a_i, \mathbf{r}, \mathbf{q}, \boldsymbol{\beta}, \beta_0)$ in (6.18);

 Sample β_i according to $p(\beta_{m,n} | \mathbf{a}, b, d)$ in (6.11), (6.12) or (6.13) corresponding to different continuity patterns;

 Sample q_i according to $p(q_{m,n}^k | \mathbf{a}, e_k, f_k)$ in (6.14) corresponding to different sparsity patterns;

end for

 Sample β_0 according to $p(\beta_0 | \mathbf{a}, \mathbf{r}, b_n, d_n)$ in (6.19).

end while

$\hat{\mathbf{a}}$ and $\hat{\beta}_0$ are to be obtained by the empirical mean of the samples of $\mathbf{a}$ and β_0 in the last few iterations, respectively.

successfully in the scenario of relatively weak continuity. Comparing to the l_1 -norm based convex optimization approaches, the proposed method can achieve high-resolution in the scenarios of low SNR signals with only a few measurements, and is free of tedious parameter selection since the method itself is a data-driven learning process. However, there are still some limitations:

1) *Dependence on the basic assumption:* The performance of the proposed imaging algorithm depends on the correctness of the basic assumption that the scatterers of the target are located continuously in the target scene. The larger the deviation of the true target scene from the assumption, the worse the algorithm will perform especially under heavy noise or small sampling ratio situation. The limitation due to weak continuity is demonstrated and explained in Fig. 6.13 (h) in section 6.5.4. It is remarkably to notice, however, that this dependence limitation will not lead to degraded performance when the signal has a high SNR, since the continuity structure is imposed statistically on the Bayesian prior. This argument has been validated empirically in Fig. 6.13 (d) where the target scene is successfully recovered.

2) *Computation burden:* Since the inference of model parameters is carried out by the Gibbs sampler which inherently requires sequential sampling and several dozens of iterations, the computational burden is much heavier than that of the

conventional RDA. It is easy to show from Algorithm 4 that the computational complexity of the proposed method is in the order of $O(L \times MN \times N_{Maxiter})$. For comparison, the computational complexities of the BP used in most of the sparse representation based ISAR [45, 48] and SAR [92, 94] imaging methods, the BCS used in [89] and the proposed method are summarized in Table 6.2, where the BP is typically implemented by the CVX toolbox³ and solved by interior-point algorithms [64]. Note that the computational complexity of the BP method depends on its implementation. It is possible to speed up the BP method by other implementations such as alternating direction method of multipliers (ADMM) with a computational complexity of $O((MN)^2 \times L)$ [50] at the cost of sacrificing the recovery accuracy. As demonstrated later, the proposed method converges quickly, $N_{Maxiter}$ used in our

Table 6.2: Comparison of the computational complexity

Algorithm	Proposed	BP [64]	BCS [16]
Complexity	$O(L \times MN \times N_{Maxiter})$	$O((MN)^3)$	$O(L^3 \times N_{Maxiter})$

method is much smaller than L , i.e., $N_{Maxiter} \ll L$. Therefore, the computational complexity of the proposed algorithm can be written as $O(L \times MN)$, and is smaller than that of the BP. The computational complexity of the BP is commonly smaller than that of the BCS, since L is much smaller than MN . In common compressive sensing settings, L is usually selected from $MN/10$ to MN . Therefore, we generally have $L^2 \gg MN$ and the computational complexity of the BCS is larger than that of the proposed method. Although the computational complexity of the method is relatively smaller than that of the BP and the BCS, it is still much larger than that of the RDA.

3) *Other limitations:* Since the formulated imaging problem is under-determined, the proposed imaging algorithm cannot deal with complex target scene where scatterers are extremely dense. Another limitation is that the algorithm proposed currently has no ability of de-speckling which may be critical for certain applications.

³Available at website: <http://cvxr.com/cvx/>

6.5 Experiments and Performance Comparisons

In this section, we first analyze the behavior of the maximal coherence of dictionary $\mathbf{D}$ versus the number of the random samples in the range domain. Then, synthetic and practical data experiment results of the ISAR images produced by the proposed algorithm and other existing methods are studied and compared in this section to show the merits of our proposed algorithm. At the same time, the performance of the proposed method with respect to the target sparsity is evaluated. The parameters of the proposed algorithm used in the experiments are given as follows: $b = d = b_n = d_n = 10^{-6}$, $(e_0, f_0) = (1/L, 1 - 1/L)$, $(e_1, f_1) = (1/L, 1/L)$, $(e_2, f_2) = (1 - 1/L, 1/L)$, the initial value of β is set to be $1/(\mathbf{D}^H \mathbf{r})$, $\mathbf{q} = \mathbf{1}$, $\beta_0 = 1/\text{var}(\mathbf{r}) \times 10^{-2}$, the maximum iteration number, $N_{Maxiter}$, in the Gibbs sampler is 50 and the mean of the last 5 samples are computed as the estimate of $\mathbf{a}$.

6.5.1 Dictionary Coherence Evaluation

Before conducting the experiments, the behavior of the maximal mutual coherence of the dictionary $\mathbf{D}$ versus the number of random samples is investigated firstly by Monte Carlo method with 100 independent trials. The maximal mutual coherence of $\mathbf{D}$ versus the number of random samples is evaluated when $B/F_c = 0.15, 0.2$ and $N = 50$. The choices of $B/F_c = 0.15$ and $N = 50$ are to be consistent with the parameters used in section 6.5.2. The choice of $B/F_c = 0.2$ is to show that dictionary with a low maximal mutual coherence can be achieved by designing the radar system with proper values of B and F_c . The maximal mutual coherence of the random Gaussian dictionary is also provided as a comparison. As shown in Fig. 6.5(a), when $B/F_c = 0.15$, maximal mutual coherence of the $\mathbf{D}$ is similar to that of the random Gaussian dictionary when 20 measurements in the range domain are used. However, when the number of measurements decreases, the maximal mutual coherence becomes smaller than that of the random Gaussian dictionary. As shown both in Fig. 6.5(a) and (b), the maximal mutual coherence of the dictionary increases as the number of the measurements decreases. The maximal coherence of

Figure 6.5: Maximal mutual coherence versus the number of measurements (a) $B/F_c = 0.15$, (b) $B/F_c = 0.2$.

dictionary $\mathbf{D}$ is always smaller than that of the random Gaussian dictionary when $B/F_c = 0.2$ as shown in Fig. 6.5(b). It is concluded that the dictionary constructed in this chapter is not worse than the random Gaussian matrix which has been found to satisfy the RIP with a high probability.

6.5.2 Synthetic Experiments

To favor the continuity patterns exploited, we assume that scatterers are on-grid and locate continuously in range and cross-range domains. Suppose the target is 500 km away from the Radar in the direction of the LOS and has no translational motion. The target rotates with a constant velocity of $\omega = 0.1$ rad/s. In total, $N = 50$ pulses are collected. The parameters of the simulated radar system are given in Table. 6.3.

Since 50 pulses within a dwell time of 0.5s are used, the total rotated angle θ is 2.9° . Therefore, the cross-range resolution $\frac{\lambda}{2\theta}$ is 0.3 m. Suppose the target scene of interest is of size $4 \times 15 \text{ m}^2$ with a resolution of $0.1 \times 0.3 \text{ m}^2$. The scatterers are present

Table 6.3: System parameters for the simulated data

Carrier frequency F_c	10 GHz
Bandwidth B	1.5 GHz
Range resolution $R_r = C/(2B)$	0.1 m
Pulse repetition frequency (PRF)	100 Hz
Cross-range resolution D_r	0.3 m

in the center of the target scene and their locations are shown in Fig. 6.6, where 57 scatterers are present and located continuously along both range and cross-range domains. The amplitudes of the scatterers follow a complex Gaussian distribution with a mean of 2.0 and a variance of 0.5. The received signal is corrupted by the complex white Gaussian noise with a power of 0.2. The resulted signal to noise ratio is approximately 8.5dB. Measurements are obtained by sampling randomly in the range domain and measurements of one half, one fourth and one eighth of the full data set, i.e., 20, 10 and 5 measurements are used in the following synthetic experiments. The results of the RDA [91], BP as used in [45, 48, 92] and the Bayesian compressive sensing (BCS)⁴ [17, 89] are also provided for comparison.

Figure 6.6: The scatterer locations

Figure 6.7 gives the obtained ISAR images from different methods with one half of the full measurements. The noiseless target scene is shown in Fig. 6.7(a) where the scatterers are clustered together showing a strong continuity. The image by

⁴Available at website: <http://people.ee.duke.edu/~lcarin/BCS.html>

Figure 6.7: ISAR images achieved by different methods: (a) The target scene, (b) The RDA with full measurements; (c) The RDA using SPF; (d) The BP using SPF; (e) The BCS using SPF; (f) The proposed method using SPF.

RDA with full measurements is given in Fig. 6.7(b) where the target image is clear with slight background noise. However, when one half of the measurements in the

range frequency domain (randomly selected) are used by RDA, the image becomes blurred by the increased background noise shown in Fig. 6.7 (c). The reconstructed images with SPF by BP, BCS and the proposed method are given in Fig. 6.7(d)-(f). The result of BP in Fig. 6.7(d) is much better than that of the RDA, but is subjected to a careful selection of the regularization parameter for the algorithm. Both BCS and the proposed method are free of parameter selection. However, BCS does not perform well, as shown in Fig. 6.7(e), because of its difficulties in obtaining satisfactory image by using a small number of the measurements. The diminishing of some scatterers in Fig. 6.7 (e) may be caused by the noise suppression process of the BCS to keep a relatively sparse and clear target scene. It is observed that some of the scatterers with low amplitudes are eliminated by the BP in Fig. 6.7(d). In contrast, since the proposed method exploits the structure of the signal and enforces the continuity of the scatterers, the scatterers with relatively low amplitudes can be preserved. As shown in Fig. 6.7(f), every scatterer in the target scene is preserved and the background noise is effectively removed. Figure 6.8 shows the performances of different methods when the number of the measurements decreases to $1/4$ and $1/8$ of the original one. Since BCS cannot work under those settings, we only provide the results obtained by RDA, BP and the proposed method. As shown in Fig. 6.8(a) and (b), the RDA method shows worsen performance as the number of the measurements used decreases and fails when $1/8$ of measurements is used. Both BP and the proposed method with SPFs using $1/4$ of the full measurements perform well as shown in Fig. 6.8(c) and (e), where the proposed method outperforms BP slightly by removing the weak artificial points around the target while preserving some weak scatterers. When the number of the measurements decreases to $1/8$ of the original one, the result of BP is unsatisfactory in Fig. 6.8(d) where more than $2/3$ of the scatterers are missing with many strong artificial points around. In comparison, the result of the proposed method is much better as shown in Fig. 6.8(f) where most of the scatterers are preserved with fewer artificial points.

The convergence curve of the Gibbs sampler based on the simulated data is given in Fig. 6.9 with 10 independent experiments. The normalized reconstruction error is used to diagnose the convergence of the method. For the ISAR image,

Figure 6.8: ISAR image comparison in terms of the number of SPF. (a) The RDA using 1/4 of measurements; (b) The RDA using 1/8 of measurements; (c) The BP using 1/4 of measurements; (d) The BP using of 1/8 measurements; (e) The proposed method using of 1/4 measurements; (f) The proposed method using of 1/8 measurements.

since we are more interested in the absolute amplitude of the scatterer than the complex one, the normalized reconstruction error of the absolute amplitude vector $\bar{\mathbf{a}} = [|a_{1,1}| \cdots |a_{M,N}|]^T$ is computed by:

$$\left\| \frac{\bar{\mathbf{a}}}{\|\bar{\mathbf{a}}\|_2} - \frac{\hat{\mathbf{a}}}{\|\hat{\mathbf{a}}\|_2} \right\|_2 \quad (6.20)$$

where $\hat{\mathbf{a}}$ is the absolute sample value of the sparse coefficient in each iteration of the Gibbs sampler. The noise power added in each experiment is 0.2 and one half of the measurements, i.e., 20 measurements in the range frequency domain, are used. As shown in Fig. 6.9, the MCMC converges fast and the relative reconstruction error remains roughly unchanged after about 25 iterations. Therefore, a small number of iterations is sufficient to accurately represent the full posterior on all model parameters. Note that there is a remaining convergence error even after 30 iterations because of the mechanism of the Gibbs sampler and the random noise. Since the final estimate is the average of the last 5 samples of the Gibbs sampler, the error of the final estimate is expected to be similar or smaller than the convergence error, which is also demonstrated in Fig. 6.10(a) that when 20 random measurements are used under a noise of power 0.2, the error of the final estimate given by the proposed method is 0.1772.

Figure 6.9: The converge curve.

Since the RDA does not have the ability to suppress either the additive noise or the noise from undersampling, we simply compare the performance of the BP with that of the proposed algorithm in terms of normalized mean square error (NMSE). The NMSE is computed as the mean of the normalized reconstruction error defined in (6.20) from 50 independent experiments with $\hat{\mathbf{a}}$ replaced by the absolute value of the final estimate. Figures 6.10(a) and (b) give the NMSE of the reconstructed

images by the proposed algorithm and the BP method with respect to the number of measurements and the noise power, respectively. It is seen from Figs. 6.10(a) and (b), the proposed method achieves smaller NMSE than the BP method due to the exploitation of the continuity structure of the target scene. The CPU time

Figure 6.10: The NMSEs: (a) against the number of measurements (noise power =0.2), (b) against the noise power with 10 measurements in the range domain.

consumed by the proposed algorithm and the BP method against the number of measurements is listed in Table 6.4 for comparison, in which each value is evaluated based on the average of 5 trials. In each trial, measurements of the same number are randomly selected from the range domain. Note that 50 iterations are performed for the Gibbs sampler. Parameters are the same as those used for Fig. 6.10(a). It

Table 6.4: The required CPU time (s)

Measurements	5	10	15	20	25
Proposed	85.1	206.3	310.5	418.7	538.7
BP	164.7	616.9	1312.6	2423.9	3608.4

is observed from Table 6.4 that the proposed method is much faster than the BP method, especially when the number of measurements used is relatively large. It has been shown in section 6.4.2 that the computational complexity of the proposed

method is in the order of $O(L \times MN \times N_{Maxiter})$ which increases linearly with the number of the measurements, as verified in Table 6.4. However, the computational time for the BP method increases quadratically with the number of measurements when it is implemented by CVX toolbox.

6.5.3 Performance evaluation with respect to the target sparsity

To evaluate the performance of the proposed method with respect to the target sparsity, we design and conduct the following simulations. The parameters of radar system are the same as those shown in Table 6.3. A ratio of 1/4 of the full measurements in the range domain is used. Other parameters remain the same as those used in Fig. 6.7, except the target scene. We increase successively the number of the nonzero scatterers, i.e., the degree of the sparsity, in the following simulations, and evaluate the NMSE of the results obtained by the proposed method. The results are given in the Fig. 6.11. The number of the nonzero scatters is 7×7 , 11×11 , 13×13 and 15×15 for each column in Fig. 6.11. The degree of the target sparsity is computed and shown below each column along with the NMSE for each image obtained by the proposed method. The average NMSE obtained by 10 independent trials is shown in Fig. 6.12. As shown in Fig. 6.12, the performance of the proposed method decreases as the degree of the target sparsity increases. We emphasize that the degree of sparsity at which a correct target image can be recovered depends highly on the coherence of the dictionary. The lower the maximal coherence of the dictionary, the higher the sparsity degree can be allowed. Therefore, it is expected to achieve a higher degree of the target sparsity if we design a dictionary with a lower coherence.

6.5.4 Experiment with data set “Mig-25”

The proposed algorithm is also tested by a widely used synthetic data set “Mig-25” downloaded from the homepage of V. C. Chen⁵. The Radar system model

⁵Available at website: <http://airborne.nrl.navy.mil/vchen/tftsa.html>

Figure 6.11: Performance of the proposed method versus the degree of the target sparsity. The first row presents the target scenes with increased degree of target sparsity; the second row shows the images obtained by the RDA; the third row presents the images obtained by the proposed method; and the last row indicates the degree of the target sparsity and the NMSE of the results obtained by the proposed method.

Figure 6.12: The NMSE versus the degree of the target sparsity.

parameters are given in Table 6.5. The number of points in the range domain is 64

Table 6.5: System model parameters for data “Mig-25”

Carrier frequency F_c	9 GHz
Bandwidth B	512 MHz
Range resolution $R_r = C/(2B)$	0.293 m
Pulse repetition frequency (PRF)	15 KHz

and 32 pulses are collected. The additive noise follows a complex white Gaussian distribution with a variance of 5. The signal to noise ratio (SNR) in this case equals to 7.5dB. The results obtained by different methods using different proportions of measurements are given in Fig. 6.13. Note that the results in Figs 6.7 and 6.8 are presented in the axes of range and cross-range with the unit of meters since the scatterer positions can be easily calculated by the known rotating velocity of the target. However, since the information of the rotating velocity is unavailable for “Mig-25” and the data “Yak-42” used in the following subsections, the results in the rest of the chapter will be presented with axes labelled by “Range cell” and “Doppler cell”. The result of the RDA with full measurements in Fig. 6.13(a) shows some continuity in the range domain and a weak continuity in the Doppler (cross-range) domain of the simulated target. The RDA method performs well in noisy environment, as shown in Fig. 6.13(e). However, the performance severely

Figure 6.13: Comparison of ISAR images against number of SPFs achieved by the (a) RDA with full measurements; (b) RDA using 1/2 of measurements; (c) BP using 1/2 of measurements; (d) proposed method using 1/2 of measurements; (e) RDA with full measurements ($SNR = 7.5\text{dB}$); (f) RDA using 1/4 of measurements; (g) BP using 1/4 of measurements; (h) proposed method using 1/4 of measurements.

degrades when $1/2$ and $1/4$ measurements are used, as shown in Figs. 6.13(b) and (f), especially in Fig. 6.13(f) where target cannot be observed. The images given by the BP and the proposed method are much better, as shown in Figs. 6.13(c) and (d) using $1/2$ measurements. In comparison of Figs. 6.13(c) and (d) with Fig. 6.13(a), it seems that the result of the proposed method is slightly better than that of BP in term of removing the artificial points outside the target region and preserving the weak scatterers. Neither of the methods can produce an acceptable result with $1/4$ of the measurements. But a plane profile can be seen from the image obtained by the proposed method in Fig. 6.13(h). However, the noise components in the target region are preserved and are taken wrongly as the signal components by the proposed algorithm since it always enforces a continuity inside the target region. Comparing to Fig. 6.13(h), the noise components inside the target region in Fig. 6.13(g) are efficiently removed by the BP which does not use any continuity assumption. However, we can hardly observe a plane from the image in Fig. 6.13(g) where only some strong target scatterers are preserved with many comparable artificial points around.

6.5.5 Practical Data Experiments with data set “Yak-42”

The data used in this section is “Yak-42” plane data set with the Radar system parameters given in Table 6.1 in section 6.3.1. The number of points in the range domain is 128 and the number of pulses is 32 within a dwell time of 1.28s. The RDA image with full measurements are given in Fig. 6.14 where a relatively strong continuity of the plane fuselage and a weak continuity around the head and wings of the plane can be observed. Since the data size of “Yak-42” is relatively large, we only perform 35 Gibbs sampling iterations and the estimate of $\mathbf{a}$ is given by the empirical mean of the samples in the last 5 iterations.

The results of “Yak-42” plane by different methods are given in Figs. 6.15 and 6.16. In Fig. 6.15, the results obtained by different methods with various numbers of SPFs are compared. The effect of under-sampling becomes extremely severe in Fig.

Figure 6.14: The ISAR image of Yak-42 airplane obtained by the RDA with full measurements.

6.15(a) and (b). In Fig. 6.15(b), the head and wings of the target are nearly covered by the noise. The target images given by both the BP and proposed method using one half of the measurements are satisfactory. In the image obtained by the BP in Fig. 6.15(c), some weak scatterers on the target region are missing and several weak artificial points are around. The image obtained by the proposed method in Fig. 6.15(e) is relatively better, where the artificial points outside the target region are almost removed and the continuity property of the target image is well preserved. The performances of both the proposed method and the BP using 1/4 of measurements degrade. Some scatterers on the target are removed along with the noise by both the BP and the proposed method. The part connecting the fuselage and the head of the target is lost in Fig. 6.15(f). The image given by BP in Fig. 6.15(d) seems to be too sparse where too many scatterers on the target region are missing comparing to that shown in Fig. 6.15(f).

Figure 6.16 compares the results of different methods using one half of the measurements with noise powers of 150 and 300, respectively. The corresponding SNRs are 11.3dB and 5.3dB, respectively. The results of the RDA with full measurements are shown in Figs. 6.16(a) and (e), respectively. The plane profile can still be clearly

Figure 6.15: ISAR images versus the number of SPFs obtained by the (a) RDA using 1/2 of measurements; (b) RDA using 1/4 of measurements; (c) BP using 1/2 of measurements; (d) BP using 1/4 of measurements; (e) proposed method using 1/2 of measurements; (f) proposed method using 1/4 of measurements.

recognized. However, when one half of the measurements are used, the head and wings of the plane can hardly be seen from the results of RDA in Figs. 6.16(b) and (f). In Fig. 6.16(f), only a part of the fuselage is seen. Although the result of the

BP with a SNR of 11.3dB shown in Fig. 6.16(c) preserves most of the strong scatterers, the artificial points, which are comparable to those of the weak scatterers, are still present on the entire target scene. The result of the proposed method in Fig. 6.16(d) seems to be better with a clean background, although some scatterers on the head and the lower wing are missing. The result of BP becomes poor when SNR is 5.3dB, as shown in Figure. 6.16(g) where only several strong scatterers can be seen and others are immersed in the remaining artificial points. However, the proposed method can still work and provide in Fig. 6.16(h) a rough but clear profile of the plane with the continuity among parts of the plane preserved.

In addition to those visual result, we further use image entropy and the correlation value [104] with the reference image as two criteria to quantitatively evaluate the quality of the recovered images after random sampling. In the correlation criterion, the gray level image obtained by the RDA using the original dataset with full measurements, given in Fig. 6.14, is used as the reference image and is denoted by $\mathbf{X}_o$. Suppose that the obtained image by different methods is denoted as $\mathbf{X}_r$, the correlation is then defined as

$$\text{Corr}(\mathbf{X}_r, \mathbf{X}_o) = \frac{|\langle \text{vec}(\mathbf{X}_r), \text{vec}(\mathbf{X}_o) \rangle|}{\|\text{vec}(\mathbf{X}_r)\|_2 \|\text{vec}(\mathbf{X}_o)\|_2},$$

which evaluates the similarity between the recovered image and the reference image, where $\text{vec}(\mathbf{X})$ denotes the function which vectorizes the matrix $\mathbf{X}$ and $\langle \cdot, \cdot \rangle$ denotes the inner product. The higher the correlation value a method can provides, the better the method preserves the information of the target. The entropy of the image, given by $\text{Entropy}(\mathbf{X}_r) = -\sum \mathbf{p} \cdot \log 2(\mathbf{p})$, where $\mathbf{p}$ is the histogram of the recovered gray level image, measures the focusing quality of the obtained ISAR image. The more the image is focused, the smaller its entropy is. The correlation and the entropy values of the obtained images by different methods under different SNRs and sampling ratios are summarized in Table 6.6, where the best results among different methods are indicated in bold type. As indicated in Table 6.6, the results obtained by the proposed method maintain a higher correlation with the reference image and lower image entropy than those given by the BP.

Table 6.6: Performance evaluation by **Corr** and **Entropy**

SNR (dB)	10				5	
Sampling ratio	1/2		1/4		1/2	
Algorithm	BP	Proposed	BP	Proposed	BP	Proposed
<i>Corr</i>	0.8139	0.8567	0.7579	0.8159	0.7689	0.8239
<i>Entropy</i>	1.9326	0.2456	1.1568	0.1848	1.7106	0.2307

6.5.6 Discussion on the Results

The proposed method shows a great potential in removing the noise outside the target region based on the assumption that target scatterers are continuously located in a particular region. As long as the target scene well fits the assumption, the proposed method will greatly outperform those imaging algorithms based on conventional sparsity assumption, which is validated by the results shown in Figs. 6.7 and 6.8. If weak continuity of the target occurs, such as the continuity exhibited in dataset “Mig-25”, and there are relatively heavy noise components inside target region due to undersampling, the proposed method tends to give a profile of the target as shown in Fig. 6.13(h), where detailed information inside the target region is lost. Comparing to the ISAR image achieved by BP, another advantage of the proposed method is to be free of regularization parameter selection. All the parameters and hidden variables including the noise power can be learned directly from the data.

6.6 Conclusion

The conventional CS-based ISAR imaging methods only assume that scatterers on target scene are sparse or compressible without considering any particular signal structure. In contrast, a new ISAR imaging method within the Bayesian framework is proposed in this chapter to obtain an enhanced target image by exploiting the structural continuity of the scatterers on the target scene. To impose the sparsity on the signal, the amplitude of each scatterer is assumed to follow a Bernoulli-Gaussian distribution. Then a fully correlated prior is imposed to further enforce

the continuity by assuming that scatterers follow certain continuity pattern. By exploiting the structural continuity of target scene, the proposed algorithm requires fewer measurements in range domain, which results in a simplified radar system transmitting only a small number of random sparse probing frequencies, and achieves effective denoising outside target region and higher resolution at the expense of relatively heavy computational complexity.

Figure 6.16: ISAR images comparison with different noise powers. (a) Image achieved by RDA under noise ($SNR = 11.3\text{dB}$); (b) Image achieved by RDA ($SNR = 5.3\text{dB}$); (c) Image achieved by RDA by using 1/2 of measurements ($SNR = 11.3\text{dB}$); (d) Image achieved by RDA using 1/2 of measurements ($SNR = 5.3\text{dB}$); (e) Image achieved by BP by using 1/2 of measurements ($SNR = 11.3\text{dB}$); (f) Image achieved by BP using 1/2 of measurements ($SNR = 5.3\text{dB}$); (g) Image achieved by proposed method using 1/2 of measurements ($SNR = 11.3\text{dB}$); (h) Image achieved by proposed method using 1/2 of measurements ($SNR = 5.3\text{dB}$).

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, we have focused on the theory and applications of the structured sparse representation, especially the structure of block sparsity. The main contributions of this thesis are to re-analyze the theoretical reconstruction performance of BOGA, develop two harmonic tonal detectors based on BOGA, develop a reconstruction algorithm for hierarchical sparse signal and propose a novel ISAR imaging algorithm based on the structured sparse representation to exploit the continuity of the target scene.

We have studied the recovery conditions and approximation recovery error bound of the algorithm BOGA to recover the block-sparse signals in noisy environment. Derivation results in this thesis show that the stable recovery conditions of the BOGA can be relaxed and the representation error bound can be much tighter under the block-coherence defined in [32] for signals with block-sparsity. The theoretical results obtained provide a generalization of those reported by Tseng [13] for block-sparse signal and serve as a complement of the BOGA reported by Eldar [32] for noisy signals. The simulation results also show that the BOGA method outperforms the OGA method in terms of smaller representation error, signal recovery error, as well as higher support recovery probability if the signal is of the structure of block sparsity.

One direct application of the BOGA is to recover the harmonic signal, since under the designed dictionary, harmonic tonals shows a block-sparse structure. Each harmonic set can be regarded as an active block. Motivated by the success of BOGA in recovering the signal with block sparsity, two BOGA based harmonic tonal detectors in low SNR environment are proposed in Chapter 4. Theoretically, both detectors use integrated signal energy with limited prior information. Theoretical analysis of the test statistics and the detection performances based on the theory of order statistics are given comprehensively under the orthogonal dictionary. Significant improvements on detection performance in low SNR environments are shown over the conventional detectors by extensive simulations.

We have proposed a novel recovery algorithm for signal with hierarchical sparsity, i.e., a natural generation of block sparsity, under a Bayesian framework. We first model each component of the sparse coefficient vector by a Gaussian distributed variable with zero mean and then construct the variance by using the product of two classes of hidden variables controlling the block-sparsity and the internal sparsity, respectively. Under the proposed probabilistic model, we adopt the VB method to recover the coefficient vector and other hidden unknown variables from the noise corrupted data. The proposed method avoids the tedious regularization parameter selection procedure required by those based on convex optimization and learns the block-sparsity and the internal sparsity automatically from the input data. Numerical simulation and experimental results show that the proposed method outperforms those recently reported recovery methods for hierarchical sparse signals.

Finally, we have proposed a novel structured sparse representation based ISAR imaging algorithm to obtain enhanced target images, where the continuity of target scene is naturally exploited by imposing a two dimensional cluster structure over the scatterers. Under a simplified radar system which transmits the sparse probing frequency signal, we formulate the ISAR imaging into an under-determined linear inverse scattering problem and analyze the maximum coherence behavior of our constructed dictionary. Then, we impose a correlated prior to statistically encourage the continuity structure of the scatterers in the target region and infer all the hidden

variables by Gibbs sampler. Since the proposed method is a data-driven learning process, the tedious parameter tuning process required by the convex optimization based approaches can be avoided. Both the synthetic and the experimental results demonstrate that the proposed algorithm can achieve substantial improvements in the scenarios of limited measurements and low SNR compared with other reported algorithms for ISAR imaging problems.

7.2 Future work

In this section, we present some research directions which can be explored in the future.

The stability conditions and the representation error of BOGA for block sparse signal are obtained under the assumption that the noise is bounded. Statistical results under unbounded random noise are not included in our thesis. Actually, the reconstruction conditions of BOGA under random noise case are already studied in [66]. However, there is still a big gap between the simulation and the theoretical results. Tighter theoretical performance bounds are possible to be derived.

The theoretical performance of two BOGA based harmonic detectors are analyzed under orthogonal dictionary. The greedy pursuit-type algorithm in such a situation is simplified to a procedure of picking the first several largest block-norms of $\mathbf{D}^H \mathbf{x}$ and then the order statistic theory can be used. However, the story would change to be more difficult if the dictionary is non-orthogonal, which is of much interest. The performances of the proposed BOGA based harmonic tonal detectors are to be explored under the case of non-orthogonal dictionary. Yet, the presented detectors require the prior knowledge of the maximum harmonic order, which determines the size of each sub-dictionary in the block sparse model. Another problem remaining is to theoretically determine the stopping criterion to yield the optimal detection performance.

Due to high computational complexity, the proposed ISAR imaging algorithm in Chapter 6 seems to be less applicable to the SAR imagery which involves a huge

amount of data to be processed since usually an extensive target scene is of interest in SAR. An efficient variational inference which exploits the structure of the signal for large-scale Bayesian compressive sensing problem will be investigated in future research for alleviating the computational complexity. Since speckle noise widely exists in SAR/ISAR, we are to investigate the possibility of Bayesian compressive sensing technique to deal with it.

This thesis mainly focuses on the model of single measurement vector (SMV). Multiple measurement vectors (MMV) are commonly available in most real applications, such as electroencephalography (EEG) source localization, direction-of-arrival (DOA) estimation and multi-channel SAR imaging, and so on. It is assumed in most literature [105–107] that the signal support for different measurements is identical. In this sense, MMV model corresponds to a special case of the block sparsity pattern and can be efficiently dealt with by the block sparse recovery methods. However, the support of the signal may involve with measurements, that is, signal support may share atoms for parts of the measurements or changes continuously with time. This is true for some applications, such as the DOA or multi-channel SAR imaging of a moving target. To automatically learn the common sparsity patterns of the measurements and make a correct partition of the measurements according to the learned patterns are to be developed with a great importance.

Appendix A

Proof of Lemma 1 in Chapter 3

In this appendix, the detailed derivations for **Lemma 1** in Chapter 3 is provided.

Proof of Lemma1:

Let $\mathbf{D}_K$ be the concatenation of K column-blocks from $\mathbf{D}$ and $\mathbf{M}[i, j] = \mathbf{D}_K^H[i] \mathbf{D}_K[j]$,

where $i, j \in \{1, \dots, K\}$. For any vector $\mathbf{b} \in \mathcal{C}^{Kd}$, we have

$$\begin{aligned}
 \mathbf{b}^H \mathbf{D}_K^H \mathbf{D}_K \mathbf{b} &= \sum_{i=1}^K \sum_{j=1}^K \mathbf{b}^H[i] \mathbf{M}[i, j] \mathbf{b}[j] \\
 &= \sum_{i=1}^K \mathbf{b}^H[i] \mathbf{M}[i, i] \mathbf{b}[i] + \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \mathbf{b}^H[i] \mathbf{M}[i, j] \mathbf{b}[j] \\
 &\stackrel{(1)}{\geq} \sum_{i=1}^K (1 - (d-1)\nu) \|\mathbf{b}[i]\|_2^2 - \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K |\mathbf{b}^H[i] \mathbf{M}[i, j] \mathbf{b}[j]| \\
 &\stackrel{(2)}{\geq} (1 - (d-1)\nu) \|\mathbf{b}\|_2^2 - \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \|\mathbf{b}[i]\|_2 \|\mathbf{M}[i, j] \mathbf{b}[j]\|_2
 \end{aligned} \tag{A.1}$$

where (A.1.(1)) is obtained by using lemma 2.2 in [1] , (A.1.(2)) follows from $\sum_{i=1}^K \|\mathbf{b}[i]\|_2^2 = \|\mathbf{b}\|_2^2$ and the application of Cauchy-Schwarz inequality to the sec-

ond item. Then, we have

$$\begin{aligned}
 \mathbf{b}^H \mathbf{D}_K^H \mathbf{D}_K \mathbf{b} &\stackrel{(1)}{\geq} (1 - (d-1)\nu) \|\mathbf{b}\|_2^2 - \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \|\mathbf{b}[i]\|_2 \|\mathbf{b}[j]\|_2 d\mu_B \\
 &= (1 - (d-1)\nu) \|\mathbf{b}\|_2^2 - d\mu_B \sum_{s=1}^{K-1} \sum_{i=1}^K \|\mathbf{b}[i]\|_2 \|\mathbf{b}[(i+s)_K]\|_2 \quad (\text{A.2}) \\
 &\stackrel{(2)}{\geq} (1 - (d-1)\nu) \|\mathbf{b}\|_2^2 - d\mu_B \sum_{s=1}^{K-1} \|\mathbf{b}\|_2^2 \\
 &= (1 - (d-1)\nu - (K-1)d\mu_B) \|\mathbf{b}\|_2^2
 \end{aligned}$$

where $(x)_K$ is the operation of x modulo K , (A.2.(1)) follows from $\|M[i, j] \mathbf{b}[j]\|_2 \leq d\mu_B \|\mathbf{b}[j]\|_2$ which can be obtained by lemma 2.2 in [1] and definition 1. Equation (A.2.(2)) is obtained by applying the Cauchy-Schwarz inequality to the inner product $\sum_{i=1}^K \|\mathbf{b}[i]\|_2 \|\mathbf{b}[(i+s)_K]\|_2$. The procedures of A.2 are the same as (66), (67) and (68) in [32].

Similarly, we can prove

$$\mathbf{b}^H \mathbf{D}_K^H \mathbf{D}_K \mathbf{b} \leq (1 + (d-1)\nu + (K-1)d\mu_B) \|\mathbf{b}\|_2^2.$$

Then we have

$$(1 - (d-1)\nu - (K-1)d\mu_B) \|\mathbf{b}\|_2^2 \leq \|\mathbf{D}_K \mathbf{b}\|_2^2 \leq (1 + (d-1)\nu + (K-1)d\mu_B) \|\mathbf{b}\|_2^2.$$

Appendix B

Detection Performance Analysis in Chapter 4

It is emphasized that the derivations are based on the assumption that the dictionary is orthogonal, therefore variables in set $\left\{\left\|\mathbf{D}[i]^H \mathbf{v}\right\|_2^2\right\}_{0 \leq i \leq L}$ are independent. Performance analysis under overcomplete dictionary is more general but extremely hard, which is still under consideration.

B.1 Analysis on architecture 1

B.1.1 PDF of T_{arch1} under H_0 for fixed K'

For simplicity, we assume that the values, $\left\{\left\|\mathbf{D}[i]^H \mathbf{v}\right\|_2^2\right\}_{0 \leq i \leq L}$, are ordered by

$$\left\|\mathbf{D}[1]^H \mathbf{v}\right\|_2^2 \geq \left\|\mathbf{D}[2]^H \mathbf{v}\right\|_2^2 \geq \cdots \geq \left\|\mathbf{D}[L]^H \mathbf{v}\right\|_2^2.$$

Under H_0 , $\mathbf{x} = \mathbf{v}$, the first step is,

$$T_1 = \max_i \left\|\mathbf{D}[i]^H \mathbf{r}_0\right\|_2^2 = \left\|\mathbf{D}[1]^H \mathbf{v}\right\|_2^2. \quad (\text{B.1})$$

The residual signal at this step is

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{v} - \hat{\mathbf{x}}_1 = \mathbf{v} - \mathbf{D}[1]\hat{\mathbf{b}}[1] \\ &= (\mathbf{I} - \mathbf{D}[1]\mathbf{D}[1]^\dagger)\mathbf{v}. \end{aligned} \quad (\text{B.2})$$

For step 2,

$$\begin{aligned}
 T_2 &= \max_i \|\mathbf{D}[i]^H \mathbf{r}_1\|_2^2 \\
 &= \max_i \|\mathbf{D}[i]^H (\mathbf{I} - \mathbf{D}[1]\mathbf{D}[1]^\dagger) \mathbf{v}\|_2^2 \\
 &= \max_i \|\mathbf{D}[i]^H \mathbf{v} - \mathbf{D}[i]^H \mathbf{D}[1]\mathbf{D}[1]^\dagger \mathbf{v}\|_2^2 \\
 &\stackrel{A}{=} \max_i \|\mathbf{D}[i]^H \mathbf{v}\|_2^2,
 \end{aligned} \tag{B.3}$$

where (B.3.A) holds because of the orthogonality of the Fourier basis, i.e.,

$$\mathbf{D}[i]^H \mathbf{D}[j] = \begin{cases} N\mathbf{I}_M & i = j, \\ \mathbf{0} & i \neq j. \end{cases} \tag{B.4}$$

Since the BOGA never selects a previously chosen block, we have $2 \leq i \leq L$. Then (B.3.A) becomes

$$T_2 = \max_i \|\mathbf{D}[i]^H \mathbf{v}\|_2^2 = \|\mathbf{D}[2]^H \mathbf{v}\|_2^2,$$

which means that T_2 is the second largest value in $\left\{ \|\mathbf{D}[i]^H \mathbf{v}\|_2^2 \right\}_{0 \leq i \leq L}$. Similarly, it can be found that $T_{K'}$ is the K' -th largest value in $\left\{ \|\mathbf{D}[i]^H \mathbf{v}\|_2^2 \right\}_{0 \leq i \leq L}$.

Since $\mathbf{v} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$, $\mathbf{D}[i]^H \mathbf{v}$ becomes an M dimensional vector following complex white Gaussian distribution with zero-mean and variance $\sigma^2 \mathbf{D}[i]^H \mathbf{D}[i] = N\sigma^2 \mathbf{I}_M$. $\|\mathbf{D}[i]^H \mathbf{v}\|_2^2$ is then a variable following the scaled χ_{2M}^2 with $2M$ degree of freedom (dof), i.e.,

$$\frac{\|\mathbf{D}[i]^H \mathbf{v}\|_2^2}{N\sigma^2/2} \sim \chi_{2M}^2.$$

Since all variables in $\left\{ \|\mathbf{D}[i]^H \mathbf{v}\|_2^2 \right\}_{0 \leq i \leq L}$ are independent, the test statistic T_{arch1} is then the sum of first $\hat{K}$ largest observations in a sample of independent and identical distributed (IID) observations from a scaled χ_{2M}^2 distribution.

It is known that the distribution of the sum of l largest observations in a sample of IID observations from a gamma distribution is a linear function of the gamma distribution functions [108]. Let us define

$$g_n(x) = x^{n-1} e^{-x} / \Gamma(n) \quad x > 0 \tag{B.5}$$

to represent the gamma distribution with n degrees of freedom. Suppose that $\{x_1, x_2, \dots, x_L\}$ are L samples generated from $g_n(x)$ where x'_l denotes the l -th largest value in the L observations. Let $Z_l = \sum_{i=1}^l x'_i$ denotes the sum of l largest observations. As stated in [108], the density function f_l of Z_l follows

$$f_l(x) = \frac{L}{\Gamma(n)} \binom{L-1}{l} \sum_{r=0}^{L-l-1} \sum_{j=0}^{(n-1)l} \sum_{k=0}^{(n-1)r} (-1)^r c_{jl} c_{kr} \times \binom{L-l-1}{r} \Gamma(j+k+n) (l+1+r)^{-j-k-n} g_{rjk}^*(x) \quad (\text{B.6})$$

where

$$g_{rjk}^*(x) = \sum_{s=0}^{nl-j-1} a_s g_{nl-j-s}(x) + \alpha_r^{-1} \sum_{s=0}^{j+k+n-1} b_s g_{j+k+n-s}\left(\frac{x}{\alpha_r}\right)$$

with

$$\begin{aligned} \alpha_r &= \frac{l}{1+r+l} \\ a_s &= (-\alpha_r)^s \binom{j+k+n+s-1}{s} (1-\alpha_r)^{-j-k-n-s} \\ b_s &= \alpha_r^{-s} (-1)^{nl-j} \binom{nl-j+s-1}{s} \left(\frac{1}{\alpha_r} - 1\right)^{-nl+j-s} \\ \binom{L}{l} &= \frac{L!}{(L-l)!l!} \end{aligned}$$

and c_{jk} can be computed recursively from

$$\begin{aligned} c_{jk} &= \frac{k^j}{j!}, \quad j \leq n-1 \\ c_{jk} &= 0, \quad j \geq (n-1)k \\ c_{jk} &= \sum_{m=0}^{n-1} \frac{1}{m!} c_{(j-m)(k-1)}, \quad n \geq j \geq (n-1)k \end{aligned}$$

where $k > 1$ and $m!$ computes the factorial of m . Based on the density function of Z_l , the upper p -th percentile of the distribution of Z_l for certain values of L , l and n can be pre-computed off-line for further searching.

It is found that $\frac{T_{arch1}}{N\sigma^2/2}$ under H_0 is the sum of K' largest order statistics in a sample of L observations from χ_{2M}^2 , which is a special case of gamma distribution.

Based on (B.6), the density function of $T'_{arch1} = \frac{T_{arch1}}{N\sigma^2/2}$ under null hypothesis is given by the density function of $2Z_{K'}$, i.e.,

$$f_{T'_{arch1}}(x) = 1/2 f_{K'}(2x). \quad (\text{B.7})$$

Therefore the PFA is given by

$$P_{FA} = \int_{\frac{2\gamma_{arch1}}{N\sigma^2}}^{\infty} f_{T'_{arch1}}(x) dx. \quad (\text{B.8})$$

B.1.2 PDF of T_{arch1} under H_1 for fixed K'

The detection probability is derived under the assumption that BOGA selects the correct signal blocks in the first K steps and that $K' \geq K$. Under H_1 , $\mathbf{x} = \mathbf{D}\mathbf{b} + \mathbf{v}$. When $K' = K$, under the assumption that BOGA selects the correct signal blocks in the first K steps, the test statistic $T_{arch1} = \sum_{l=1}^K T_l = \sum_{l=1}^K \|N\mathbf{b}[l] + \mathbf{D}[l]^H \mathbf{v}\|_2^2$. It is then easy to show that $T'_{arch1} = \frac{T_{arch1}}{N\sigma^2/2}$ follows noncentral chi-squared distribution with noncentrality parameter $\lambda = \frac{N\|\mathbf{b}_0\|_2^2}{\sigma^2/2}$ and degree $2MK$. The density function is given by

$$f_{\chi_{2MK}^2(\lambda)}(x) = \frac{1}{2} \left(\frac{x}{\lambda}\right)^{\frac{KM-1}{2}} \exp\left[-\frac{1}{2}(x+\lambda)\right] I_{KM-1}(\sqrt{\lambda x}). \quad (\text{B.9})$$

Then the PD of architecture 1 is given by

$$P_D = \int_{\gamma}^{\infty} f_{\chi_{2MK}^2(\lambda)}(x) dx = Q_{\chi_{2MK}^2(\lambda)}(\gamma) \quad (\text{B.10})$$

where $Q_{\chi_m^2}$ represents the right-tail probability for a χ_m^2 variable with a dof m and $\gamma = \frac{2\gamma_{arch1}}{N\sigma^2}$.

When $K' > K$, the distribution of T_{arch1} is more complex. Rewrite T_{arch1} into

two parts:

$$\begin{aligned}
 T'_{arch1} &= \frac{T_{arch1}}{N\sigma^2/2} = \sum_{l=1}^{K'} \frac{T_l}{N\sigma^2/2} \\
 &= \sum_{l=1}^K \frac{\|N\mathbf{b}[l] + \mathbf{D}[l]^H \mathbf{v}\|_2^2}{N\sigma^2/2} + \sum_{l=K+1}^{K'} \frac{\|\mathbf{D}[l]^H \mathbf{v}\|_2^2}{N\sigma^2/2} \\
 &= T_{arch1}^1 + T_{arch1}^2.
 \end{aligned} \tag{B.11}$$

As previously discussed, the density function of T_{arch1}^1 is $f_{\chi_{2KM}^2(\lambda)}(x)$ given in (B.9) and T_{arch1}^2 is the sum of $K' - K$ largest order statistics in a sample of $L - K$ observations from χ_{2M}^2 . Therefore, the density function of T_{arch1}^2 is $f_{T_{arch1}^2} = 1/2 f_{K'-K}(2x)$ with $f_{K'-K}$ given by (B.6). Since T_{arch1}^1 and T_{arch1}^2 are independent, the density function of T_{arch1} is the convolution of density functions of T_{arch1}^1 and T_{arch1}^2 . Then the density function of T'_{arch1} is given by

$$f_{T'_{arch1}}(y) = \int f_{\chi_{2KM}^2(\lambda)}(x) f_{T_{arch1}^2}(y-x) dx. \tag{B.12}$$

Therefore the PD of architecture 1 when $K' > K$ is given by

$$P_D = \int_{\gamma}^{\infty} f_{T'_{arch1}}(y) dy \tag{B.13}$$

with $\gamma = \frac{2\gamma_{arch1}}{N\sigma^2}$.

Equations (B.8), (B.10) and (B.13) describe the theoretical detection performance of the detection architecture 1 with fixed K' under the assumption that the BOGA selects the correct signal blocks in the first K steps.

B.1.3 PDF of T_{arch1} under H_0 for Random K'

When K' is determined randomly by stopping criterion $\|\mathbf{r}_l\|_2^2 \leq \beta N\sigma^2$, it is an integer from the set $\{0, 1, 2, 3, \dots\}$ and its PMF depends on the parameter β . For easy representation, we denote $C_k(\beta) = \Pr(K' = k)$ with $k \in \{0, 1, 2, 3, \dots\}$.

Based on the stopping criterion of the BOGA, if we have $r_0 \leq \beta N\sigma^2$, then $K' = 0$. Therefore,

$$C_0(\beta) = \Pr(K' = 0) = \Pr(\|\mathbf{r}_0\|_2^2 \leq \beta N\sigma^2).$$

Under H_0 hypothesis, $\|\mathbf{r}_0\|_2^2 = \|\mathbf{v}\|_2^2$, so $\|\mathbf{r}_0\|_2^2$ follows a scaled χ_{2N}^2 . We have

$$C_0(\beta) = \Pr(\|\mathbf{r}_0\|_2^2 \leq \beta N \sigma^2) = \Pr\left(\frac{\|\mathbf{r}_0\|_2^2}{\sigma^2/2} \leq 2\beta N\right) = 1 - Q_{\chi_{2N}^2}(2\beta N). \quad (\text{B.14})$$

To have $K' = 1$, it is required that $\|\mathbf{r}_0\|_2^2 > \beta N \sigma^2$ and $\|\mathbf{r}_1\|_2^2 \leq \beta N \sigma^2$,

$$C_1(\beta) = \Pr(\|\mathbf{r}_1\|_2^2 \leq \beta N \sigma^2, \|\mathbf{r}_0\|_2^2 > \beta N \sigma^2). \quad (\text{B.15})$$

Since the events $\{\|\mathbf{r}_0\|_2^2 > \beta N \sigma^2\}$ and $\{\|\mathbf{r}_1\|_2^2 \leq \beta N \sigma^2\}$ are dependent, the theoretical expressions of C_k can not be found when $k > 0$. The density function of $T_{arch1} = \sum_{l=1}^{K'} T_l$ under the condition that $K' = k$, denoted as $f'_{T_{arch1}|K'=k}$, is no longer be the density of the sum of k largest order statistics. It can be clearly shown in the following equation

$$\begin{aligned} & \Pr(T_{arch1} < x | K' = k) \\ = & \Pr\left(\sum_{l=1}^k \|\mathbf{D}[l]^H \mathbf{v}\|_2^2 < x \mid \|\mathbf{r}_0\|_2^2 > N\beta\sigma^2, \right. \\ & \left. \dots, \|\mathbf{r}_{k-1}\|_2^2 > N\beta\sigma^2, \|\mathbf{r}_k\|_2^2 < N\beta\sigma^2\right). \end{aligned} \quad (\text{B.16})$$

Since events $\{\|\mathbf{r}_l\|_2^2 > N\beta\sigma^2\}_{0 \leq l \leq k-1}$ and $\left\{\sum_{l=1}^k \|\mathbf{D}[l]^H \mathbf{v}\|_2^2 < x\right\}$ are dependent,

$$\Pr(T_{arch1} | (K' = k) < x) \neq \Pr(T_{arch1} < x).$$

Unfortunately, the theoretical expressions of $f'_{T_{arch1}|K'=k}$ can not be found either.

Suppose that C_k and $f'_{T_{arch1}|K'=k}$ are known, the distribution of T_{arch1} can be found by marginalizing K' :

$$f'_{T_{arch1}}(x) = \sum_{K'} f'_{T_{arch1}|K'=k}(x) \Pr(K' = k). \quad (\text{B.17})$$

Therefore the PFA of architecture 1 for random K' is given by

$$P_{\text{FA}} = \int_{\frac{2\gamma_{arch1}}{N\sigma^2}}^{\infty} f'_{T_{arch1}}(x) dx. \quad (\text{B.18})$$

To find the threshold for architecture 1 when K' is random, we estimate C_k and $f'_{T_{arch1}|K'=k}$ by Monte Carlo method.

B.2 Analysis on architecture 2

B.2.1 K' is fixed

Let us assume that we know the approximated number of fundamental frequencies which is used as the value of K' in architecture 2. Suppose in each step the threshold used is γ_l for $1 \leq l \leq K'$. Under H_0 hypothesis, if BOGA stops in the first step, as shown in (B.1.1), $T_{arch2} = T_1 = \left\| \mathbf{D}[1]^H \mathbf{v} \right\|_2^2$ is the largest value selected from set $\left\{ \left\| \mathbf{D}[i]^H \mathbf{v} \right\|_2^2 \right\}_{1 \leq i \leq L}$. The PDF of $\frac{T_{arch2}}{N\sigma^2/2}$ is given by $1/2f_1(2x)$ in (B.6). So the PFA in this step is given by

$$P_{FA,1} = \frac{1}{2} \int_{\frac{2\gamma_1}{N\sigma^2}}^{\infty} f_1(2x) dx = Q_{f_1} \left(\frac{\gamma_1}{N\sigma^2} \right)$$

where $Q_{f_l} \left(\frac{\gamma_l}{N\sigma^2} \right) = \int_{\frac{\gamma_l}{N\sigma^2}}^{\infty} f_l(x) dx$, for $1 \leq l \leq K'$, is the right-tail probability of a random variable with the PDF f_l . If $T_1 \leq \gamma_1$, $T_{arch2} = T_1 + T_2$ is created. Similarly, T_{arch2} is the sum of first two largest values in $\left\{ \left\| \mathbf{D}[i]^H \mathbf{v} \right\|_2^2 \right\}_{1 \leq i \leq L}$. Therefore the PDF of T_{arch2} when BOGA is stopped at the second step is given by $1/2f_2(2x)$ in (B.6). The PFA of T_{arch2} at this step is

$$\begin{aligned} P_{FA,2} &\stackrel{A}{=} \Pr \left(T_1 < \frac{2\gamma_1}{N\sigma^2}, T_2 + T_1 \geq \frac{2\gamma_2}{N\sigma^2} \right) \\ &\stackrel{B}{\leq} \Pr \left(T_1 < \frac{2\gamma_1}{N\sigma^2} \right) \Pr \left(T_1 + T_2 \geq \frac{2\gamma_2}{N\sigma^2} \right) \\ &= Q_{f_2}(1 - Q_{f_1}). \end{aligned} \tag{B.19}$$

Since events $\{T_1 < \frac{2\gamma_1}{N\sigma^2}\}$ and $\{T_1 + T_2 \geq \frac{2\gamma_2}{N\sigma^2}\}$ in (B.19.A) are negatively dependent, (B.19.B) holds.

Similarly, the PFA of T_l can be bounded by

$$P_{FA,l} \leq (1 - Q_{f_{l-1}}) \cdots (1 - Q_{f_1}) Q_{f_l}. \tag{B.20}$$

After K' steps, the final PFA of architecture 2 is bounded by

$$\begin{aligned} P_{FA} &= \sum_{l=1}^{K'} P_{FA,l} \\ &\leq Q_{f_1} + (1 - Q_{f_1}) Q_{f_2} + \cdots + (1 - Q_{f_1}) \cdots (1 - Q_{f_{K'-1}}) Q_{f_{K'}}. \end{aligned} \tag{B.21}$$

When $K' = 1$, the equality of (B.21) holds strictly.

B.2.2 K' is random

K' will be a random integer with its PMF given by C_k when it is determined by the stopping criterion that $\|\mathbf{r}_l\|_2^2 \leq \beta N \sigma^2$. When $K' = k$, T_{arch2} may have k possible expressions depending on the step index at which the BOGA stops

$$T_{arch2} \Big|_{K'=k} = \begin{cases} T_1 \Big|_{K'=k} \\ T_1+T_2 \Big|_{K'=k} \\ \vdots \\ \sum_{l=1}^k T_l \Big|_{K'=k} \end{cases}. \quad (\text{B.22})$$

From the last line in (B.22), it is noted that when the BOGA stops at the k -th step, $T_{arch2} \Big|_{K'=k}$ has the same expression as $T_{arch1} \Big|_{K'=k}$ in section (B.1.3). Once C_k and the density functions $f'_{T_{arch1} \Big|_{K'=k}} = f'_{T_1+T_2+\dots+T_k \Big|_{K'=k}}$ are estimated by the Monte Carlo method in (B.1.3), they are readily used in this part. Apart from $f'_{T_1+T_2+\dots+T_k \Big|_{K'=k}}$, we still need to evaluate the other $k-1$ density functions $f'_{T_1 \Big|_{K'=k}}, f'_{T_1+T_2 \Big|_{K'=k}}, \dots$ to make a decision in each step of architecture 2. Luckily, we can prove that

$$\begin{aligned} f'_{T_1 \Big|_{K'=k}} &= f'_{T_1 \Big|_{K'=1}} \\ f'_{T_1+T_2 \Big|_{K'=k}} &= f'_{T_1+T_2 \Big|_{K'=2}} \\ &\vdots \\ f'_{T_1+T_2+\dots+T_{k-1} \Big|_{K'=k}} &= f'_{T_1+T_2+\dots+T_{k-1} \Big|_{K'=k-1}}. \end{aligned} \quad (\text{B.23})$$

Therefore, all the density functions used in architecture 2 are readily estimated by $f'_{T_1+T_2+\dots+T_k \Big|_{K'=k}}$ in (B.1.3). We prove the first equation in (B.23) as follows, and the proofs of other equations can be made in a similar way.

Proof of (B.23) : Since

$$\begin{aligned} \Pr \left(T_1 < x \Big|_{K'=k} \right) &= \Pr \left(\left\| \mathbf{D}[1]^H \mathbf{v} \right\|_2^2 < x \mid \|\mathbf{r}_0\|_2^2 > N\beta\sigma^2, \right. \\ &\quad \left. \dots, \|\mathbf{r}_{k-1}\|_2^2 > N\beta\sigma^2, \|\mathbf{r}_k\|_2^2 < N\beta\sigma^2 \right) \end{aligned} \quad (\text{B.24})$$

and $\mathbf{D}[1]^H \mathbf{v}$ is orthogonal to $\mathbf{r}_k$ with $k \geq 1$, event $\left\{ \left\| \mathbf{D}[1]^H \mathbf{v} \right\|_2^2 < x \right\}$ is only dependent

on event $\{\|\mathbf{r}_0\|_2^2 < N\beta\sigma^2\}$ and independent on events $\{\|\mathbf{r}_k\|_2^2 < N\beta\sigma^2\}_{k \geq 1}$. So we have

$$\begin{aligned} & \Pr(T_1 < x \mid K' = k) \\ &= \Pr(T_1 < x \mid \|\mathbf{r}_0\|_2^2 > N\beta\sigma^2) \\ &= \Pr(T_1 < x \mid \|\mathbf{r}_0\|_2^2 > N\beta\sigma^2, \|\mathbf{r}_1\|_2^2 < N\beta\sigma^2) \cdot \\ &= \Pr(T_1 < x \mid K' = 1) \end{aligned}$$

With the density functions and C_k , we derive the PFA of architecture 2 as follows. Suppose that $\|\mathbf{r}_l\|_2^2 \leq \beta N\sigma^2$ and the BOGA stops at step $K' = k$. According to the detection architecture 2, detection may end at each step l for $l \leq k$. When detection ends at $l = 0$, there is no false alarm. When $l = 1$, $T_{arch2} = T_1$ follows $f'_{T_1|K'=k}(x) = f'_{T_1|K'=1}(x)$. The false alarm in this step is given by

$$P_{FA,1|K'=k} = \int_{\frac{2\gamma'_1}{N\sigma^2}}^{\infty} f'_{T_1|K'=1}(x) dx = Q_{f'_1}\left(\frac{2\gamma'_1}{N\sigma^2}\right) \quad (\text{B.25})$$

where $Q_{f'_l}$ is the right tail function of the density function of T_{arch2} at the l -th step. When $l = 2$, $T_{arch2} = T_1 + T_2$ follows $f'_{T_1+T_2|K'=2}(x)$, the PFA at this step is bounded by

$$\begin{aligned} P_{FA,2|K'=k} &\leq \Pr\left(T_1 < \frac{2\gamma'_1}{N\sigma^2} \mid K'=2\right) \times \Pr\left(T_2 + T_1 \geq \frac{2\gamma'_2}{N\sigma^2} \mid K'=2\right) \\ &= Q_{f'_2}(1 - Q_{f'_1}). \end{aligned} \quad (\text{B.26})$$

Similarly, at step $l' = k$, the PFA is

$$P_{FA,k|K'=k} \leq \{Q_{f'_1} + (1 - Q_{f'_1})Q_{f'_2} + \cdots + (1 - Q_{f'_1}) \cdots (1 - Q_{f'_{k-1}})Q_{f'_k}\} \quad (\text{B.27})$$

The final PFA bound of architecture 2 under a random K' is

$$\begin{aligned} P_{FA} &= \sum_k P_{FA,k|K'=k} \Pr(K' = k) \\ &\leq c_1(\beta)Q_{f'_1} + c_2(\beta)\left(Q_{f'_1} + (1 - Q_{f'_1})Q_{f'_2}\right) + \cdots \\ &\quad + c_l(\beta)\{Q_{f'_1} + (1 - Q_{f'_1})Q_{f'_2} + \cdots + (1 - Q_{f'_1}) \cdots (1 - Q_{f'_{k-1}})Q_{f'_k}\} \\ &\quad + \cdots \end{aligned} \quad (\text{B.28})$$

Since the exact PFA of architecture 2 can not be given theoretically, there is no way to find the optimal set of thresholds for all the steps of the BOGA. One practical way

is to compute the threshold γ_l or γ'_l in the l -th step directly based on the distribution of T_l under a given false alarm P_{FA}^g . The bounds of final PFA by (B.21) and (B.28) of architecture 2 under different selections of K' can be accordingly recomputed to be:

$$\begin{aligned} P_{\text{FA}}^F &\leq P_{\text{FA}}^g + (1 - P_{\text{FA}}^g) P_{\text{FA}}^g + \cdots + (1 - P_{\text{FA}}^g) \cdots (1 - P_{\text{FA}}^g) P_{\text{FA}}^g \\ &= 1 - (1 - P_{\text{FA}}^g)^{K'} \end{aligned} \quad (\text{B.29})$$

with

$$P_{\text{FA}}^g = Q_{f_l} \left(\frac{\gamma_l}{N\sigma^2} \right). \quad (\text{B.30})$$

and

$$P_{\text{FA}}^R \leq c_1(\beta) P_{\text{FA}}^g + c_2(\beta) \left(1 - (1 - P_{\text{FA}}^g)^2 \right) + \cdots + c_l(\beta) \left(1 - (1 - P_{\text{FA}}^g)^l \right) + \cdots \quad (\text{B.31})$$

with

$$P_{\text{FA}}^g = Q_{f'_l} \left(\frac{2\gamma'_l}{N\sigma^2} \right) \quad (\text{B.32})$$

where P_{FA}^F , P_{FA}^R and P_{FA}^g represent the probability of false alarm at fixed K' , random K' determined by the stopping criterion $\|\mathbf{r}_l\|_2^2 \leq \beta N\sigma^2$ and the given false alarm at each step of the BOGA.

B.3 Detection Probability of GLRT

In this part, the theoretical detection probability of the GLRT detector is derived under the prior that one harmonic set is present with unknown fundamental frequency ω . The data follows linear model $\mathbf{x} = \mathbf{D}_\omega \mathbf{b}_\omega + \mathbf{v}$. The logarithm of the GLRT is then given by $\ln \max_{\omega} \left\{ p(\mathbf{x}; \hat{\mathbf{b}}_\omega, \omega, H_1) / p(\mathbf{x}; H_0) \right\}$.

Using the GLRT for linear model [109], the test statistic is

$$T_{\text{GLRT}} = \max_{\omega} \frac{\|\mathbf{D}_\omega^H \mathbf{x}\|_2^2}{N\sigma^2/2}. \quad (\text{B.33})$$

Suppose there are L possible fundamental frequencies, the size of the searching range of ω is L . The PD and PFA are given by

$$P_D = Q_{\chi_{2M}^2(\lambda)}(\gamma_{GLRT}), \quad (\text{B.34})$$

$$\begin{aligned} P_{FA} &= \Pr \left(\max_{\omega} \frac{\|\mathbf{D}_{\omega}^H \mathbf{x}\|_2^2}{N\sigma^2/2} \geq \gamma_{GLRT}; H_0 \right) \\ &= 1 - \left(1 - Q_{\chi_{2M}^2}(\gamma_{GLRT}) \right)^L. \end{aligned} \quad (\text{B.35})$$

with $\lambda = 2N\|\mathbf{b}_{\omega}\|_2^2/\sigma^2$ and γ_{GLRT} is the threshold determined by the false alarm probability. For harmonics with one fundamental frequency, this GLRT detector is optimal. When multiple fundamental frequencies are present in the signal, this detector can be used to successively detect harmonic block corresponding to each fundamental frequency with the detection probabilities determined by the SNRs of harmonic blocks using (B.34).

When the harmonic structure is not considered, (B.33) turns into the conventional single tone detector [109]:

$$T_s = \max_{\omega} \frac{\|\mathbf{d}_{\omega}^H \mathbf{x}\|_2^2}{N\sigma^2/2} = \max_{\omega} P(\omega) \quad (\text{B.36})$$

where $\mathbf{d}_{\omega}$ is one atom in Fourier basis corresponding to ω and $P(\omega)$ is the periodogram evaluated at frequency ω . The PD and PFA are given by

$$P_D = Q_{\chi_2^2(\lambda)}(\gamma_s), \quad (\text{B.37})$$

$$P_{FA} = 1 - \left(1 - Q_{\chi_2^2}(\gamma_s) \right)^L. \quad (\text{B.38})$$

where $\lambda = 2N\|a\|_2^2/\sigma^2$, a is the complex amplitude of the tone corresponding to fundamental frequency ω and γ_s is the threshold.

B.4 Detection Probability of Energy Detector (ED)

Suppose the only information known is that signal follows the linear model (4.11). Without using the sparsity prior, the ED is then given by [109]

$$T_{ED} = \frac{\|\mathbf{D}^H \mathbf{x}\|_2^2}{N\sigma^2/2}. \quad (\text{B.39})$$

The detection performance can be given by

$$P_D = Q_{\chi^2_{2ML}}(2N\|\mathbf{b}_0\|_2^2/\sigma^2)(\gamma_{ED}), \quad (\text{B.40})$$

$$P_{FA} = Q_{\chi^2_{2ML}}(\gamma_{ED}), \quad (\text{B.41})$$

where γ_{ED} is the threshold, L is the number of possible fundamental frequencies.

The detection probability can be very low if L is large.

Appendix C

Maximal Coherence of the Dictionary $\mathbf{D}$ in Chapter 6

The approximate maximal coherence of the dictionary $\mathbf{D}$ with full measurements is derived in this Appendix. When full measurements are used, the number of measurements $L = MN$. The mutual coherence between atom $\mathbf{d}_k$ and $\mathbf{d}_{k'}$, denoted as $h(k, k')$, where $k \neq k'$, can be computed as

$$h(k, k') = \frac{|\mathbf{d}_k^H \mathbf{d}_{k'}|}{MN}. \quad (\text{C.1})$$

Substituting (6.6) into (C.1), we have

Table C.1: The maximal coherence

B/F_c	0.04	0.06	0.08	0.1	0.12	0.14	0.15	0.16	0.18	0.2
$N = 50$	0.2738	0.2991	0.2848	0.2118	0.1875	0.1576	0.1422	0.1382	0.1287	0.1096
$N = 100$	0.2800	0.1868	0.1383	0.1088	0.0909	0.0780	0.0731	0.0700	0.0647	0.0598

$$h(k, k') = \frac{1}{MN} \left| \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \exp \left\{ -j \frac{4\pi f_m}{c} \left((y_{2,k} - y_{2,k'}) + \frac{(y_{1,k} - y_{1,k'}) |\Omega_{eff}|}{PRF} n \right) \right\} \right|. \quad (\text{C.2})$$

Let $\Delta y_2 = (y_{2,k} - y_{2,k'})$ and $\Delta y_1 = \frac{(y_{1,k} - y_{1,k'}) |\Omega_{eff}|}{PRF}$, it can be easily found that

$$\begin{aligned} \Delta y_2 &= l_2 \frac{c}{2B}, & -M + 1 &\leq l_2 \leq M - 1, \\ \Delta y_1 &= l_1 \frac{\lambda}{2N}, & -N + 1 &\leq l_1 \leq N - 1. \end{aligned}$$

Substituting the above equations into (C.2), after some direct derivations, we have

$$\begin{aligned}
 h(k, k') &= \frac{1}{MN} \left| \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \exp \left\{ -j \frac{4\pi f_m}{c} \left(l_2 \frac{c}{2B} + l_1 \frac{\lambda}{2N} n \right) \right\} \right| \\
 &= \frac{1}{MN} \left| \sum_{n=0}^{N-1} \exp \left\{ -j \frac{4\pi}{c} (F_c - 2\Delta f) \frac{\lambda}{2N} l_1 n \right\} \frac{\sin \left(\frac{2\pi B}{c} (l_2 \frac{c}{2B} + l_1 \frac{\lambda}{2N} n) \right)}{\sin \left(\frac{2\pi \Delta f}{c} (l_2 \frac{c}{2B} + l_1 \frac{\lambda}{2N} n) \right)} \right| \quad (C.3)
 \end{aligned}$$

where $\Delta f = B/M$. If M is large enough, that is Δf is extremely small, then the last item in (C.3) can be simply approximated by a **sinc** function, and $2\Delta f$ in the first term in (C.3) can be ignored. Therefore, (C.3) can be written as

$$h(k, k') \approx \frac{1}{N} \left| \sum_{n=0}^{N-1} \exp \left\{ -j 2\pi \frac{l_1}{N} n \right\} \text{sinc} \left(\pi \left(l_2 + \frac{B}{F_c} \frac{l_1}{N} n \right) \right) \right|. \quad (C.4)$$

As seen from (C.4), the coherence is determined by N , B/F_c , l_1 and l_2 . In the following, the effect of l_1 and l_2 on the value of h is discussed.

(i) When $l_1 = l_2 = 0$, h evaluates the coherence of one atom with itself, and $h = 1$.

(ii) When $l_1 = 0$ and $l_2 \neq 0$, we have

$$h \approx \frac{1}{N} \left| \sum_{n=0}^{N-1} \text{sinc}(\pi l_2) \right| = 0$$

(iii) When $l_1 \neq 0$, the analyses of h become complicated. Given a fixed N , h highly depends on the value of B/F_c .

Generally, signal of X-band is used for ISAR imaging with F_c around 8–10GHz and B around 500–1000MHz. Therefore, we consider in this chapter $0.04 \leq \frac{B}{F_c} \leq 0.2$. Since close form of h over B/F_c is difficult to derive, we present maximal value of h for two different atoms numerically in Table C.1 when $N = 50$ and 100. As indicated by Table C.1, the maximal coherence decreases with the ratio B/F_c in the range 0.08–0.2. And the maximal coherence for $N = 100$ is almost one half of that for $N = 50$ when $0.08 \leq \frac{B}{F_c} \leq 0.2$. Therefore, we can always design $B/F_c = 0.2$ for radar system to guarantee a low maximal coherence in dictionary **D**.

Author's Publications

- (1) **Lu Wang**, Lifan Zhao, Guoan Bi, Chunru Wan, and Lei Yang, “Enhanced ISAR Imaging by Exploiting the Continuity of the Target Scene,” *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 52, no. 9, pp. 5736-5750, Sep. 2014.
- (2) **Lu Wang**, Lifan Zhao, Guoan Bi and Chunru Wan, “Hierarchical Sparse Signal Recovery by Variational Bayesian Inference,” *IEEE Signal Processing Letter*, vol, 21, no, 1, pp. 110-113, Jan. 2014.
- (3) **Lu Wang**, Guoan Bi, Chunru Wan and Xiaolei Lv, “Improved Stability Conditions of BOGA for Noisy Block-Sparse Signals,” *Signal Processing*, vol. 91, no. 11, pp. 2567-2574, Nov. 2011.
- (4) **Lu Wang**, Chunru Wan, Shenghong Li and Guoan Bi, “Harmonic Tonal Detectors based on the BOGA,” Accepted for publication in *Signal Processing*.
- (5) **Lu Wang**, Lifan Zhao, Guoan Bi and Chunru Wan, “Sparse Representation based ISAR Imaging Using Markov Random Field,” Submitted to *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS)*.
- (6) **Lu Wang**, Lifan Zhao, Guoan Bi, Chunru Wan and Haijian Zhang, “Novel Wideband DOA Estimation based on Bayesian Compressive Sensing with Dirichlet Process Prior,” Submitted to *IEEE Transactions on Wireless Communication (TWC)*.

- (7) Lifan Zhao, **Lu Wang**, Guoan Bi and Lei Yang, “An Autofocus Technique for High-resolution Inverse Synthetic Aperture Radar Imagery,” Accepted for publication in *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*.
- (8) Lifan Zhao, Guoan Bi, **Lu Wang** and Haijian Zhang, “An Improved Auto-Calibration Algorithm Based on Sparse Bayesian Learning Framework,” *IEEE Signal Processing Letter*, vol. 20, no. 9, pp. 889-892, Sep. 2013.
- (9) **Lu Wang**, Lifan Zhao, Guoan Bi and Liren Zhang, “ISAR Imaging by Exploiting the Continuity of Target Scene,” Accepted by *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.
- (10) **Lu Wang**, Lifan Zhao and Guoan Bi, “Structured Sparse Representation based ISAR Imaging,” Accepted by *International Radar Symposium(IRS)*, 2014.
- (11) **Lu Wang**, Guoan Bi, “Harmonic signal recovery and order estimation based on cascaded sparse processing,” *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2171-2174, 19-23 May 2013.
- (12) **Lu Wang**, Guoan Bi and Chunru Wan, “Detector for tonals with harmonics,” *IEEE Region 10 Conference (TENCON)*, pp. 1-5, 19-22 Nov. 2012.
- (13) **Lu Wang**, Guoan Bi and Xiaolei Lv, “Analysis of sidelobe effects and resolution of PTFT,” *IEEE International Conference on Control and Automation (ICCA)*, pp. 1049-1054, 19-21 Dec. 2011.

References

- [1] D. Donoho, M. Elad, and V. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [2] X. Lv, G. Bi, and C. Wan, “Block orthogonal greedy algorithm for stable recovery of block-sparse signal representations,” *Signal Processing*, vol. 86, no. 4, pp. 511–532, Apr. 2010.
- [3] B. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Sci. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
- [4] I. Gorodnitsky and B. 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, Mar. 1997.
- [5] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Apr. 1998.
- [6] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.
- [7] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [8] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, pp. 4655–4666, Dec. 2007.

REFERENCES

- [9] D. Needell and R. Vershynin, “Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit,” *Found. Comput. Math.*, vol. 9, no. 12, pp. 317–334, Dec. 2009.
- [10] M. Elad and A. Bruckstein, “A generalized uncertainty principle and sparse representation in pairs of bases,” *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.
- [11] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [12] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [13] P. Tseng, “Further results on stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 888–899, Feb. 2009.
- [14] J. Fuchs, “Recovery of exact sparse representations in the presence of bounded noise,” *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.
- [15] M. Tipping, “Sparse bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Jan. 2001.
- [16] D. Wipf and B. Rao, “Sparse bayesian learning for basis selection,” *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [17] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.
- [18] M. Girolami, “A variational method for learning sparse and overcomplete representations,” *Neural Comput.*, vol. 13, no. 11, pp. 2517–2532, Nov 2001.

REFERENCES

- [19] S. Babacan, R. Molina, and A. Katsaggelos, “Bayesian compressive sensing using laplace priors,” *IEEE Transactions on Image Processing*, vol. 19, pp. 53–63, Jan. 2010.
- [20] D. Wipf, J. Palmer, and B. Rao, “Perspectives on sparse bayesian learning,” in *Advances in Neural Information Processing Systems 16*, 2003.
- [21] M. A. Figueiredo, “Adaptive sparseness for supervised learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1150–1159, 2003.
- [22] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, “An iterative bayesian algorithm for sparse component analysis in presence of noise,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4378–4390, 2009.
- [23] C. Soussen, J. Idier, D. Brie, and J. Duan, “From bernoulli-gaussian deconvolution to sparse signal restoration,” *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4572–4584, 2011.
- [24] M. Beal, “Variational algorithms for approximate bayesian inference,” Ph.D. dissertation, University College of London, 2004.
- [25] L. He, H. Chen, and L. Carin, “Tree-structured compressive sensing with variational bayesian analysis,” *IEEE Signal Process. Lett.*, vol. 17, no. 3, pp. 233–236, Mar. 2010.
- [26] L. He and L. Carin, “Exploiting structure in wavelet-based bayesian compressive sensing,” *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, Sep. 2009.
- [27] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.
- [28] A. Dremeau, C. Herzet, and L. Daudet, “Boltzmann machine and mean-field approximation for structured sparse decompositions,” *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3425–3438, 2012.

REFERENCES

- [29] Y. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [30] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *J. Roy. Stat. Soc. Series B*, vol. 68, pp. 49–67, 2006.
- [31] X. Lv, G. Bi, and C. Wan, “The group lasso for stable recovery of block-sparse signal representations,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1371–1382, Apr. 2011.
- [32] Y. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010.
- [33] M. Stojnic, F. Parvaresh, and B. Hassibi, “On the reconstruction of block-sparse signals with an optimal number of measurements,” *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
- [34] Z. Zhang and B. Rao, “Extension of sbl algorithm for recovery of block sparse signals with intra-block correlation,” *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.
- [35] L. Yu, H. Sun, J. Barbot, and G. Zheng, “Bayesian compressive sensing for cluster structured sparse signals,” *Signal Processing*, vol. 92, pp. 259–269, Jan. 2012.
- [36] L. Yu, J. Barbot, G. Zheng, and H. Sun, “Compressive sensing for cluster structured sparse signals: variational bayes approach,” in *Technical report*. Cergy-Pontoise, France, 2012.
- [37] J. Huang, T. Zhang, and D. Metaxas, “Learning with structured sparsity,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 1–8.
- [38] T. Peleg, Y. Eldar, and M. Elad, “Exploiting statistical dependencies in sparse representations for signal recovery,” *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2286–2303, 2012.

REFERENCES

- [39] P. J. Garrigues and B. A. Olshausen, “Learning horizontal connections in a sparse coding model of natural images,” in *Advances in Neural Info. Proc. Sys. (NIPS)*, vol. 60, no. 5, Dec. 2008, pp. 505–512.
- [40] V. Cevher, M. F. Duarte, C. Hegde, and R. G. Baraniuk, “Sparse signal recovery using markov random fields,” in *Advances in Neural Info. Proc. Sys. (NIPS)*, Dec. 2008.
- [41] R. Gribonval and E. Bacry, “Harmonic decomposition of audio signals with matching pursuit,” *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–110, 2003.
- [42] J. Fuchs, “Extension of the global matched filter to structured groups of atoms: Application to harmonic signals,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*. IEEE, May 2011, pp. 3996–3999.
- [43] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. Eldar, “C-Hilasso: A collaborative hierarchical sparse modeling framework,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 4183–4198, Sep. 2011.
- [44] J. Friedman, T. Hastie, and R. Tibshirani, “A note on the group lasso and a sparse group lasso,” in *Technical report*. Department of Statistics, Stanford University, 2010.
- [45] L. Zhang, M. Xing, C. Qiu, J. Li, J. Sheng, Y. Li, and Z. Bao, “Resolution enhancement for inversed synthetic aperture radar imaging under low snr via improved compressive sensing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3824–3838, Oct. 2010.
- [46] G. Li, H. Zhang, X. Wang, and X. Xia, “Isar 2-d imaging of uniformly rotating targets via matching pursuit,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1838–1846, Apr. 2012.
- [47] G. Xu, M. Xing, L. Zhang, Y. Liu, and Y. Li, “Bayesian inverse synthetic aperture radar imaging,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1150–1154, Nov. 2011.

REFERENCES

- [48] H. Wang, Y. Quan, M. Xing, and S. Zhang, "Isar imaging via sparse probing frequencies," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 451–455, May 2011.
- [49] J. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [50] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.
- [51] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
- [52] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*. IEEE, April 2008, pp. 3869–3872.
- [53] D. Wipf and S. Nagarajan, "Iterative reweighted l_1 and l_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, April 2010.
- [54] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," *J. Constr. Approx.*, vol. 13, p. 57C98, 1997.
- [55] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *In Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, vol. 1. IEEE, Nov. 1993, pp. 40–44.
- [56] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

REFERENCES

- [57] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec. 2010.
- [58] J. Tropp and S. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.
- [59] M. S. Lewicki, T. Sejnowski, and H. Hughes, “Learning overcomplete representations,” *Neural Computation*, vol. 12, pp. 337–365, 1998.
- [60] S. D. Babacan, S. Nakajima, and M. N. Do, “Bayesian group-sparse modeling and variational inference,” *IEEE Transactions on Signal Processing*, submitted for publication.
- [61] B. Chen, J. Paisley, and L. Carin, “Sparse linear regression with beta process priors,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 1234–1237.
- [62] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, “Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images,” *IEEE Trans. Image Processing*, vol. 21, pp. 130–144, Jan. 2012.
- [63] J. Vila and P. Schniter, “Expectation-maximization bernoulli-gaussian approximate message passing,” in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, 2011, pp. 799–803.
- [64] S. Chatterjee, D. Sundman, M. Vehkapera, and M. Skoglund, “Projection-based and look-ahead strategies for atom selection,” *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 634–647, Feb. 2012.
- [65] T. Blumensath and M. Davies, “Sampling theorems for signals from the union of finite-dimensional linear subspaces,” *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.

REFERENCES

- [66] Z. Ben, Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1032–1046, 2011.
- [67] L. Gray and D. Greeley, “Source level model for propeller blade rate radiation for the world’s merchant fleet,” *J. Acoust. Soc. Am.*, vol. 67, no. 2, pp. 516–522, 1980.
- [68] J. Terry, A. Crampton, and C. Talbot, “Passive sonar harmonic detection using feature extraction and clustering analysis,” in *OCEANS, Proceedings of MTS/IEEE*, 2005, pp. 2760–2766, Vol. 3.
- [69] J. Lourens and J. du Prcez, “Passive sonar ml estimator for ship propeller speed,” *IEEE J. Oceanic Eng.*, vol. 23, no. 4, pp. 448–453, 1998.
- [70] R. Nielsen, *Sonar Signal Processing*. Artech House, 1991.
- [71] R. Wagstaff, “The awsum filter: A 20 db gain fluctuation based processor,” *IEEE J. Oceanic Eng.*, vol. 22, no. 1, pp. 110–118, 1997.
- [72] C. Wan, J. Goh, and H. Chee, “Optimal tonal detectors based on the power spectrum,” *IEEE J. Oceanic Eng.*, vol. 25, no. 4, pp. 504–552, 2000.
- [73] Q. Wang and C. Wan, “A novel cfar tonal detector using phase compensation,” *IEEE J. Oceanic Eng.*, vol. 30, no. 4, pp. 2146–2157, 2005.
- [74] L. Scharf and B. Friedlander, “Matched subspace detector,” *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2146–2157, 1994.
- [75] A. Yang, J. Wright, Y. Ma, and S. Sastry, “Feature selection in face recognition: A sparse representation perspective,” UC Berkeley Tech, Tech. Rep., 2007.
- [76] R. Vidal and Y. Ma, “A unified algebraic approach to 2-d and 3-d motion segmentation and estimation,” *Journal of Mathematical Imaging and Vision*, vol. 25, no. 3, pp. 403–421, 2006.

REFERENCES

- [77] M. Mishali and Y. Eldar, “Blind multi-band signal reconstruction: compressed sensing for analog signal,” *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, 2009.
- [78] L. Wang, G. Bi, C. Wan, and X. Lv, “Improved stability conditions of boga for noisy block-sparse signals,” *Signal Processing*, vol. 91, no. 11, pp. 2567–2574, 2011.
- [79] M. Duarte and R. Baraniuk, “Spectral compressive sensing,” [online] Available: <http://dsp.rice.edu/sites/dsp.rice.edu/files/publications/journal-article/9999/scs-acha.pdf>, 2012.
- [80] H. Rauhut, K. Schnass, and P. Vandergheynst, “Compressed sensing and redundant dictionaries,” *IEEE Trans. Info. Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.
- [81] J. Tabrikian, S. Dubnov, and Y. Dickalov, “Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 12, no. 1, pp. 76–87, 2004.
- [82] A. Koretz and J. Tabrikian, “Maximum a posteriori probability multiple-pitch tracking using the harmonic model,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2210–2221, 2011.
- [83] M. Duarte, M. Davenport, M. Wakin, and R. Baraniuk, “Sparse signal detection from incoherent projections,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP’06)*. Toulouse, 2006, pp. III 305–308.
- [84] D. Tzikas, A. Likas, and N. Galatsanos, “The variational approximation for bayesian inference,” *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
- [85] B. Jorgensen, *Statistical properties of the generalized inverse Gaussian distribution*. New York: Springer-Verlag, 1982.

REFERENCES

- [86] L. Zhao, G. Bi, L. Wang, and H. Zhang, “An improved auto-calibration algorithm based on sparse bayesian learning framework,” *IEEE Signal Process. Lett.*, vol. 20, pp. 889–892, Sep. 2013.
- [87] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311–4322, Nov. 2006.
- [88] V. Patel, G. Easley, J. Healy, D.M., and R. Chellappa, “Compressed synthetic aperture radar,” *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 244–254, Apr. 2010.
- [89] J. Xu, Y. Pi, and Z. Cao, “Bayesian compressive sensing in synthetic aperture radar imaging,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 2–8, Jan. 2012.
- [90] G. Zhao, Z. Wang, Q. Wang, G. Shi, and F. Shen, “Robust isar imaging based on compressive sensing from noisy measurements,” *Signal Process.*, vol. 92, no. 1, pp. 120–129, Jan. 2012.
- [91] J. Walker, “Range-doppler imaging of rotating objects,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 16, no. 1, pp. 23–52, Jan. 1980.
- [92] M. Tello Alonso, P. Lopez-Dekker, and J. Mallorqui, “A novel strategy for radar imaging based on compressive sensing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4285–4295, 2010.
- [93] M. Çetin and W. Karl, “Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization,” *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 623–631, Sep. 2001.
- [94] N. Onhon and M. Cetin, “A sparsity-driven approach for joint sar imaging and phase error correction,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2075–2088, Sep. 2012.

REFERENCES

- [95] N. Dobigeon, A. Basarab, D. Kouame, and J. Tourneret, “Regularized bayesian compressed sensing in ultrasound imaging,” in *Proc. IEEE 20th European Signal Process. Conf. (EUSIPCO)*. Bucharest, Romania, Aug. 2012, pp. 2600–2604.
- [96] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [97] D. Baron, M. Duarte, S. Sarvotham, M. Wakin, and R. Baraniuk, “Distributed compressed sensing,” *[online] Available: <http://dsp.rice.edu/cs/DCS112005.pdf>*, 2005.
- [98] C. La and M. Do, “Tree-based orthogonal matching pursuit algorithm for signal reconstruction,” in *IEEE Int. Conf. on Image Process. (ICIP)*. Atlanta, GA, Oct. 2006, pp. 1277–1280.
- [99] M. Duarte, M. Wakin, and R. Baraniuk, “Wavelet-domain compressive signal reconstruction using a hidden markov tree model,” in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*. Las Vegas, NV, Apr. 2008, pp. 5137–5140.
- [100] F. Berizzi, E. Mese, E. Diani, and M. Martorella, “High-resolution isar imaging of maneuvering targets by means of the range instantaneous doppler technique: modeling and performance analysis,” *IEEE Trans. Image Process.*, vol. 10, no. 12, pp. 1880–1890, Dec. 2001.
- [101] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [102] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, “Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images,” *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130 – 144, Jan. 2012.

REFERENCES

- [103] C. Quinsac, N. Dobigeon, A. Basarab, J. Tourneret, and D. Kouame, “Bayesian compressed sensing in ultrasound imaging,” in *Proc. IEEE Int. Workshop Comput. Adv. in Multi-Sensor Adaptive Process. (CAMSAP)*. San Juan, Puerto Rico, Dec. 2011, pp. 101–104.
- [104] L. Yang, M. Xing, L. Zhang, J. Sheng, and Z. Bao, “Entropy-based motion error correction for high-resolution spotlight sar imagery,” *IET Radar, Sonar Navigation*, vol. 6, no. 7, pp. 627–637, July 2012.
- [105] D. Wipf and B. Rao, “An empirical bayesian strategy for solving the simultaneous sparse approximation problem,” *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.
- [106] L. Sun, J. Liu, J. Chen, and J. Ye, “Efficient recovery of jointly sparse vectors,” in *NIPS*, 2009, pp. 1812–1820.
- [107] K. Lee, Y. Bresler, and M. Junge, “Subspace methods for joint sparse recovery,” *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.
- [108] K. Alam and K. Wallenius, “Distribution of a sum of order statistics,” *Scand. J. Statis.*, vol. 6, no. 3, pp. 123–126, 1979.
- [109] S. Kay, *Fundamentals of statistical signal processing: Detection Theory*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall PTR: Prentice-Hall Signal Processing Series, ch.7, sec.7.6-7.7, 1998.