香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

A Primal-Dual Approach to Robust Covariance Estimation in the Presence of Non-Gaussian Interferences

LIU Ye
刘晔

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Computer and Information Engineering
The Chinese University of Hong Kong, Shenzhen
January 2025

Thesis Assessment Committee

Professor YIN Feng(Chair)

Professor LUO Zhiquan (Thesis Supervisor)

Doctor PU Wenqiang (Thesis Co-supervisor)

Professor SHEN Kaiming (Committee Member)

Professor WAI Hoi To (Examiner from CUHK)

Professor FU Xiao (External Examiner)

Abstract

of thesis entitled:
A Primal-Dual Approach to Robust Covariance Estimation in the Presence of Non-Gaussian Interferences

Submitted by LIU Ye

for the degree of Master of Philosophy

at The Chinese University of Hong Kong, Shenzhen in January 2025

Robust covariance estimation is critical for array processing in complex environments, where the non-uniform power distribution of multi-source interference induces non-Gaussian and heavy-tail signal statistics. Traditional methods relying on Gaussian assumptions suffer performance degradation in such scenarios. To address this challenge, we propose a cardinality-constrained variance (CCV) model that explicitly models sparse spatial power distribution in the variance domain, corresponding to the average power of signal sources. The CCV establishes direct connections with physically interpretable parameters (e.g., the maximum number of dominant interferers or sparsity level of interference sources), thereby eliminating the need for a sensitive hyper parameter tune by traditional penalty based approaches. This fundamentally differs from conventional sparsity constraints on instantaneous signal amplitudes (e.g., ℓ_0/ℓ_1-norm regularization), the CCV directly imposes cardinality constraint on the variance domain, thereby capturing the statistical essence of the

interference power, with physically meaningful control parameters.

We formulated the robust covariance estimation within a maximum likelihood estimation (MLE) framework, incorporating the proposed CCV constraint. However, the resulting optimization problem inherits the non-convexity from the likelihood objective and combinatorial complexity from the cardinality constraint, making conventional gradient-based methods inapplicable. To overcome this, we first develop a convex relaxation on CCV constraint. By introducing an auxiliary variable, we transform the discontinuous constraint into a continuous feasible domain. Subsequently, leveraging the Majorization-Minimization (MM) framework, we derived a convex upper bound of the non-convex objective, and construct a sequence of convex surrogates subproblems. This sequence is theoretically guaranteed to linearly converge to the stationary points. Meanwhile, by deriving closed-form primal updates under Karush-Kuhn-Tucker (KKT) conditions, we eliminate the inner-loop iterations, and finally establish a single-loop primal-dual algorithm.

The key contributions of this work are twofold: (1) the introduction of the variance-centric CCV model to characterize non-uniform interference power distributions, establishing the connection between sparse signal recovery and statistical covariance estimation; and (2) a theoretically grounded optimization framework with convergence guarantees to stationary points, integrated with a single-loop primal update scheme that achieves computational efficiency without inner iterations.

摘要

中文标题

阵列信号处理中的稳健协方差估计对复杂环境下的干扰抑制具有至关重要的作用。当多源干扰信号在空间上呈现非均匀功率分布时，其统计特性往呈现重尾非高斯特性，导致传统高斯假设下的估计方法性能显著下降。为此，本文提出基数约束方差模型（Cardinality-Constrained Variance, CCV），通过方差域的基数约束来建模干扰源空间功率分布的稀疏特性。CCV 通过直接关联物理可解释参数（如主导干扰源最大数量），从根本上规避了传统惩罚方法中敏感超参的人工选择。相较于传统方法对瞬时幅值施加 ℓ_0/ℓ_1 范数正则化的稀疏约束，CCV 方法将约束拓展至方差域，因此从统计本质上以物理可解释参数来刻画多源干扰的非均匀特性。

本研究在最大似然估计框架下面临双重挑战：目标函数的非凸性与基数约束的组合复杂性。这使得传统梯度求解方法在此问题上失效。为此，本文提出分阶段求解策略：首先对 CCV 约束实施凸松弛，通过引入辅助变量将组合型离散约束转换至连续可行域；随后基于最大化-最小化（Majorization-Minimization, MM）框架来构造非凸目标函数的凸上界逼近，构建序列凸优化子问题。该序列理论上以线性速度收敛至驻点。与此同时，本文通过 KKT 条件（Karush-Kuhn-Tucker）推导原始变量的显式解析解，从而避免内层迭代，最终建立单循环的原始-对偶迭代算法。

本文的贡献体现在两方面：首先，提出基于方差域基数约束（CCV）的干扰

功率表征模型，在信号统计特征上直接刻画稀疏约束机制；其次，针对非凸目标与基数约束，提出了具有驻点收敛理论保证的凸松弛序列优化框架，通过闭式解更新机制实现单循环的原始对偶迭代，从而消除内层迭代，提升计算效率。

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Luo Zhi-Quan, for providing me with the opportunity to explore the fields of engineering and optimization and taught the value of research.

I am grateful to Dr. Pu Wenqiang for his guidance, patience, and support throughout my academic journey. His mentorship has been crucial in helping me overcome challenges and strengthen my research confidence.

I would like to extend my gratitude to Professor Feng Yin, who has known me since my undergraduate studies and has been a constant source of encouragement.

I feel fortunate to have had the chance to get to know my lab mates: Zhang Jiawei, Mao Jingwei, Ren Shuyi, Liang Hao, Xiao Jiancong, Li Yingru, Tang Liping, Liu Guangyi, Chen Congliang, Liu Xiang, Wang Wenxuan, Wang Boyuan, Meng Qingyan, Liu Liangqi, Li Ying, and Li Kai, as well as Dimitry Rybin and Sergei Kudria. I also want to acknowledge my peers from other labs, Qin Zeyu, Wang Yinhao, and Xiao Peijun. I cherish the time I spent with all of you, especially in the beginning when I was learning the ropes. I deeply appreciate your willingness to generously share your knowledge and expertise. The memories we made together outside the lab – from sports to sharing meals –will be forever treasured.

I am profoundly grateful to my friend, Zhao Heming, for her unwavering support and for being there to witness my growth over these years.

Finally, I would like to thank my parents for their endless love and support.

Contents

Abstract i

摘要 iii

Acknowledgement v

Contents vi

List of Figures ix

List of Tables x

1 Introduction 1

1.1 Covariance Estimation in Signal Processing 1

1.2 Contributions 2

1.3 Organization 3

2 Fundamentals of Interference Suppression 5

2.1 Beamforming Problem 5

2.2 Interference Signal Modeling with Statistical Constraints 8

2.2.1 Steering Vector: The Angular Encoder 10

2.2.2 Dictionary Matrix 10

2.2.3 Cardinality-Constrained Variance (CCV) 11

2.2.4 Spiky Coefficient Vector . . . 14
2.2.5 Representing Interference . . . 15
2.3 Covariance Structure and Parametrization . . . 16

3 Problem Formulation: Robust Maximum Likelihood 22
3.1 Robust Elliptical Prior as Objective Function . . . 22
3.1.1 Heavy-Tail Behavior Induced by Spiky Coefficients . . . 23
3.1.2 Heavy-Tailed Behavior Under Limited Samples . . . 26
3.1.3 The Robust Elliptical Prior . . . 27
3.1.4 The Robust Maximum Likelihood Objective . . . 28
3.2 Constraints Formulation . . . 29
3.2.1 Cardinality-Constraint Variance . . . 29
3.2.2 Model-Induced Constraint . . . 30
3.3 Robust Covariance Estimation with CCV . . . 31
3.3.1 Optimization Challenges . . . 31

4 Convex Relaxation 33
4.1 Convex Relaxation on Cardinality Variance Constraint . . . 33
4.1.1 Convexify CCV with Box Variable . . . 33
4.1.2 Simplify: One-Side Box . . . 34
4.1.3 Enhancement: Log-Domain Transformation . . . 35
4.2 Majorization-Minimization Relaxation on the Objective . . . 36
4.2.1 The Majorization-Minimization Framework . . . 36
4.2.2 Surrogate Subproblem . . . 38

5 Single Loop Primal-Dual Update Algorithm 47
5.1 Primal-Dual Method . . . 47
5.2 Update Rules . . . 49

5.2.1 Primal Update 49
5.2.2 Dual Update 50

6 Numerical Results **52**
6.1 Experimental Setup 52
6.2 Convergence 53
6.3 Computational Complexity 55
6.4 Beam Pattern 59

Bibliography **62**

List of Figures

4.1 Illustration of Majorization-Minimization 38

6.1 Raw range-amplitude data with interference. 54

6.2 Angle-Doppler response of entire intereference subspace 55

6.3 Raw Angle-Doppler Response, Demonstrating the Interference Power Spikiness and its Dominance to Target Reflection 56

6.4 Convergence of Different Algorithms 57

6.5 Beam patterns of Different Estimators, Limited Snapshots 59

6.6 Beam patterns of Different Estimators, Sufficient Snapshots 60

List of Tables

2.1 Comparison of sparsity constraints in statistics 14

3.1 Comparison of sparsity constraints in optimization aspect 30

6.1 Experimental Parameters for the Moving Array Platform 53

6.2 Theoretical Convergence Rates 54

6.3 Theoretical Per-Iteration Complexity 58

6.4 Empirical per Iteration Time (seconds) 58

Notations

$\|\cdot\|$	norm
$\lvert\cdot\rvert$	cardinality of set
$\cdot^{\mathsf{H}}$	hermitian
$\cdot^{\mathsf{T}}$	transpose
$\mathrm{tr}(\cdot)$	trace of a matrix
$\det(\cdot)$	determinant of a matrix
$\otimes$	Kronecker product
$\mathrm{tr}(\cdot)$	trace of a matrix
$\mathbf{A} \succ \mathbf{0}$	matrix $\mathbf{A}$ is positive-definite
$\mathbf{A} \succeq \mathbf{0}$	matrix $\mathbf{A}$ is semi-definite
$\mathbb{E}$	expectation
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{R}$
$\sim$	follows distribution of
$p(\cdot)$	probability density function
$P(\cdot)$	probability of event
$\mathbb{R}$	real space
$\mathbb{C}$	complex space
$\mathrm{diag}(\mathbf{A})$	vector with its elements given by diagonal entries of $\mathbf{A}$
$\mathrm{Diag}(\mathbf{A})$	diagonal matrix with its elements given by diagonal entries of $\mathbf{A}$

Abbreviations

AWGN	Additive white Gaussian noise
CCV	Cardinality-Constrained Variance
CPI	Coherent processing interval
DOA	Direction of arrival
MIMO	Multiple Input Multiple Output
MLE	Maximum likelihood estimation
MM	Majorization-minimization
MVDR	Minimum Variance Distortless Response
SINR	Signal-to-Interference-plus-Noise Ratio
SR	Sparse recovery

Chapter 1

Introduction

1.1 Covariance Estimation in Signal Processing

Covariance estimation, as a fundamental statistical tool with wide-ranging applications, plays a critical role in signal processing. In array signal processing, where systems employ sensor arrays to capture spatial/temporal signal characteristics, it enables applications such as beamforming [Capon, 1969; Van Veen et al., 1997; Li et al., 2003; Vorobyov et al., 2003], direction-of-arrival (DOA) estimation [Schmidt, 1986; Stoica and Sharman, 1990] and optimal array design [Van Trees, 2002]. Its utility extends to wireless communications, particularly in multiple input multiple output (MIMO) channel capacity analysis [Telatar, 1999; Tse and Viswanath, 2005]. Beyond signal processing, the methodology demonstrates cross-disciplinary impact through applications in financial risk modeling [Ruppert and Matteson, 2011; Rubinstein, 2002] and genomic analytics [Schäfer and Strimmer, 2005].

Beamforming enables selective amplification of signals from specified directions while suppressing interference and noise. For instance, consider a wireless base station receiving a target user's signal at a specific angle: by adaptively adjusting antenna phase shifts (i.e. beamforming weights), the system enhances the desired signal

and suppress interference from other directions. This filtering capability fundamentally relies on precise estimation of the interference and noise covariance matrix, a statistical profile of the electromagnetic environment that is typically unknown in practice.

Challenges in real-world beamforming deployment emerge from two key factors: *insufficient snapshots* for covariance estimation and *non-Gaussian interference patterns.* These issues cause numerical instability in estimators and severely limit the effectiveness of traditional methods grounded in Gaussian assumptions.

1.2 Contributions

This thesis addresses robust covariance estimation for interference suppression under limited sample size and non-Gaussian interference environments. The key contributions are:

Modeling

- **Variance-domain sparse interference characterization:** Proposed the *Cardinality-Constrained Variance (CCV)* model that imposes sparsity constraints on second-order statistics rather than instantaneous signal amplitudes, different from conventional ℓ_0/ℓ_1-norm regularization approaches.

- **Physical-interpretable control:** The proposed method differs from ℓ_1/ℓ_0 penalty methods in parameter selection. Many existing methods rely on empirical hyperparameter tunning, while our formulation uses physically interpretable parameters, specifically, the dominant interference source count and power significance thresholds obtained from expert knowledge.

- **Verification of Heavy-Tailed Property:** Theoretically verified that the received signal generated with an uneven power profile satisfying Definition 2

exhibits heavy-tailed characteristics (Theorem 1). This provides justification for selecting heavy-tailed prior distributions to model the non-Gaussian interference in the scale of this work.

Optimization

- **Convex relaxation with convergence guarantees:** Developed a Majorization-Minimization (MM)-based convex relaxation with convergence guarantee to stationary points, addressing the non-convexity of the log-likelihood objective. The proposed MM algorithm has linear convergence rate.
- **Combinatorial constraint transformation:** Relaxed the combinatorial CCV constraint into a convex continuous constraint and resolved the non-convexity and discontinuous issue of the constraint.
- **Single-loop Primal-Dual update:** Derived a close-form primal updates for the optimization subproblems via KKT conditions, thereby eliminating the need for nested iterative procedures and achieved an inner-loop-free algorithm architecture.

1.3 Organization

Chapter 2 introduced the concept of interference and the basic signal model. Then, it defines the spiky signal, including the cardinality-constrained variance (CCV), which is to characterize the non-Gaussian interference signal.

Chapter 3 showed the derivation of heavy-tailed prior of the interference characterized by Chapter 2, and then demonstrates why traditional Gaussian assumptions fail in this case. At the end, we presented the formulation of the robust covariance estimation optimization problem.

Chapter 4 introduces the relaxation of the combinatorial CCV constraint, and built the sequential approximation of the original objective function.

Chapter 5 introduces the development of the primal-dual update scheme to solve the relaxed subproblem obtained in Chapter 4.

Chapter 6 evaluates the proposed algorithm through convergence behavior and beamforming performances.

□ End of chapter.

Chapter 2

Fundamentals of Interference Suppression

This chapter establishes the foundational concepts necessary for understanding the interference suppression problem addressed in this work. We begin by introducing the beamforming problem within a relevant application scenario, highlighting the fundamental motivation for interference suppression and covariance estimation. Subsequently, we will introduce the modeling of interference, which provides insights into the structure of the interference-plus-noise covariance matrix. This structure is helpful and informative, as it will be employed as a structural constraint in the maximum likelihood estimation (MLE) method described in subsequent chapters. Additionally, we will review several approaches to interference covariance estimation, with a detailed discussion of our approach in Chapter 3.

2.1 Beamforming Problem

In various wireless communication and array signal processing applications, extracting signals of interest from complex and noisy environments is crucial. A primary

factor degrading signal quality in such scenarios is interference. Sensor array overcome this through spatial filtering, leveraging angular diversity even for spectrally close sources.

Sensor arrays are ubiquitous in daily life. Consider a speech enhancement system designed to isolate a desired speaker's voice in a crowded room. The received signal at an array of microphones will be a mixture of the target speaker's voice, background noise, and interfering speech from non-target person. Similarly, in radar and sonar applications, weak echoes from targets are often masked by strong clutter returns (e.g., ground clutter in radar, sea clutter in sonar). Additionally, systems may encounter intentional jamming signals or unintentional interference from other sources. Moreover, in wireless communication networks, co-channel interference poses a significant challenge to achieving high spectral efficiency and reliable communication links.

Many applications require the systems' capabilities on focusing their sensing or transmission on specific spatial directions ,while minimizing signals from undesired directions. This directional focus is essential for enhancing system performance.

To address these challenges, beamforming techniques are developed. Beamforming algorithms filter received signals from an array of sensors, constructively combining signals from a desired direction while suppressing those from other locations. A key metric for evaluating beamforming's effectiveness in interference suppression is the Signal-to-Interference-plus-Noise Ratio (SINR), defined as:

$$\text{SINR} = \frac{\text{Power of Desired Signal}}{\text{Power of Interference} + \text{Power of Noise}}$$

Mathematically, the beamformer's SINR can be expressed as:

$$\text{SINR}(\mathbf{w}) = \frac{\sigma_s^2 |\mathbf{w}^{\text{H}} \boldsymbol{\phi}(\theta_s)|^2}{\mathbf{w}^{\text{H}} \mathbf{R} \mathbf{w}}$$

where M is the number of array elements, $\mathbf{w} \in \mathbb{C}^M$ is the beamforming weight vector, $\boldsymbol{\phi}(\theta_s) \in \mathbb{C}^M$ is the steering vector toward the desired direction θ_s, σ_s^2 is the power of the desired signal and $\mathbf{R}$ is the interference-plus-noise covariance matrix.

There are three classical beamforming frameworks: the delay-and-sum method aligns signal phases through time delays [Van Trees, 2002]; the minimum variance distortionless response (MVDR) technique calculates phase-shift weights using the covariance matrix inverse to suppress unwanted components [Capon, 1969]; and subspace methods decompose the covariance matrix's eigenvalues to isolate signal components from interference [Li et al., 2003].

Minimum Variance Distortionless Response (MVDR) beamformer [Capon, 1969] is a well-known beamforming solution derived from SINR optimization. The MVDR problem is formulated as:

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \text{SINR}(\mathbf{w}) \\ \text{subject to} \quad & \mathbf{w}^{\text{H}} \boldsymbol{\phi}(\theta_s) = 1 \end{aligned}$$

The solution to this optimization problem is:

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{R}^{-1}\boldsymbol{\phi}(\theta_s)}{\boldsymbol{\phi}^{\text{H}}(\theta_s)\mathbf{R}^{-1}\boldsymbol{\phi}(\theta_s)}$$

The MVDR solution finds the weight vector $\mathbf{w}_{\text{MVDR}}$ that maintains listening capability from θ_s, mathematically translating to maximizing this SINR ratio.

A critical quantity for implementing the MVDR beamformer is the interference-plus-noise covariance matrix, $\mathbf{R}$. In practice, this matrix is often unknown and must be estimated from received data. Therefore, the accuracy of covariance estimation significantly impacts the MVDR beamformer's performance. Subsequent sections will explore methods for estimating $\mathbf{R}$ and address challenges in complex operational en-

vironments.

2.2 Interference Signal Modeling with Statistical Constraints

Before diving into the structure of interference-plus-noise covariance matrix, $\mathbf{R}$, we first introduce the concept of interference and noise.

Thermal Noise is the primary unstructured noise component. It originates from the random motion of charge carriers in conductive materials. Think of the background hiss you hear from a radio even when no station is tuned in. It sets a fundamental noise in received signals.

In modeling, thermal noise $\mathbf{n} \in \mathbb{C}^M$ is typically assumed as additive white Gaussian noise (AWGN), specifically a multivariate zero-mean isotropic complex Gaussian distribution: $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$, where its power is denoted by σ_n^2.

Interferences Sources in signal processing, refer to any unwanted signals that disrupt desired signals, thereby degrading system performance. Such interference originates from diverse sources including environmental reflections, intentional jamming, and unintended cross-user emissions. The distinct physical origins of interference sources necessitate their separate consideration from thermal noise in system modeling.

The focus of this work is on echo-like interference, characterized by its energy dominance over target signals. Such interference primarily occurs when the system receives strong reflections of the transmitted signal. In radar systems, for instance, the transmitted signal can be scattered or reflected by terrain features, artificial structures, or meteorological phenomena (e.g., rain and clouds), resulting in a strong power return that substantially exceeds target returns. Similarly, in sonar applications, sea

surface scattering can cause significant interference, thereby degrading detection performance in **the signal-to-interference-plus-noise ratio (SINR)**.

A notable characteristic of interference is its **spatial energy distribution**. Unlike thermal noise, which can be considered as uniform across all directions, interference often exhibits a non-uniform distribution in space. This means that the energy of interference is concentrated in specific directions while being minimal in others. As a result, systems that rely on uniform interference assumptions may struggle to perform effectively.

Similarly, in wireless communication systems, co-channel interference from other users can dominate certain spatial directions, such as the locations of interfering transmitters relative to the receiver. If multiple users are sharing the same communication channel, their signals may interfere with each other, with the strength of the interference varying depending on their spatial arrangement.

The aforementioned application scenarios share a common observation: the spatial energy distribution of interference often exhibits a heavy-tailed profile, where a small subset of directions dominates most interference power [Nikias and Shao, 1995; Richards et al., 2005; Van Trees, 2002]. This characteristics introduces challenges for covariance estimation, including ill-conditioning of the interference-plus-noise matrix $\mathbf{R}$ and the consequent numerical instability. These issues are particularly acute in scenarios with limited snapshots, as in our case.

Building upon the conceptual framework of interference mechanisms established earlier, this subsection develops a parametric interference model incorporating directional signal components. We first define the steering vector - a mathematical construct characterizing antenna array responses to signal wavefronts impinging from specific incidence angles. Subsequently, we introduce the steering dictionary $\mathbf{\Phi}$, a predefined knowledge structure derived from array system parameters, such as platform location, element spacing and operating frequency, which enables representa-

tion of interference signals as weighted linear combinations of steering vectors. Finally, through analytical derivation, we establish the mathematical structure of the interference-plus-noise covariance matrix **R**.

2.2.1 Steering Vector: The Angular Encoder

Consider a Uniform Linear Array (ULA) with M identical elements evenly spaced by d. When a narrowband signal arrives at the array from incidence angle θ, which is the angle between signal propagation direction and the array surface. The phase difference between adjacent elements due to wavefront propagation is $2\pi d \sin\theta/\lambda$, where λ is the wavelength (for carrier frequency f, $\lambda = c/f$, c is wave speed). The steering vector $\boldsymbol{\phi}(\theta) \in \mathbb{C}^M$ captures these phase shifts:

$$\boldsymbol{\phi}(\theta) = \begin{bmatrix} 1 \\ e^{-j2\pi d \sin\theta/\lambda} \\ \vdots \\ e^{-j2\pi(M-1)d \sin\theta/\lambda} \end{bmatrix}.$$

Each $\boldsymbol{\phi}(\theta)$ uniquely encodes wavefronts from incident angle θ, establishing the steering vector as a bijective mapping from angular space to signal subspace.

Overall, the resulting phase shift is fully characterized by two fundamental factors: the incident angle θ, and the array's physical parameters, specifically, the array geometric configuration and operating frequency.

2.2.2 Dictionary Matrix

Considering the angular range, denoted as Θ, of potential interference sources, we can map the angular space to the signal subspace through a steering matrix $\boldsymbol{\Phi}$. By discretizing the potential angular region Θ into grid points $\{\theta_1, \cdots, \theta_i \cdots, \theta_{N_c}\}$, it

allows constructing a steering vector dictionary matrix:

$$\mathbf{\Phi}(\Theta) = [\boldsymbol{\phi}(\theta_1), ..., \boldsymbol{\phi}(\theta_{N_c})] \in \mathbb{C}^{M \times N_c}.$$

The completeness of the dictionary matrix is determined by the relationship between grid density N_c and array size M: for smaller arrays where $M \ll N_c$, $\mathbf{\Phi}$ becomes overcomplete, while in dense arrays designed for high-resolution applications with $M \gg N_c$, the matrix transitions to an undercomplete regime. Different from conventional orthogonal basis decompositions, the steering vectors $\{\boldsymbol{\phi}_i\}$ are inherently non-orthogonal and exhibit mutual coherence.

By discretizing the angular domain through uniformly slicing in the $\sin\theta$ space into N_c grid points, we establish a discrete angular representation system. The grid number N_c determines the angular resolution of the specific array, which refers to the minimum angular separation where two sources remain distinguishable through their steering vectors.

For notational convenience, the steering matrix $\mathbf{\Phi}(\Theta)$ will be $\mathbf{\Phi}$ in subsequent chapters and sections.

2.2.3 Cardinality-Constrained Variance (CCV)

In array processing, strong reflections typically exhibit these physical characteristics: spatial sparsity of dominant sources, energy concentration in certain directions, and amplitude extremes. These three properties originates from the fact that only a few objects (e.g., building, vehicles) generate strong reflections and the power of dominant reflections usually overwhelms the insignificant reflections and thermal noise. Also, given the distinct material properties of each source, the reflection capabilities vary among objects. Magnified by the multipath during propagation, the amplitude coefficient usually exhibit a heavy-tailed behavior.

The meaning of strong interference inherently relates to power distribution. This motivates our focus on variance constraints rather than direct signal constraints, as variance presents the average power distribution over a coherent processing interval. Therefore, we define:

Definition 1: Cardinality-Constrained Variance

An element-wise independent random vector $\mathbf{x} \in \mathbb{C}^{n\times 1}$ has ***cardinality-constrained variance*** **(CCV)** if its variance vector $\mathbf{v} = \left[v_1, \cdots, v_n\right]^\top$ (where $v_i = \mathrm{Var}(x_i)$) satisfies:

$$\sum_{i=1}^{n} \mathbb{I}(v_i > \zeta) \leq s,$$

where $\mathbb{I}(\cdot)$ denotes the *indicator function* that returns 1 if the argument condition holds (i.e., $v_i > \zeta$) and 0 otherwise. Here:

- $\frac{1}{2}V_{\text{max}} \leq \zeta \leq V_{\text{max}}$ is a predefined threshold distinguishing significant variances from negligible ones, where V_{max} is the maximum of the variance, which is physically limited by the maximum power the receiver can receive.
- $0 < s \ll n$ specifies the maximum number of components allowed to exhibit substantial variance.

The cardinality-constrained variance (CCV) selects s significant reflective sources among N_{c} directions with threshold ζ, aiming to describe the sparsity of reflection sources. The CCV property will be used to model the sparsity in the upcoming problem formulation. The power dominance is to address that the majority of the power is contributed by significant sources. The power-law decay is describing the statistical property, aiming to model the spiky nature of observations. In the next chapter, we will demonstrate how the spikiness of coefficient results in outliers in observations, especially when number of snapshots is limited. In interference mitigation, CCV enforces sparsity **in the variance domain rather than the amplitude domain**.

Comparison to ℓ_0 and ℓ_1 Norm Constraints

While sparse or heavy-tail signal modeling typically employs ℓ_0 or ℓ_1 constraints on instantaneous amplitudes [Candes et al., 2008], our cardinality-constrained variance (CCV) approach differs by regularizing the second-order statistics. The key distinctions are:

- **Physical interpretability**: The CCV approach preserves physical consistency by regulating variance statistics $\mathbb{E}[|x_i|^2]$, which corresponds the average signal power over time, rather than instantaneous amplitudes. This accommodates interference scenarios where sources exhibit stochastic instantaneous amplitudes while maintaining stable average radiation power. In contrast, ℓ_0 and ℓ_1 norm constraints impose artificial sparsity on instantaneous snapshots, which is too restrictive.

- **Alignment with beamforming objective**: Beamforming performance fundamentally depends on second-order statistics through the covariance matrix $\mathbf{R}$. CCV's directly constraint on variance $\{v_i\}$ reduces error from estimating $\mathbf{x} \in \mathbb{C}^{N_\mathrm{c}}$ prior to covariance calculation.

- **Transformed complex-valued to real-valued**: Constraining on variance converts the original complex-valued estimation into real-valued optimization problem, therefore reducing the number of parameters to be estimated from $2N_\mathrm{c}$ to N_c ($\mathbf{x} \in \mathbb{C}^{N_\mathrm{c}} \rightarrow \mathbf{v} \in \mathbb{R}^{N_\mathrm{c}}$).

While the cardinality constraint generally induces non-convex optimization, it will be later shown that our interference modeling enables closed-form solutions. This special property transcends the general NP-hard nature of cardinality constraints.

Table 2.1: Comparison of sparsity constraints in statistics

Property	ℓ_0	ℓ_1	CCV
Mathematical Form	$\|\mathbf{x}\|_0 \leq s$	$\|\mathbf{x}\|_1 \leq \varepsilon$	$\sum \mathbb{I}(v_i > \zeta) \leq s$
Statistical Order	First	First	Second
Constraint Target	Instantaneous amplitudes	Instantaneous amplitudes	Average power

2.2.4 Spiky Coefficient Vector

Beyond directional characteristics, signal amplitude constitutes another critical parameter. The amplitude of interference signals originates from complex physical interactions. Taking reflective interference as an example, it depends on multiple factors including the object's reflectivity, propagation attenuation, and multipath effects. In practice, multiple interference sources usually coexist. Consider N_c interferers located at distinct angles $\{\theta_i\}_{i=1}^{N_\mathrm{c}}$.

To simplify the analysis, we model this physical process through a complex amplitude coefficient vector: $\mathbf{x} = \left[x_1, \cdots, x_{N_\mathrm{c}}\right]^\mathrm{T}$, where $\mathbf{x}$ is a random vector representing the signal strength from 1 to N_c directions. $\mathbf{x}$ is assumed to be zero-mean, with variance of each component being $v_i = \mathbb{E}[|x_i|^2]$, and it's also assumed to be component-wise independence $\mathbb{E}[x_i x_j] = 0, \forall i \neq j$, variance v_i is directly linked to the average power of the $i-$th component and therefore, is inherently bounded by the receiver limitations, $V_{\max} = \max_i v_i$.

To rigorously characterize the behavior of the coefficient $\mathbf{x}$, we define the **spiky coefficient vector** as follows:

Definition 2: Spiky Coefficient Vector

A random vector $\mathbf{x} \in \mathbb{C}^n$ is **spiky** if it satisfies:

1. **(Cardinality-Constrained Variance)** There exists a threshold $\zeta > 0$

and a $s \ll n$, such that $\mathcal{I}_s = \{i | v_i > \zeta\}$ satisfies:

$$|\mathcal{I}_s| \leq s. \tag{2.1}$$

2. **(Energy Dominance)** The significant components dominate over insignificant components and thermal noise:

$$\sum_{i \in \mathcal{I}_s} a v_i \geq \kappa \left(\sum_{j \notin \mathcal{I}_s} a v_j + \sigma_n^2 \right), \tag{2.2}$$

where $a = \|\boldsymbol{\phi}_i\|^2, \forall i, \boldsymbol{\Phi}(\boldsymbol{\Theta})$ and $\kappa \geq 1$ quantifies the dominance ratio.

3. **(Power-Law Tail Behavior)** Each significant component satisfies for $C > 0, \alpha > 0$:

$$P(|x_i| > t) \geq C_i t^{-\alpha}, \forall t \geq t_0. \tag{2.3}$$

where $t_0 > 0$ is a threshold for tail region.

2.2.5 Representing Interference

The composite interference signal $\mathbf{y} \in \mathbb{C}^M$ received by the M-element array is commonly described as a linear model [Ward, 1998; Guerci, 2014],:

$$\mathbf{y} = \sum_{i=1}^{N_c} x_i \boldsymbol{\phi}(\theta_i) + \mathbf{n} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n}.$$

where x_i is a complex coefficient of the i-th interferer, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$, an additive thermal noise.

2.3 Covariance Structure and Parametrization

Given interference components, we analyze the structure of the interference-plus-noise covariance matrix. Denote $\mathbf{y}_i \in \mathbb{C}^M$ as $\mathbf{y}_i = \sum_{i\in\mathcal{I}_s} v_i \boldsymbol{\phi}_i \boldsymbol{\phi}_i^{\mathsf{H}}$The composite signal $\mathbf{y} = \mathbf{y}_i + \mathbf{n}$ consists of dominant interference and thermal noise $\mathbf{n} \in \mathbb{C}^M$, with covariance matrix:

$$\mathbf{R} = \mathbb{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}] \tag{2.4}$$

where $\mathbb{E}[\cdot]$ denotes expectation and $(\cdot)^{\mathsf{H}}$ represents Hermitian transpose. Under the independence assumption between interference and noise components, the covariance matrix decomposes as:

$$\mathbf{R} = \mathbf{R}_{\mathrm{i}} + \sigma_n^2 \mathbf{I}_M \tag{2.5}$$

where $\mathbf{R}_{\mathrm{i}} \triangleq \mathbb{E}[\mathbf{y}_i \mathbf{y}_i^{\mathsf{H}}]$ is the interference covariance matrix and $\sigma_n^2 \mathbf{I}_M \in \mathbb{R}^{M\times M}$ represents white noise covariance with power σ_n^2.

Specifically, building on the interference model established in Section 2.2, we consider N_{c} independent interference sources:

$$\mathbf{R}_{\mathrm{i}} = \sum_{k=1}^{N_{\mathrm{c}}} \mathbb{E}[|x_i|^2] \boldsymbol{\phi}_i \boldsymbol{\phi}_i^{\mathsf{H}} = \sum_{k=1}^{N_{\mathrm{c}}} v_i \boldsymbol{\phi}_i \boldsymbol{\phi}_i^{\mathsf{H}} \tag{2.6}$$

where $v_i \triangleq \mathbb{E}[|x_i|^2]$ denotes the power of the i-th interference source, and $\boldsymbol{\phi}_i \triangleq \boldsymbol{\phi}(\theta_i) \in \mathbb{C}^M$ is the steering vector at direction θ_i. This leads to a matrix representation:

$$\mathbf{R} = \mathbf{\Phi}\mathbf{V}\mathbf{\Phi}^{\mathsf{H}} + \sigma_n^2\mathbf{I}_M \tag{2.7}$$

$$\mathbf{\Phi} \triangleq [\boldsymbol{\phi}_1 \ \cdots \ \boldsymbol{\phi}_{N_\mathrm{c}}] \in \mathbb{C}^{M\times N_\mathrm{c}} \tag{2.8}$$

$$\mathbf{V} \triangleq \mathrm{Diag}(v_1, \ldots, v_{N_\mathrm{c}}) \in \mathbb{R}^{N_\mathrm{c}\times N_\mathrm{c}} \tag{2.9}$$

$\mathbf{R}_\mathrm{i}$ consists of N_c rank-1 components of the form $v_i\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathsf{H}}$. Consequently, the rank of $\mathbf{R}_\mathrm{i}$ is generally less than N_c when $N_\mathrm{c} < M$ since the vectors $\{\boldsymbol{\phi}_i\}$ are linearly dependent. Therefore, $\mathbf{R}_\mathrm{i}$ is a low-rank matrix. Finally, the full-rank noise covariance $\sigma_n^2\mathbf{I}_M$ ensures that the matrix $\mathbf{R}$ is positive definite.

This structural decomposition reveals that the covariance matrix $\mathbf{R}$, which nominally contains $O(M^2)$ unknown parameters, can be parameterized by only $N_\mathrm{c}+1$ key parameters: N_c interference powers and $\{\sigma_k^2\}_{k=1}^{N_\mathrm{c}}$ noise power σ_n^2.

To summarize, the beamforming problem reduced to recovering the interference power profile, given the observed snapshots. While theoretically sound, practical implementations face conditioning challenges. The matrix $\mathbf{R}$ becomes ill-posed given the spikiness of interference power distribution. These conditions amplify estimation errors in covariance inversion, particularly affecting the MVDR beamformer:

$$\mathbf{w}_\mathrm{MVDR} \propto \mathbf{R}^{-1}\boldsymbol{\phi}(\theta_\mathrm{desired}), \tag{2.10}$$

where accurate $\mathbf{R}$ estimation is critical for preserving desired signals while nulling interference.

Existing Methods

Covariance matrix estimation methods can be broadly categorized into two paradigms: **model-free** approaches (without explicit model assumptions) and **model-dependent**

methods (rely on model-based or optimization frameworks). While this work focuses on robust covariance estimation for beamforming applications, we first provide a concise overview of general methodologies before delving into robustness-specific techniques.

Model-Free

Given a set of samples $\{\mathbf{y}_1, \cdots, \mathbf{y}_{N_s}\}$, the sample covariance matrix (SCM), defined as

$$\hat{\mathbf{R}}_{\text{SCM}} \triangleq \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbf{y}_k \mathbf{y}_k^{\text{H}}, \tag{2.11}$$

serves as the asymptotically unbiased maximum likelihood estimator under Gaussian assumptions. While theoretically consistent for $N_s \gg M$, its practical effectiveness degrades with limited snapshot. In such cases, the SCM becomes ill-conditioned, leading to numerical instability during matrix inversion—a critical issue for adaptive beamforming where weight vectors depend on $\hat{\mathbf{R}}_{\text{SCM}}^{-1}$.

To address these limitations, regularization strategies are often employed. Diagonal loading technique [Li et al., 2003; Vorobyov et al., 2003], which modifies the SCM as

$$\hat{\mathbf{R}}_{\text{DL}} = \hat{\mathbf{R}}_{\text{SCM}} + \beta \mathbf{I}, \tag{2.12}$$

where $\beta > 0$ is the loading factor, thresholds the minimum eigenvalue of $\hat{\mathbf{R}}_{\text{DL}}$, thereby stabilizing its inverse. Diagonal loading also provides robustness under uncertainty as shown in [Vorobyov et al., 2003]. The performance of diagonal loading is highly sensitive to the selection of β.

Another method is to exploit the low-rank nature of the interference covariance subspace [Stoica and Nehorai, 1989]. By constructing a modified estimator $\hat{\mathbf{R}_{\text{LR}}} = \hat{\mathbf{R}}_{\text{SCM}} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\text{H}}$, where $\mathbf{U}$ spans the dominant interference subspace, $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements being the power of corresponding basis. How-

ever, obtaining the subspace matrix is computationally expensive, given the matrix decomposition process. Therefore, it is not optimal for scenarios with limited sample size and requires real-time adaptation. There are also alternative regularization approaches, include shrinkage estimation [Ledoit and Wolf, 2004] and factor models [Fan et al., 2008].

Model-Dependent

Model-driven approaches achieve dimensionality reduction through physics-informed parametric modeling by exploiting scenario-specific covariance structures (e.g., spatial correlations, interference distributions). By representing the covariance matrix as an analytical function of these physical parameters (e.g., AoA, path loss coefficients), they transform high-dimensional matrix estimation into structured parameter identification. This physics framework maintains computational tractability while preserving interpretability through causal relationships between physical mechanisms and statistical properties, is considered useful when training samples are limited.

Furthermore, in terms of estimation itself, model-driven covariance estimation methods can be categorized by their objective functions into two frameworks, probabilistic and statistical optimization.

A probabilistic modeling framework constructs estimators based on statistical distribution assumptions. Employing a Gaussian prior is a frequent choice, primarily due to: 1) the prevalence of the central limit theorem in real-world phenomena, and 2) the mathematical tractability it offers. However, similar to model-free approaches, regularization is still necessary when dealing with limited sample sizes. One common approach is to pose regularization, when using Gaussian prior for the likelihood function [Banerjee et al., 2008; Friedman et al., 2008; Yuan and Lin, 2007].

While the optimal prior distribution is scenario-dependent, robustness is often a key consideration. Real-world signals frequently exhibit characteristics like impul-

sive noise, outliers, or sudden unexpected events. These imperfections lead to heavy tails in the signal distribution, meaning that extreme values are more frequent than predicted by a Gaussian model. Therefore, heavy-tailed PDFs are commonly used to model them. The Laplace distribution [Cui et al., 2022; Xie et al., 2023] is one such example; its sharp peak and heavy tails make it less susceptible to being skewed by outliers, leading to more stable estimates. Gaussian mixtures [Wipf and Rao, 2004; Sadler et al., 1995; Lin et al., 2013] provide a more flexible alternative, capable of capturing complex noise distributions by combining several Gaussian distributions with varying means and variances, allowing for a better fit to the data. Elliptical distributions [Goldstein et al., 2018; Sun et al., 2016] offer a broader class of distributions, including Gaussian and Laplace, that share the property of having elliptical contours; this property facilitates the development of robust estimation algorithms that are less sensitive to deviations from normality.

Besides probabilistic framework, the other approach focuses on minimizing statistically defined estimation errors, such as the Mean Squared Error (MSE), while incorporating model information as constraints. The most famous approach within this framework is sparse regularization, exemplified by the Least Absolute Shrinkage and Selection Operator (LASSO) [Tibshirani, 1996; Friedman et al., 2008; Candes and Tao, 2006]. These models often impose ℓ_1-norm constraint on the variables being estimated. This constraint encourages solutions with many zero-valued components, thus retaining only the most significant features. Imposing an ℓ_1-norm constraint is equivalent to assuming a Laplace prior on the variables [Tibshirani, 1996].

While sparse regularization methods like LASSO promote parsimony via ℓ_1 norm constraint, they suffer from systematic shrinkage bias, distorting estimates of large coefficients. Their performance further deteriorates under correlated predictors, where they arbitrarily select variables from groups, harming interpretability. The equiva-

lence of ℓ_1-regularization to a Laplace prior assumes unrealistic independent, identically distributed coefficients, failing to adapt to structured or hierarchical data.

Heavy-tailed alternatives, though robust to outliers, face computational trade-offs. Laplace priors lack flexibility for modeling dependencies (e.g., grouped sparsity), while Gaussian mixtures (GMMs), though multi-modal, suffer from EM training instability and poor scalability. Elliptical distributions rely heavily on accurate covariance estimation, which falters with limited data or misspecified correlations.

□ End of chapter.

Chapter 3

Problem Formulation: Robust Maximum Likelihood

This chapter establishes the mathematical framework for the robust covariance estimation. First, we demonstrate how spiky coefficients induce heavy-tailed observation. Second, we present the challenge of outliers in finite sample scenarios. Finally, we propose a robust maximum likelihood formulation, denoted as problem $\mathcal{P}_1$ in (3.9). This formulation integrates a robust elliptical density function as a prior distribution, along with cardinality-constrained variance (CCV) and other constraints.

3.1 Robust Elliptical Prior as Objective Function

The Need for Robustness

Conventional signal processing methods relying on Gaussian assumptions reveal limitations in real-world applications. While mathematically tractable, particularly enabling closed-form covariance matrix estimation through maximum likelihood methods, its effectiveness breaks down in practical interference environments for two key reasons: model mismatch and sensitivity to outliers.

First, the fundamental model mismatch. As demonstrated in section 2.2, real-world interference exhibits non-Gaussian properties. For instance, heavy-tailed clutter in radar systems caused by terrain reflections and sudden noise bursts in wireless communications. These phenomena violate the Gaussian prior assumption, leading to estimation errors of model mismatch.

Second, sensitivity to outliers and snapshot limitations. Gaussian models intrinsically underestimate the probability of extreme events due to its rapid tail decay. This makes Gaussian-based estimators numerically unstable, especially with limited number of samples; and vulnerable to outlier corruption. As shown in [Bishop and Nasrabadi, 2006; Murphy, 2012], these limitations translates to reduced robustness, such as increased false alarm in target detection task.

Therefore, we require a density function that simultaneously satisfies two essential criteria: **maintaining mathematical tractability** to facilitate maximum likelihood estimation, while being **capable of describing the non-uniform power distributions** to capture the inherent spatial non-uniformity of interference fields.

3.1.1 Heavy-Tail Behavior Induced by Spiky Coefficients

As mentioned in section 2.2, spiky coefficient vector $\mathbf{x}$ induces heavy-tailed behavior in the received signal $\mathbf{y}$. To formalize this relationship, we first characterize heavy-tailed distribution through tail probability decay:

We define the heavy-tailness of a random vector according to its norms probability tail behavior:

Definition 3: Heavy-Tailed Random Vector

A random vector $\mathbf{v} \in \mathbb{C}^M$ is **heavy-tailed** if there exists constants $C > 0, \alpha >$

0 and a threshold $t_0 > 0$, such that:

$$P(\|\mathbf{v}\| > t) \geq Ct^{-\alpha}, \quad \forall t \geq t_0. \tag{3.1}$$

where $t_0 > 0$ is the threshold for tail region.

Recall the definition 2 of spiky coefficient vector:

1. **Cardinality-Constrained Variance** At most s components have variances exceeding a significance threshold $\zeta > 0$: $\sum_{i=1}^{N_c} \mathbb{I}(v_i > \zeta) \leq s$.

2. **Energy Dominance** The energy of significant components in the significance index set $\mathcal{I}_s = \{i | v_i > \zeta\}$ dominates residual and noise: $\sum_{i \in \mathcal{I}_s} av_i \geq \kappa \left(\sum_{j \notin \mathcal{I}_s} av_j + \sigma_n^2 \right)$, where $a = \|\boldsymbol{\phi}_i\|^2, \forall i$ is the norm of a column in steering matrix $\boldsymbol{\Phi}$, $\kappa \geq 1$ is the dominance ratio, σ_n^2 is the noise power.

3. **Power-Law Tail Behavior** Each significant component $x_i (i \in \mathcal{I}_s)$ exhibits heavy-tailed behavior: $P(|x_i| > t) \geq C_i t^{-\alpha_i}, \forall t \geq t_0 \quad \exists C_i > 0, \alpha_i > 0$. $\mathcal{I}_s$ is the index set of significant components.

Since $\mathbf{x}$ is spiky and each component of it is i.i.d., the heavy-tailed property of its significant component x_i ($i \in \mathcal{I}_s$) implies that

Theorem 1: Sufficient Condition of Heavy-Tailed Observation

If $\mathbf{x} \in \mathbb{C}^{N_c}$ satisfies Definition 2, with $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N_c}$, $\|\boldsymbol{\phi}_i\|^2 = a, \forall i$, $\mathbf{n} \in \mathbb{C}^M \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$, there exists $C > 0, \alpha_y > 0$ such that the observation $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n}$ is heavy-tailed with tail probability of:

$$P(\|\mathbf{y}\| > t) \geq Ct^{-\alpha_y}. \tag{3.2}$$

Proof. Decompose the signal $\mathbf{y}$ into significant and residual components:

$$\mathbf{y} = \underbrace{\sum_{i \in \mathcal{I}_s} \boldsymbol{\phi}_i x_i}_{\mathbf{y}_{\text{sig}}} + \underbrace{\sum_{j \notin \mathcal{I}_s} \boldsymbol{\phi}_j x_j + \mathbf{n}}_{\mathbf{y}_{\text{res}}}.$$

By the reverse triangular inequality:

$$\|\mathbf{y}\| \geq \|\mathbf{y}_{\text{sig}}\| - \|\mathbf{y}_{\text{res}}\|.$$

Therefore, $\forall t > 0$,

$$P(\|\mathbf{y}\| > t) \geq P(\|\mathbf{y}_{\text{sig}}\| > t + \|\mathbf{y}_{\text{res}}\|)$$

Residual tail control From the energy dominance assumption,

$$a \sum_{i \in \mathcal{I}_s} v_i \geq \kappa \left(a \sum_{j \notin \mathcal{I}_s} v_j + \sigma_n^2 \right)$$

Calculate the residual energy bound:

$$\mathbb{E}[\|\mathbf{y}_{\text{res}}\|^2] = a \sum_{j \notin \mathcal{I}_s} v_j + \sigma_n^2 \leq \frac{a}{\kappa} \sum_{i \in \mathcal{I}_s} v_i$$

Apply Markov's inequality:

$$P(\|\mathbf{y}_{\text{res}}\| > t) \leq \frac{\mathbb{E}[\|\mathbf{y}_{\text{res}}\|^2]}{t^2} \leq \frac{a \sum_{i \in \mathcal{I}_s} v_i}{\kappa t^2} \leq \frac{C_r}{t^2}$$

where $C_r = \frac{asV_{\max}}{\kappa}$, $V_{\max}$ is the maximum of v_i.

Significant tail behavior By Energy Dominance and $|\mathcal{I}_s| \leq s$:

$$asV_{\max} \geq a \sum_{i \in \mathcal{I}_s} v_i \geq \kappa \left(a \sum_{j \notin \mathcal{I}_s} v_j + \sigma_n^2 \right) \geq \kappa \sigma_n^2$$

Hence:

$$V_{\max} \geq \frac{\kappa \sigma_n^2}{as}$$

Let the most significant component index as $i_0 = \arg\max_{i \in \mathcal{I}_s} v_i$, and it has a tail probability:

$$P\left(|x_{i_0}| > t/\sqrt{a}\right) \geq C_{i_0}(t/\sqrt{a})^{-\alpha_0}$$

Take $\eta > 0$, combining together:

$$\begin{aligned} P(\|\mathbf{y}\| > t) &\geq P(\|\mathbf{y}_{\text{sig}}\| > t + \|\mathbf{y}_{\text{res}}\|) \\ &\geq P\left(\|\mathbf{y}_{\text{sig}}\| > t(1+\eta)\right) \cdot P\left(\|\mathbf{y}_{\text{res}}\| \leq \eta t\right)) \\ &\geq P\left(|x_{i_0}| > \frac{t(1+\eta)}{\sqrt{a}}\right) \cdot P\left(\|\mathbf{y}_{\text{res}}\| \leq \eta t\right)) \\ &\geq C_{i_0} \cdot a^{\alpha_0/2} \cdot [t(1+\eta)]^{-\alpha_0} \cdot (1 - \frac{C_r}{\eta^2 t^2}) \end{aligned}$$

Take $t \geq \sqrt{\frac{2C_r}{\eta^2}}$, therefore, $(1 - \frac{C_r}{\eta^2 t^2}) \geq \frac{1}{2}$ So we have

$$P(\|\mathbf{y}\| > t) \geq Ct^{-\alpha}$$

where $C = \frac{C_{i_0} \cdot a^{\alpha_0/2}(1+\eta)^{-\alpha_0}}{2}$, $\alpha = \alpha_0$. □

3.1.2 Heavy-Tailed Behavior Under Limited Samples

Heavy-tailed distributions differ fundamentally from Gaussians in finite-sample regimes. Let $\{\mathbf{y}_k\}_{k=1}^{N_s}$ be heavy-tailed observations satisfying $P(\|\mathbf{y}\| > t) \geq Ct^{-\alpha}$, and $\{\mathbf{z}_k\}_{k=1}^{N_s}$ be Gaussian with variance σ^2. Define the extreme value as $Y_{\max} = \max_k \|\mathbf{y}_k\|$ and $Z_{\max} = \max_k \|\mathbf{z}_k\|$.

Extreme values scale differently: for heavy-tailed data, the maximum $\mathbb{E}[Y_{\max}] \leq K_1 N_s^{1/\alpha}$ grows polynomially with sample size N_s, while Gaussian extremes $\mathbb{E}[Z_{\max}] \approx \sqrt{2\sigma^2 \log N_s}$ increase only logarithmically [Coles et al., 2001]. This gap explains why

wireless channels ($\alpha \in (2, 4)$) show stronger outlier amplification than Gaussian models predict.

Outlier persistence follows power-law dynamics. Define threshold violations using indicators:

$$\mathbf{1}_k(t) := \mathbb{I}(\|\mathbf{y}_k\| > t), \quad \hat{P}_{N_s}(t) = \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbf{1}_k(t)$$

The expected outlier count $\mathbb{E}[N_{\text{out}}] = N_s \hat{P}_{N_s}(t)$ diverges sharply:

- Heavy-tailed: $\mathbb{E}[N_{\text{out}}] \geq C N_s t^{-\alpha}$ (polynomial decay)
- Gaussian: $\mathbb{E}[N_{\text{out}}] \leq N_s e^{-t^2/(2\sigma^2)}$ (exponential decay)

Therefore, in a finite-sample regime, we can conclude that the heavy-tailed data behaves differently from Gaussian data. Also, the more frequent appearance of outlier in data is a consequence of the heavier tail, and dealing with outlier will be important, this also explained why Gaussian fails.

3.1.3 The Robust Elliptical Prior

In this work, we chose to model the interference with the robust elliptical density, which is a heavy-tailed density function of the following form:

$$f(\mathbf{y}) \propto \det(\mathbf{R})^{-1} \left(\mathbf{y}^{\text{H}} \mathbf{R}^{-1} \mathbf{y} \right)^{-M}, \tag{3.3}$$

where M is the number of array elements, and also signal length.

This robust elliptical density belongs to the family of multivariate Cauchy-type distributions, characterized by its heavy-tails. This property enhances robustness against outliers and non-Gaussian interference, which are common in practical radar or sonar environments.

This prior is essentially a **Cauchy-like prior,** characterized by its tail decay $\sim (\|\mathbf{y}\|_2)^{-M}$. The robust elliptical density inherently suppresses large deviations from the nominal covariance structure without requiring explicit sparsity constraints.

To model heavy-tailed interference, there are substantial existing studies. Among them, one easy to think about is the multivariate Laplace prior enforces sparsity through an ℓ_1 -norm penalty [Bishop and Nasrabadi, 2006]:

$$f_{\text{Laplace}}(\mathbf{y}) \propto \exp\left(-\lambda\sqrt{\mathbf{y}^{\text{H}}\mathbf{R}^{-1}\mathbf{y}}\right), \tag{3.4}$$

where $\lambda > 0$ is a regularization parameter, which is decided by human, based on experience and belief. While effective for sparse signal recovery, this formulation suffers from sensitivity to the choice of λ, which lacks a direct physical interpretation and requires case-specific calibration. Alternatively, Gaussian mixture models (GMMs) attempt to capture complex interference statistics through a weighted superposition of multivariate normal components:

$$f_{\text{GMM}}(\mathbf{y}) = \sum_{p=1}^{P} \alpha_p \cdot \mathcal{CN}(\mathbf{y}; \mathbf{0}, \mathbf{R}_p), \tag{3.5}$$

where α_p represent mixing coefficients ($\sum_p \alpha_p = 1$) and $\mathcal{CN}(\cdot)$ denotes the complex normal distribution. Despite their flexibility, GMMs face practical challenges due to overparameterization (requiring estimation of P covariance matrices $\mathbf{R}_p$) and ill-conditioned covariance estimates under limited training samples, particularly in high-dimensional systems.

3.1.4 The Robust Maximum Likelihood Objective

Given the use of elliptical distributions for interference modeling, we derive the corresponding maximum likelihood estimation (MLE) problem for the covariance matrix.

For received samples $\{\mathbf{y}_1, \cdots, \mathbf{y}_{N_s}\}$, with N_s snapshots following the elliptical distribution in (3.3), the log-likelihood function (omitting constant terms) takes the form:

$$-\frac{1}{2}\left[\log|\det(\mathbf{R})| + \frac{M}{N_s}\sum_{k=1}^{N_s}\log(\mathbf{y}_k^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{y}_k).\right] \tag{3.6}$$

The MLE problem therefore reduces to minimizing the negative log-likelihood, yielding the optimization objective:

$$\mathcal{L}(\mathbf{R}) = \frac{1}{2}\left[\log|\det(\mathbf{R})| + \frac{M}{N_s}\sum_{k=1}^{N_s}\log(\mathbf{y}_k^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{y}_k).\right] \tag{3.7}$$

where $\mathbf{R}$ denotes the covariance matrix to be estimated.

3.2 Constraints Formulation

3.2.1 Cardinality-Constraint Variance

Among the three conditions defining spiky coefficient (as specified in Definition 2, the cardinality-constrained variance (CCV) is adopted as the primary criterion to characterize spikiness.

The CCV is formally expressed as:

$$\sum_{i=1}^{N_c}\mathbb{I}(v_i > \zeta) \leq s,$$

where $\zeta > 0$ represents the significance threshold, and s is an integer (between 0 and N_c) which controls the sparsity level of variances.

The cardinality constraint is widely used in sparse optimization, feature selection [Tropp, 2004; Abdi, 2013], or resource allocation problems [Biswas and Barman, 2018], with the core objective of restricting significant deviations from a predefined

threshold to a small subset of variables. For instance, in financial portfolio optimization, cardinality constraint is frequently employed to limit the number of actively held stocks under resource constraints [Rubinstein, 2002; Anagnostopoulos and Mamanis, 2011; Gao and Li, 2013].

The key methodologies for addressing cardinality constraint includes mathematical programming formulations, including convex relaxation, where the non-convex indicator function $\mathbb{I}(\cdot)$ is replaced with convex constraints such as ℓ_1 norm approximation [Tropp, 2004]; Mixed-integer programming (MIP), which introduces binary variables to explicitly model the indicator function, thereby reformulating the problem into tractable forms [Bertsimas and Shioda, 2009; Cui et al., 2013; Abdi, 2013]; Practical implementations often integrate projection-based strategies to maintain feasibility, where solutions are iteratively projected onto the constraint set through thresholding operations Bian and Chen [2020]. Additionally, there are also heuristic methods such as genetic algorithm and tabu search [Chang et al., 2000; Woodside-Oriakhi et al., 2011].

Table 3.1: Comparison of sparsity constraints in optimization aspect

Property	ℓ_0	ℓ_1	CCV
Mathematical Form	$\|\mathbf{x}\|_0 \leq s$	$\|\mathbf{x}\|_1 \leq \varepsilon$	$\sum \mathbb{I}(v_i > \zeta) \leq s$
Convexity	Non-convex	Convex	Non-convex

3.2.2 Model-Induced Constraint

The physical interference model in Section 2.2.5 establishes the fundamental signal constraint:

$$\mathbf{y} = \sum_{i=1}^{N_c} x_i \boldsymbol{\phi}_i + \mathbf{n} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n}, \tag{3.8}$$

where $\boldsymbol{\phi}_i \in \mathbb{C}^{N_c}$ denotes interference steering vector from i- source and $\{x_i\}$ are zero-mean complex amplitude coefficients with variance $\mathbb{E}[|x_i|^2] = v_i$.

This give a structural constraint:

$$\mathbf{R} = \boldsymbol{\Phi} \underbrace{\text{diag}(v_1, ..., v_{N_c})}_{\mathbf{V}} \boldsymbol{\Phi}^{\text{H}} + \sigma_n^2 \mathbf{I},$$

3.3 Robust Covariance Estimation with CCV

Problem 1: Original problem $\mathcal{P}_1$

The robust covariance estimation with CCV $\mathcal{P}_1$ is formulated as:

$$\min_{\mathbf{R},\mathbf{v}} \quad \mathcal{L}(\mathbf{R}) \tag{3.9}$$

$$\text{s.t.} \quad \mathbf{R} = \boldsymbol{\Phi}\mathbf{V}\boldsymbol{\Phi}^{\text{H}} + \sigma_n^2 \mathbf{I}, \tag{3.10}$$

$$\mathbf{V} = \text{diag}(\mathbf{v}),\ \mathbf{v} = [v_1, \cdots, v_{N_c}]^{\text{T}}, v_i > 0, \forall i, \tag{3.11}$$

$$\sum_{i=1}^{N_c} \mathbb{I}(v_i > \zeta) \leq s. \tag{3.12}$$

where $\frac{1}{2}V_{\max} < \zeta < V_{\max}$ denotes the significant threshold, $s \ll N_c$.

3.3.1 Optimization Challenges

The formulated problem in 3.9 presents fundamental challenges that intertwine physical modeling with computational complexity.

Nonconvex Objective Function The log-likelihood objective $\mathcal{L}(\mathbf{R})$ is convex with respect to the covariance matrix $\mathbf{R}$, but is non-convex with respect to the variance $\{v_i\}$, which is the variable to be solved.

Attempting direct optimization with first-order condition leads to a circular de-

pendency captured by the fixed-point equation:

$$\mathbf{R} = \frac{M}{N_{\mathrm{s}}} \sum_{k=1}^{N_{\mathrm{s}}} \frac{\mathbf{y}_k \mathbf{y}_k^{\mathsf{H}}}{\mathbf{y}_k^{\mathsf{H}} \mathbf{R}^{-1} \mathbf{y}_k}, \tag{3.13}$$

which is known as the Tyler's estimator [Tyler, 1987].

It is hard to directly derive $\mathbf{R}$ through this fixed-point relationship, especially with insufficient snapshots $N_{\mathrm{s}} \ll M$.

Combinatorial Cardinality Constraints The cardinality constraint $\sum \mathbb{I}(v_i > \zeta) \leq s$ is a combinatorial constraint that is non-convex, discontinuous, and NP-hard.

□ End of chapter.

Chapter 4

Convex Relaxation

4.1 Convex Relaxation on Cardinality Variance Constraint

To address the discontinuity and non-convexity caused by the cardinality constraint $\sum \mathbb{I}(v_i > \zeta) \leq s$, we propose a continuous relaxation framework by introducing an auxiliary binary variables to model the cardinality constraint. The constraint is subsequently convexified and transformed according to the specific structure of the robust covariance estimation.

4.1.1 Convexify CCV with Box Variable

Given $V_{\max} = \max_i \ v_i$, the the maximum receivable power at the receiver end, the significance threshold $\frac{1}{2}V_{\max} < \zeta < V_{\max}$, the feasible region of v_i can be partitioned into two intervals:

$$v_i \in \begin{cases} (0, \zeta], \\ (\zeta, V_{\max}]. \end{cases}$$

Initially, the status of v_i within this binary-valued interval is denoted by the indicator function $\mathbb{I}(\cdot)$. To achieve convex relaxation, it is common practice to introduce a Bernoulli variable that take value of 0 and 1. Instead of using a pure binary variable, we introduce a continuous box variable $y_i \in [0, 1]$. This serves as a relaxation of the binary constraint.

Therefore, the original cardinality constraint $\sum_{i=1}^{N_c} \mathbb{I}(v_i > \zeta) \leq s$ can be relaxed and rewritten in terms of y_i:

$$\sum_{i=1}^{N_c} y_i \leq s. \tag{4.1}$$

Regarding the binary selection, it can be formulated as:

$$v_i \leq V_{\max}^{y_i} \cdot \zeta^{1-y_i}. \tag{4.2}$$

This inequality encodes the following logical implication:

$$\begin{cases} y_i = 0, \Rightarrow 0 < v_i \leq \zeta, & \text{(insignificant)} \\ y_i = 1, \Rightarrow \zeta < v_i \leq V_{\max}, & \text{(significant components).} \end{cases}$$

4.1.2 Simplify: One-Side Box

$y_i \leq 1$ appears natural for binary emulation, but this constraint is redundant in this problem.

Assume there exists $y_i > 1$ in a feasible solution. Substituting into 4.2:

$$v_i \leq V_{\max}^{y_i} \cdot \zeta^{1-y_i} = V_{\max} \cdot \left(\frac{V_{\max}}{\zeta}\right)^{y_i - 1}. \tag{4.3}$$

While $\frac{1}{2}V_{\max} < \zeta < V_{\max}$, we have $1 < \frac{V_{\max}}{\zeta} < 2$. For $y_i > 1$, we have:

$$V_{\max} \leq V_{\max} \cdot \left(\frac{V_{\max}}{\zeta}\right)^{y_i - 1}. \tag{4.4}$$

Since $v_i \leq V_{\max}$ is an inherit physical restriction, so, the right-hand side would always be true. Therefore, the constraint $y_i \leq 1$ is proven redundant and can be removed.

4.1.3 Enhancement: Log-Domain Transformation

In previous section, we construct inequality (4.2), to bridge the original optimization variable v_i and the auxiliary variable y_i. To better solve the problem, we further transform the relaxed constraint into logarithmic domain, which yields:

$$\log v_i \leq a y_i + b, \tag{4.5}$$

where

$$a := \log\left(\frac{V_{\max}}{\zeta}\right) \in (0, \log 2),$$
$$b := \log \zeta \in (\log V_{\max} - \log 2,\ \log V_{\max}).$$

This is a beneficial transform for two reasons:

- In signal processing applications, power levels are often analyzed in decibel (dB), therefore, making the logarithmic scale a natural choice. While the linear relaxation $v_i \leq V_{\max} \cdot y_i + \zeta \cdot (1 - y_i)$ is also capable of describing the binary selection at point $y_i \in \{0, 1\}$, such formulation lacks synergy with logarithmic power constraint.

- Moreover, the logarithmic term $\log v_i$ is a more convenient form for solving the

problem compared to v_i, as we will see in the section of deriving subproblems.

Problem 2: Problem with Relaxed log-CCV: $\mathcal{P}_2$

$$
\begin{aligned}
\underset{\mathbf{R},\mathbf{v}}{\text{minimize}} \quad & \mathcal{L}(\mathbf{R}) \\
\text{subject to} \quad & \mathbf{R} = \mathbf{\Phi}\mathbf{V}\mathbf{\Phi}^{\text{H}} + \sigma_n^2\mathbf{I}, \\
& \mathbf{V} = \text{diag}(\mathbf{v}),\ \mathbf{v} = \left[v_1, \cdots, v_{N_c}\right]^{\text{T}}, \\
& \begin{cases} \log v_i \le a \cdot y_i + b, \\ \displaystyle\sum_{i=1}^{N_c} y_i \le s, \end{cases} \quad \text{(relaxed log-CCV constraints)} \\
& v_i > 0, y_i \ge 0, \forall i.
\end{aligned} \tag{4.6}
$$

where $a \in (0, \log 2), b \in (\log V_{\max} - \log 2, \log V_{\max})$.

4.2 Majorization-Minimization Relaxation on the Objective

After handling the CCV constraint, we focus on minimizing the objective via the MM framework.

4.2.1 The Majorization-Minimization Framework

Majorization-Minimization (MM) is a technique to minimize complex objectives (e.g., non-convex or non-smooth) by iteratively constructing **surrogate functions** that are both simpler and locally aligned with the original function.

The intuition is as follows: starting from an initial feasible point $\mathbf{x}_0$, we build a surrogate function at $\mathbf{x}_k$ that:

- Touches the original objective at $\mathbf{x}_k$ (shares the same function value and gradient),

- Lies strictly above the objective globally (acts as an upper bound).

By minimizing this surrogate, we obtain $\mathbf{x}_{k+1}$, and repeat until convergence to a stationary point.

If a surrogate sequence satisfies Assumption 1, can guaranteed stationary convergence, as shown in [Razaviyayn et al., 2013]. These requirements can be translated into conditions:

Assumption 1:

Given the original function $f(\mathbf{x})$, the surrogate $g(\mathbf{x}|\mathbf{x}_t)$ must satisfy to guarantee convergence to stationary points:

- **Function Matching**:

$$g(\mathbf{x}_t|\mathbf{x}_t) = f(\mathbf{x}_t) \tag{4.7}$$

- **Global Upper Bound**:

$$g(\mathbf{x}|\mathbf{x}_t) \geq f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X} \tag{4.8}$$

- **Gradient Consistency**:

$$\nabla_{\mathbf{x}} g(\mathbf{x}_t|\mathbf{x}_t) = \nabla_{\mathbf{x}} f(\mathbf{x}_t) \tag{4.9}$$

where $\mathcal{X}$ is the feasible region.

The process can be represented as Algorithm 1

Algorithm 1 Majorization-Minimization (MM) Algorithm

Input: Objective $f(\mathbf{x})$, initial point $\mathbf{x}_0$, tolerance ϵ
Output: Approximate minimizer $\mathbf{x}^*$
Initialize $t = 0$
repeat
    Majorization: Construct surrogate $g(\mathbf{x}|\mathbf{x}_t)$ satisfying Assumption 1
    Minimization: Update $\mathbf{x}_{t+1} \leftarrow \arg\min_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x}|\mathbf{x}_t)$
    $t \leftarrow t + 1$
until Convergence criteria met
return $\mathbf{x}^* = \mathbf{x}_t$

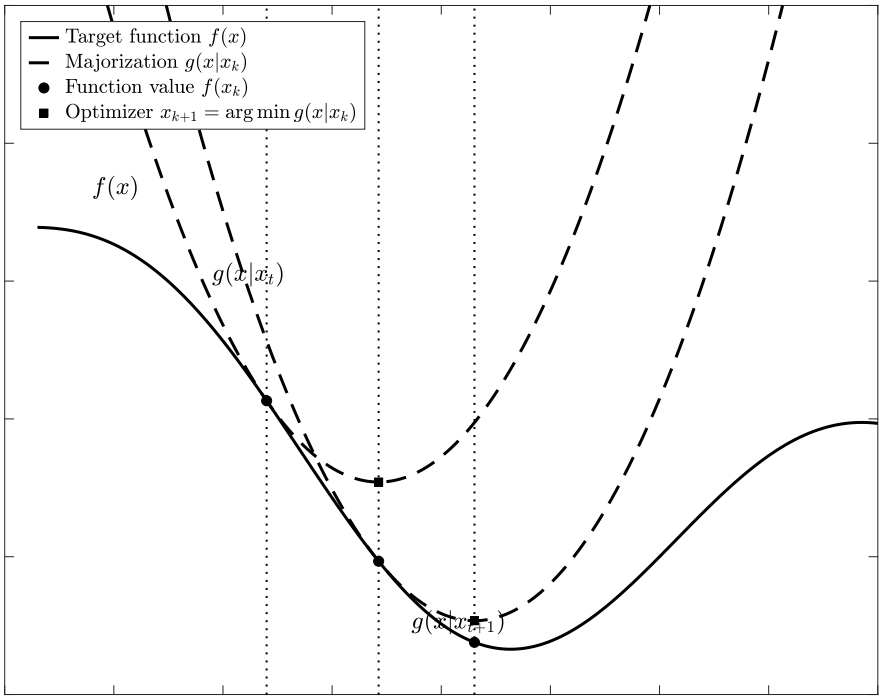

Figure 4.1: Illustration of Majorization-Minimization

4.2.2 Surrogate Subproblem

Define a set of coefficients:

Definition 4: Coefficients for time step t

For each time step $t \in \{1, \cdots, T\}$, given the iterate $\mathbf{v}_t$, define the coefficient components:

$$\mathbf{V}_t \triangleq \operatorname{diag}(\mathbf{v}_t) \tag{4.10}$$

$$\mathbf{X}_t \triangleq \left(\mathbf{V}_t^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi}\right)^{-1}, \tag{4.11}$$

$$\mathbf{M}_t \triangleq \frac{M}{N_s}\sum_{k=1}^{N_s}\frac{\mathbf{y}_k\mathbf{y}_k^{\mathsf{H}}}{\mathbf{y}_k^{\mathsf{H}}\mathbf{R}_t^{-1}\mathbf{y}_k} \tag{4.12}$$

$$\mathbf{Y}_t \triangleq \mathbf{\Phi}^{\mathsf{H}}\mathbf{M}_t\mathbf{\Phi}, \tag{4.13}$$

$$\beta_{t,i} \triangleq \left[\mathbf{X}_t^{-1} - \sigma_n^{-4}\mathbf{X}_t^{-1}\mathbf{Y}_t\mathbf{X}_t^{-1}\right] ii > 0, \quad \forall i \in 1, ..., N_c. \tag{4.14}$$

Lemma 1: Surrogate Function Construction

For time steps $t = 1, ..., T$, under Assumption 1, the surrogate function $g(\mathbf{v}|\mathbf{v}_t)$ of original objective $\mathcal{L}(\mathbf{R})$ is derived as:

$$g(\mathbf{v}|\mathbf{v}_t) = \boldsymbol{\beta}_t^{\mathsf{T}}\mathbf{v}^{-1} + \sum_{i=1}^{N_c}\log v_i \tag{4.15}$$

where $\boldsymbol{\beta}_t$ is calculated by Definition 4.

Proof. This is a proof sketch, with detailed derivations provided in the Appendix at the end of this chapter.

We upper bound the two terms of $\mathcal{L}(\mathbf{R})$: $\log\det(\mathbf{R})$ and $\sum_k \log(\mathbf{y}_k^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{y}_k)$.

1. With concave property of $\log\det(\cdot)$:

$$\log\det(\mathbf{R}) \leq \operatorname{tr}\left(\mathbf{X}_t\mathbf{V}^{-1}\right) + \log\det\mathbf{V} + \text{const.}$$

2. With first-order Taylor expansion:

$$\sum_k \log(\mathbf{y}_k^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{y}_k) \leq \text{tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right)$$

With Woodbury equality,

$$\mathbf{R}^{-1} = \sigma_n^{-2}\mathbf{I} - \sigma_n^{-4}\mathbf{\Phi}(\mathbf{V}^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathsf{H}}$$

$$\text{tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) = \sigma_n^{-2}\text{tr}(\mathbf{M}_t) - \sigma_n^{-4}\text{tr}\left(\mathbf{\Phi}^{\mathsf{H}}\mathbf{M}_t\mathbf{\Phi}\mathbf{X}\right)$$

Combining the bounds for these two terms, the function $g(\mathbf{v}|\mathbf{v}_t)$ can be expressed as:

$$\begin{aligned} g(\mathbf{v}|\mathbf{v}_t) = & \underbrace{\text{tr}\left(\mathbf{X}_t\mathbf{V}^{-1}\right) + \log\det\mathbf{V}}_{\text{Log-det term}} \\ & \underbrace{-\sigma_n^{-4}\text{tr}\left(\mathbf{X}_t\mathbf{Y}_t\mathbf{X}_t\mathbf{V}^{-1}\right) + \sigma_n^{-2}\text{tr}(\mathbf{M}_t)}_{\text{Trace term}} + \text{const.} \end{aligned} \tag{4.16}$$

□

Now we check whether Assumption 1 holds:

- Function matching: at $\mathbf{v} = \mathbf{v}_t$:

$$\begin{aligned} g(\mathbf{v}_t|\mathbf{v}_t) &= \text{tr}(\mathbf{X}_t\mathbf{V}_t^{-1}) + \log\det\mathbf{V}_t + \sigma_n^{-2}\text{tr}(\mathbf{M}_t) \\ &\quad - \sigma_n^{-4}\text{tr}\left(\mathbf{X}_t\mathbf{Y}_t\mathbf{X}_t\mathbf{V}_t^{-1}\right) \\ &= \mathcal{L}(\mathbf{R}|\mathbf{v}_t) \end{aligned}$$

- Global upper bound: the Jensen's inequality and Taylor's expansion in deriva-

tion would ensure

$$g(\mathbf{v}|\mathbf{v}_t) \geq \mathcal{L}(\mathbf{R}|\mathbf{v}), \quad \forall \mathbf{V} \succ 0$$

- Gradient consistency: Differentiating $g(\mathbf{v}|\mathbf{v}_t)$ at $\mathbf{v}_t$:

$$\begin{aligned}\nabla g(\mathbf{v}_t|\mathbf{v}_t) &= -\mathbf{V}_t^{-1}\mathbf{X}_t\mathbf{V}_t^{-1} + \mathbf{V}_t^{-1} \\ &\quad - \sigma_n^{-4}\mathbf{\Phi}\mathbf{X}_t\mathbf{Y}_t\mathbf{X}_t\mathbf{\Phi}^{\mathsf{H}} \\ &= \nabla\mathcal{L}(\mathbf{R}|\mathbf{v}_t)\end{aligned}$$

Change of Variable

The surrogate $g(\mathbf{v}|\mathbf{v}_t)$ is concave over $\mathbf{v}$, leading to the non-convexity of the t-subproblem. Therefore, we define a new variable $u_i = \frac{1}{v_i}$, and the t-th subproblem is:

Problem 3: The t-th subproblem

$$\begin{aligned}\min_{\mathbf{u},\mathbf{y}} \quad & g_u(\mathbf{u} \mid \mathbf{u}_t) := \boldsymbol{\beta}_t^{\mathsf{T}}\mathbf{u} - \sum_{i=1}^{N_c} \log u_i \\ \text{s.t.} \quad & -\log u_i \leq a y_i + b \\ & \sum y_i \leq s, \\ & 0 \leq y_i, 0 < u_i, \quad \forall i.\end{aligned} \tag{4.17}$$

where $\boldsymbol{\beta}_t$ is calculated in the way defined in Definition 4

Note that this problem becomes **strongly convex** with respect to $\mathbf{u}$.

With lemma 1, we proposed Algorithm 2 as sequential approximation of solving $\mathcal{P}_2$.

Algorithm 2 MM for Robust Covariance Estimation with log-Relaxed CCV

1: Initialize $\mathbf{u}_0 \succ \mathbf{0}$, $t = 0$
2: **while** not converged **do**
3: Compute $\boldsymbol{\beta}_t$ via current estimate $\mathbf{v}_t$ by Definition 4
4: Solve for $\mathbf{u}_{t+1}$, which is the solution of $\mathcal{P}_3$.
5: Update $\mathbf{u}_t \leftarrow \mathbf{u}_{t+1}$ (element-wise inverse)
6: $t \leftarrow t + 1$
7: **end while**

Appendix: Derivation of Majorizing Surrogate Function

Majorization-Minimization Framework

The MM algorithm constructs a surrogate function $g(\mathbf{v}|\mathbf{v}_t)$ that satisfies:

$$g(\mathbf{v}_t|\mathbf{v}_t) = \mathcal{L}(\mathbf{R}|\mathbf{v}_t) \quad \text{(Function matching)} \tag{4.18}$$

$$g(\mathbf{v}|\mathbf{v}_t) \geq \mathcal{L}(\mathbf{R}|\mathbf{v}),\ \forall \mathbf{v} \succ 0 \quad \text{(Global upper bound)} \tag{4.19}$$

$$\nabla g(\mathbf{v}_t|\mathbf{v}_t) = \nabla \mathcal{L}(\mathbf{R}|\mathbf{v}_t) \quad \text{(Gradient consistency)} \tag{4.20}$$

where $\mathcal{L}(\mathbf{v})$ is the original non-convex objective. At each iteration, we solve:

$$\mathbf{v}_{t+1} = \arg\min_{\mathbf{v}\succ 0} g(\mathbf{v}|\mathbf{v}_t) \tag{4.21}$$

Log-Determinant Term

Given the covariance matrix $\mathbf{R} = \mathbf{\Phi V \Phi}^{\mathsf{H}} + \sigma_n^2 \mathbf{I}_M$, we apply Sylvester's identity:

$$\log\det(\mathbf{R}) = \log\det\left(\mathbf{\Phi V \Phi}^{\mathsf{H}} + \sigma_n^2 \mathbf{I}_M\right) \tag{4.22}$$

$$= \log\det\left(\mathbf{\Phi}^{\mathsf{H}} \mathbf{\Phi V} + \sigma_n^2 \mathbf{I}_{N_c}\right) \quad \text{(Sylvester's identity)} \tag{4.23}$$

$$\tag{4.24}$$

where $N_c \ll M$ is the reduced dimension.

Factorize (4.23) as:

$$\log\det\mathbf{R} = \log\det\left(\sigma_n^2(\mathbf{V}^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi})^{-1}\right) + \log\det\mathbf{V} + N_c \log\sigma_n^2 \tag{4.25}$$

$$= \log\det\mathbf{X}^{-1} + \log\det\mathbf{V} + N_c \log\sigma_n^2 \tag{4.26}$$

where $\mathbf{X} \triangleq \mathbf{V}^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi} \succ \mathbf{0}$, and $\mathbf{X} \in \mathbb{C}^{N_c}$.

Using the concave property of $\log\det(\cdot)$, we construct a tight upper bound: for any $\mathbf{X}, \mathbf{X}_t \succ 0$:

$$\log\det\mathbf{X}^{-1} \leq -\mathrm{tr}\left(\mathbf{X}_t(\mathbf{X} - \mathbf{X}_t^{-1})\right) - \log\det\mathbf{X}_t^{-1} \tag{4.27}$$

Substituting (4.27) into (4.26) and discarding constants:

$$\log\det(\mathbf{R}) \leq \mathrm{tr}\left(\mathbf{X}_t\mathbf{V}^{-1}\right) + \log\det\mathbf{V} + \text{const.} \tag{4.28}$$

where $\mathbf{X}_t \triangleq (\mathbf{V}_t^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi})^{-1}$.

Trace Term

For the trace term $\frac{M}{N_s}\sum_{k=1}^{N_s}\log(\mathbf{y}_k^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{y}_k)$, denote $z_k = \mathbf{y}_k^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{y}_k$ and $z_{k,t} = \mathbf{y}_k^{\mathsf{H}}\mathbf{R}_t^{-1}\mathbf{y}_k$. We rewrite the trace term as:

$$\frac{M}{N_\mathrm{s}}\sum_{i=1}^{N_\mathrm{s}}\log z_k$$

Since the $\log(\cdot)$ is concave, we have first-order Taylor expansion at the current iteration point $z_{k,t}$:

$$\begin{aligned}\log z_k &\le \log z_{k,t} + \frac{1}{z_{k,t}}(z_k - z_{k,t})\\ &= \frac{z_k}{z_{k,t}} - 1 + \log z_{k,t}\end{aligned}$$

Therefore, we have:

$$\frac{1}{N_s}\sum_{k=1}^{N_s}\log z_k \le \sum_{k=1}^{N_s}\frac{z_k}{N_s z_{k,t}} + \text{const.} \tag{4.29}$$

Define the weighting matrix:

$$\mathbf{M}_t \triangleq \frac{M}{N_s}\sum_{k=1}^{N_s}\frac{\mathbf{y}_k\mathbf{y}_k^{\mathsf{H}}}{\mathbf{y}_k^{\mathsf{H}}\mathbf{R}_t^{-1}\mathbf{y}_k} \succcurlyeq 0 \tag{4.30}$$

This converts (4.29) to:

$$\text{RHS of (4.29)} = \mathrm{tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) + \text{const.} \tag{4.31}$$

Woodbury Identity

Using $\mathbf{R}^{-1} = \sigma_n^{-2}\mathbf{I} - \sigma_n^{-4}\mathbf{\Phi}(\mathbf{V}^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathsf{H}}$, we expand:

$$\mathrm{tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) = \sigma_n^{-2}\mathrm{tr}(\mathbf{M}_t) - \sigma_n^{-4}\mathrm{tr}\left(\mathbf{\Phi}^{\mathsf{H}}\mathbf{M}_t\mathbf{\Phi}\mathbf{X}\right) \tag{4.32}$$

where $\mathbf{X} \triangleq (\mathbf{V}^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi})^{-1}$.

Linear Approximation

A first-order expansion of $\mathrm{tr}(\mathbf{Y}_t\mathbf{X})$ at $\mathbf{X}_t \triangleq (\mathbf{V}_t^{-1} + \sigma_n^{-2}\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi})^{-1}$ yields:

$$\mathrm{tr}(\mathbf{Y}_t\mathbf{X}) \geq \mathrm{tr}(\mathbf{Y}_t\mathbf{X}_t) - \mathrm{tr}\left(\mathbf{X}_t\mathbf{Y}_t\mathbf{X}_t(\mathbf{V}^{-1} - \mathbf{V}_t^{-1})\right) \tag{4.33}$$

where $\mathbf{Y}_t \triangleq \mathbf{\Phi}^{\mathrm{H}}\mathbf{M}_t\mathbf{\Phi}$. Substituting (4.33) into (4.32) gives:

$$\mathrm{tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) \leq \sigma_n^{-2}\mathrm{tr}(\mathbf{M}_t) - \sigma_n^{-4}\mathrm{tr}\left(\mathbf{X}_t\mathbf{Y}_t\mathbf{X}_t\mathbf{V}^{-1}\right) + \text{const.} \tag{4.34}$$

Combined Surrogate Function

Merging (4.28) and (4.34), we obtain:

$$\begin{aligned} g(\mathbf{v}|\mathbf{v}_t) = & \underbrace{\mathrm{tr}\left(\mathbf{X}_t\mathbf{V}^{-1}\right) + \log\det\mathbf{V}}_{\text{Log-det term}} \\ & \underbrace{-\sigma_n^{-4}\mathrm{tr}\left(\mathbf{X}_t\mathbf{Y}_t\mathbf{X}_t\mathbf{V}^{-1}\right) + \sigma_n^{-2}\mathrm{tr}(\mathbf{M}_t)}_{\text{Trace term}} + \text{const.} \end{aligned} \tag{4.35}$$

Diagonal Case Simplification

Assume $\mathbf{V} = \mathrm{diag}(v_1, \ldots, v_{N_c}) \succ 0$. The surrogate function simplifies to:

$$g(\mathbf{v}|\mathbf{v}_t) = \sum_{i=1}^{N_c}\left(\frac{\beta_{t,i}}{v_i} + \log v_i\right) + \text{const.} \tag{4.36}$$

where $\beta_{t,i} \triangleq [\mathbf{X}_t^{-1} - a^2\mathbf{X}_t^{-1}\mathbf{Y}_t\mathbf{X}_t^{-1}]_{ii} > 0$.

Chapter 5

Single Loop Primal-Dual Update Algorithm

Building on the convex relaxation framework from Chapter 4, we propose a single-loop primal-dual algorithm to solve for the t-th subproblem 3. After deriving the Lagrangian, with an alternative update strategy, we have a closed-form update for both primal variables. The dual update is by gradient assent rule obtained by KKT conditions.

5.1 Primal-Dual Method

The **primal-dual** method is an iterative framework for solving optimization problems with constraints, especially when strictly enforcing feasibility (e.g., satisfying all constraints at every step) is computationally expensive or difficult. Instead, the method allows temporary violations of constraints but punishes them with a price.

To quantify the violations, **dual** variables are introduced, while **primal** variables (the original decision variables) remain responsible for optimizing the original objective function. Over iterations, this mechanism guides the solution toward feasibility

while optimizing the original objective.

The primal dual method operates on two classes of variables alternately during iterations. When the dual variables are fixed, primal variables are updated to simultaneously pursue objective optimization and incorporate penalty adjustments based on current violation costs. Conversely, when the primal variables are held constant, dual updates evaluate the latest solution's constraint violations and adjust penaltites accordingly.

A common choice for dual variables is Lagrange multipliers, which link the primal problem to its dual through the Lagrangian function. Ideally, at convergence, the solution satisfies the KKT conditions.

For the t-th subproblem $\mathcal{P}_3$, directly solving the KKT system of the Lagrangian function associated with the surrogate $g_u(\mathbf{u}|\mathbf{u}_t)$ is difficult. To address this, we propose a hybrid method.

- The local constraint: $-\log u_i \leq ay_i + b$ is incorporated into Lagrangian framework with dual variable $\lambda_i, \forall i$.
- The global constraint $\sum_{i=1}^{N_c} y_i \leq s$ is enforced via a projection step, ensuring primal feasibility.

The partial Lagrangian function is defined as:

$$\begin{aligned} L(\mathbf{u}, \mathbf{y}, \boldsymbol{\lambda}, \mu) = \boldsymbol{\beta}_t^{\mathrm{T}} \mathbf{u} - \sum_{i=1}^{N_c} \log u_i + \sum_{i=1}^{N_c} \lambda_i (-\log u_i - a y_i - b) \\ \text{s.t.} \sum_{i=1}^{N_c} y_i \leq s. \end{aligned} \tag{5.1}$$

where $a = \log \frac{V_{\max}}{\zeta} > 0, b = \log \zeta$. $\mathbf{u}, \mathbf{y}$ are primal variables, and $\boldsymbol{\lambda} = [\lambda_i, \cdots, \lambda_{N_c}] \succeq \mathbf{0}$ is the dual variable. The requirements of $u_i > 0, y_i \geq 0$ are implicitly satisfied with proper initialization.

5.2 Update Rules

5.2.1 Primal Update

The primal updates are updated by solving the first-order optimality condition $\nabla_{\mathbf{u}} L = \mathbf{0}$. Solving $\partial L / \partial u_i = 0$ yields the closed-form expression for i-th element:

$$u_i^{t+1} = \frac{1 + \lambda_i^t}{\beta_i^t}. \tag{5.2}$$

The update directly corresponds to the variance parameter $v_i = \frac{1}{u_i}$, ensuring that

$$v_i^{t+1} = \frac{\beta_i^t}{1 + \lambda_i^t} \leq \beta_i^t.$$

Here, the dual variable $\lambda_i^t \geq 0$ acts as an **adaptive shrinkage factor**. If the local constraint $-\log u_i^t \leq a y_i^t + b$ is violated (i.e., v_i^t exceeding $e^{a y_i^t + b}$), the term λ_i^t increases, shrinking v_i^{t+1} to enforce feasibility. Conversely, if the constraint is satisfied, λ_i^t decreases, relaxing the shrinkage effect. Compared to the case when local constraint is taken away (i.e., $\lambda_i^t = 0$), the solution would reduce to $v_i^{t+1} = \beta_i^t$.

For the significance indicator, the other primal variable y_i, recall that it was originally introduced to approximate whether v_i exceeds a significant threshold. In another word, y_i^{t+1} should also reflect on how λ_i^{t+1} violates the local constraint. To update y_i^{t+1}, we first compute an intermediate variable reflecting the constraint tightness via complementary slackness $\lambda_i \left(-\log u_i - a y_i - b\right) = 0$:

$$z_i^{t+1} = \max(0, \frac{-(\log u_i^t + b)}{a}), \tag{5.3}$$

Here, $z_i^{t+1} > 0$ indicates the significance of v_i^{t+1}. To enforce the global constraint

$\sum_i y_i \leq s$, we project z_i^{t+1} as:

$$y_i^{t+1} = \begin{cases} z_i^{t+1} & \text{if } \sum_{j=1}^{N_c} z_j^{t+1} \leq s \\ \frac{s}{\sum z_j^{t+1}} z_i^{t+1} & \text{otherwise} \end{cases}$$

5.2.2 Dual Update

From the complementary slackness $\lambda_i \left(-\log u_i - a y_i - b\right) = 0$, we define the dual residual $r_i^t = -\log u_i^t - a y_i^t - b$, which quantifies local constraint violation. The dual update follows a gradient ascent scheme:

$$\lambda_i^{t+1} = \max\left(0, \lambda_i^t + \alpha r_i^t\right)$$

where $\alpha > 0$ is a step size.

Therefore, we propose the following primal-dual algorithm for solving the t-th subproblem of $\mathcal{P}_3$:

Algorithm 3 Primal-Dual Update for Solving t-th subproblem of $\mathcal{P}_3$

Given $\mathbf{u}^t \succ \mathbf{0}$, $\boldsymbol{\lambda}^t = \mathbf{0}$, N_c
Compute: $\boldsymbol{\beta}_t$ with $\mathbf{u}_t$ by Definition 4
for $i = 1 : N_c$ **do**
    Primal update: $u_i^{t+1} = \frac{1+\lambda_i^t}{\beta_i^t}$
    Compute: $z_i^{t+1} = \max(0, -\frac{\log u_i^{t+1} + b}{a})$
    Project: $\mathbf{y}^{t+1} = \frac{s\mathbf{z}^{t+1}}{\sum z_i^{t+1}}$
end for
for $i = 1 : N_c$ **do**
    Dual update: $\lambda_i^{t+1} = \max\left(0, \lambda_i^t + \alpha(-\log u_i^{t+1} - a y_i^{t+1} - b)\right)$
end for

Combining with the primal-dual solution on t-th subproblem, we propose the following single-loop algorithm for solving the problem with relaxed CCV $\mathcal{P}_2$:

Algorithm 4 Single Loop Primal-Dual Algorithm for Solving $\mathcal{P}_2$

1: Initialize $\mathbf{u}^0 > 0$, $\boldsymbol{\lambda}^0 = \mathbf{0}$, $t = 0$
2: **while** not converge **do**
3: compute $\mathbf{u}_{t+1}, \mathbf{y}_{t+1}, \boldsymbol{\lambda}_{t+1}$ according to Algorithm 3
4: $t \leftarrow t + 1$
5: **end while**
6: Return $\mathbf{u}^*$.

Chapter 6

Numerical Results

This chapter evaluates the proposed robust elliptical covariance estimation algorithm through three aspects: convergence behavior, computational and time efficiency, and its performance in target detection tasks in the regime of space and time adaptive processing.

6.1 Experimental Setup

We simulate a moving array platform with MATLAB Phased Array Toolbox, comprising M array elements, with N pulses transmitted within a single Coherent Processing Interval (CPI). These elements form a uniform linear array (ULA) with a half-wavelength ($\lambda/2$) spacing to prevent spatial aliasing. The platform moves at a constant velocity v. Table 6.1 details the parameters used in the experiment.

The simulations were performed in MATLAB 2024b, using a MacBook Pro with a 1.4 GHz Quad-Core Intel Core i5 processor, 16 GB 2133 MHz LPDDR3 RAM, and Intel Iris Plus Graphics 645 (1536 MB).

Figure 6.1 shows the unprocessed range-amplitude profile where interference power decays exponentially with range ($\propto R^{-4}$). A target was injected in range cell 71,

Table 6.1: Experimental Parameters for the Moving Array Platform

Parameter	Symbol	Value
Number of array elements	M	12
Number of pulses per CPI	N	32
Platform height	H	1000 m
Range of detection	R	5000 m
Operating frequency	f_c	10 GHz
Pulse repetitive frequency	PRF	30 kHz
Inter-element spacing	d	$\lambda/2$
Wavelength	λ	0.03 m
Platform velocity	v	294 m/s
Normalized target Doppler frequency	-	0.21
Target range	-	1732 m
Interference-to-Noise Ratio	INR	67 dB
Signal-to-Noise Ratio	SNR	20 dB

around 1700 meters, and the target signal remains undetectable beneath the overwhelming interference.

Figure 6.2, 6.3, demonstrate joint angle-Doppler processing through an azimuth coverage -90 degree to 90 degree, and normalized Doppler frequency from -0.5 to 0.5. The clutter subspace (Fig. 6.2) has diagonal energy distribution (white dashed line) due to platform motion-induced Doppler-azimuth coupling. Raw data (Fig. 6.3) confirms discrete interference components along this ridge, completely masking the target (star marker) despite its spatial-Doppler separation.

Simulations conducted in MATLAB R2024b on MacBook Pro (1.4 GHz Core i5, 16 GB RAM).

6.2 Convergence

The proposed algorithm, as detailed in Algorithm 4, exhibits converges. We benchmark our method with the covariance estimator with the Laplace prior (LP), which is known to be equivalent to mean-square-error minimization with ℓ_1-norm penalty

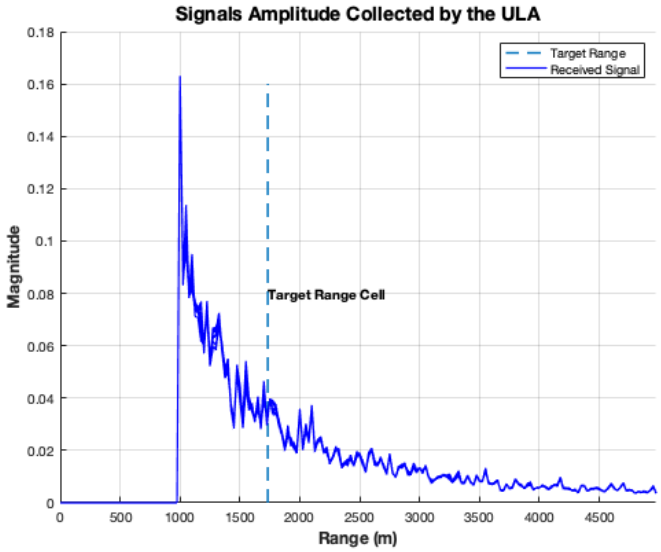

Figure 6.1: Raw range-amplitude data with interference.

[Tibshirani, 1996].

Table 6.2 summarizes the theoretical convergence rates. For LP, we list two implementations, interior-point methods (LP-int) and first-order proximal gradient methods (LP-FASTA).

Table 6.2: Theoretical Convergence Rates

Algorithm	**Iterations**
LP-int (SDPT3/Mosek)	$O(\log \frac{1}{\epsilon})$
LP-FASTA	$O(\frac{1}{\sqrt{\epsilon}})$
RE (Proposed)	$O(\log \frac{1}{\epsilon})$

Implementation details: We compare with the FASTA proximal gradient implementation (`https://github.com/tomgoldstein/fasta-matlab`). Figure 6.4 shows the convergence behavior under stopping criteria $\|g_u(\mathbf{u}_{t+1}) - g_u(\mathbf{u}_t)\|^2 / \max_t \|g_u(\mathbf{u}_t)\|^2 \leq 10^{-8}$, through $T = 100$ trials with 20 snapshots each.

In numerical experiment, usable covariance estimates can be obtained at around 10 iterations.

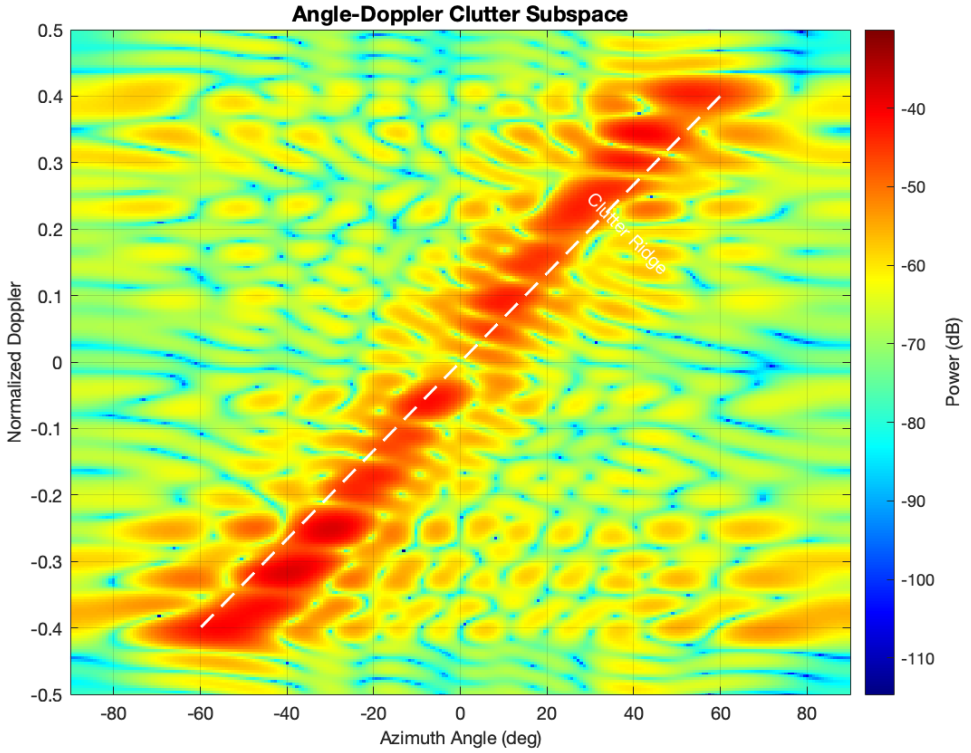

Figure 6.2: Angle-Doppler response of entire intereference subspace

6.3 Computational Complexity

Now we compare the computational demands under the configuration: signal dimension $MN = 384$ ($M = 12$ number of array elements, $N = 32$ pulses per CPI), dictionary size $N_c = 180$ (angular resolution), and snapshots $N_s \leq 50$.

We compare our robust elliptical estimator (RE) with Laplace prior-based methods (LP), including interior-point methods (LP-int, with SDPT and Mosek), first-order proximal gradient methods (LP-FASTA).

The computational complexity per iteration is summarized in Table 6.3. The *Dominant Operation* column indicates the computational bottleneck at each iteration.

Since N_s is significantly small in this configuration, it is not included in the table.

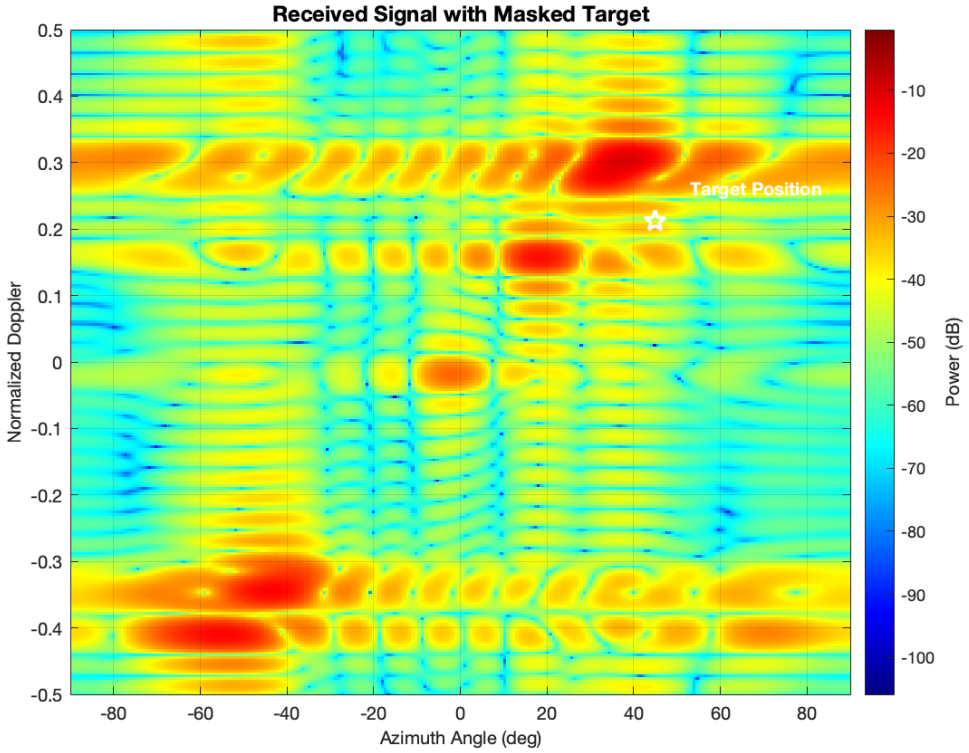

Figure 6.3: Raw Angle-Doppler Response, Demonstrating the Interference Power Spikiness and its Dominance to Target Reflection

The key observation is the computational complexity hierarchy: $O(N_c^2)$(RE) $<$ $O(MN \cdot N_c)$(FASTA) $< O(N_c^3))$(LP-int) in this problem configurations where $MN \gg N_c$.

The LP-int's cubic complexity originates for the KKT systems factorization, while Mosek accelerates SDPT3 via presolving technique. LP-FASTA reduces per-iteration complexty via gradient methods, but requires more iteration to converge, as shown in previous section.

The proposed robust elliptical estimator (RE) has $O(N_c^3)$ complexity from the matrix inversion of the intermediate $\mathbf{X}_t \in \mathbb{C}^{N_c \times N_c}$ matrix, defined as $\mathbf{X}_t := \sigma_n^2 \mathbf{\Phi}\mathbf{\Phi}^{\mathsf{H}} + \mathbf{U}_t$, where $\mathbf{U}_t$ is the diagonal matrix form of the current iterate t. The positive definiteness

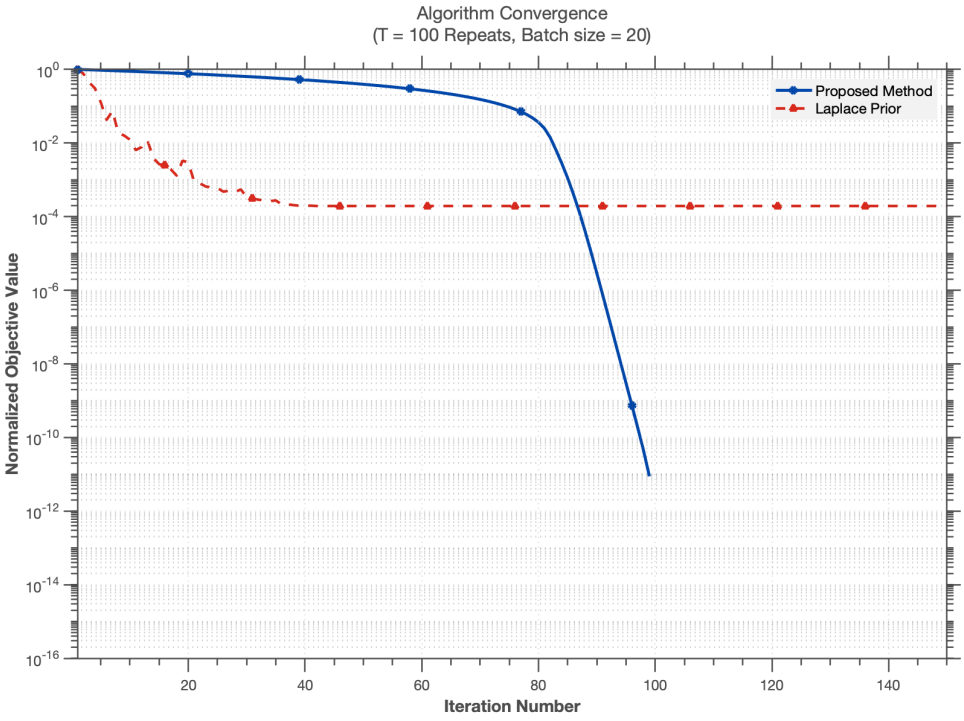

Figure 6.4: Convergence of Different Algorithms

Table 6.3: Theoretical Per-Iteration Complexity

Algorithm	**Complexity/Iteration**	**Dominant Operation**
LP-int (SDPT3)	$O(N_c^3)$	KKT factorization
LP-int (Mosek)	$O(N_c^3)$	Presolved KKT factorization
LP-FASTA	$O(MN \cdot N_c)$	Gradient projection
RE (Proposed)	$O(N_c^2)$	Cholesky decomposition with rank 1 update

of $\mathbf{X}_t$ permits efficient inversion though Cholesky decomposition. A small diagonal loading $\delta = 10^{-20}$ is added for numerical stability:

$$\mathbf{X}_t \leftarrow \mathbf{X}_t + \delta \mathbf{I}. \tag{6.1}$$

The Cholesky matrix inversion update is:

$$\mathbf{X}_t^{-1} = \mathbf{L}_t^{-\mathsf{H}}(\mathbf{L}_t^{-1}\mathbf{I}) \tag{6.2}$$

where $\mathbf{X}_t = \mathbf{L}_t\mathbf{L}_t^{\mathsf{H}}$, and $\mathbf{L}_t$ being a lower triangular matrix, which needs to be solved each time.

Also note that between successive iterations, only the diagonal matrix $\mathbf{U}_t$ changes, while $\sigma_n^2\mathbf{\Phi}\mathbf{\Phi}^{\mathsf{H}}$ remains constant. This further allows space for acceleration by precalculating the unchanged $\sigma_n^2\mathbf{\Phi}\mathbf{\Phi}^{\mathsf{H}}$ and applying rank-1 update, the practical complexity can be reduced from $O(N_c^3)$ to $O(N_c^2)$ per iteration after initial decomposition.

For fair comparison and practical efficiency, all matrix inversion of covariance matrix $\mathbf{R}_t$ is performed under Woodbury matrix inversion formula, exploiting the low-rank structure of $\mathbf{\Phi}\mathbf{\Phi}^{\mathsf{H}}$ when $MN \gg N_c$.

Table 6.4: Empirical per Iteration Time (seconds)

Algorithm	$N_s = 50$	$N_s = 20$
LP-int (SDPT3)	9.23	2.52
LP-int (Mosek)	7.67	2.53
LP-FASTA	0.01	0.01
RE (Proposed)*	0.04	0.02

Implementation was conducted in MATLAB R2024b with Mosek 9.1.9/SDPT3 4.0, using FASTA from https://github.com/tomgoldstein/fasta-matlab. RE's runtime includes initialization of precomputed terms.

6.4 Beam Pattern

How estimator recovers beam pattern is the evaluation criteria in the beamforming perspective. As shown in Fig. 6.2, the interference energy concentration along the diagonal ridge (-45° to +45° azimuth vs. -0.5 to +0.5 normalized Doppler). An ideal beam pattern should have deep nulls along the diagonal clutter ridge where interference concentrates. Also, it should maintain a distinct peak at true target coordinates.

For benchmarks, besides RE (proposed) and LP-FASTA we also include DL (diagonal loading) as well.

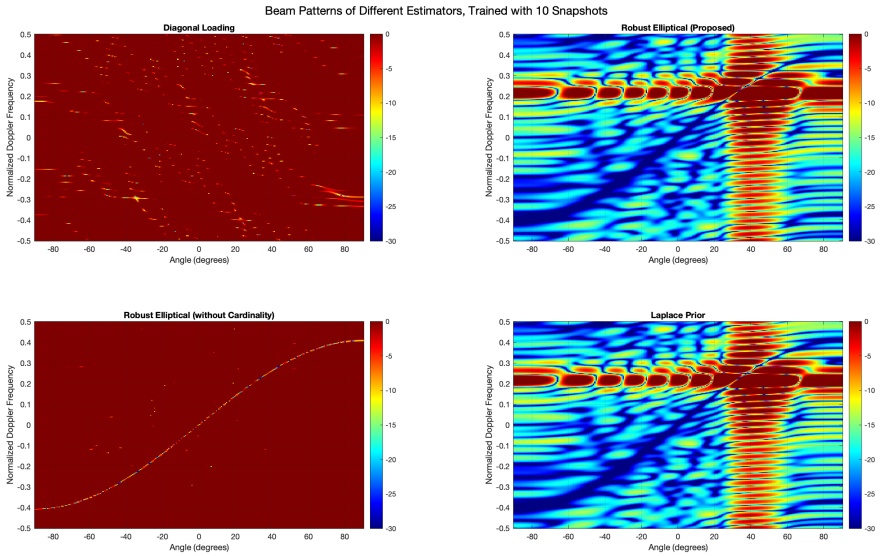

Figure 6.5: Beam patterns of Different Estimators, Limited Snapshots

When sample size is limited, figure. 6.5 (N=10 samples) Under limited sample conditions ($N = 10$, Fig. 6.5), diagonal loading (DL) fails to form sufficient nulls along

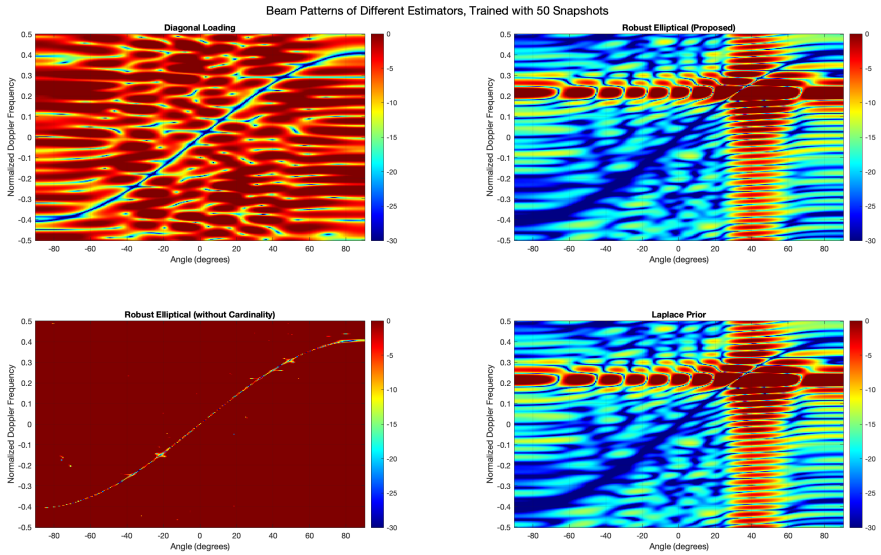

Figure 6.6: Beam patterns of Different Estimators, Sufficient Snapshots

the interference ridge, while both Laplace prior (LP) and the proposed robust elliptical (RE) estimators achieve comparable null depths. The unconstrained RE variant produces partial diagonal nulls with reduced clarity. When sample size increases to $N = 50$ (Fig. 6.6), DL shows significant performance improvement in null formation, whereas LP, RE, and unconstrained RE maintain behaviors consistent with the limited-sample regime.

Conclusion

This work presents a framework for robust covariance estimation in non-Gaussian interference environments with limited snapshots. The main results are summarized below:

Key Contributions

- **Variance-domain sparsity model**: The cardinality-constrained variance (CCV) model (§2.2.3) characterizes sparse interference power distribution through second-order statistics. It uses two physical parameters: the maximum number of dominant interferers (s) and power threshold (ζ), avoiding empirical hyperparameter tuning in traditional sparse methods.
- **Optimization framework**: We address the non-convexity of maximum likelihood estimation with CCV constraints (§3.3) through:
    1. Convex relaxation of cardinality constraints via logarithmic domain transformation (§4.1)
    2. A Majorization-Minimization (MM) sequence with closed-form primal updates (§5.2)

Future Work

- **Convergence analysis**: Analyze the theoretical convergence rate of the MM relaxation framework under the robust elliptical likelihood objective.
- **Statistical efficiency**: Derive the Cramér-Rao bound for CCV estimators and compare with ℓ_0/ℓ_1-constrained approaches to quantify performance limits.

Bibliography

Abdi, M. J. (2013). *Cardinality optimization problems*. PhD thesis, University of Birmingham.

Anagnostopoulos, K. P. and Mamanis, G. (2011). The mean–variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms. *Expert Systems with Applications*, 38(11):14208–14217.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

Bertsimas, D. and Shioda, R. (2009). Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22.

Bian, W. and Chen, X. (2020). A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty. *SIAM Journal on Numerical Analysis*, 58(1):858–883.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

Biswas, A. and Barman, S. (2018). Fair division under cardinality constraints. In *IJCAI*, pages 91–97.

Candes, E. J. and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425.

Candes, E. J., Wakin, M. B., and Boyd, S. P. (2008). Enhancing sparsity by reweighted ℓ 1 minimization. *Journal of Fourier analysis and applications*, 14:877–905.

Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418.

Chang, T.-J., Meade, N., Beasley, J. E., and Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Cui, W., Wang, T., Wang, D., and Liu, K. (2022). An efficient sparse bayesian learning stap algorithm with adaptive laplace prior. *Remote Sensing*, 14(15):3520.

Cui, X., Zheng, X., Zhu, S., and Sun, X. (2013). Convex relaxations and miqcqp reformulations for a class of cardinality-constrained portfolio selection problems. *Journal of Global Optimization*, 56(4):1409–1423.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Gao, J. and Li, D. (2013). Optimal cardinality constrained portfolio selection. *Operations research*, 61(3):745–761.

Goldstein, L., Minsker, S., and Wei, X. (2018). Structured signal recovery from nonlinear and heavy-tailed measurements. *IEEE Transactions on Information Theory*, 64(8):5513–5530.

Guerci, J. R. (2014). *Space-time adaptive processing for radar*. Artech House.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.

Li, J., Stoica, P., and Wang, Z. (2003). On robust capon beamforming and diagonal loading. *IEEE transactions on signal processing*, 51(7):1702–1715.

Lin, J., Nassar, M., and Evans, B. L. (2013). Impulsive noise mitigation in powerline communications using sparse bayesian learning. *IEEE Journal on Selected Areas in Communications*, 31(7):1172–1183.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nikias, C. L. and Shao, M. (1995). *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience.

Razaviyayn, M., Hong, M., and Luo, Z.-Q. (2013). A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153.

Richards, M. A. et al. (2005). *Fundamentals of radar signal processing*, volume 1. Mcgraw-hill New York.

Rubinstein, M. (2002). Markowitz's" portfolio selection": A fifty-year retrospective. *The Journal of finance*, 57(3):1041–1045.

Ruppert, D. and Matteson, D. S. (2011). *Statistics and data analysis for financial engineering*, volume 13. Springer.

Sadler, B. M., Giannakis, G. B., and Shamsunder, S. (1995). Noise subspace techniques in non-gaussian noise using cumulants. *IEEE transactions on aerospace and electronic systems*, 31(3):1009–1018.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280.

Stoica, P. and Nehorai, A. (1989). Music, maximum likelihood, and cramer-rao bound. *IEEE Transactions on Acoustics, speech, and signal processing*, 37(5):720–741.

Stoica, P. and Sharman, K. C. (1990). Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(7):1132–1143.

Sun, Y., Babu, P., and Palomar, D. P. (2016). Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *IEEE Transactions on Signal Processing*, 64(14):3576–3590.

Telatar, E. (1999). Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Tropp, J. A. (2004). Just relax: Convex programming methods for subset selection and sparse approximation. *ICES report*, 404.

Tse, D. and Viswanath, P. (2005). *Fundamentals of wireless communication*. Cambridge university press.

Tyler, D. E. (1987). A distribution-free m-estimator of multivariate scatter. *The annals of Statistics*, pages 234–251.

Van Trees, H. L. (2002). *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on biomedical engineering*, 44(9):867–880.

Vorobyov, S. A., Gershman, A. B., and Luo, Z.-Q. (2003). Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem. *IEEE transactions on signal processing*, 51(2):313–324.

Ward, J. (1998). Space-time adaptive processing for airborne radar. In *IEE Colloquium on Space-Time Adaptive Processing*, pages 2–1. IET.

Wipf, D. P. and Rao, B. D. (2004). Sparse bayesian learning for basis selection. *IEEE Transactions on Signal processing*, 52(8):2153–2164.

Woodside-Oriakhi, M., Lucas, C., and Beasley, J. E. (2011). Heuristic algorithms for the cardinality constrained efficient frontier. *European Journal of Operational Research*, 213(3):538–550.

Xie, D., Lu, Z.-R., Li, G., Liu, J., and Wang, L. (2023). Efficient laplace prior-based sparse bayesian learning for structural damage identification and uncertainty quantification. *Mechanical Systems and Signal Processing*, 188:110000.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.