Inference Methods for Synthetic Aperture Radar and Array Processing With Lattices

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

David Tucker, B.S.

Graduate Program in Electrical and Computer Engineering

The Ohio State University

2024

Dissertation Committee:

Lee C. Potter, Advisor

Emre Ertin

Kiryung Lee

Abstract

This dissertation addresses two distinct topics: inference from synthetic aperture radar (SAR) data and the application of lattice theory to array signal processing. First, we propose a series of methods that mitigate the effect of speckle in inference tasks involving multichannel SAR data. Coherent imaging methods like SAR are subject to speckle, and the suppression of this noise-like quality is often a prerequisite to image interpretation. We propose a technique that recovers the per-pixel multichannel SAR covariance matrix and incorporates a statistical model of speckle and *a priori* knowledge of the varieties of clutter present in the scene. In this approach, an expectation-maximization algorithm is made computationally tractable by a graph-coloring probing technique. We next address the problem of coherent change detection in repeat-pass SAR data. A Bayesian change detection approach is given that assigns prior distributions to the unobserved model variables to exploit spatial structure both in the geophysical scattering qualities of the scene and among the scene disturbances that take place between the passes. We also give a polarimetric SAR (PolSAR) despeckling method based on convolutional neural networks (CNNs). An invertible transformation involving a matrix logarithm is used to facilitate CNN processing of the PolSAR data. A residual learning strategy is adopted, in which the CNN is trained to identify the speckle component which is then removed from the corrupted image.

The latter portion of this dissertation is concerned with the application of lattice theory to array signal processing. We consider the problem of maximum likelihood parameter estimation in mixed integer linear models and provide two polynomial-time solution methods for special cases of this problem. These approaches extend the prior art by allowing for multivariate real-valued unknowns and more general linear models. We then provide a generally applicable alternative solution method that uses sphere decoding.

We next consider the design and analysis of nonuniform arrays in one, two, and three dimensions. We give simply tested necessary and sufficient conditions for an array of sensors to unambiguously determine the direction of arrival for a specified set of possible directions of arrival. The new results facilitate the design of nonuniform arrays, allowing for configurations with widely separated sensors or increased apertures without an increase in the number of sensors. The unambiguous region is shown to be a parallelotope, a property which admits simple geometric interpretation and facilitates array design.

Acknowledgments

First, I thank the faculty of Ohio State's electrical and computer engineering department. In particular, I thank my advisor Lee Potter for his mentorship and patience. I also thank Emre Ertin, Kiryung Lee, Bradley Clymer, and Phil Schniter for serving on my exam committees. I am grateful to the Center for Surveillance Research, an NSF Industry and University Cooperative Research Center (IUCRC), and AFRL for funding in part the work in this dissertation. I also thank my coauthors for their insight and collaboration. Finally, I thank my family and friends for their support.

Vita

December 2013	B.S., Electrical and Computer Engineering, The Ohio State University
August 2014–June 2016	Radar systems engineer, MDA Information Systems
September 2016–Present	Graduate research associate, The Ohio State University.

Publications

D. Tucker and L. C. Potter, "Approximate maximum likelihood estimation in split-dwell pulse-Doppler radar," in *IEEE Radar Conf.*, to be published.

D. Tucker, S. Zhao, and L. C. Potter, "Maximum likelihood estimation in mixed integer linear models," *IEEE Signal Process. Lett.*, vol. 30, pp. 1557–1561, 2023.

D. Tucker, J. N. Ash, and L. C. Potter, "SAR coherent change detection with variational expectation maximization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2163–2175, 2023.

D. Tucker, S. Zhao, R. Ahmad, and L. C. Potter, "Alias-free arrays," *IEEE Signal Process. Lett.*, vol. 29, pp. 2457–2461, 2022.

S. Zhao, R. Ahmad, L. Potter, and D. Tucker "Venc design and velocity estimation for phase-contrast MRI," US Patent 63 324 133, filed Mar. 27, 2022.

D. Tucker and L. C. Potter, "Polarimetric SAR despeckling with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

D. Tucker and L. C. Potter, "Speckle suppression in multi-channel coherent imaging: a tractable Bayesian approach," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1429–1439, 2020.

Fields of Study

Major Field: Electrical and Computer Engineering

Specialization: Signal Processing

Table of Contents

List of Tables

List of Figures

Figure **Page**

Chapter 1: Introduction

This dissertation addresses two topics: inference from synthetic aperture radar (SAR) data and the use of lattice theory in array signal processing.

1.1 Inference Methods for Synthetic Aperture Radar

In the first portion of this dissertation, we propose a series of inference methods for multichannel SAR data. SAR is a microwave sensing technology capable of producing high-resolution images of terrain and targets. The airborne or spaceborne radar system traverses a flight path, transmitting and receiving a series of high-bandwidth radio frequency (RF) pulses, most often linear frequency modulated (LFM) chirps, towards a scene of interest. In spotlight mode SAR, the antenna is steered to illuminate a fixed patch of ground continuously. The backscattered pulses are mixed with a reference signal and low-pass filtered, resulting in a set of samples of the Fourier transform of the complex scene reflectivity confined to an annular region in the spatial frequency domain. The recorded pulses or phase history can then be processed into an image.

In this dissertation, we consider multichannel SAR imaging modalities, such as polarimetric SAR (PolSAR) or interferometric SAR (InSAR), in which data are collected via multiple channels. These data are considerably more informative than

single-channel collections; the inter-channel relationships are informative of physical scattering behavior [10, 11]. These multichannel radar systems have proven critical for monitoring changes on the earth's surface, and numerous techniques employ these data for terrain classification or surface parameter estimation [12].

We use a tomographic formulation of SAR and a plane-wave assumption, permitting the expression of the sampled radar returns as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{u}, \tag{1.1}$$

in which $\boldsymbol{A}$ is a SAR imaging operator, the lexicographically-ordered 2D scene reflectivity samples are denoted by $\boldsymbol{x}$, and $\boldsymbol{u}$ represents thermal noise.

Speckle in SAR imaging is a result of the coherent interaction of reflections from objects within a resolution cell. When illuminating a homogeneous medium that is rough on the scale of the radar's central frequency, the reflectivity is modeled as a random variable with a complex normal distribution [13–15].

In Chapter 2 we give a method for multichannel coherent image recovery based on a hierarchical statistical model. This approach accommodates a reduced synthetic aperture but also suppresses speckle and thermal noise. The model presented here uses latent terrain class variables with an MRF prior to exploit both structure among coherent imaging channels and spatial correlation for improved image recovery.

In Chapter 3, we consider the problem of coherent change detection from repeat-pass SAR imagery. A Bayesian approach is formulated as an alternative to conventional window-based change detection statistics that entail losses to spatial resolution. The proposed approach assigns prior distributions to the unobserved model variables to exploit spatial structure both in the geophysical scattering qualities of the scene and among the scene disturbances that take place between the passes. Variational

expectation maximization is used to efficiently approximate the posterior distribution of the latent variables and the prior model hyperparameters. Experiments on simulated and measured interferometric SAR data pairs indicate the effectiveness of the proposed change detection method and highlight improvements over traditional window-based approaches.

In Chapter 4, we consider the problem of speckle reduction of multichannel SAR images. It is often a prerequisite to suppress speckle in some way before SAR data are considered suitable for analysis [16]. This is frequently realized by the application of a speckle filter, classically one which acts on the image locally [17]. We describe a polarimetric SAR despeckling approach that uses convolutional neural networks (CNNs). The approach operates in the matrix logarithm domain and uses a residual learning architecture that has been successful in natural image denoising. We have implemented this approach with a training set of spatially multilooked PolSAR images that are corrupted by a simulated speckle process. Further, the proposed approach provides improved despeckling over conventional filtering methods and operates with a comparably fast running time afforded by the use of CNNs.

1.2 Array Processing With Lattices

This dissertation next addresses the applications of lattices (regular arrangements of points in Euclidean space) to sensor array processing. In Chapter 5, we consider the problem of maximum likelihood (ML) parameter estimation in mixed integer linear models. That is, we consider a linear model with a measurement vector $\boldsymbol{y}$ that is related to a real-valued unknown $\boldsymbol{x}$ and an integer-valued unknown $\boldsymbol{k}$ as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{M}\boldsymbol{k} + \boldsymbol{u}, \tag{1.2}$$

where $\boldsymbol{A}$ is an encoding matrix with full column rank and $\boldsymbol{M}$ is an invertible matrix. This problem is relevant to many applications, including ranging [18–21], frequency estimation [22–24], interferometry [25], MIMO decoding [26–28], direction of arrival estimation [29], multichannel modulo sampling [30], and magnetic resonance imaging [9, 31].

In [32], a case of (1.2) is considered, in which there is a single real-valued variable and the encoding matrix meets a certain coprimeness condition. There, a polynomial-time ML solution method is given that leverages the classical Chinese remainder theorem. In Chapter 5, we provide two ML solution methods that extend polynomial-time computation to a wider range of problems. These extensions of the prior art account for both multivariate real-valued unknowns and general encoding matrices with full column rank. We then provide a solution method based on sphere decoding that applies generally to (1.2).

In Chapter 6, we focus on the use of lattice theory in the analysis and design of nonuniform sensor arrays in one, two, and three dimensions. Nonuniform array geometries are often used to extend the array aperture or to reduce mutual coupling among sensors. The work in this chapter is motivated by the use of nonuniform arrays in the direction of arrival (DoA) estimation problem. We give a necessary and sufficient condition for an array of isotropic sensor elements to be unambiguous for any specified set of DoAs. The set of unambiguous DoAs for a given array configuration is shown to be a parallelotope, a property that admits simple geometric interpretation and facilitates sensor array design. The result is an easily tested necessary and sufficient condition for arbitrary sets of DoAs, an alternative to a widely used sufficient

condition [33, Thm. 8]. We include additional properties in this chapter to aid in the design of nonuniform arrays that meet aliasing requirements.

Chapter 2: Speckle Suppression in Multichannel SAR Imaging

2.1 Introduction

An issue endemic to coherent imaging is the presence of speckle, a phenomenon which manifests as a granular quality in images and may obscure features of interest.[1] Speckle is the result of the coherent interaction of multiple reflections from within a resolution cell. It is common to account for the presence of speckle by modeling the scene reflectivity stochostically, canonically as a zero-mean Gaussian-distributed random variable [13, 14].

Also of concern for image formation and interpretation are gaps in the SAR coherent phase history, which may be caused by radio frequency jamming or interruptions due to hardware multiplexing; spectral gaps result in ringing artifacts which hinder image interpretation. The objective of this work is to estimate the multichannel scattering properties of a scene afflicted by speckle when the collection suffers from gaps in spatial frequency observations. We hypothesize that directly addressing speckle and ringing artifacts in a multichannel image formation procedure can produce results

[1] This chapter is largely drawn from [34], coauthored with Lee C. Potter ().

superior to the existing data flow of channel-by-channel imaging followed by speckle-reducing post-processing. To this end, we adopt a maximum a posteriori (MAP) estimation approach to recover the per-pixel multichannel SAR covariance with dual goals of speckle reduction and preservation of spatial resolution. Rather than placing a prior directly on the covariance, we employ a latent terrain class random vector on which a Markov random field (MRF) prior is placed. To accommodate texture within a terrain class, the per-pixel covariance within a class is assumed to be distributed as a complex inverse Wishart (CIW) matrix. To alleviate computational complexity, we use an expectation-maximization (EM) surrogate function for the MAP objective function and a probing method for estimating the block-diagonal of a matrix inverse.

2.1.1 Motivating Applications

The proposed approach is motivated by SAR imaging modalities in which data are collected via multiple channels, such as PolSAR and InSAR. Multichannel SAR data have been shown to be considerably more informative than single-channel collections because the relationships between channels capture information corresponding to physical scattering behavior [10, 11]. The information that these multichannel systems gather has proven critical for monitoring changes on the earth's surface, and numerous techniques employ these data for terrain classification and surface parameter estimation [12].

A common multichannel SAR modality is PolSAR, in which diversity of the polarizations of the transmitted and received signals provides more information about the environment than would a single-channel sensor. A pair of orthogonal polarizations are used (e.g., horizontal and vertical polarizations) to emit and receive energy,

the permutations of which yield a reflectivity vector of four complex values for each pixel. The experimental results presented here use simulated and measured PolSAR data, although the proposed approach applies to other SAR modalities and, indeed, to other multichannel coherent imaging tasks.

2.1.2 Ameliorating Speckle by Post-Processing

The existing data flow is to first form single-channel SAR images, then post-process to reduce speckle before image interpretation [16]. The speckle suppression step is frequently realized by the application of a speckle filter, classically one which acts on the image locally [17]. The simplest such despeckling technique is a boxcar filter, which suppresses speckle effectively in homogeneous areas but degrades the resolution and the fine structural details of the image [16]. Further local filtering methods include the Lee [35] and Refined Lee [36] filters, which have, among other classical PolSAR speckle filtering techniques, been reviewed and analyzed by Foucher and López-Martínez [17].

In recent years, there has been a trend toward non-local patch-based speckle reduction based on the non-local means (NLM) filter [37]. Unlike the local filtering methods, the NLM ethos is based on the nonlocal averaging of pixels across the image. Applications of this principle to PolSAR despeckling have demonstrated robust performance [38–40].

There are also many variational PolSAR speckle reduction methods, in which the underlying idea is to suppress speckle by optimizing an energy functional consisting of a regularization term and a data fidelity term. Some such methods reduce complexity

by operating only on the diagonal elements of the covariance [41], ignoring the statistical correlations between channels. In [42], a scheme for multichannel SAR speckle reduction is proposed which maps multichannel SAR covariance images to real-valued images using a homomorphic transform, allowing for the use of general-purpose denoising algorithms. In [43], a MAP PolSAR despeckling approach is proposed which uses a matricial total variation (TV) regularization term to enforce smoothness, and this approach is modified in [44] with a nonlocal TV-based regularization term. A review of the recent developments in PolSAR speckle filtering can be found in [16].

2.1.3 Regularized Imaging

The speckle present in coherent imaging is central to the problem formulation in this work. In contrast, many regularized SAR imaging approaches eschew a speckle model and opt to estimate the scene reflectivity, rather than its variance. To combat artifacts due to gaps in the SAR phase history data, many of these techniques enforce sparsity on the magnitude of the reflectivity [45, 46]. For a review of this literature, see [47–49]. Another family of sparsity-based approaches is based on sparse Bayesian learning (SBL) [50–52], in which the reflectivity is modeled as a zero-mean circularly-symmetric Gaussian, and an independent Gamma hyperprior is placed on the variance of each reflectivity element [53].

The proposed approach can also be considered in the context of other coherent image recovery methods that estimate the second moment of the image. In [54], an image recovery framework is proposed for single-channel synthetic aperture LADAR; the approach adopts a statistical model for fully-developed speckle and uses an iterative EM optimization scheme that treats reflectivity as a nuisance parameter. The

variances of the reflectivity elements are found via MAP estimation with a generalized Gaussian MRF prior. In [55], this approach is broadened into a plug-and-play framework [56, 57] which allows for regularization by way of off-the-shelf Gaussian denoising algorithms. In both [54] and [55], the forward model operator is a scaled unitary matrix, a limitation which greatly reduces computational complexity.

Rather than treating reflectivity as a nuisance parameter in the MAP variance estimation as in [54], reflectivity and its variance are jointly estimated in a single-channel SAR image recovery approach in [58]. This approach uses a classical stochastic speckle model and a TV regularization term in a joint recovery approach which also accounts for the deleterious effects of arbitrary gaps in phase history.

The speckle-aware imaging approach proposed here draws motivation from [54, 55] where the variance of the scene reflectivity is estimated in the fully-developed speckle regime. Like [54, 55], our approach considers an image recovery task that jointly forms images and combats speckle. In contrast, the literature for speckle reduction customarily considers despeckling as an image post-processing problem [38, 42, 43]. Three novel components comprise the contribution of the proposed approach. First, we extend the Bayesian modeling and estimation in [54, 55] to the multi-variate case; rather than estimate a variance at each pixel, we estimate a positive semi-definite multichannel covariance matrix. In this way, informative relationships among the channels are exploited to combat both speckle and image side-lobes from data dropouts. Second, we present a novel algorithmic component that overcomes an intractable matrix inversion encountered when extending [54, 55] to a more general class of linear measurement models; in particular, a matrix probing scheme in the new framework allows for the correlation in the random speckle process induced by the

gaps in the phase history data. Finally, we incorporate into the proposed framework a terrain class-based MRF prior which models spatial dependencies in a latent terrain class label.

2.2 Imaging Model

2.2.1 Spotlight-Mode SAR Imaging

In spotlight-mode SAR, high-bandwidth pulses, most often linear frequency modulated (LFM) chirps, are transmitted from the radar system towards the scene of interest, and the reflected pulses recorded. The received signal is mixed with a reference sent through a low-pass filter, resulting in a set of what may be considered slices from the 2-D Fourier transform of the complex scene reflectivity [59].

The single-channel SAR data collection process can be approximated by a linear mapping [59]

$$\boldsymbol{y} = \boldsymbol{A}_m \boldsymbol{x} + \boldsymbol{w}, \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{C}^N$ is the scene reflectivity, $\boldsymbol{y} \in \mathbb{C}^M$ noisy linear measurements, and $\boldsymbol{w} \in \mathbb{C}^M$ complex zero-mean white Gaussian noise with covariance $\sigma_w^2 \boldsymbol{I}_M$; each corresponds to an image ordered lexicographically as a vector, with $\boldsymbol{I}_M$ denoting an $M \times M$ identity matrix. Note that the vector $\boldsymbol{w}$ in (2.1) models additive thermal noise present at the receiver and is not related to the speckle phenomenon. Assuming SAR to be a linear shift-invariant imaging system, the forward operator $\boldsymbol{A}_m$ is characterized by the spatial frequency support of the system [60].

2.2.2 Multichannel SAR

For a SAR system with D channels, the reflectivity vectors at each pixel are

$$\boldsymbol{x}_i = [x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(D)}]^{\mathsf{T}}, \qquad i = 1, 2, \ldots, N, \tag{2.2}$$

which compose the scene reflectivity vector $\boldsymbol{x} \in \mathbb{C}^{DN}$ as

$$\boldsymbol{x} = [\boldsymbol{x}_1^{\mathsf{T}}, \boldsymbol{x}_2^{\mathsf{T}}, \ldots, \boldsymbol{x}_N^{\mathsf{T}}]^{\mathsf{T}} \tag{2.3}$$

where $(\cdot)^{\mathsf{T}}$ denotes transposition. The multichannel measurement and thermal noise vectors $\boldsymbol{y}$ and $\boldsymbol{w}$ are likewise formed by stacking their constituents.

A multichannel SAR linear system model relates the scene reflectivity to the observations as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}, \tag{2.4}$$

in which the multichannel SAR forward operator $\boldsymbol{A} \in \mathbb{C}^{DM \times DN}$ is

$$\boldsymbol{A} = \boldsymbol{A}_m \otimes \boldsymbol{I}_D, \tag{2.5}$$

where $\otimes$ is the Kronecker product operator.

2.2.3 Modeling Speckle in Multichannel SAR

Speckle in SAR imaging is a result of the coherent interaction of reflections from objects within a resolution cell. When illuminating a homogeneous medium that is rough on the scale of the radar's central frequency, the classical reflectivity model is [13, 14]:

$$p(\boldsymbol{x}_i \mid \boldsymbol{\Sigma}_i) = \mathcal{CN}(\boldsymbol{x}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_i), \qquad i = 1, 2 \ldots, N \tag{2.6}$$

where $\mathcal{CN}(\boldsymbol{x}_i; \mathbf{0}, \cdot)$ denotes a complex zero-mean circularly-symmetric Gaussian distribution [15] defined by:

$$\mathcal{CN}(\boldsymbol{x}_i; \mathbf{0}, \boldsymbol{\Sigma}_i) = \frac{1}{\pi^D |\boldsymbol{\Sigma}_i|} \exp\left\{-\boldsymbol{x}_i^{\mathsf{H}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{x}_i\right\}, \tag{2.7}$$

in which $|\cdot|$ denotes a determinant. Note that this stochastic modeling of $\boldsymbol{x}_i$ in (2.6) is to account for the speckle phenomenon in coherent imaging. Under this model, the scattering behavior within a pixel is characterized by the covariance $\boldsymbol{\Sigma}_i \in \mathbb{H}_+^D$, where $\mathbb{H}_+^D$ is the set of $D \times D$ Hermitian positive-definite matrices. This is known as the fully-developed speckle model, and most PolSAR filtering techniques are derived from its assumption.

2.2.4 Markov Random Field Prior

The proposed model stipulates that each pixel belongs to one of ℓ terrain classes (e.g., grass or scrub), a construct similar to the non-homogeneous clutter model detailed in [61]. The terrain class labels are unknown, but the statistics of each class are assumed to be known *a priori*. The latent random vector $\boldsymbol{z} \in \mathcal{Z} = \{1, 2 \ldots, \ell\}^N$ indicates the terrain class to which each pixel belongs. Given a pixel's terrain class, its covariance is conditionally distributed as a complex inverse Wishart (CIW) distribution [62], defined for $\boldsymbol{\Sigma} \in \mathbb{H}_+^D$ by

$$\mathrm{CIW}(\boldsymbol{\Sigma}_i; \boldsymbol{G}, v) = \frac{|\boldsymbol{\Sigma}_i|^{-(v+D)} \exp\left\{\mathrm{tr}(-\boldsymbol{G}\boldsymbol{\Sigma}_i^{-1})\right\} |\boldsymbol{G}|^v}{\pi^{D(D-1)/2} \prod_{i=1}^{D} \Gamma(v - i + 1)} \tag{2.8}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. The conditional distribution of $\boldsymbol{\Sigma}_i$ is given by

$$p(\boldsymbol{\Sigma_i} \mid z_i) = \begin{cases} \mathrm{CIW}(\boldsymbol{\Sigma}_i; \boldsymbol{G}_1, v_1), & z_i = 1 \\ \quad \vdots & \quad \vdots \\ \mathrm{CIW}(\boldsymbol{\Sigma}_i; \boldsymbol{G}_\ell, v_\ell), & z_i = \ell \end{cases}, \tag{2.9}$$

for $i = 1, 2, \ldots, N$, where $(\boldsymbol{G}_a, v_a)$ are the distribution parameters corresponding to class $a \in \{1, 2, \ldots, \ell\}$. Each z_i is assigned a marginal prior distribution of $p(z_i = j) = \alpha_j$ for $i = 1, 2, \ldots, N$, and $j = 1, 2, \ldots, \ell$ where $\sum_j \alpha_j = 1$. A Markov random field (MRF) prior encodes spatial correlation among neighboring z_i. Specifically, we have, for $\gamma_n(\ell - 1) + \gamma_e = 1$,

$$p(z_i \mid z_j) = \begin{cases} \gamma_n, & z_i \neq z_j \\ \gamma_e, & z_i = z_j \end{cases}$$

as the pair-wise marginal probability of z_i conditional on any of its four neighbors $z_j \in \delta_i$ [63]. The total distribution of $\boldsymbol{z}$ is given by

$$p(\boldsymbol{z}) = \frac{1}{Z} \prod_{i=1}^{N} p(z_i) \prod_{j \in \delta_i} p(z_i \mid z_j) \tag{2.10}$$

where Z is a normalizing term. A factor graph representing this probability model is shown in Fig. 2.1, where, for clarity, the MRF along the bottom of the graph is visualized as a chain. Square nodes are used to represent factors of the distribution and circular nodes denote unknown random variables.

2.3 MAP Estimation

The recovery of the per-pixel covariance from the observations is posed as a MAP estimation problem; i.e., we find

$$\widehat{\boldsymbol{\Sigma}} \in \underset{\boldsymbol{\Sigma} \in \Omega}{\arg\max}\ p(\boldsymbol{\Sigma} \mid \boldsymbol{y}) \tag{2.11}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma}_N \end{bmatrix}, \tag{2.12}$$

with a set of feasible values defined by

$$\Omega = \left\{ \boldsymbol{\Sigma} \ \mid \ \boldsymbol{\Sigma} = \mathrm{diag}_D(\boldsymbol{\Sigma}) \text{ and } \boldsymbol{\Sigma} \in \mathbb{H}_+^{DN} \right\} \tag{2.13}$$

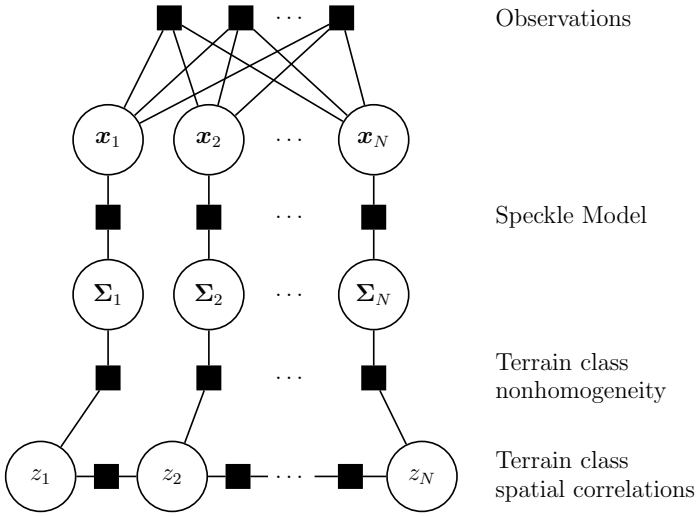


Figure 2.1: Factor graph depicting the conditional probabilities comprising the assumed signal model. The text at the right describes the physical characteristics captured in the model, depicted as factor nodes in the graph.

in which $\text{diag}_D(\boldsymbol{\Sigma})$ denotes a block-diagonal matrix of $D \times D$ blocks whose block-diagonal entries are the same as those of $\boldsymbol{\Sigma}$. From Bayes' theorem, the MAP problem is equivalent to

$$\widehat{\boldsymbol{\Sigma}} \in \underset{\boldsymbol{\Sigma} \in \Omega}{\arg\min} \ \big\{ -\log p(\boldsymbol{y} \mid \boldsymbol{\Sigma}) - \log p(\boldsymbol{\Sigma}) \big\}, \tag{2.14}$$

where [53]

$$p(\boldsymbol{y} \mid \boldsymbol{\Sigma}) = \mathcal{CN}(\boldsymbol{0}, \boldsymbol{T}), \tag{2.15}$$

with

$$\boldsymbol{T} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathsf{H}} + \sigma_w^2 \boldsymbol{I}_{DM}. \tag{2.16}$$

With (2.6), the MAP estimation optimization problem is

$$\widehat{\boldsymbol{\Sigma}} \in \underset{\boldsymbol{\Sigma} \in \Omega}{\arg\min} \ \{\log |\boldsymbol{T}| + \boldsymbol{y}^{\mathsf{H}} \boldsymbol{T}^{-1} \boldsymbol{y} - \log p(\boldsymbol{\Sigma})\}. \tag{2.17}$$

The objective function in (2.17) is non-convex and generally does not admit a solution in closed form [53]. As such, we use a surrogate for the objective function in (2.17) as is proposed in [54].

2.3.1 MAP Objective Surrogate Function

Treating reflectivity $\boldsymbol{x}$ and terrain class $\boldsymbol{z}$ as hidden variables, we form an EM surrogate for the MAP cost in (2.17) parameterized by the previous estimate $\boldsymbol{\Sigma}'$ and given by [54, 64]

$$\begin{aligned} Q(\boldsymbol{\Sigma};\boldsymbol{\Sigma}') &= -\,\mathrm{E}_{\mathsf{x},\mathsf{z}|\mathsf{y},\boldsymbol{\Sigma}'}\left[\log p(\boldsymbol{x},\boldsymbol{\Sigma},\boldsymbol{z}\mid\boldsymbol{y})\right] \\ &= -\,\mathrm{E}_{\mathsf{x},\mathsf{z}|\mathsf{y},\boldsymbol{\Sigma}'}\left[\log p(\boldsymbol{y}\mid\boldsymbol{x}) + \log p(\boldsymbol{x},\boldsymbol{\Sigma},\boldsymbol{z})\right] + \text{const.}, \end{aligned}$$

where we have used the fact that $p(\boldsymbol{y}\mid\boldsymbol{x},\boldsymbol{\Sigma},\boldsymbol{z}) = p(\boldsymbol{y}\mid\boldsymbol{x})$, and const. is used to collect terms which do not depend on $\boldsymbol{\Sigma}$. We then have

$$\begin{aligned} Q(\boldsymbol{\Sigma};\boldsymbol{\Sigma}') &= -\,\mathrm{E}_{\mathsf{x},\mathsf{z}|\mathsf{y},\boldsymbol{\Sigma}'}\Big[\log p(\boldsymbol{y}\mid\boldsymbol{x}) + \log p(\boldsymbol{x}\mid\boldsymbol{\Sigma}) \\ &\qquad + \log p(\boldsymbol{\Sigma}\mid\boldsymbol{z}) + \log p(\boldsymbol{z})\Big] + \text{const.} \\ &= -\,\mathrm{E}_{\mathsf{x},\mathsf{z}|\mathsf{y},\boldsymbol{\Sigma}'}\Big[\log p(\boldsymbol{x}|\boldsymbol{\Sigma}) + \log p(\boldsymbol{\Sigma}\mid\boldsymbol{z})\Big] + \text{const.} \\ &= -\,\mathrm{E}_{\mathsf{x}|\boldsymbol{y},\boldsymbol{\Sigma}'}\Big[\log p(\boldsymbol{x}\mid\boldsymbol{\Sigma})\Big] - \mathrm{E}_{\mathsf{z}|\boldsymbol{\Sigma}'}\Big[\log p(\boldsymbol{\Sigma}\mid\boldsymbol{z})\Big] + \text{const.}, \end{aligned} \tag{2.18}$$

where we have used that $p(\boldsymbol{z}\mid\boldsymbol{\Sigma}',\boldsymbol{y},\boldsymbol{x}) = p(\boldsymbol{z}\mid\boldsymbol{\Sigma}')$. The conditional posterior density $p(\boldsymbol{x}\mid\boldsymbol{y},\boldsymbol{\Sigma}')$ is a complex Gaussian-distributed random vector with mean and covariance [65]

$$\boldsymbol{\mu}_{\boldsymbol{\Sigma}'} = \sigma_w^{-2}\boldsymbol{C}_{\boldsymbol{\Sigma}'}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{y} \tag{2.19a}$$

$$\boldsymbol{C}_{\boldsymbol{\Sigma}'} = \left[\sigma_w^{-2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A} + (\boldsymbol{\Sigma}')^{-1}\right]^{-1}, \tag{2.19b}$$

respectively. With this, (2.18) can be simplified to

$$Q(\boldsymbol{\Sigma};\boldsymbol{\Sigma}') = \log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\boldsymbol{S}_{\boldsymbol{\Sigma}'}\boldsymbol{\Sigma}^{-1}\right) - \mathrm{E}_{\boldsymbol{z}|\boldsymbol{\Sigma}'}\left[\log p(\boldsymbol{\Sigma} \mid \boldsymbol{z})\right] + \text{const.}, \tag{2.20}$$

where $\boldsymbol{S}_{\boldsymbol{\Sigma}'} = \boldsymbol{C}_{\boldsymbol{\Sigma}'} + \boldsymbol{\mu}_{\boldsymbol{\Sigma}'}\boldsymbol{\mu}_{\boldsymbol{\Sigma}'}^{\mathsf{H}}$. Evaluation of $p(\boldsymbol{z} \mid \boldsymbol{\Sigma}')$ over all possible realizations of $\boldsymbol{z}$ in (2.20) is prohibitively costly. As such, we will approximate $p(\boldsymbol{z} \mid \boldsymbol{\Sigma}')$ as a product of its marginal probabilities; that is, assume $p(\boldsymbol{z} \mid \boldsymbol{\Sigma}') \approx \prod_{i=1}^{N} p(z_i \mid \boldsymbol{\Sigma}')$. In this case, the surrogate in (2.20) is approximated as

$$Q(\boldsymbol{\Sigma};\boldsymbol{\Sigma}') \approx \log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\boldsymbol{S}_{\boldsymbol{\Sigma}'}\boldsymbol{\Sigma}^{-1}\right) - \sum_{i=1}^{N}\sum_{k=1}^{\ell} p(z_i = k|\boldsymbol{\Sigma}')\log p(\boldsymbol{\Sigma}_i \mid z_i) + \text{const.} \tag{2.21}$$

The marginal distributions $p(z_i \mid \boldsymbol{\Sigma}')$, for $i = 1, 2, \ldots, N$, are approximated using the sum-product algorithm [65, 66]. The function in (2.21) decouples by pixel, and the ith block of its minimizer is

$$\widehat{\boldsymbol{\Sigma}}_i \triangleq \underset{\boldsymbol{\Sigma}_i \in \mathbb{H}_+^D}{\arg\min}\ \log|\boldsymbol{\Sigma}_i| + \operatorname{tr}(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{S}_{\boldsymbol{\Sigma}'})_i) - \sum_{k=1}^{\ell} p(z_i = k \mid \boldsymbol{\Sigma}')\log p(\boldsymbol{\Sigma}_i \mid z_i) \tag{2.22}$$

$$= \frac{(\boldsymbol{S}_{\boldsymbol{\Sigma}'})_i + \sum_{k=1}^{\ell} p(z_i = k|\boldsymbol{\Sigma}')\boldsymbol{G}_k}{1 + \sum_{k=1}^{\ell} p(z_i = k|\boldsymbol{\Sigma}')(v_k + D)}. \tag{2.23}$$

In (2.23), we have used the fact that the objective in (2.22) has a unique minimizer over $\mathbb{H}_+^D$ [67].

2.3.2 Evaluating the Conditional Posterior Covariance

To evaluate (2.21), it is necessary to compute the conditional posterior covariance matrix $\boldsymbol{C}_{\boldsymbol{\Sigma}'}$. Its direct, explicit evaluation is both memory-intensive and computationally prohibitive, involving as it does the inversion of a $DN \times DN$ matrix. For scaled unitary $\boldsymbol{A}_m$, $\boldsymbol{C}_{\boldsymbol{\Sigma}'}$ is block-diagonal, making its inversion tractable for practical image sizes. This simplifying assumption, which is made in [54, 55], is met when

imaging sidelobes are not present. To accommodate more general imaging scenarios, first note that we need only compute the block-diagonal of $\boldsymbol{C}_{\boldsymbol{\Sigma}'}$ because $\boldsymbol{C}_{\boldsymbol{\Sigma}'}$ appears in (2.21) only via the trace of its product with the block-diagonal $\boldsymbol{\Sigma}^{-1}$.

With this in mind, we consider computing the D-block-diagonal of the inverse of a Hermitian matrix, adapting an approach from Tang and Saad [68]. Their method exploits sparsity to ease the computational burden of computing the diagonal of the inverse of a symmetric and positive-definite matrix. In describing its adaptation, we adopt generic notation: let $\boldsymbol{B} \in \mathbb{H}_+^{DN}$ denote a known positive-definite Hermitian matrix from which we seek to approximate $\mathrm{diag}_D(\boldsymbol{B}^{-1})$.

Consider $\boldsymbol{B}$ to be composed of blocks as

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1N} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{N1} & \boldsymbol{B}_{N2} & \cdots & \boldsymbol{B}_{NN} \end{bmatrix} \tag{2.24}$$

where $\boldsymbol{B}_{ik} \in \mathbb{C}^{D\times D}$. This method relies on $\boldsymbol{B}^{-1}$ having relatively few "non-trivial" blocks. This is to say that $\mathcal{S}_\epsilon(\boldsymbol{B}^{-1})$ is sparse, where the sparsified block-support operator $\mathcal{S}_\epsilon(\cdot) : \mathbb{C}^{DN\times DN} \to \{0,1\}^{N\times N}$ is defined by

$$(\mathcal{S}_\epsilon(\boldsymbol{B}^{-1}))_{ik} = \begin{cases} 1, & \text{if } \|(\boldsymbol{B}^{-1})_{ik}\|_F^2 > \epsilon \\ 0, & \text{otherwise.} \end{cases} \tag{2.25}$$

for a small constant $\epsilon > 0$. The sparsified $\boldsymbol{B^{-1}}$, for which entries outside the sparsified block-support are set to zero, is denoted by $(\boldsymbol{B}^{-1})_\epsilon$.

To proceed, note the following relation:

$$\mathrm{diag}_D(\boldsymbol{B}^{-1}) = \mathrm{diag}_D(\boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}})\,\mathrm{diag}_D^{-1}(\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}}), \tag{2.26}$$

which holds trivially for orthogonal $\boldsymbol{V}$. In this estimation scheme, we exploit the approximately sparse structure of $\boldsymbol{B}^{-1}$ to construct a tall, thin $\boldsymbol{V} \in \{0,1\}^{DN\times DP}$ for

which (2.26) holds with $P << N$. In [68], $\boldsymbol{V}$ is called a "probing" matrix. Here we decompose the probing matrix $\boldsymbol{V}$ as

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1^{\mathsf{T}} & \boldsymbol{V}_2^{\mathsf{T}} & \cdots & \boldsymbol{V}_N^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \tag{2.27}$$

where

$$\boldsymbol{V}_i = \begin{bmatrix} \boldsymbol{V}_{i1} & \boldsymbol{V}_{i2} & \cdots & \boldsymbol{V}_{iP} \end{bmatrix}, \qquad i = 1, 2, \ldots, N \tag{2.28}$$

and $\boldsymbol{V}_{ij} \in \{0,1\}^{D \times D}$. The construction of the probing matrix rests on the following proposition, stated in [68] for $D = 1$:

Proposition 1. *Given Hermitian positive-definite $\boldsymbol{B} \in \mathbb{H}_+^{DN}$, let $\boldsymbol{V} \in \{0,1\}^{DN \times DP}$ be full-rank with $P \leq N$, then*

$$\operatorname{diag}_D(\boldsymbol{B}^{-1}) = \operatorname{diag}_D(\boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}}) \operatorname{diag}_D(\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}}) \tag{2.29}$$

holds if $\boldsymbol{V}_i\boldsymbol{V}_k^{\mathsf{T}} = \boldsymbol{0}$ for all (i,k) for which $i \neq k$ and $(\boldsymbol{B}^{-1})_{ik} \neq \boldsymbol{0}$.

A probing matrix which satisfies Proposition 1 can be formed by first constructing a graph in which each node corresponds to a pixel in the image, and the corresponding adjacency matrix is the sparse support of $\boldsymbol{B}^{-1}$. We are able to form a suitable $\boldsymbol{V}$ by coloring the nodes of this graph. Graph coloring is known to be an NP-hard problem, and, as is common practice, we rely on a heuristic to find a graph coloring with few colors. A well-known greedy graph coloring algorithm is summarized in Algorithm 1 [68], and further details on graph coloring algorithms can be found in [69].

Algorithm 1 Greedy Graph Coloring Algorithm

Input: Graph corresponding to $N \times N$ matrix
Output: Colors corresponding to nodes of the graph

for $i = 1$ **to** N **do**
  Set $\text{Color}(i) = 0$
end for
for $i = 1$ **to** N **do**
  Set $\text{Color}(i) = \min\{k > 0 \mid k \neq \text{Color}(l)\ \forall\ l \in \text{Adjacent}(i)\}$
end for
return Color

As is done for the estimation of the diagonal of a matrix inverse in [68], a probing matrix can be formed from the colored nodes. Specifically, $\boldsymbol{V}$ is constructed as

$$\boldsymbol{V}_{ik} = \begin{cases} \boldsymbol{I}_D, & \text{Color}(i) = k \\ \boldsymbol{0}, & \text{otherwise} \end{cases}. \tag{2.30}$$

In general the support of $\boldsymbol{B}^{-1}$ is not known, and it must be estimated from the known $\boldsymbol{B}$. The support of the sparsified inverse can be estimated as

$$\mathcal{S}_\epsilon(\boldsymbol{B}^{-1}) \approx \mathcal{S}_\epsilon(\boldsymbol{B}^q) \tag{2.31}$$

where $q \in \mathbb{N}$ is small [68].

Next, the matrix $\boldsymbol{G} = \boldsymbol{B}^{-1}\boldsymbol{V}$ is computed by solving the linear systems

$$\boldsymbol{B}\boldsymbol{g}_i = \boldsymbol{v}_i, \qquad i = 1, 2, \ldots, DP, \tag{2.32}$$

from which we get $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_{DP}]$. The linear systems in (2.32) can by solved efficiently and in parallel by a Krylov subspace method such as the preconditioned conjugate gradient method. Finally, the desired output is found as $\text{diag}_D(\boldsymbol{B}^{-1}) \approx \text{diag}_D(\boldsymbol{G}\boldsymbol{V}^\mathsf{T})$. The estimation procedure detailed above is summarized in Algorithm 2.

Algorithm 2 Matrix Inverse Block-Diagonal Estimation Probing Scheme

Input: $\boldsymbol{B} \in \mathbb{H}_+^{DN}$

Output: $\mathrm{diag}_D((\boldsymbol{B}^{-1})_\epsilon)$

1: Color the graph associated with adjacency matrix $\mathcal{S}_\epsilon(\boldsymbol{B}^q)$ using Algorithm 1
2: Construct $\boldsymbol{V}$ per (2.30)
3: **for** $i = 1$ **to** DP **do**
4: Solve $\boldsymbol{B}\boldsymbol{g}_i = \boldsymbol{v}_i$ (e.g. with the preconditioned conjugate gradient method)
5: **end for**
6: Set $\boldsymbol{G} \leftarrow [\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{DP}]$
7: Set $\mathrm{diag}_D((\boldsymbol{B}^{-1})_\epsilon) \leftarrow \mathrm{diag}_D(\boldsymbol{G}\boldsymbol{V}^{\mathsf{T}})$
8: **return** $\mathrm{diag}_D((\boldsymbol{B}^{-1})_\epsilon)$

This method relies on coloring a graph corresponding to the sparse support of $\boldsymbol{B}^{-1}$ with few colors. Physically, the P distinct colors used to label the graph partition the image into regions for which the speckle decorrelates, thereby averting the complexity of computing the full conditional posterior covariance. Rather than inverting the $DN \times DN$ matrix in (2.19b), the graph coloring with P colors allows us to compute the block-diagonal elements of the inverse by solving DP linear systems, where P is much smaller than N.

2.3.3 Summary of MAP Estimation Approach

Beginning with the matched filter estimate of reflectivity $\boldsymbol{x}$,

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathsf{H}}\boldsymbol{y}, \tag{2.33}$$

an unbiased initial estimate of $\boldsymbol{\Sigma}_i$ is given by

$$\widetilde{\boldsymbol{\Sigma}}_i = \frac{\widehat{\boldsymbol{x}}_i\widehat{\boldsymbol{x}}_i^{\mathsf{H}}}{\sum_k |(\boldsymbol{A}_m^{\mathsf{H}}\boldsymbol{A}_m)_{ki}|^2} - \sigma_w^2\boldsymbol{I}_D\sum_k |(\boldsymbol{A}_m)_{ki}|^2, \tag{2.34}$$

for $i = 1, 2, \ldots, N$. We observe improved performance by applying a small diagonal loading to the unbiased estimate and re-scaling the off-diagonal elements as

$$(\widetilde{\boldsymbol{\Sigma}}_i)_{kj}^{(\text{new})} \leftarrow \frac{|\sum_l h_{i,l}(\widetilde{\boldsymbol{\Sigma}}_l)_{kj}|}{\sqrt{\sum_l h_{i,l}(\widetilde{\boldsymbol{\Sigma}}_l)_{kk} \sum_l h_{i,l}(\widetilde{\boldsymbol{\Sigma}}_l)_{jj}}}(\widetilde{\boldsymbol{\Sigma}}_i)_{kj} \tag{2.35}$$

where $h_{i,l}$ are the weights of a Gaussian kernel with a bandwidth of one pixel [42].

The iterative EM scheme for approximating a solution to (2.17) is summarized in Algorithm 3.

Algorithm 3 Iterative MAP Estimation Scheme

Input: $\boldsymbol{y}, \boldsymbol{A}, \sigma_w, N_s, \epsilon_s, \{(\boldsymbol{G}_i, v_i, \alpha_i)\}_{i=1}^{\ell}, \gamma$
Output: $\widehat{\boldsymbol{\Sigma}}$

1: Set $\widehat{\boldsymbol{x}} \leftarrow \boldsymbol{A}^{\mathsf{H}}\boldsymbol{y}$
2: **for** $i = 1$ **to** N **do**
3: Set $\widetilde{\boldsymbol{\Sigma}}_i \leftarrow \frac{\widehat{\boldsymbol{x}}_i\widehat{\boldsymbol{x}}_i^{\mathsf{H}} - \sigma_w^2 \boldsymbol{I}_D \sum_k |(\boldsymbol{A}_m)_{ki}|^2}{\sum_k |(\boldsymbol{A}_m^{\mathsf{H}}\boldsymbol{A}_m)_{ki}|^2}$
4: Set $(\boldsymbol{\Sigma}_i)_{kj}^{(0)} \leftarrow \frac{|\sum_l h_{il}(\widetilde{\boldsymbol{\Sigma}}_l)_{kj}|}{\sqrt{\sum_l h_{il}(\widetilde{\boldsymbol{\Sigma}}_l)_{kk} \sum_l h_{il}(\widetilde{\boldsymbol{\Sigma}}_l)_{jj}}}(\widetilde{\boldsymbol{\Sigma}}_i)_{kj}$
5: Apply diagonal loading to $(\boldsymbol{\Sigma}_i)^{(0)}$
6: **end for**
7: Set $k \leftarrow 0$
8: **while** $k < N_s$ and $\epsilon < \epsilon_s$ **do**
9: Compute $\boldsymbol{\mu}_{\boldsymbol{\Sigma}'} = \sigma_w^{-2}\boldsymbol{C}_{\boldsymbol{\Sigma}'}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{y}$
10: Estimate $\text{diag}_D(\boldsymbol{C}_{\boldsymbol{\Sigma}^{(k)}})$ with Algorithm 2
11: Set $\text{diag}_D(\boldsymbol{S}_{\boldsymbol{\Sigma}^{(k)}}) \leftarrow \text{diag}_D(\boldsymbol{C}_{\boldsymbol{\Sigma}^{(k)}} + \boldsymbol{\mu}_{\boldsymbol{\Sigma}^{(k)}}\boldsymbol{\mu}_{\boldsymbol{\Sigma}^{(k)}}^{\mathsf{H}})$
12: **for** $i = 1$ **to** N **do**
13: Estimate $p(z_i|\boldsymbol{\Sigma}^{(k)})$ using the sum-product algorithm
14: $\boldsymbol{\Sigma}_i^{(k+1)} \leftarrow \frac{(\boldsymbol{S}_{\boldsymbol{\Sigma}^{(k)}})_i + \sum_{q=1}^{\ell} p(z_i=q|\boldsymbol{\Sigma}^{(k)})\boldsymbol{G}_q}{1+\sum_{q=1}^{\ell} p(z_i=q|\boldsymbol{\Sigma}^{(k)})(v_q+D)}$
15: **end for**
16: Set $\epsilon \leftarrow \frac{\|\boldsymbol{\Sigma}^{(k+1)} - \boldsymbol{\Sigma}^{(k)}\|_F}{\|\boldsymbol{\Sigma}^{(k)}\|_F}$
17: Set $k \leftarrow k + 1$
18: **end while**
19: **return** $\boldsymbol{\Sigma}^{(k)}$

The complexity of this iterative MAP estimation scheme in Algorithm 3 is dominated by the solving of linear systems in step 10. If we use the preconditioned

conjugate gradient (PCG) method and assume that the matrix vector product in (2.32) can be computed with NUFFT-based methods [70], step 10 requires

$$\mathcal{O}(\eta D^2 NP \log(DN)) \tag{2.36}$$

flops, where η is a constant which depends on the conditioning of $\boldsymbol{C}_{\boldsymbol{\Sigma}^{(k)}}$ [71, 72]. The benefits of this method rely on P being much smaller than N, which is the case when $\boldsymbol{C}_{\boldsymbol{\Sigma}^{(k)}}$ is approximately sparse and its corresponding graph requires relatively few colors. Computing $\boldsymbol{C}_{\boldsymbol{\Sigma}^{(k)}}$ explicitly using Gauss-Jordan elimination requires $\mathcal{O}(D^3N^3)$ flops and is generally non-feasible. If $P \ll N$, Algorithm 3 presents significant reductions in complexity, e.g. in the experiments that follow, we have that $P/N \approx 1.4 \times 10^{-3}$. To provide an indication of the computing burden in practice, note that an image chip in the measured data experiments of Section IV required 190.7 seconds of computation time in the MATLAB (R2018a) environment with an 3.20GHz Intel(R) Core(TM) i5-6500 CPU.

The proposed method relies on a terrain class-based prior to express prior beliefs about the structure of the scene. For such a prior it is necessary to have knowledge of the types of terrain present in the scene and their statistical properties. As such, this method requires training data.

In experiments with simulated data, we can define the statistical properties of terrain classes and ensure that the simulated data adheres to them. In application to measured data, it is necessary to estimate the statistics of the terrain classes through auxiliary data.

2.4 Experimental Results

This section provides an evaluation, using both synthetic and measured PolSAR data, of the proposed method and its performance compared to conventional processing chains. The conventional processing chains used for comparison involve applying two despeckling approaches to $\widetilde{\mathbf{\Sigma}}$, the biased empirical covariance matrix computed from the matched filter images formed channel by channel. The first, a simple yet commonplace boxcar filtering approach, is a convolutional filter with a small rectangular kernel (here a 3×3 window) of equal weights [73]. The other is MuLoG (MUltichannel LOgarithm with Gaussian denoising), a state-of-the-art multichannel SAR despeckling procedure that facilitates the use of Gaussian denoising algorithms on multichannel SAR data and has demonstrated impressive despeckling performance [42]. In each experiment, we use the default MuLoG parameters in the provided code with block-matching and 3D filtering (BM3D) [74] as the embedded Gaussian denoiser. The BM3D denoising algorithm has shown remarkable results in additive Gaussian noise removal from natural images, and performed notably well when embedded in the MuLoG framework in [42].

2.4.1 Evaluation of Algorithm Performance

The two properties we consider desirable in coherent image recovery are speckle reduction and preservation of spatial resolution. To evaluate the degree of speckle reduction we use the conventional equivalent number of looks (ENL), defined by [75]:

$$\mathrm{ENL} = \frac{\mu^2}{\sigma^2}, \tag{2.37}$$

where the empirical mean μ and empirical standard deviation σ are computed within a homogeneous region. A higher ENL indicates greater speckle reduction. As ENL

is a single channel metric, we compute it on the intensity (the squared amplitude) image for each channel.

To measure edge preservation, we use edge-preservation degree based on the ratio of average (EPD-ROA), defined by

$$\text{EPD-ROA} = \frac{\sum_{i=1}^{K} |I_{d1}(i)/I_{d2}(i)|}{\sum_{i=1}^{K} |I_{s1}(i)/I_{s2}(i)|}, \tag{2.38}$$

in which I_{d1} and I_{d2} are adjacent intensity values of a despeckled image along a particular direction, and I_{s1} and I_{s2} are the corresponding adjacent intensity values of a reference image [76]. EPD-ROA values close to one indicate good edge preservation. As with ENL, we measure the EPD-ROA of each channel's intensity. The EPD-ROA is measured in two orientations, one considering the immediate horizontal neighbor and one the vertical.

Additionally, a loss metric is used to measure fidelity to the simulated data: the normalized root mean square error (NRMSE), defined by

$$\text{NRMSE}(\widehat{\boldsymbol{\Sigma}}; \boldsymbol{\Sigma}) = \frac{\|\widehat{\zeta}\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F}{\|\boldsymbol{\Sigma}\|_F} \tag{2.39}$$

for estimate $\widehat{\boldsymbol{\Sigma}}$, reference $\boldsymbol{\Sigma}$, and least-squares scaling constant $\widehat{\zeta}$ found as

$$\widehat{\zeta} = \arg\min_{\zeta} \left\{ \|\zeta\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F \right\} \tag{2.40}$$

2.4.2 PolSAR data

Full PolSAR data comprise four complex channels corresponding to the pairs of transmission/reception modes. Assuming a basis of horizontal and linear polarizations, each pixel then has four reflectivity elements $x_i^{HH}, x_i^{HV}, x_i^{VH}, x_i^{VV}$, where H and V refer to the horizontal and vertical polarization states, respectively. The

first superscript character refers to the transmitting polarization and the second to the receiving polarization. It is common to assume a reciprocal medium in which $x_i^{HV} = x_i^{VH}$, meaning that each pixel can be represented by just three reflectivity elements [77]. The reflectivity at pixel i is then

$$\boldsymbol{x}_i = \left[x_i^{(1)}, \quad x_i^{(2)}, \quad x_i^{(3)}\right]^{\mathsf{T}}, \tag{2.41}$$

the elements of which correspond to the HH, HV, and VV channels, respectively. To visualize the experimental results, we use the span, a measure of total backscattering power defined for fully polarimetric data by [78]

$$\text{span}(\boldsymbol{\Sigma}_i) = (\boldsymbol{\Sigma}_i)_{11} + 2(\boldsymbol{\Sigma}_i)_{22} + (\boldsymbol{\Sigma}_i)_{33}. \tag{2.42}$$

2.4.3 Simulated Data Results

In this section, we generate simulated PolSAR data and use it to evaluate the performance of the proposed framework. In our simulations, we begin with focused SAR imagery and use a clustering method to segment the image into ℓ terrain classes, yielding a terrain class labeling vector $\boldsymbol{z} \in \{1, 2, \ldots, \ell\}^N$. Within each class, the mean covariance is determined and used as the scale parameter $\boldsymbol{G}_i$ for that class. From the labeling $\boldsymbol{z}$ and the terrain class distribution parameters which have been determined from the data, a ground truth covariance can be had by drawing a realization from the CIW distribution corresponding to each pixel. In this manner, measured SAR data are used to generate realistic simulated data which adhere to the proposed terrain class model. From the ground truth covariance, the reflectivity can be drawn as a realization of the random vector

$$\boldsymbol{x}_i \sim \mathcal{CN}(\boldsymbol{x}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_i), \qquad i = 1, 2, \ldots, N. \tag{2.43}$$

Note that the image reflectivity at pixel i is a random quantity by virtue of the speckle model, and per Goodman [13],a Gaussian distribution is assumed. The imaging process including additive thermal noise is then simulated according to (2.4), yielding a set of noisy linear measurements.

Dual-pol SAR Simulated Data Result

In a first experiment, dual-pol SAR data collected by the Sentinel-1 mission [79] are segmented into $\ell = 3$ terrain classes using k-means clustering, providing a ground truth terrain class labeling. Within each terrain class, the mean covariance is determined and assigned as the scale parameter $\boldsymbol{G_i}$ for that class with a degree of freedom $v_i \in [4, 10]$. A ground truth covariance is then drawn as a realization of $\boldsymbol{\Sigma}_i \sim \mathrm{CIW}(\boldsymbol{\Sigma}_i; \boldsymbol{G}_{z_i}, v_{z_i})$, and speckle is simulated by drawing a realization of $\boldsymbol{x}_i \sim \mathcal{CN}(\boldsymbol{x}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_i)$, for $i = 1, 2, \ldots, N$. Finally, the linear observations are simulated according to (2.4) with thermal noise added to simulate a signal-to-noise ratio (SNR) of 16 dB. The synthetic aperture is reduced to simulate a scenario in which 14.1% of pulses are unobserved in the pattern $\{+9.0\%, -6.3\%, +43.8\%, -7.8\%, +33.2\%\}$, where $+$ denotes observations and $-$ gaps, an interruption pattern to be repeated in each subsequent experiment.

A matched filter estimate of $\boldsymbol{x}$ is found as in (2.33), from which a biased estimate of $\boldsymbol{\Sigma}$ is determined as in (2.34,2.35). This serves both as an initial estimate for the proposed iterative method and as an input to despeckling. In this section, we compare a conventional SAR processing chain (i.e. the despeckling an empirical covariance from conventionally-formed PolSAR imagery) to the proposed iterative MAP estimation scheme.

The ground truth span of our simulated dual-pol SAR image is depicted in Fig. 2.2a. The biased empirical covariance estimate from the linear measurements is shown in Fig. 2.2b; the result of its processing by boxcar despeckling is in Fig. 2.2c and by MuLog with BM3D in Fig. 2.2d.

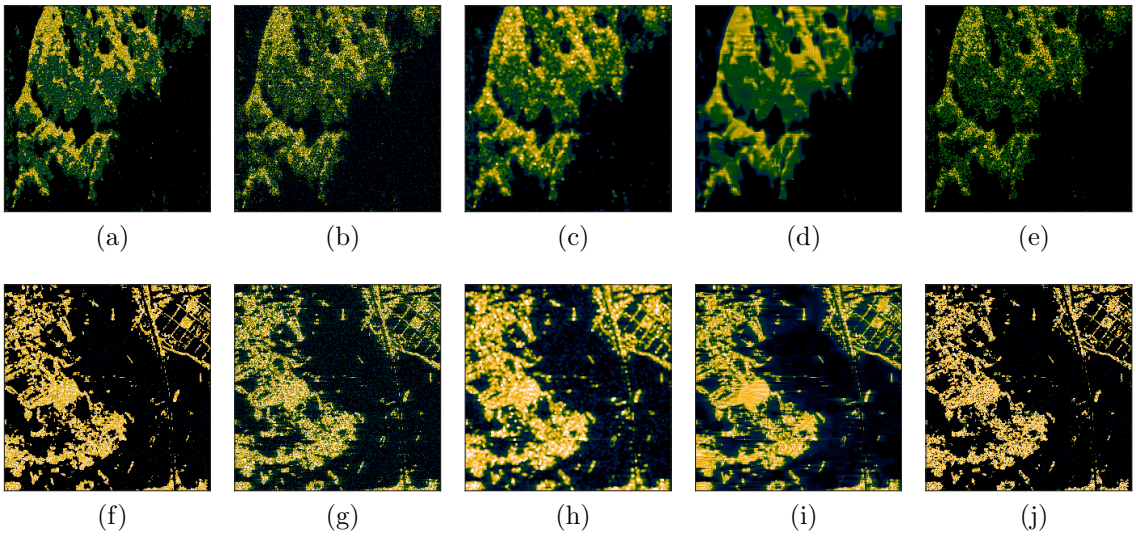


Figure 2.2: Simulated SAR image estimation results: (a,f) the span of the ground truth covariance, (b,g) the single-look empirical covariance, (c,h) the boxcar-filtered estimate, (d,i) the MuLoG-filtered estimate, and (e,j) the MAP estimate with the top row involving dual-pol data and the bottom fully-polarimetric data.

The MAP estimate by the proposed method is found by using Algorithm 3. In our experiments we use equal marginal prior probabilities for each terrain class. For the conditional marginal probabilities of the terrain class variables, we use $\gamma_e/\gamma_n = 1.5$, a ratio which was found to work well and is used in each subsequent experiment. The preconditioned conjugate gradient method is used to solve the linear systems in Algorithm 2, and 10 iterations of the sum-product algorithm are used to estimate $p(z_i \mid \mathbf{\Sigma}'_i)$, for $i = 1, 2, \ldots, N$, in step 13 of Algorithm 3. The span of the resulting

Table 2.1: Recovery metrics from simulated fully polarimetric SAR data experiment.

Method	NRMSE	ENL	EPD-ROA (H)	EPD-ROA (V)
Boxcar	0.928	5.153	0.1305	0.1170
MuLoG	0.874	112.399	0.1663	0.1610
MAP	0.736	33.4412	0.687	0.6349

estimate is in Fig. 2.2e, and the recovery metrics for each method are shown in Table 2.1 where the ENL and EPD-ROA measurements have been averaged across channels.

Degraded spatial resolution is readily apparent in the boxcar-filtered result, paired with meager despeckling performance. The MuLoG-filtered results are spatially smooth and evince impressive despeckling performance that is both visually-apparent and supported by its ENL. The ENL measured on the MAP result is more than ten times that of the boxcar-processed image, indicating strong speckle suppression, but not to the degree supplied by MuLoG; yet, the image texture qualitatively appears to be better preserved by the MAP approach than by either filtering method. As the ground truth is not itself homogeneous, even within a terrain class, ENL alone renders an incomplete portrait of speckle suppression. Note as well that the MAP approach fares better in NRMSE than the filtering approaches, and in edge-preservation, it yields EPD-ROA values significantly closer to unity than either filtering approach in both the orientations measured.

Fully Polarimetric SAR Simulated Data Experiment

This simulation procedure is repeated using fully polarimetric SAR data from the Pi-SAR sensor [80]. Here two terrain classes are used with equal marginal prior

probabilities on each terrain class and $\gamma_e/\gamma_n = 1.5$. The span images of the results from the various methods from this fully polarimetric simulation are shown in Fig. 2.2, and the recovery metrics are listed in Table 2.1. The quantitative results in this case are similar to those from the dual-pol data: the MAP estimate provides speckle reduction with superior edge preservation and reduced reconstruction error.

2.4.4 Measured Data Results

The proposed approach is next tested on measured PolSAR data collected by the Advanced Detection Technology Sensor (ADTS), an airborne Ka band SAR sensor operated by MIT Lincoln Laboratory. Further details and sample imagery from the sensor can be found in [61]. Once again we analyze image recovery performance under a reduced-aperture regime, now in the absence of ground truth covariance for reference.

It is first necessary to estimate the statistics of the terrain classes which comprise the scenes used for this analysis in order to inform the prior model. To do so, we will use a corpus of PolSAR data to learn the distribution parameters for our prior model, which we will then use to test image recovery performance on other data of a similar terrain composition.

In [61], Novak et al. segment ADTS PolSAR images into four region classes and reported the mean covariance of each class. Four types of clutter regions were considered in this analysis: shadows, grass, mixed scrub, and trees. Each bin of the segmentation contained 2000 to 8000 pixels (equivalent to an area of several hundred square meters) [61]. We segment a corpus of twelve 128×128 pixel image chips using a maximum likelihood classifier in which was assume that the terrain classes

Table 2.2: Stockbridge ADTS data recovery metrics.

Method	ENL	EPD-ROA (H)	EPD-ROA (V)
Boxcar	2.581	0.711	0.782
MuLoG	7.207	0.669	0.746
MAP	70.923	0.842	0.883

are CIW-distributed with scale matrices corresponding to the empirical terrain class results reported in [61] and degrees of freedom $v_i = 4$, for $i \in \{1, 2, 3, 4\}$. Once this training data was segmented into the $\ell = 4$ terrain classes, we determine the maximum likelihood CIW parameters for each class. These distribution parameters are then used to inform the prior model used to test recovery performance.

As in the simulated data experiments, the matched filter estimate of the reflectivity is found as in (2.33), from which a biased estimate of the polarimetric covariance is found for each pixel as in (2.34,2.35). Note that, unlike in the simulated data experiments, speckle is not simulated as it is assumed to be present in the data. The biased empirical covariance estimates for two sample images shown in Fig. 2.3b and Fig. 2.3g; the results of two speckle filtering methods, boxcar and MuLoG, are shown alongside them in Fig. 2.3. This procedure is applied to a set of twelve 128×128 images, the mean recovery metrics from which are reported in Table 2.2.

The proposed MAP inference method is carried out on the measured data with $\gamma_e/\gamma_n = 1.5$ and equal marginal prior probabilities on the terrain class. The recovered span images from the proposed inference method are shown in Fig. 2.3e and Fig. 2.3j for two representative image chips. Absent ground truth, we have plotted in Fig. 2.3a and Fig. 2.3f the span of the matched filter-based empirical covariances found as

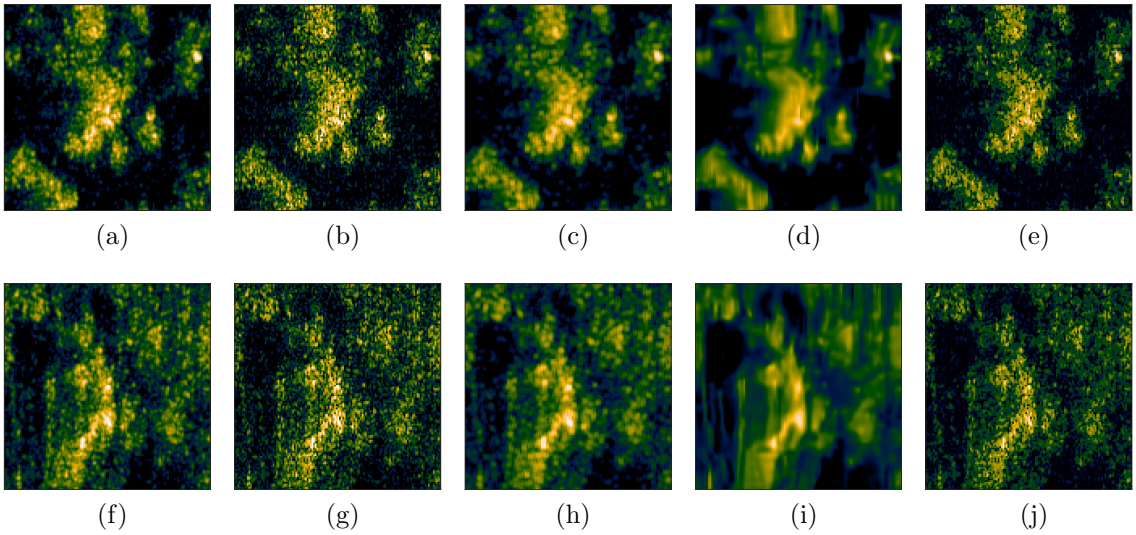


Figure 2.3: Measured fully polarimetric Ka-band SAR data estimation results from two image chips: (a,f) the span of the full-data empirical covariance, (b,g) the single-look empirical covariance, (c,h) the boxcar-filtered estimate, (d,i) the MuLoG-filtered estimate, and (e,j) the MAP estimate. High-quality color graphics of reconstruction results have been made available as supplementary material.

in (2.34) using the uninterrupted measurements, to which spatial averaging has been applied using a 3×3 Gaussian kernel. These filtered full-data images are also used as references to compute the EPD-ROA metrics shown in Table 2.2.

The ENL metrics in Table 2.2, indicate that the proposed approach provides greater speckle suppression in homogeneous areas than the boxcar filter or MuLoG. Note that in the simulated data experiments, it was found that the MuLoG-filtered results had a higher ENL metric than the proposed method.

The EPD-ROA values are evidence of better edge preservation by the MAP approach than either of the filtering techniques. Considering the full-data span images to be imperfect references, the textures seem qualitatively to have been recovered more faithfully by the MAP approach.

2.4.5 Discussion

The proposed Bayesian method provides a principled balance between competing objectives when computing multichannel coherent images from under-sampled data. On one hand, we wish to maintain resolution, reduce imaging side-lobes, and preserve texture, and, on the other hand, we seek to suppress speckle and thermal noise. The proposed hierarchical model provides this balance, while exploiting correlations across channels. For example, the balance with respect to preserving texture arises from the CIW scale matrix: e.g., a larger determinant encourages richer texture at the expense of speckle reduction. In contrast, the existing processing chain performs despeckling that is agnostic to gaps in phase history data, to typical image texture, and to thermal noise power. Traditionally, speckle is ameliorated by averaging: using local sums, as in boxcar or Gaussian filtering, or by summing similar patches found throughout an image, as in BM3D. By tuning the filtering, such as increasing the filter support beyond a 3-by-3 neighborhood, any level of despeckling may be achieved, but at the expense of coarser image resolution and blurred image texture. The balancing of objectives via the Bayesian imaging procedure is evident in the numerical results presented in this section; for example, in the synthetic data experiment reported in Table 2.3, MuLoG provides more image smoothing, and hence a larger ENL, but at the expense of texture and resolution. The specific statistical model presented here uses a mixture model generated from terrain classes; for each class, a CIW scale matrix provides a prior, and the algorithm applies the mixture density without any explicit image segmentation required. The scale matrices are estimated from reference images and applied to mission images. Variations of the proposed framework might also be implemented with an MRF prior model [54], a TV regularization term [43,44],

Table 2.3: Recovery metrics from simulated dual-pol SAR data experiment.

Method	NRMSE	ENL	EPD-ROA (H)	EPD-ROA (V)
Boxcar	0.811	9.148	0.4958	0.535
MuLoG	0.752	504.272	0.457	0.491
MAP	0.682	125.940	0.914	0.925

or other Gibbs priors [81]. Further, an implicit prior could be used in a plug-and-play framework [55, 56], e.g. those in the form of neural network-based denoisers [82–84].

2.5 Conclusion

2.5.1 Summary

We have proposed a new method for multichannel coherent image recovery based on a hierarchical statistical model which not only accommodates a reduced synthetic aperture but also suppresses speckle and thermal noise. The model presented here uses latent terrain class variables with an MRF prior to exploit both structure among coherent imaging channels and spatial correlation for improved image recovery.

The algorithm to produce the MAP estimate based on this model uses a surrogate scheme from [54,55], extending it for use with multiple channels in a reduced aperture regime by the novel application of a matrix inverse block-diagonal estimation scheme. This approach, tested on synthetic and measured PolSAR data, has been shown to suppress speckle and preserve spatial resolution in the recovery of multichannel coherent images.

2.5.2 Future Work

Future work might involve learning of the terrain classes from the noisy data, rather than requiring auxiliary data. The approach in this chapter assumes that the fully developed speckle model is followed throughout the scene. Instead, a mixture model might be used that accounts for areas in which man-made media are present that do not follow the speckle model.

Chapter 3: Coherent Change Detection in SAR Imaging

3.1 Introduction

An important application of SAR is the detection of changes to the complex scene reflectivity that occur between repeated collections of the same terrain.[2] In coherent change detection (CCD), both the magnitude and the phase components of SAR images are used to identify subtle changes in the scene, such as those caused by the movement of vehicles over grass [86], foot traffic [87], or land subsidence [88].

The complex correlation coefficient between a pair of SAR images (denoted time-1 and time-2) at a given pixel with index i is [89]

$$\gamma_i e^{j\phi_i} = \frac{\mathrm{E}\left[y_i^{(1)}\overline{y}_i^{(2)}\right]}{\sqrt{\mathrm{E}\left[|y_i^{(1)}|^2\right]\mathrm{E}\left[|y_i^{(2)}|^2\right]}}, \tag{3.1}$$

where $y_i^{(k)} \in \mathbb{C}$ is the value of pixel i in the time-k image, $\mathrm{E}[\cdot]$ denotes expectation, and $\overline{\ }$ denotes complex conjugation. The interferometric phase ϕ_i in (3.1) depends on the terrain topography as well as the displacement between the flight paths of the two passes. The magnitude of the correlation coefficient, denoted by γ_i and referred to as the coherence, is sensitive to the temporal changes in complex reflectivity that are of interest in CCD. Environmental effects can cause decorrelation, the degree of

[2]This chapter is largely drawn from [85], coauthored with Joshua N. Ash and Lee C. Potter ().

which will depend on the time between passes and the types of scatterers present [86]. The coherence is also influenced by a number of other factors: thermal noise in the SAR system, the co-registration process of the images, the mismatch in acquisition geometries between the two passes, and the presence of volume scatterers in the scene [87, 90].

Change detection in SAR has relied largely on the computation of local statistics within sliding estimation windows, e.g., [91–94]. The coherence, for example, is locally estimated as

$$\widehat{\gamma}_i = \frac{\left|\sum_j y_j^{(1)} \overline{y}_j^{(2)}\right|}{\sqrt{\sum_j \left|y_j^{(1)}\right|^2 \sum_j \left|y_j^{(2)}\right|^2}}, \tag{3.2}$$

in which each summation occurs over a small neighborhood or window of image samples [89, 91]. Man-made disturbances may result in small, isolated fluctuations in the scene coherence, for which a commensurately small estimation window is desirable. However, the use of a small estimation window to localize fine-scale scene changes leads to noisy coherence estimates and poor detection performance [86]. Other window-based CCD methods rely on likelihood ratio tests [93, 95] or involve Stokes parameters [96]. Rather than computing a statistic within a sliding rectangular window, locally-adaptive directional windows are used in [97] for despeckling and coherence estimation.

Various alternatives to window-based coherence estimation have been proposed. In [98], the two-dimensional discrete wavelet packet transform is used to adapt to the spatially nonhomogeneous character of the scene coherence. This approach allows for high resolution coherence estimation that is not biased by the interferometric phase. Nonlocal methods, based on the nonlocal means filter [37], have also been used to estimate both the coherence and the interferometric phase, e.g., the nonlocal

interferometric SAR (NL-InSAR) estimator [99]. Nonlocal methods exploit redundant patterns in the image, selecting pixels to combine in estimation based on the similarity of the patches that surround them. NL-InSAR and similar approaches [100, 101] estimate the coherence in a manner that may mitigate the biases and loss of resolution that are caused by scene heterogeneity in window-based approaches. However, these methods may be ill-suited to change detection scenarios with small, isolated regions of change.

Bayesian methods have been useful in the problem of joint SAR imaging and change detection from interrupted collections [63, 102]. These approaches incorporate prior information about the scene and the physical properties of change into Bayesian models which mitigate the impact of interruptions to the SAR data collections. In both [63, 102], Markov random field (MRF) priors are used to model the spatial sparsity and clustering tendencies of the scene changes. Another class of prior models called gamma MRFs was introduced to model the local dependencies among the variances of time-frequency coefficients of audio sources [103]. Gamma MRFs enforce positive correlations between neighboring variance elements and have the convenient property that they are conditionally conjugate to a Gaussian likelihood with unknown variance.

In this work we propose a Bayesian approach to SAR change detection that enables fine-resolution detection of temporal disturbances between pairs of co-registered SAR images. An MRF prior models the spatial dependencies of the scene changes, which are anticipated to appear in clusters. Gamma MRFs are used to model the spatial dependencies of the unknown mean backscatter powers of the image pair. These prior models are incorporated into a Bayesian framework from which the joint

posterior distribution of the latent variables is estimated using variational expectation maximization (EM) [104, 105], a framework which also allows for the iterative estimation of the model parameters. In addition, a 2-D polynomial regression based on the local phase differences is used to account for the spatially nonhomogeneous interferometric phase.

The proposed Bayesian approach assigns prior distributions to the latent model variables, exploiting spatial structure in the scene changes as well as in the scattering properties of the scene. This approach does not suffer from the loss of spatial resolution and fine-scale details as do conventional window-based SAR CCD methods. The proposed approach has the added benefit that it estimates the prior model hyperparameters from the data, rather than requiring manual tuning. In experiments using both simulated data and measured interferometric SAR images, this approach is shown to yield improved detection performance compared to the conventional window-based statistics used for SAR CCD.

3.2 Background

3.2.1 Variational Inference

Of primary interest in Bayesian inference is the estimation of the posterior distribution of the latent random vector $\boldsymbol{x}$ given the observed random vector $\boldsymbol{y}$. The idea behind variational inference is to approximate the posterior with a distribution $q(\boldsymbol{x})$ that minimizes the Kullback-Leibler (KL) divergence between the two distributions.

The KL divergence between $q(\boldsymbol{x})$ and the posterior distribution is given by [104]

$$D_{\mathrm{KL}}(q \parallel p) = \int q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p(\boldsymbol{x} \mid \boldsymbol{y}\,; \Theta)}\, d\boldsymbol{x} \tag{3.3}$$

$$= \log p(\boldsymbol{y}\,; \Theta) + \int q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p(\boldsymbol{x}, \boldsymbol{y}\,; \Theta)}\, d\boldsymbol{x} \tag{3.4}$$

$$= \log p(\boldsymbol{y}\,; \Theta) - \mathcal{L}(q\,; \Theta), \tag{3.5}$$

where Θ denotes a set of fixed parameters. Note that $\mathcal{L}(q\,; \Theta)$ is a lower bound on the log evidence, $\log p(\boldsymbol{y}\,; \Theta)$. Because the evidence $p(\boldsymbol{y}\,; \Theta)$ is constant with respect to the variational distribution q, maximizing $\mathcal{L}(q\,; \Theta)$ is equivalent to minimizing the KL divergence in (3.4). Optimization of $\mathcal{L}(q\,; \Theta)$ over the space of all candidates for $q(\boldsymbol{x})$ is generally infeasible, and the search is restricted to a family of distributions. This commonly involves a mean field assumption, in which a particular factorization for the variational distribution is postulated, e.g.,

$$q(\boldsymbol{x}) = \prod_{i=1} q_{x_i}(x_i). \tag{3.6}$$

By fixing the other factors, the q_{x_i} which maximizes $\mathcal{L}(q\,; \Theta)$ can be found as [65]

$$\log q_{x_i}(x_i) = \langle \log p(\boldsymbol{y}, \boldsymbol{x}\,; \Theta) \rangle_{q(\boldsymbol{x}_{-i})} + \text{const.} \tag{3.7}$$

where $\langle \cdot \rangle_{f(x)}$ denotes expectation under the probability distribution $f(x)$, and $\boldsymbol{x}_{-i}$ denotes the variables other than x_i. A solution is found by cycling through the factors and updating them according to (3.7), using the current estimates for the other factors.

In variational EM, variational inference approximates the expectation step (E-step) of the EM algorithm [64], and like exact EM, the fixed model parameters are also estimated [105]. The maximization step (M-step) of variational EM for the

update of model parameter Θ_i is given by

$$\arg\max_{\Theta_i} \langle \log p(\boldsymbol{y}, \boldsymbol{x}\,; \Theta) \rangle_{q(\boldsymbol{x})}. \tag{3.8}$$

Note that unlike in exact EM, the variational M-step uses an approximate expectation to perform parameter optimization.

3.2.2 Gamma Markov Random Fields

Gamma MRFs were first used to model the dependencies among the variances of audio signal time-frequency elements [103, 106], making use of the fact that the signal energies often change slowly over both axes. These models have also been used for spectral unmixing problems [107, 108]. In this chapter, we use gamma MRFs to model the local dependencies of the mean backscatter powers of interferometric SAR image pairs in order to improve change detection performance.

A key feature of the gamma MRF defined in [103] is the introduction of auxiliary random variables which do not necessarily correspond to quantities of interest. Instead, these auxiliary variables are present to enforce positive correlations among the primary variables while preserving the conditional conjugacy of the prior model.

A gamma MRF models a joint probability distribution on $\boldsymbol{r} \in \mathbb{R}_+^{|\mathcal{V}|}$ and the auxiliary variables $\boldsymbol{z} \in \mathbb{R}_+^{|\mathcal{V}^z|}$ by defining a bipartite undirected graph between $\boldsymbol{r}$ and $\boldsymbol{z}$, with $\mathcal{V}$ and $\mathcal{V}^z$ denoting sets of indices on which $\boldsymbol{r}$ and $\boldsymbol{z}$ are respectively defined.

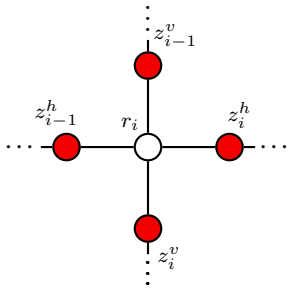


Figure 3.1: The neighborhood structure of a 4-connected gamma MRF showing the vertical and horizontal neighbors for r_i. The auxiliary variables are shown in red.

The joint probability of $\boldsymbol{r}$ and $\boldsymbol{z}$ is expressed as [103]

$$p(\boldsymbol{r}, \boldsymbol{z}\,; \boldsymbol{a}) = \frac{1}{Z(\boldsymbol{a})}\pi(\boldsymbol{r}, \boldsymbol{z}\,; \boldsymbol{a}) \tag{3.9}$$

$$= \frac{1}{Z(\boldsymbol{a})}\prod_{i\in\mathcal{V}} \exp\left\{ -\big(1 + \sum_{j\in S_i^r} a_{i,j}\big)\log r_i \right\} \times \prod_{j\in\mathcal{V}^z} \exp\left\{ (-1 + \sum_{i\in S_j^z} a_{i,j})\log z_j - \sum_{i\in S_j^z} \frac{a_{i,j} z_j}{r_i} \right\}, \tag{3.10}$$

in which $\pi(\cdot\,; \boldsymbol{a})$ denotes the non-normalized probability density, $Z(\boldsymbol{a})$ is a normalizing factor or partition function, and S_i^r and S_j^z contain the indices of the neighboring elements of r_i and z_j, respectively. The hyperparameter $a_{i,j}$ determines the coupling strength between r_i and z_j for $i \in \mathcal{V}$ and $j \in S_i^r$.

The neighborhood structure of a 4-connected gamma MRF is illustrated in Fig. 3.1, in which the neighboring auxiliary nodes (in red) are shown for a given primary variable r_i. The auxiliary variables, denoted $\boldsymbol{z}^h$ (horizontal) and $\boldsymbol{z}^v$ (vertical), are present between the immediately adjoining elements of $\boldsymbol{r}$ in the horizontal and vertical directions.

3.3 Coherent Change Detection

3.3.1 Bayesian Change Detection Model

We assume that the time-1 image of a repeat-pass interferometric SAR image pair conditionally follows a complex circularly-symmetric Gaussian distribution:

$$p(y_i^{(1)} \mid r_i^{(1)}) = \mathcal{CN}(y_i^{(1)}; 0, r_i^{(1)}), \qquad i \in \mathcal{V}, \tag{3.11}$$

where $\mathcal{V}$ is a set of pixel indices on a 2-D planar grid. The latent variable $r_i^{(1)}$ is referred to as the mean backscatter power of the time-1 image at pixel i. The mean backscatter power is a geophysical quantity that is dependent on the structural and electromagnetic properties of the scattering elements within the scene [86]. The model in (3.11) is common for distributed target environments where many statistically identical discrete scatterers occupy a given resolution cell [13, 109].

We next define the conditional distribution on the time-2 image as the Gaussian mixture

$$\begin{aligned} p(y_i^{(2)} \mid y_i^{(1)}, r_i^{(2)}, c_i\,; \phi_i, \gamma) &= \delta(c_i - 1)\mathcal{CN}(y_i^{(2)}; 0, r_i^{(2)}) \\ &+ \delta(c_i + 1)\mathcal{CN}(y_i^{(2)}; \gamma e^{-j\phi_i} y_i^{(1)}, (1-\gamma^2) r_i^{(2)}) \end{aligned} \tag{3.12}$$

where $\delta(\cdot)$ is the Kronecker delta function, ϕ_i is the unknown interferometric phase at pixel i, and $c_i \in \{-1, 1\}$ is a change indicator random variable. Here, $c_i = 1$ indicates a temporal change at pixel i, and $c_i = -1$ indicates the absence of change. The unknown parameter $\gamma \in (0, 1)$ determines the degree of change allowable within the no-change state. Note that (3.12) is a temporal change model similar to that in [94]; (3.12) differs by omitting thermal noise terms and allowing the mean backscatter power to differ between passes.

The interferometric phase $\boldsymbol{\phi}$ is often assumed to be constant over the small neighborhoods that are conventionally used to compute change statistics. However, if $\boldsymbol{\phi}$ is not constant within the estimation window, the sample coherence estimator will be biased [98]. For methods operating on a pixel-wise basis, the phase must be estimated, either by means of external digital elevation models (DEMs) or from the radar data. In this work, we estimate the spatially varying interferometric phase from the radar data by modeling the interferometric phase with a low-order two-dimensional polynomial.

To account for the spatial correlations between the neighboring change indicator variables, we define an MRF prior distribution for $\boldsymbol{c}$. In particular, this chapter focuses on the following prior distribution for the change indicator variables [110]

$$p(\boldsymbol{c} \mid \chi) = \frac{1}{Z(\chi)} \exp\left\{ \sum_{i \in \mathcal{V}} \left(\chi_0 c_i + \frac{\chi}{2} \sum_{j \in S_i^c} c_i c_j \right) \right\}, \tag{3.13}$$

in which χ_0 is assumed to be known *a priori* and S_i^c contains the neighboring elements of c_i for $i \in \mathcal{V}$. Values of $\chi_0 < 0$ enforce the notion that change pixels are less likely than no-change pixels. Larger values of the interaction parameter χ imply greater spatial correlation among the change indication variables, and when $\chi = 0$, the elements of $\boldsymbol{c}$ are mutually independent. In this chapter, we use a 4-connected neighborhood structure for the change indicator variables.

Further, we assign gamma MRF priors to $\boldsymbol{r}^{(1)}$ and $\boldsymbol{r}^{(2)}$ that model the local dependencies of the mean backscatter power within the scene. This quantity may vary across the scene due to errors in the antenna pointing accuracy or spatial variations in the geophysical qualities of the scene. The structure of these spatial variations is meant to be captured by the gamma MRFs. The mean backscatter power of a given pixel may also differ between passes due to radiometric miscalibrations or scene

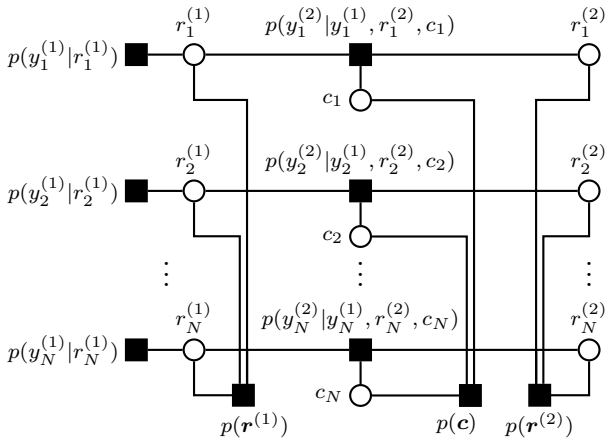


Figure 3.2: Factor graph representation of the proposed change detection model. The squares denote factor nodes and the circles variable nodes.

disturbances that occur between the collections [93]. We assume that the hyperparameters $a_{i,j}$ are uniform across and between both images and that a single hyperparameter a need be learned for the gamma MRFs. We use a 4-connected neighborhood structure for the gamma MRFs as depicted in Fig. 3.1. A factor graph representation of the proposed change detection model is shown in Fig. 3.2, in which square nodes denote factors and round nodes represent variables.

3.3.2 Inference

The joint posterior density of the latent variables is given by

$$\begin{aligned} p(\boldsymbol{r}^{(1)}, \boldsymbol{r}^{(2)}, \boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \boldsymbol{c} \mid \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}\,; \Theta) &\propto p(\boldsymbol{c}\,; \chi) \\ &\times p(\boldsymbol{r}^{(1)}, \boldsymbol{z}^{(1)}\,; a)\, p(\boldsymbol{r}^{(2)}, \boldsymbol{z}^{(2)}\,; a) \\ &\times \prod_{i \in \mathcal{V}} p(y_i^{(1)} \mid r_i^{(1)})\, p(y_i^{(2)} \mid y_i^{(1)}, r_i^{(2)}, c_i\,; \phi_i, \gamma), \end{aligned} \tag{3.14}$$

where $\Theta = \{\gamma, \chi, \boldsymbol{\phi}, a\}$ is the set of model parameters, and $\mathcal{V}$ is the set of pixel indices in the image. While exact computation of the posterior distribution in (3.14)

is computationally prohibitive, the marginal posteriors of the latent variables can be efficiently approximated by using variational EM on the factor graph of Fig. 3.2. In doing so, we use the following mean field approximation for the posterior:

$$q = \prod_{i \in \mathcal{V}} q_{r_i^{(1)}}(r_i^{(1)}) q_{r_i^{(2)}}(r_i^{(2)}) q_{c_i}(c_i) \prod_{j \in \mathcal{V}^z} q_{z_j^{(1)}}(z_j^{(1)}) q_{z_j^{(2)}}(z_j^{(2)}), \tag{3.15}$$

the factors of which are cyclically updated until convergence. Note that, due to the mean-field assumption, the computational complexity of a cycle of factor updates is linear in the number of pixels—$\mathcal{O}(|\mathcal{V}|)$. For brevity, the subscripts of the variational factors may be omitted when clarity is not impacted.

Next we describe the updates to the variational factors in (3.15). From (3.7) and (3.14), the factor associated with $r_i^{(1)}$ is updated as $q(r_i^{(1)}) = \mathcal{IG}(r_i^{(1)}; \alpha_{r,i}^{(1)}, \beta_{r,i}^{(1)})$ with

$$\alpha_{r,i}^{(1)} = 1 + a|S_i^r| \tag{3.16a}$$

$$\beta_{r,i}^{(1)} = |y_i^{(1)}|^2 + \sum_{j \in S_i^r} a\langle z_j^{(1)} \rangle, \tag{3.16b}$$

where $\mathcal{IG}(\cdot\,; \alpha, \beta)$ is the inverse gamma distribution's probability density function defined in Appendix A. The factor associated with $r_i^{(2)}$ is updated as $q(r_i^{(2)}) = \mathcal{IG}(r_i^{(2)}; \alpha_{r,i}^{(2)}, \beta_{r,i}^{(2)})$ with

$$\alpha_{r,i}^{(2)} = 1 + a|S_i^r| \tag{3.17a}$$

$$\beta_{r,i}^{(2)} = \frac{\langle \delta(c_i + 1) \rangle |y_i^{(2)} - \gamma e^{-j\phi_i} y_i^{(1)}|^2}{(1 - \gamma^2)} + \langle \delta(c_i - 1) \rangle |y_i^{(2)}|^2 + \sum_{j \in S_i^r} a\langle z_j^{(2)} \rangle. \tag{3.17b}$$

The time-k auxiliary variable factors for $k \in \{1, 2\}$ are given by

$$q(z_j^{(k)}) = \mathcal{G}(z_j^{(k)}; \alpha_{z,j}^{(k)}, \beta_{z,j}^{(k)}), \tag{3.18}$$

with

$$\alpha_{z,j}^{(k)} = a|S_j^z| \tag{3.19a}$$

$$\beta_{z,j}^{(k)} = \sum_{i \in S_j^z} \left\langle \frac{a}{r_i^{(k)}}, \right\rangle \tag{3.19b}$$

where $\mathcal{G}(\cdot\,; \alpha, \beta)$ is the gamma probability density function defined in Appendix A.

From (3.7) and (3.14), the factors associated with the change bits are updated using

$$\begin{aligned} &\log q(c_i = -1) = -\chi_0 - \log\left(1 - \gamma^2\right) \\ &- \left\langle \frac{|y_i^{(2)} - \gamma e^{-j\phi_i} y_i^{(1)}|^2}{r_i^{(2)}(1-\gamma^2)} \right\rangle - \chi \sum_{j \in S_i^c} \langle c_j \rangle + \text{const.} \end{aligned} \tag{3.20a}$$

and

$$\log q(c_i = 1) = \chi_0 - \left\langle \frac{|y_i^{(2)}|^2}{r_i^{(2)}} \right\rangle + \chi \sum_{j \in S_i^c} \langle c_j \rangle + \text{const.} \tag{3.20b}$$

Further details regarding these updates are in Appendix A.

3.3.3 Parameter Estimation

Now we describe the procedures used to estimate the unknown model parameters. Two of the model parameters, γ and χ, are updated in a straightforward manner via the M-step in (3.8). The estimate for the interferometric phase $\boldsymbol{\phi}$ is updated using a 2-D polynomial regression based on the local phase differences, rather than a conventional M-step update. The gamma MRF hyperparameter governing the local dependencies of the mean backscatter powers is estimated using contrastive divergence [111].

Gamma MRF Hyperparameter

We assume that a single hyperparameter, denoted by a, governs the prior models for the mean backscatter power of both the time-1 and time-2 images. This hyperparameter can then be learned from a single image as a prerequisite to the other parameter estimation or inference tasks. Because variational Bayesian methods have been shown to poorly estimate gamma MRF hyperparameters [112], we use a contrastive divergence approach. Contrastive divergence is an iterative approximate maximum likelihood method that uses Markov chain Monte Carlo (MCMC) samples to approximate the intractable log-partition function [111]. The approach described in Algorithm 4 is similar to one used in [103] for the estimation of the gamma MRF hyperparameters in audio source modeling, though, Algorithm 4 uses variational inference rather than MCMC samples to approximate the expectations in the contrastive divergence gradients. The variational inference procedure used to facilitate this approximation is described in Algorithm 5. This revised approach results in similar performance to [103] in our experiments, though it is faster as it requires fewer MCMC samples.

Interferometric Phase

We adopt a 2-D polynomial model for the local interferometric phase differences and, as in [113], find a vector of polynomial coefficients $\boldsymbol{g}$ by solving a weighted least squares problem:

$$\widehat{\boldsymbol{g}} = \arg\min_{\boldsymbol{g}} \left\| \boldsymbol{P}\left(\begin{bmatrix} \boldsymbol{\phi}_{\Delta 1}^{\mathsf{T}} & \boldsymbol{\phi}_{\Delta 2}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} - \begin{bmatrix} \boldsymbol{B}_{\Delta 1}^{\mathsf{T}} & \boldsymbol{B}_{\Delta 2}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \boldsymbol{g} \right) \right\|_2^2,$$

where $\boldsymbol{\phi}_{\Delta 1}$ and $\boldsymbol{\phi}_{\Delta 2}$ are respectively the differences of the interferometric phase image along the first and second image dimensions; the matrices $\boldsymbol{B}_{\Delta 1}$ and $\boldsymbol{B}_{\Delta 2}$ are formed

Algorithm 4 (`cd_gmrf_normal`) Contrastive divergence on a gamma MRF model with Gaussian observations.

Input: Observations $\boldsymbol{y}$.

Fixed parameters: number of contrastive divergence steps N_{cd}, number of MCMC cycles N_{mc}, and step sizes $\boldsymbol{\lambda} \in \mathbb{R}_+^{N_{\text{cd}}}$
Initialize hyperparameter estimate $\widehat{a}$
for $i \in \{1, \ldots, N_{\text{cd}}\}$ **do**
  $q \leftarrow$ `vi_gmrf_normal` $(\boldsymbol{y}, a)$
  Draw $\boldsymbol{r}^{(0)}$, $\boldsymbol{z}^{(0)}$ from q
  for $k \in \{1, \ldots, N_{\text{mc}}\}$ **do**
    Draw $\boldsymbol{z}^{(k)}$ from $p(\boldsymbol{z} \mid r^{(k-1)}; a)$
    Draw $\boldsymbol{r}^{(k)}$ from $p(\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{z}^{(k)}; a)$
  end for
  Draw y_i' from $\mathcal{CN}(y_i; 0, r_i^{(N_{\text{mc}})})$ for $i \in \mathcal{V}$
  $q' \leftarrow$ `vi_gmrf_normal` $(\boldsymbol{y}', a)$
  $\widehat{a} \leftarrow \widehat{a} - \lambda_i \big(\big\langle \frac{\partial \log \pi(\boldsymbol{r}, \boldsymbol{z} \mid a)}{\partial a} \big\rangle_{q'} - \big\langle \frac{\partial \log \pi(\boldsymbol{r}, \boldsymbol{z} \mid a)}{\partial a} \big\rangle_{q} \big)$
end for

Output: Estimated hyperparameter $\widehat{a}$

Algorithm 5 (`vi_gmrf_normal`) Variational inference on a gamma MRF model with complex Gaussian observations .

Input: Observations $\boldsymbol{y}$, gamma MRF hyperparameter a

Initialize $q_{\boldsymbol{r}}$
repeat
  $q_{\boldsymbol{z}} \leftarrow \mathcal{G}(\cdot\,; \boldsymbol{\alpha}_z, \boldsymbol{\beta}_z)$ with $\alpha_{z,j} = a|S_j^z|$ and $\beta_{z,j} = a \sum_{i \in S_j^z} \langle r_i^{-1} \rangle$ for $j \in \mathcal{V}^z$
  $q_{\boldsymbol{r}} \leftarrow \mathcal{IG}(\cdot\,; \boldsymbol{\alpha}_r, \boldsymbol{\beta}_r)$ with $\alpha_{r,i} = 1 + |S_i^r|$ and $\beta_{r,i} = |y_i|^2 + a \sum_{j \in S_i^r} \langle z_j \rangle$ for $i \in \mathcal{V}$
until Terminated

Output: $q = q_{\boldsymbol{z}} q_{\boldsymbol{r}}$

by taking the corresponding differences in the rows of the polynomial design matrix $\boldsymbol{B}$, each column of which defines the spatial distribution of a polynomial term. The matrix $\boldsymbol{P}$ is a row-pruning operator which excludes areas of temporal change from the computation of the polynomial coefficients. In particular, the pruning operation excludes elements with an estimated posterior probability of change above a threshold,

after which the element is considered unreliable. A threshold of 0.5 is used for the computed results, and performance was not found to be overly sensitive to the choice of threshold.

Once the polynomial coefficients are estimated, the method of least squares can be used to estimate the constant phase offset. The interferometric phase estimate is thus given by

$$\widehat{\boldsymbol{\phi}} = \boldsymbol{B}\widehat{\boldsymbol{g}} + \widehat{\phi}_0 \tag{3.21}$$

where $\widehat{\phi}_0 \in [-\pi, \pi)$ is the constant phase offset estimate. The interferometric phase estimate is updated iteratively throughout the inference procedure along with the M-step parameter updates.

Coherence Parameter

Using (3.8), the updated estimate for γ is given by

$$\widehat{\gamma} = \arg\max_{\gamma} \left\langle \log p(\boldsymbol{y}^{(2)} \mid \boldsymbol{y}^{(1)}, \boldsymbol{c}, \boldsymbol{r}^{(2)}; \boldsymbol{\phi}, \gamma) \right\rangle \tag{3.22}$$

$$= \arg\max_{\gamma} \left\{ \sum_{i \in \mathcal{V}} -\langle \delta(c_i + 1) \rangle \log(1 - \gamma^2) - \langle \delta(c_i + 1) \rangle \left\langle \frac{\gamma^2 |y_i^{(1)}|^2 + |y_i^{(2)}|^2 - \gamma b_i}{r_i^{(2)}(1 - \gamma^2)} \right\rangle \right\}, \tag{3.23}$$

in which

$$b_i = 2\mathrm{Re}\{\overline{y}_i^{(2)} y_i^{(1)} e^{-j\phi_i}\}. \tag{3.24}$$

By setting to zero the derivative of the objective in (3.23) with respect to γ, we have a third-order polynomial,

$$\begin{aligned} &-\left(\sum_{i\in\mathcal{V}} 2\langle\delta(c_i+1)\rangle\right)\gamma^3+\left(\sum_{i\in\mathcal{V}}\left\langle\frac{b_i\delta(c_i+1)}{r_i^{(2)}}\right\rangle\right)\gamma^2 \\ &+\left(\sum_{i\in\mathcal{V}}\langle\delta(c_i+1)\rangle\left(2-2\left\langle\frac{|y_i^{(1)}|^2+|y_i^{(2)}|^2}{r_i^{(2)}}\right\rangle\right)\right)\gamma \\ &+\left(\sum_{i\in\mathcal{V}}\left\langle\frac{b_i\delta(c_i+1)}{r_i^{(2)}}\right\rangle\right)=0, \end{aligned} \tag{3.25}$$

that may be used to find the optimal value for γ.

Change Indicator MRF Hyperparameter

Due to the difficulty of assessing the partition function in (3.13), given by

$$Z(\chi)=\sum_{\boldsymbol{c}\in\{-1,1\}^{|\mathcal{V}|}}\exp\left\{\sum_{i\in\mathcal{V}}\chi_0 c_i+\frac{\chi}{2}\sum_{j\in S_i^c}c_i c_j\right\}, \tag{3.26}$$

we use the following mean field approximation [114]:

$$Z_{\text{MF}}(\chi)=\prod_{i\in\mathcal{V}}\sum_{c_i\in\{-1,1\}}\exp\left\{\chi_0 c_i+\chi\sum_{j\in S_i^c}c_i\langle c_j\rangle\right\}. \tag{3.27}$$

Note that (3.27) is a latent variable version of the approximation used in the psuedo-likelihood of Besag [115], i.e., equivalent if $\langle c_j\rangle$ in (3.27) is replaced with c_j. With the mean field partition approximation and (3.8), the M-step update to χ is given by

$$\widehat{\chi}=\arg\max_{\chi}\left\{-\log Z_{\text{MF}}(\chi)+\sum_{i\in\mathcal{V}}\chi_0\langle c_i\rangle+\chi\sum_{j\in S_i^c}\langle c_i\rangle\langle c_j\rangle\right\}, \tag{3.28}$$

which can be solved using Newton's method.

3.3.4 Summary

The proposed coherent change detection approach is summarized in Algorithm 6.

Algorithm 6 The proposed variational EM change detection algorithm

Input: Observations: $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}$

Initialize $q_{\boldsymbol{r}}$, $q_{\boldsymbol{c}}$, $\widehat{\boldsymbol{\phi}}$, $\widehat{\chi}$, $\widehat{\gamma}$
$\widehat{a} \leftarrow$ `cd_gmrf_normal` $(\boldsymbol{y}^{(1)})$
repeat
  repeat
    Update $q_{\boldsymbol{z}^{(1)}}$ according to (3.19a) and (3.19b)
    Update $q_{\boldsymbol{z}^{(2)}}$ according to (3.19a) and (3.19b)
    Update $q_{\boldsymbol{r}^{(1)}}$ according to (3.16a) and (3.16b)
    Update $q_{\boldsymbol{r}^{(2)}}$ according to (3.17a) and (3.17b)
    Update $q_{\boldsymbol{c}}$ according to (3.20a) and (3.20b)
  until Terminated
  Update $\widehat{\boldsymbol{\phi}}$ according to (3.21)
  Update $\widehat{\gamma}$ according to (3.23)
  Update $\widehat{\chi}$ according to (3.28)
until Terminated

Output: Approximate change posterior distribution $q_{\boldsymbol{c}}$

3.4 Experiments

In this section, we evaluate the proposed change detection approach using both simulated and measured interferometric SAR image pairs. The analysis includes comparisons to conventional window-based change detection as well as to coherence estimates from NL-InSAR [99], a more advanced approach that uses nonlocal techniques.

3.4.1 Simulated Data Experiments

Interferometric SAR data are first simulated according to four example scenarios that are intended to provide a variety of test conditions for the evaluation of change detection performance. The scenarios use various configurations of the underlying simulation parameters that may impact change detection: the interferometric phase,

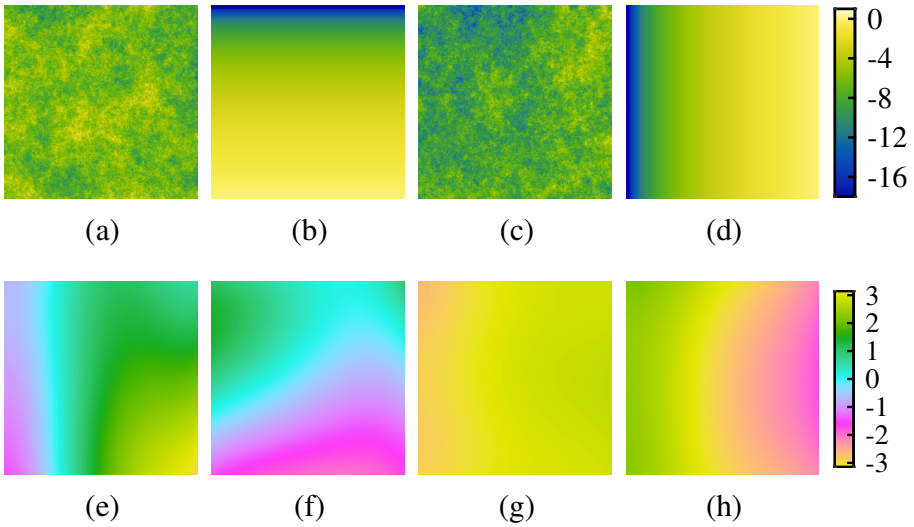


Figure 3.3: Values used to simulate Scenarios 1–4 (columns 1–4, respectively). (a–d) Normalized mean backscatter power (in dB). (e–h) Interferometric phase (in radians).

mean backscatter power, coherence, and change maps (i.e., the arrangement of temporal disturbances in the scene).

To provide an empirical basis for the change maps used to simulate the data, we use ground truth temporal change maps derived from the Onera satellite change detection dataset [116]. The change maps derived from the Onera change detection dataset for the four scenarios appear in the first column of Fig. 3.5 as binary images with black pixels indicating a change between the two passes. Scenarios 1 and 3 use mean backscatter power images that have been realized by a Gibbs sampling procedure on gamma MRF models with $a = 20$ and $a = 16$, respectively. Scenarios 2 and 4 follow linear ramps that range in power from 0 dB to 30 dB. In each case the time-1 and time-2 images have the same mean backscatter power, the patterns for which are

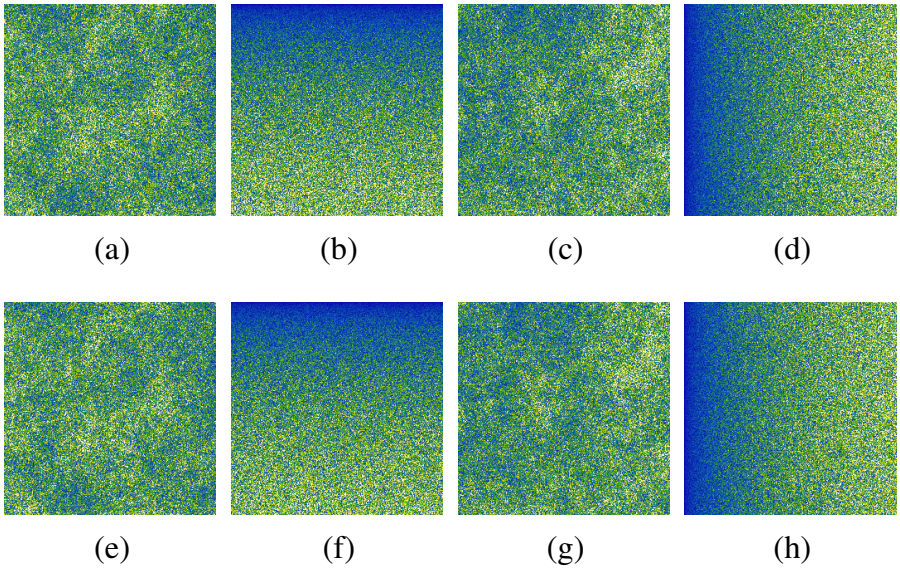


Figure 3.4: Magnitude images for Scenarios 1–4 (columns 1–4, respectively). (a–d) Time-1 magnitude images. (e–h) Time-2 magnitude images.

shown in the first row of Fig. 3.3. The underlying interferometric phase data are two-dimensional polynomials which have been fit to interferogram phases obtained from image pairs in the Gotcha challenge problem dataset [2]. The phase images used to simulate each scenario are shown in the second row of Fig. 3.3. Interferometric SAR image pairs of size 256×256 pixels are then drawn according to the change model in (3.11) and (3.12) using γ values of 0.55, 0.65, 0.75, and 0.85 for Scenarios 1–4, respectively. The magnitudes of the resulting image pairs are shown in Fig. 3.4.

The change detection results from the four scenarios are shown in Fig. 3.5. The second and third columns of Fig. 3.5 show the sample coherence maps computed with 3×3 and 5×5 windows, respectively. The fourth column shows the coherence estimates from NL-InSAR [99], computed using the default parameters in the publicly

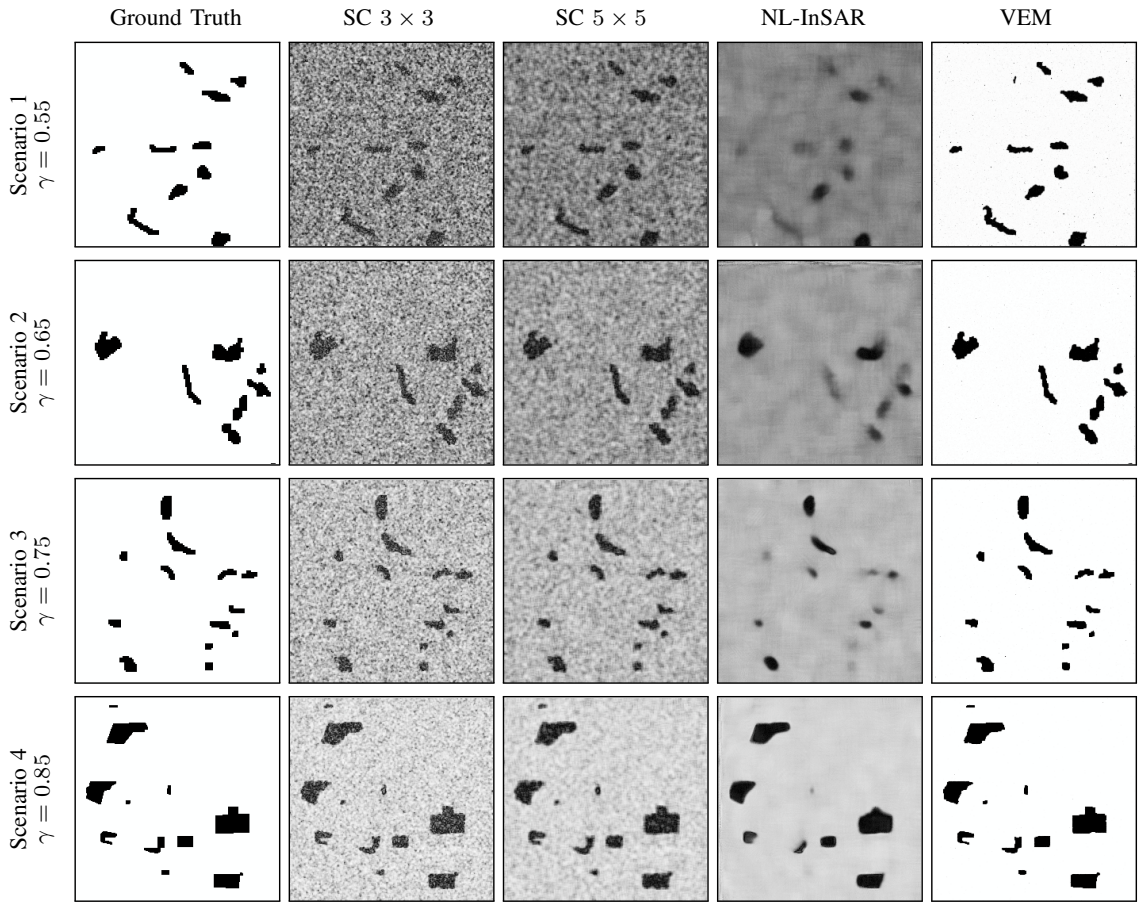


Figure 3.5: Comparison of change detection results on simulated data Scenarios 1–4 (rows 1–4, respectively) with various coherence values γ. The first column has the true change maps with change pixels in black, and the remaining columns show the change detection results from the compared methods. The last column shows the results of the proposed inference approach with black indicating a high probability of change.

available MATLAB code.[3] Shown in the fifth column of Fig. 3.5, the results from the proposed method express the estimated posterior probability of change at a given pixel, with darker pixels indicating a higher probability of change. The results from the proposed method are computed according to Algorithm 6 with the variational

[3] https://www.charles-deledalle.fr/pages/nlinsar.php

factors $q_{\boldsymbol{r}^{(k)}}$ for $k \in \{1, 2\}$ initialized to the maximum likelihood estimates computed in 9×9 neighborhoods and a maximum of 35 iterations of variational EM. The change posterior estimates are initialized to 0.5 for $i \in \mathcal{V}$, and the interferometric phase parameters are initialized to the interferogram phases.

The receiver operating characteristic (ROC) curves for the four simulated data scenarios are shown in Fig. 3.6, in which the curves labeled "SC" correspond to the sample coherence computed with the specified window size and "VEM" refers to the proposed variational EM approach. The ROC curves show the proposed method to outperform the others with respect to reduced false alarm rates for a given probability of detection. The improved performance of the proposed approach is particularly apparent from the low-coherence simulations using Scenarios 1–2 (i.e., the first two rows in Fig. 3.5) in which many false alarms are visible in the coherence estimates. Qualitatively, the NL-InSAR estimates obscure the edges of the change regions and in some cases render small, isolated regions of change difficult to observe.

In these experiments, both NL-InSAR and the proposed approach are significantly slower than the classical window-based CCD methods, which averages 2.16 ms of computing time: NL-InSAR runs in 65.49 s and the proposed approach in 22.57 s. All experiments are computed in MATLAB R2020a with a 3.2 GHz Intel Core i5-6500 CPU.

As a further point of comparison, we compute the threshold values that, when applied to the coherence estimates, minimize the empirical error rate in the four simulated scenarios depicted in Fig. 3.5. The error rates for the coherence estimates thresholded in this manner are reported in Table 3.1. The error rates for the variational EM approach are computed using the maximum *a posteriori* (MAP) estimates

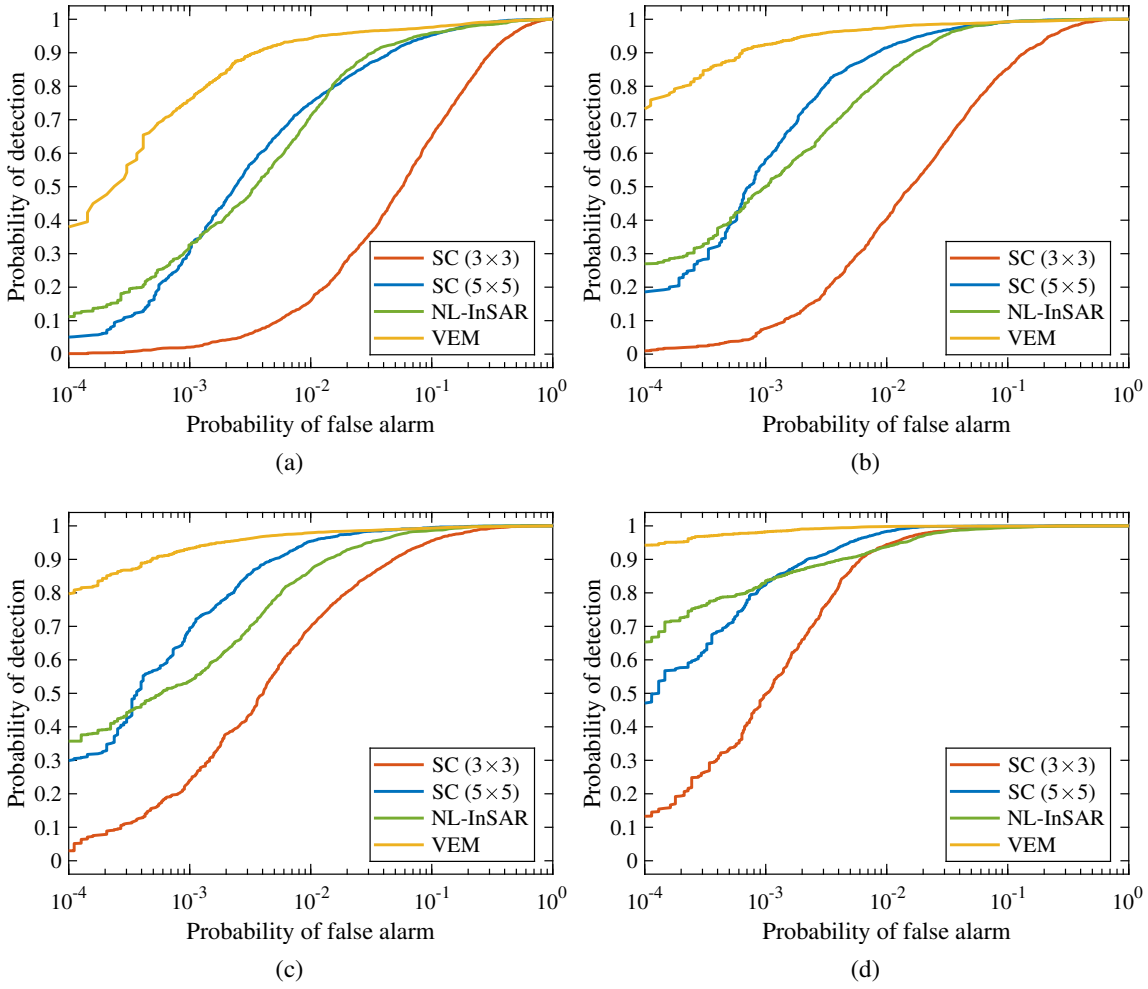


Figure 3.6: Receiver operating characteristic (ROC) curves for simulated data experiments. (a) Scenario 1 ($\gamma = 0.55$). (b) Scenario 2 ($\gamma = 0.65$). (c) Scenario 3 ($\gamma = 0.75$). (d) Scenario 4 ($\gamma = 0.85$).

of the change indicator variables. These results show the variational EM approach to improve upon the minimum error rates available from the coherence-based detectors.

Another series of experiments focuses on the impact of γ while keeping the other ground truth parameters constant. The change detection results computed using the compared methods on Scenario 4 with γ values of 0.55, 0.65, 0.75, and 0.85

Table 3.1: Error rates from the simulated data experiments using Scenarios 1–4. The error rates for the coherence estimates are computed using a threshold which minimizes the error rate.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
SC (3×3)	0.039	0.037	0.020	0.012
SC (5×5)	0.018	0.011	0.008	0.008
NL-InSAR	0.021	0.016	0.013	0.010
VEM	0.007	0.004	0.004	0.002

are shown in Fig. 3.7 with the minimum error rates reported in Table 3.2. Since the coherence quantifies the degree of pass-to-pass self-similarity between no-change pixels, we observe the expected result that all methods perform better for larger values of γ—with notably fewer numbers of false alarms. However, for all values of γ, the proposed VEM method appears to suppress substantially more false alarms while also faithfully representing fine details of the change regions. For example, with $\gamma = 0.55$, the "c"-shaped change region (identified by the red arrows in Fig. 3.7) is largely obscured by the window-based methods and NL-InSAR. However, the high contrast between the changed and unchanged regions of the VEM approach enables both fine localization of the change and identification of the change shape. Such high resolution change detection is important in forensic applications seeking to perform high resolution tracking or shape-based classification of a changed object (e.g., vehicle classification).

Next, the recovery of fine detail is evaluated by reusing the simulation parameters of Scenario 1 with the change map replaced by one consisting of square change regions with side lengths of ten pixels, placed randomly in the scene with at least a three pixel margin between them. The minimum error rates from the sample coherence using

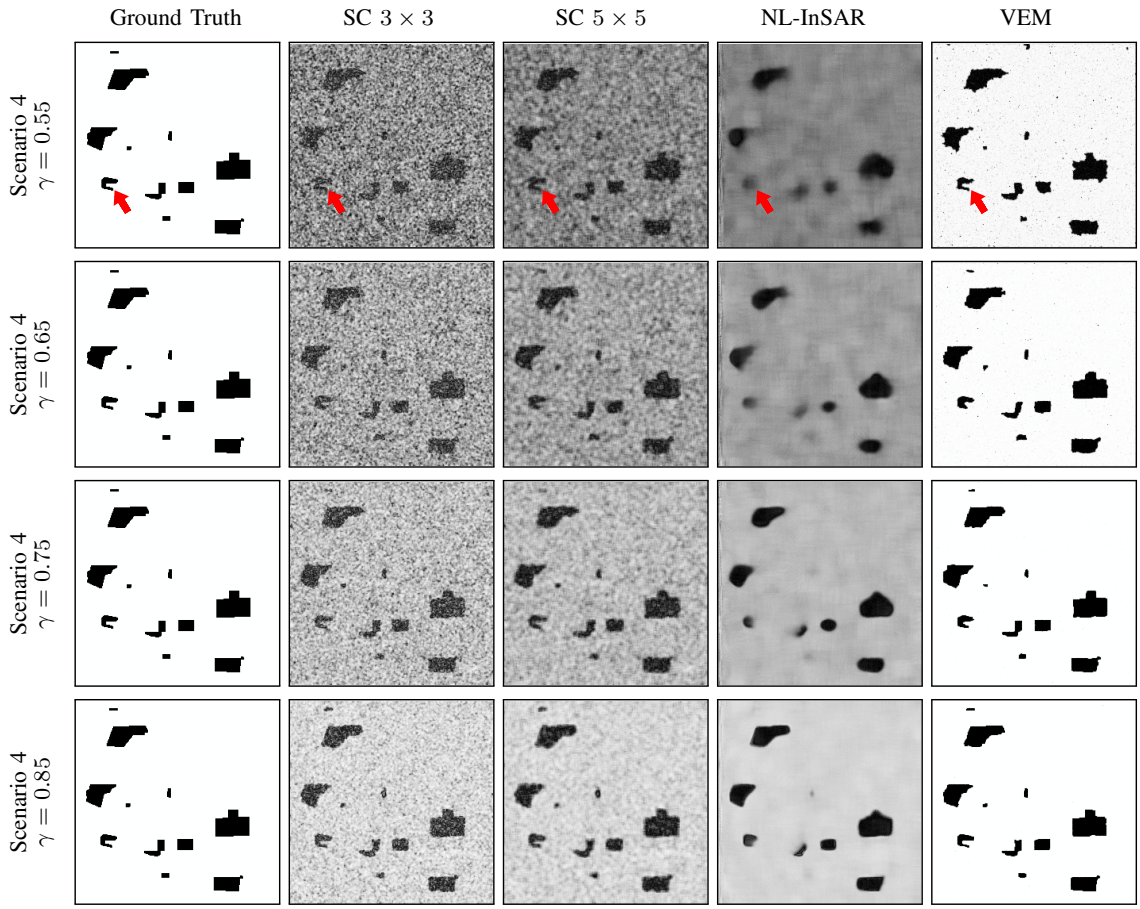


Figure 3.7: Comparison of change detection results from Scenario 4 with γ values of $0.55, 0.65, 0.75$, and 0.85 depicted in rows 1–4, respectively. The other simulation parameters are kept constant. The first column has the true change maps with change pixels in black, and the remaining columns show the change detection results from the compared methods. The last column shows the results of the proposed inference approach with black indicating a high probability of change.

various window sizes are plotted in Fig. 3.8 as the curve labeled "SC". The x-axis corresponds to the number of pixels on each side of the square estimation window used to compute the sample coherence. The "U"-shaped error curve of the SC estimator highlights competing effects resulting from different window sizes. On the left of the figure, small window sizes produce poor coherence estimates, resulting in larger change

Table 3.2: Error rates from the simulated data experiments using Scenario 4 with various values for γ. The error rates for the coherence estimates are computed using a threshold which minimizes the error rate.

	$\gamma = 0.55$	$\gamma = 0.65$	$\gamma = 0.75$	$\gamma = 0.85$
SC (3×3)	0.062	0.047	0.027	0.012
SC (5×5)	0.023	0.013	0.009	0.008
NL-InSAR	0.021	0.017	0.015	0.010
VEM	0.011	0.005	0.003	0.002

detection error rates. Increasing the window size improves the coherence estimate; however, when the window becomes too large, it captures both change and non-change pixels. This degrades the coherence estimate and affects the ability to localize fine details, which contributes to the increased SC error rates seen for large window sizes on the right side of the figure. The error rate of the MAP estimate from the variational EM approach is shown to be lower than that of the sample coherence detectors using any of the window sizes considered in Fig. 3.8. This result demonstrates superior fidelity of the proposed approach to the structure of the underlying change maps used in these experiments without the need to select a window size.

3.4.2 Measured Data Experiments

Next we evaluate the performance of the proposed change detection algorithm using measured interferometric data from two SAR systems, one X-band and the other C-band. Unlike the simulated measurements, these data were produced under non-ideal conditions and lack precise pixel-level ground truth necessary for quantitative evaluation. Therefore, we limit the discussion to qualitative assessments of change detection performance.

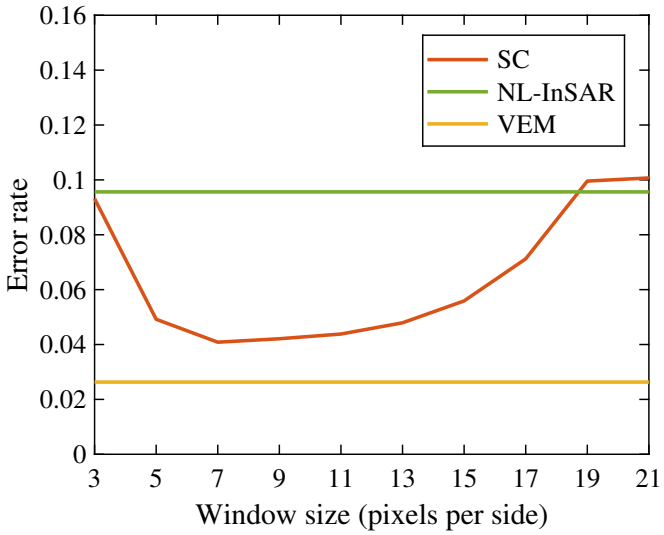


Figure 3.8: Error rates of a sample coherence estimator plotted against window size. The error rates are computed using thresholds which minimize the error rates at each window size. The x-axis corresponds to the pixels per side of the square window used to compute the sample coherence. Also shown are the error rate from the NL-InSAR coherence estimates and the MAP change estimates from the proposed VEM approach. The underlying change map consists of randomly-placed square change regions, and the simulation assumes a coherence of $\gamma = 0.55$.

Gotcha Change Detection Data

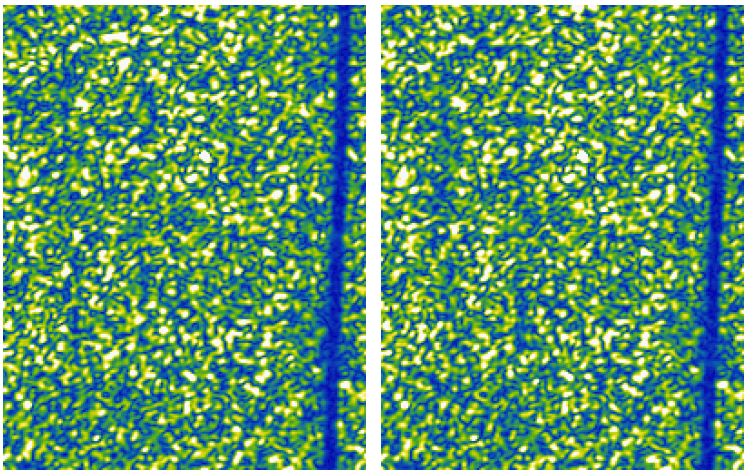
We first evaluate the proposed change detection algorithm using a co-registered pair of measured HH-polarized SAR images from the X-band Gotcha change detection dataset [2]. The images have a sample spacing of 0.2 m, and they are from data that was collected on the same day and coherently aligned to a single reference. We use corresponding portions of the HH-polarization images `FP1020` and `FP1024` as the time-1 and time-2 images, respectively. The region observed in these sub-images is roughly that contained in the black rectangle in Fig. 3.9. The 256×200 pixel sub-images feature a paved path running along the right side of the scene and, to the

(a) (b)

Figure 3.9: Optical satellite images [1] of a region observed in the Gotcha change detection challenge problem dataset [2]. (a) The observed region in the measured data experiment is shown framed by a black box. (b) The hinged grate shown here is identified by the blue arrow in (a).

left of it, a grassy area where foot traffic is known to have occurred in the interim between the two collections. The time-1 and time-2 magnitude images are shown in Fig. 3.10, from which the foot traffic is largely unobservable.

The change detection results from the compared methods are shown in Fig. 3.11. Each of the coherence estimates show the path to be a low coherence region, evidence of a temporal disturbance—and thus a false alarm, since actual changes were not present there. According to the proposed approach, the region occupied by the hinged metal grate, identified by the blue arrow in Fig. 3.9, is the only spot along the path that has a high probability of change. This hinged grate extends roughly the width



(a) (b)

Figure 3.10: Magnitude images from the Gotcha dataset [2]. (a) Time-1 (FP1020). (b) Time-2 (FP1024).

of the path as shown in Fig. 3.9. The proposed method departs from the coherence estimates, in that it properly does not indicate change along the rest of the path. Each method shows clear evidence of the foot traffic in the grass, though the proposed VEM method results in a notably more contiguous change region about the track.

Sentinel-1 Change Detection Data

We also compare the change detection methods using a pair of SAR images formed with data from the C-band SAR systems of the Sentinel-1 constellation of satellites [5]. We use the Sentinel-1 images to detect the effects of a wildfire in the southern foothills of the Troodos Mountains in Cyprus that began on July 3, 2021.[4] The images had a pixel spacing of 2.3 m in range and 14.0 m in azimuth before they were downsampled by a factor of two in range. The time-1 image in the pair was collected before the

[4] https://emergency.copernicus.eu/mapping/list-of-components/EMSR515

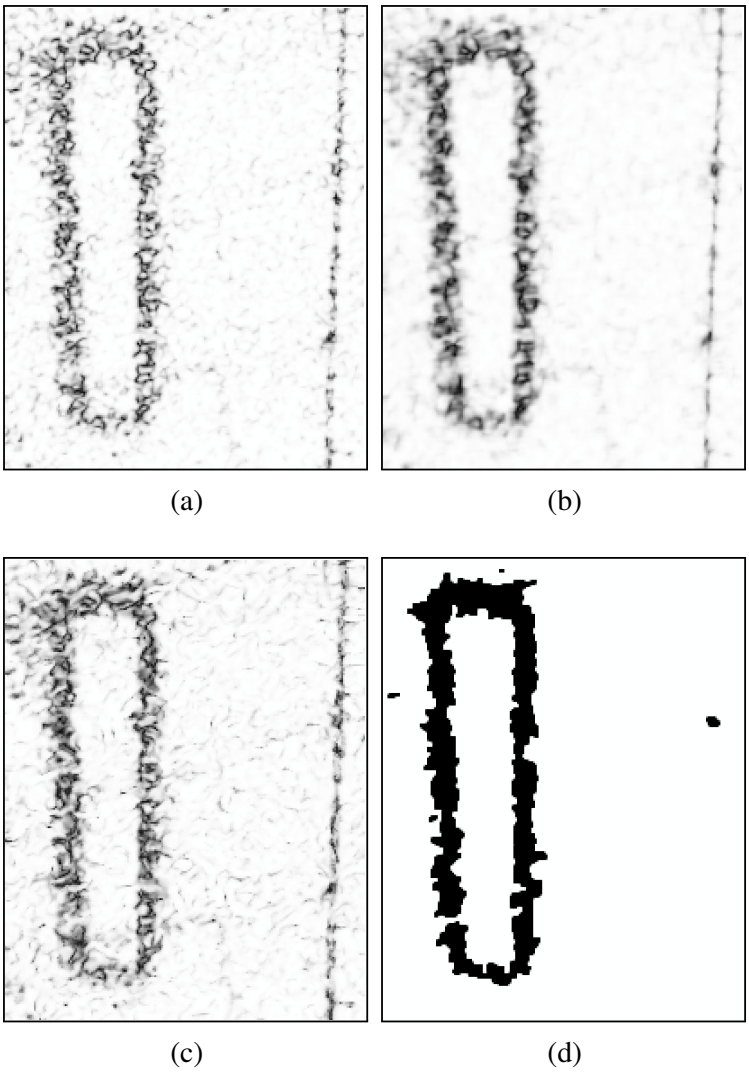


Figure 3.11: Change detection results from the X-band Gotcha dataset. (a) Sample coherence with 3×3 window. (b) Sample coherence with 5×5 window. (c) NL-InSAR coherence estimate. (d) Proposed VEM inference approach with black indicating a high probability of change.

fire, on June 30, 2021, and the time-2 image after, on July 6, 2021. The impacted area, as determined by optical image analysis [3], is shown in yellow on the map in Fig. 3.12.

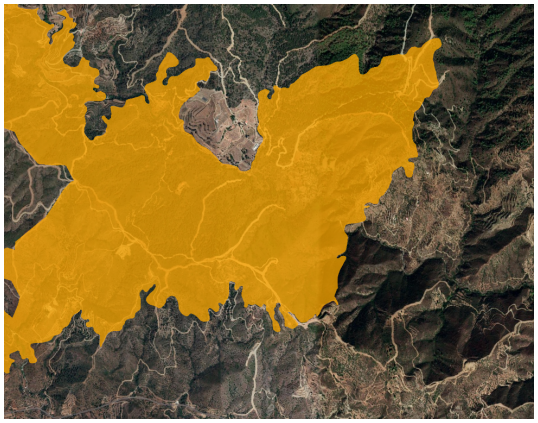


Figure 3.12: Optical image of the Troodos Mountains in Cyprus with a yellow overlay depicting the area impacted by a wildfire that began on July 3, 2021 [3,4].

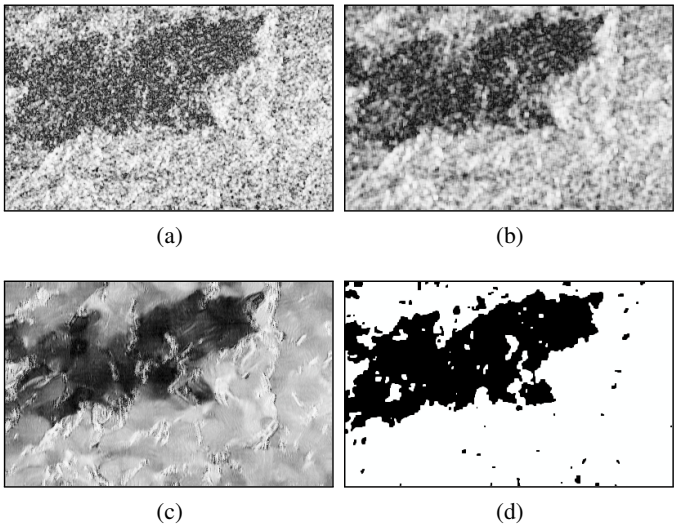


Figure 3.13: Change detection results from C-band Copernicus Sentinel SAR data [5] collected over the region of Cyprus shown in Fig. 3.12. (a) Sample coherence with 3×3 window. (b) Sample coherence with 5×5 window. (c) NL-InSAR coherence estimate. (d) Proposed VEM inference approach with black indicating a high probability of change.

The change detection results from the Cyprus Sentinel-1 data are shown in Fig. 3.13. The results from each method depict the extent of the burned area to be broadly in agreement with the estimate from the optical data. As seen in the earlier experiments, the proposed variational EM method exhibits greater contrast between the change and no-change regions. The results from NL-InSAR and the proposed approach both indicate small regions free of temporal change from within the burn region shown in Fig. 3.12, though, without precise ground truth knowledge of the temporal changes, it is difficult to speculate on the accuracy of these departures from the optical data.

3.5 Conclusion

3.5.1 Summary

In this chapter, we have introduced a Bayesian approach to SAR change detection that enables fine-resolution detection of temporal disturbances between pairs of co-registered SAR images. The proposed approach assigns prior distributions to the unobserved model variables that exploit spatial structure both in the scattering qualities of the scene and among the scene disturbances that occur between the passes. The proposed modeling advances the use of gamma MRFs for SAR change detection, using them to capture positive spatial correlations in the mean backscatter power elements. Further, a 2-D polynomial regression is used to account for the spatially nonhomogeneous interferometric phase.

The prior models are incorporated into a Bayesian framework, on which variational EM is used to approximate the posterior distribution of the latent variables and the prior model hyperparameters. In the Bayesian approach, the posterior probability of

change is directly inferred, rather than thresholding the coherence estimates obtained by local, or nonlocal, spatial averaging. To optimize the gamma MRF coupling hyperparameter, variational inference is used to efficiently approximate expectations in the contrastive divergence gradient.

The proposed approach does not suffer from losses of spatial resolution and fine-scale details as do conventional window-based CCD methods. In experiments using both simulated data and measured interferometric SAR image pairs, the new approach is shown to yield improved detection performance compared to the conventional window-based statistics used for SAR CCD.

3.5.2 Future Work

The approach of this chapter may be extended to form a joint imaging and change detection framework that is intended to mitigate the effects of interrupted data collections, as in [63]. A Bayesian approach similar to that proposed here could be used to compensate for missing data via prior distributions on the change variables, the scene reflectance, and the interferometric phase.

Chapter 4: Polarimetric SAR Despeckling

4.1 Introduction

Polarimetric SAR (PolSAR) systems use polarization diversity in the emitted and received electromagnetic waves to provide data that are considerably more informative than single-channel collections as the inter-channel relationships are informative of physical scattering behavior [16, 117, 118].[5] Multichannel radar imaging modalities such as PolSAR have proven beneficial for tasks such as terrain classification [120], target recognition [10], and surface parameter estimation [121, 122].

The speckle phenomenon lends a noise-like, granular character to imagery which can impede interpretation and the completion of automated tasks [13]. The reduction of speckle, i.e. despeckling, improves both the visual quality of the image and the prospects for tasks such as segmentation and target recognition based on the data [36].

It is often necessary to suppress speckle before SAR data are considered suitable for analysis [16], a prerequisite realized by the application of a speckle filter, classically one which acts on the image locally [17]. Notable SAR despeckling algorithms have been proposed using local filtering [35, 76, 123], patch-based filtering [38, 39, 77, 124], region-growing [125], and variational methods [42–44].

[5]This chapter is largely drawn from [119], coauthored with Lee C. Potter ().

Deep learning techniques that involve the use of hierarchical structures to extract feature representations from data have been responsible in recent years for state-of-the-art performance in numerous image processing tasks. Convolutional neural networks (CNN), flexible deep learning architectures composed of cascades of convolutions and other operations, have been used with great success in tasks such as super-resolution [126,127], object detection [128,129], and denoising [130,131]. Since 2017, there has been a trend towards deep learning in single-channel SAR despeckling, e.g. [132–135]. CNNs have been applied to PolSAR data tasks [136] such as super-resolution [137,138], land use classification [139,140], and target recognition [141,142]. Yet, there have been few CNN-based methods for PolSAR speckle reduction [143–146].

The need for large datasets in deep learning presents a hindrance to SAR despeckling applications because the speckle-free images which might be used in training are not observable. The issue has been managed in single-channel SAR despeckling by using optical images as clean intensity data [133, 134], a practice which does not have an obvious equivalent for PolSAR images. Temporal multilooking has also been used to approximate speckle-free intensity images [132,147,148]. However, the series of co-registered multi-temporal images used in these approaches may not be readily available in the quantity necessary for deep learning, particularly from fully polarimetric collections. Other approaches eschew the need for clean training data and instead learn to complete the despeckling task in a self-supervised manner using only noisy images [149–151].

PolSAR despeckling is further complicated by the nature of the data as images of complex-valued Hermitian positive definite polarimetric covariance matrices. Despeckling methods must respect the Hermitian positive definite constraint and manage

the complex-valued off-diagonal elements of the polarimetric covariance matrices. The phases of the off-diagonal elements hold valuable information, and the polarimetric characteristics that they capture ought to be preserved by any filtering process.

The approach presented in this chapter extends the single-channel SAR CNN-based speckle filtering in [132] to PolSAR despeckling, yielding a deep learning framework which estimates from PolSAR data the underlying polarimetric covariance matrices. With a residual learning approach [83, 152], a CNN recovers not the filtered image itself, but the speckle component which is to be removed from the corrupted image. Drawing from [42], a bijective transform enables real-valued CNN processing such that the network outputs represent valid covariance estimates. As training data, we use spatially-multilooked PolSAR images in lieu of unobservable ground truth covariance data and simulate speckle to form corrupted training images. We use both simulated and measured polarimetric SAR data to evaluate the proposed approach and introduce a suite of quantitative indexes to evaluate the competing objectives of speckle suppression and preservation of scene features and polarimetric information. The proposed approach markedly reduces speckle and faithfully preserves fine structural detail in comparison to state-of-the-art methods. A further benefit of the proposed approach is the fast running time due to the use of CNNs.

The remainder of the paper is organized as follows. Section 4.2 provides background information and a review of the relevant literature. Section 4.3 is a summary of the proposed despeckling method. Section 4.4 outlines an experimental procedure to compare the proposed method to existing methods. Section 4.5 presents and discusses the experimental results. Section 4.6 gives concluding remarks.

4.2 Background

4.2.1 Principles of Polarimetric SAR

The single-look scattering properties observed by a polarimetric SAR system at a given pixel can be described by a 2×2 complex scattering matrix given by

$$\boldsymbol{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}, \tag{4.1}$$

where S_{hv} denotes the scattering element corresponding to the horizontal transmitting and vertical receiving polarizations. For a monostatic radar operating in a reciprocal medium, $\boldsymbol{S}$ is symmetric and can be vectorized according to the lexicographic or Pauli basis as

$$\boldsymbol{s} = \begin{bmatrix} S_{hh} & \sqrt{2}S_{hv} & S_{vv} \end{bmatrix}^{\mathsf{T}} \tag{4.2}$$

$$\boldsymbol{\kappa} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & 2S_{hv} \end{bmatrix}^{\mathsf{T}}, \tag{4.3}$$

respectively, wherein the superscript T denotes a matrix transpose. When the scattering medium is rough on the scale of the radar's central wavelength, it is common to model a scattering vector as following a complex zero-mean circularly-symmetric Gaussian distribution [13], given by

$$p(\boldsymbol{s} \mid \boldsymbol{\Sigma}) = \frac{1}{\pi^3 |\boldsymbol{\Sigma}|} \exp\left\{-\boldsymbol{s}^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{s}\right\}, \tag{4.4}$$

in which $|\cdot|$ denotes a determinant. In (4.4), $\boldsymbol{\Sigma} \in \mathbb{H}_+^3$ is the covariance matrix $\boldsymbol{\Sigma} = \mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^{\mathsf{H}}\right]$ with the superscript H denoting a complex conjugate transpose and $\mathbb{H}_+^D$ denoting the set of $D \times D$ Hermitian positive definite matrices. The equivalent for the Pauli scattering vector is the coherency matrix, $\boldsymbol{T} = \mathrm{E}\left[\boldsymbol{\kappa}\boldsymbol{\kappa}^{\mathsf{H}}\right]$. The stochastic scattering model in (4.4), often referred to as fully-developed speckle, accounts for the

speckle phenomenon in coherent imaging. When the model is valid, the covariance matrix $\boldsymbol{\Sigma}$ contains the important polarimetric information of the imaged surface.

Multilook PolSAR observations can be represented using the L-look sample covariance matrix given by

$$\boldsymbol{Z} = \frac{1}{L}\sum_{t=1}^{L} \boldsymbol{s}(t)\boldsymbol{s}^{\mathsf{H}}(t) \tag{4.5}$$

where $\boldsymbol{s}(t)$ for $t = 1, \ldots, L$ are independent realizations of the scattering vector. These multiple "looks" are most often taken from a small window centered upon the pixel of interest. The goal of PolSAR despeckling is to recover the unobserved covariance $\boldsymbol{\Sigma}$ from the sample covariance matrix $\boldsymbol{Z}$.

Under the fully-developed speckle model, the L-look sample covariance matrix follows a complex Wishart distribution, i.e. $\boldsymbol{Z} \sim \mathrm{CW}(\boldsymbol{Z}; L^{-1}\boldsymbol{\Sigma}, L)$, where the complex Wishart density is defined for $\boldsymbol{X} \in \mathbb{H}_+^D$ as [153]

$$\mathrm{CW}(\boldsymbol{X}; \boldsymbol{G}, v) = \frac{|\boldsymbol{X}|^{v-D} \exp\{-\operatorname{tr}(\boldsymbol{G}^{-1}\boldsymbol{X})\}}{|\boldsymbol{G}|^{v}\, \widetilde{\Gamma}_D(v)} \tag{4.6}$$

with scale matrix $\boldsymbol{G}$, degrees of freedom v, $\operatorname{tr}(\cdot)$ denoting the trace function, and $\widetilde{\Gamma}_D(\cdot)$ the complex multivariate gamma function.

4.2.2 Polarimetric SAR Data Notation

In describing the statistics of PolSAR, we have considered the behavior of a single pixel. We now adopt a notation to describe a covariance image wherein a D-channel $N_1 \times N_2$ pixel covariance image is represented by a fourth-order tensor $\boldsymbol{Z} \in \mathbb{C}^{D\times D\times N_1\times N_2}$. We denote by $\boldsymbol{Z}_i \in \mathbb{H}_+^D$ the covariance matrix with pixel indices $i = (i_1, i_2)$, and denote by $\boldsymbol{Z}^{j_1 j_2} \in \mathbb{C}^{N_1\times N_2}$ the spatial image corresponding to the (j_1, j_2)th entry of the covariance matrix.

4.2.3 Related Work

PolSAR despeckling has been an active area of research since the early 1990s. The early despeckling methods relied on local window filtering, in which each element of the despeckled image is estimated from the values of nearby pixels within a small window. The simplest and most common such technique is the boxcar filter, a convolutional filter with a small window of equally-weighted elements. In 1983, the sigma filter was developed by Lee [35], offering a simple and effective despeckling approach for single-channel SAR. The sigma filter operates by averaging nearby pixels, excluding those differing significantly in intensity. The improved sigma filter was developed to rectify deficiencies related to bias and blurring in its predecessor [154] and later was extended for use with PolSAR data [123].

Vasile *et al.* [125] developed the intensity-driven adaptive neighborhood (IDAN) filter, which uses a region-growing approach to group pixels belonging to the same statistical populations. Adaptive neighborhoods are formed with respect to the intensity information (the diagonal terms of the coherency matrix), and grouped pixels are given equal weight in the estimation of the coherency matrix at each pixel.

In recent years, patch-based methods have been used in PolSAR despeckling. Rather than selecting pixels with similar intensities, such methods compare the patches about pixels to determine relative importance. The nonlocal means (NLM) filter, developed by Buades *et al.* [37] for the denoising of natural images, is responsible for popularizing these methods, and patch-wise descendants of the NLM filter have proven effective in PolSAR despeckling [38, 39, 124].

Variational methods, which pose the estimation as an optimization problem, are also used for speckle reduction. Some variational despeckling methods, such as [41],

process only the diagonal elements of the coherency or covariance matrices, ignoring the information in the cross-channel correlations. Other works formulate an objective function which operates on the full covariance matrix, such as the variational model in [43] which introduces a matricial total variation (TV) prior term to encourage spatial smoothness. This approach is augmented in [44] with a nonlocal regularization term leading to improved despeckling performance. Like [43, 44], the proposed approach operates on the full polarimetric covariance matrix so as to consider the polarimetric information of the off-diagonal elements.

Another variational despeckling methodology, the multichannel logarithm with Gaussian denoising (MuLoG) framework [42], facilitates the use of single-channel denoisers designed to remove additive white Gaussian noise (AWGN). This is a considerable benefit as the bulk of work in image denoising has been dedicated to the removal of AWGN.

Considering the success of deep learning in natural image denoising [83, 130], its application to SAR despeckling is natural, e.g. [132–134, 147]. Among them, SAR-CNN in [132] adapts the denoising CNN (DnCNN) framework [83] to single-channel SAR data by way of a log transform. A residual learning strategy [152] is used that treats the output of the CNN as a residual image, i.e. the difference between the noisy observation and the latent clean image. Residual learning and batch normalization [155] have been shown to both accelerate training and improve denoising performance on natural images [83].

Yet there are comparatively few CNN-based despeckling methods for PolSAR. The MuLoG [42] framework allows for the straight-forward application of deep learning to multichannel SAR despeckling through the techniques developed for AWGN

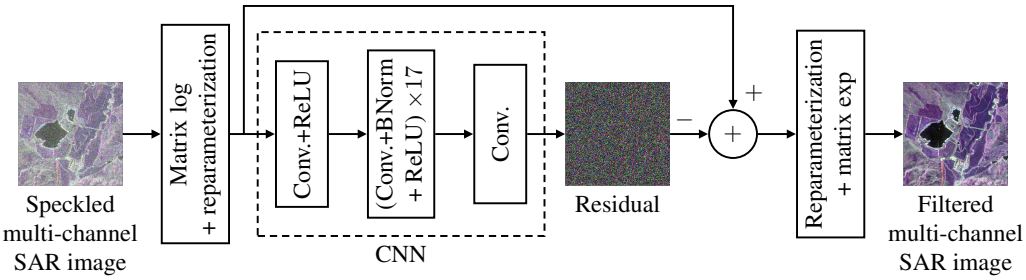


Figure 4.1: Proposed despeckling framework.

denoising of natural images, e.g. [83,130,131]. Indeed, MuLoG has been studied with deep learning-based denoisers trained on both natural images [156] and measured SAR images [145]. A related study [147] compares the performance of SAR-CNN to MuLoG when paired with either natural image-based DnCNN denoisers or DnCNN denoisers trained on SAR images. In [143], an iterative approach to PolSAR despeckling is studied that uses a CNN architecture intended for image super-resolution. This approach is limited in that it only considers the diagonal elements of the covariance, ignoring the information of the off-diagonal elements. That shortcoming is addressed in a revised approach [144] which uses a CNN to derive speckle correction terms based on all covariance elements.

4.3 Proposed Despeckling Approach

In this section, we summarize a proposed CNN-based approach to polarimetric SAR despeckling that operates in the matrix logarithm domain. The proposed approach, referred to as "MS-CNN", is depicted graphically in Fig. 4.1.

4.3.1 Approach Details

The PolSAR despeckling problem is complicated by the Hermitian positive definite constraint on $\boldsymbol{\Sigma}$, which for variational methods requires a potentially challenging optimization over the open cone of Hermitian positive definite matrices. This difficulty can be mitigated, as in [42], with a matrix logarithm transform, defined for $\boldsymbol{A} \in \mathbb{H}_+^D$ having eigendecomposition $\boldsymbol{A} = \boldsymbol{U} \operatorname{diag}(\lambda_1, \cdots, \lambda_D)\boldsymbol{U}^{\mathsf{H}}$ by

$$\log \boldsymbol{A} = \boldsymbol{U} \operatorname{diag}(\log \lambda_1, \cdots, \log \lambda_D)\boldsymbol{U}^{\mathsf{H}}. \tag{4.7}$$

Note that when applied to an image $\boldsymbol{Z}$, the $\log(\cdot)$ operator denotes pixel-wise application of the matrix logarithm to each $\boldsymbol{Z}_i$.

The matrix logarithm of a Hermitian positive definite covariance matrix is a Hermitian matrix. By filtering in the matrix log domain, we need only ensure that the output of the CNN remains Hermitian, which can be accomplished by operating on only the non-redundant elements of the matrix log-transformed covariance matrices. To do so, we define $\mathcal{V}(\cdot) : \mathbb{C}^{3\times3\times N_1\times N_2} \to \mathbb{R}^{9\times N_1\times N_2}$, a bijective operator which maps a complex fourth-order tensor to a nine channel real-valued image. The channels of the reparameterized image correspond to the non-redundant real and imaginary elements of a 3×3 Hermitian matrix.

As depicted in Fig. 4.1, the input is an L-look sample covariance image of $N_1 \times N_2$ pixels, $\boldsymbol{Z}$. At each pixel, $\boldsymbol{Z}_i$ represents a corrupted measurement of the unobserved covariance $\boldsymbol{\Sigma}_i$. For $L < 3$, it is necessary to ensure $\boldsymbol{Z}_i$ is full rank at each pixel, for which we adopt a technique used in [42] involving a minor rescaling of the off-diagonal elements. The covariance image undergoes a matrix log transformation at each pixel and is reparameterized using $\mathcal{V}(\cdot)$. The transformed image is then fed to

the CNN, and the output of the CNN is interpreted as a multichannel residual image. A despeckled estimate of the covariance image is given by subtracting the residual in the matrix log domain:

$$\widehat{\boldsymbol{\Sigma}} = \exp\left\{\mathcal{V}^{-1}\left(\mathcal{V}(\log \boldsymbol{Z}) - f_{\boldsymbol{\theta}}(\mathcal{V}(\log \boldsymbol{Z}))\right)\right\}, \tag{4.8}$$

where $f_{\boldsymbol{\theta}}(\cdot)$ is a CNN using the parameter set $\boldsymbol{\theta}$. With the matrix logarithm and this reparameterization, we ensure that the despeckled covariance matrix estimates are Hermitian positive definite.

To train the network, we assume the availability of a suitable proxy for speckle-free ground truth covariance data. A training pair $(\boldsymbol{x}, \boldsymbol{y})$ is given by $\boldsymbol{x} = \mathcal{V}(\log \boldsymbol{\Sigma})$ and $\boldsymbol{y} = \mathcal{V}(\log \boldsymbol{Z})$, where $\boldsymbol{Z}$ is an L-look sample covariance image drawn using the underlying covariance image $\boldsymbol{\Sigma}$ according to (4.4) and (4.5).

4.3.2 Network Architecture

The CNN architecture depicted within the dotted lines in Fig. 4.1 is that of [83], where the CNN is used to denoise natural images. The first layer is composed of 64 filters of size $3 \times 3 \times 9$. Next is a layer of rectified linear units (ReLUs) followed by 17 repetitions of convolutional layers with 64 filters of size $3 \times 3 \times 64$, batch normalization layers, and ReLUs. The final convolutional layer has nine filters of size $3 \times 3 \times 64$, leading to a nine channel output image. The use of 17 hidden layers was adopted here for the PolSAR despeckling problem; further expansion or contraction of the network failed to improve despeckling performance in our experiments.

We use the following loss function to form the deep learning objective

$$\ell(\boldsymbol{\theta}) = \sum_{m=1}^{M} \left\| \mathrm{vec}\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}^{(m)}) - (\boldsymbol{y}^{(m)} - \boldsymbol{x}^{(m)})\big) \right\|_1, \tag{4.9}$$

optimized with respect to the CNN parameter set $\boldsymbol{\theta}$, with the sum in (4.9) over M pairs from the training set. In our experiments, we found that an ℓ_1 loss function provided improved training performance over the ℓ_2 loss used with optical images in [83]; similar results were reported in [147] for single-channel SAR data.

4.3.3 Approach Summary

Thus, to extend deep learning to PolSAR despeckling, we adopt the residual network architecture used for natural images in [83] and single-channel SAR in [132]. In order to estimate the complex-valued covariance matrices, we employ the matrix logarithm and a reparameterization of the covariance matrix to an unconstrained set of real-valued parameters. Simulated speckle [13] is used to construct a corpus of training images. And, to better conform to the physics of the reparameterized variables, we use an ℓ_1 loss in training the network.

4.4 Experimental Procedure

In this section we detail experimental procedures to implement the algorithm described in Section 4.3 and compare its despeckling performance to existing methods. We also introduce a set of indexes for the quantitative assessment of polarimetric despeckling performance.

4.4.1 Training

In place of ground truth covariance information in the training set, we use PolSAR data from the L-band UAVSAR sensor [6, 157], images which have undergone a 36-look spatial averaging and equivalent downsampling. The noncoherent averaging coarsens the resolution in a compromise that reduces speckle so as to approximate the

unobservable ground truth covariance data. This method to generate clean covariance images relies on two assumptions which are likely to be violated at times: neighboring pixels combined by multilooking share statistical properties, and the neighboring observations are conditionally independent given the true covariance [158, 159]. This approach, while convenient and simple, may in its coarsening of image resolution diminish the value of the data for the most demanding of applications [151]. Further, the fundamental problem remains that we rely on an approximation to truly speckle-free images. We use a training set of 91490 such ground truth covariance images of size 40×40 that have been excised from 14 different collections. Ground truth images from two separate collections are later used in model validation.

The corrupted images used for training are L-look sample covariance images that are formed from clean images according to (4.4) and (4.5). For each number of looks, a separate CNN is trained using 60 epochs of stochastic gradient descent with weight decay of 10^{-3}, momentum of 0.9, batch size of 64, and a learning rate that begins at 0.1 and reduces to 10^{-4} during training. The network weights are initialized as in [160].

4.4.2 Conventional Despeckling Methods

We compare the proposed method to three well-established speckle filters: the simple boxcar filter [73], the extended sigma filter [123], and the IDAN filter [125]. The tuning parameters for these despeckling methods were selected as follows. For the boxcar filter, we use a 5×5 window. Two parameters are used to tune the performance of the extended sigma filter: the sigma value ξ, which controls the range in which pixels are considered similar, and a window size. In our experiments, we

use $\xi = 0.7$ for single-look data and $\xi = 0.9$ for four-look data, as is suggested in [123]. In either case, we use a filtering window of size 9×9. The IDAN filter uses a tuning parameter which determines the maximum number of pixels included in a neighborhood, for which we select $N_{\max} = 50$ as in [125].

4.4.3 MuLoG Despeckling Framework

We also compare results from the MuLoG framework [42], which represents the state-of-the-art in multichannel SAR despeckling. The MuLoG framework poses a maximum *a posteriori* (MAP) optimization problem that is solved by the plug-and-play variant of the alternating direction method of multipliers (ADMM) [56], an approach which facilitates the use of single-channel AWGN denoising operators. In our experiments, we use the MATLAB implementation[6] of MuLoG and adopt three choices for the denoising operator: the familiar block-matching 3D (BM3D) filter [74] used with MuLoG in [42], a DnCNN denoiser [83] trained on natural images, and a DnCNN denoiser trained on SAR images. We set the tuning parameter in MuLoG to $\beta = 1 + 2/L$ and use six iterations of ADMM, as suggested in [42].

For the DnCNN trained on natural images, we use a set of 14 pre-trained models from `https://github.com/cszn/DnCNN`; these have been trained for noise standard deviations of $\chi = \{\frac{10}{255}, \frac{15}{255}, \ldots, \frac{75}{255}\}$ on sets of natural images [83]. The range of the training images is $[0, 1]$, and, due to the nonlinearity of DnCNN, we use a rescaling operator to match this range, defined by

$$\boldsymbol{\Psi}_{\boldsymbol{v}_0}(\boldsymbol{v}) = \frac{\boldsymbol{v} - \mathcal{R}_{0.3\%}(\boldsymbol{v}_0)}{\mathcal{R}_{99.7\%}(\boldsymbol{v}_0) - \mathcal{R}_{0.3\%}(\boldsymbol{v}_0)}, \tag{4.10}$$

[6] `https://www.charles-deledalle.fr/pages/mulog.php`

where $\mathcal{R}_q(\cdot)$ returns the q^{th} quantile of its input, and we have used the quantile values suggested in [156]. We use as the denoising operator a function composition that for input $\boldsymbol{v}$ is equal to

$$\boldsymbol{\Psi}_{\boldsymbol{v}}^{-1}(\boldsymbol{g}_{\sigma^2}(\boldsymbol{\Psi}_{\boldsymbol{v}}(\boldsymbol{v}))) \tag{4.11}$$

where $\boldsymbol{g}_{\sigma^2}$ denotes a denoising operator for noise power σ^2 with σ selected according to the tuning parameter β as

$$\sigma = \underset{\sigma \in \chi}{\arg\min} \left| \frac{\beta^{-1/2}}{\mathcal{R}_{99.97\%}(\boldsymbol{v}) - \mathcal{R}_{0.3\%}(\boldsymbol{v})} - \sigma \right|. \tag{4.12}$$

We refer to MuLoG used with the pre-trained DnCNNs as "MuLoG-CNN."

We also consider in the MuLoG framework a DnCNN trained with images from the training set used for the proposed approach. To form ground truth images appropriate for use with MuLoG, each patch is matrix log-transformed, decorrelated, then each channel is rescaled using (4.10), yielding nine single-channel images from each patch. Models are trained using this data for additive noise with standard deviations of $\{0.04, 0.08, \ldots, 0.24\}$. We use 50 epochs of stochastic gradient descent with weight decay of 0.001, a momentum of 0.9, a batch size of 64, and a learning rate beginning at 0.1 and reducing to 10^{-4} during training. The denoisers trained on PolSAR data can then be applied in the same manner as those trained on natural images, i.e. as described in (4.11) and (4.12). We refer to MuLoG with the DnCNNs trained in this manner as "MuLoG-CNN-P."

4.4.4 PolSAR Despeckling Performance Assessment

We describe in this section methods used to assess despeckling performance with consideration for speckle reduction, preservation of scene features, and preservation of polarimetric information. Assessment indexes are defined with respect to a covariance

image $\widehat{\boldsymbol{\Sigma}}$ of $N_1 \times N_2$ pixels that has been estimated from an L-look sample covariance image $\boldsymbol{Z}$ with reference to a true underlying covariance image $\boldsymbol{\Sigma}$. The symbol η is used to denote the set of pixel locations in the image.

Loss Functions

We use various loss functions in the simulated data experiments to assess the fidelity of the estimated covariance image to the true underlying image and the preservation of polarimetric information. The first is the normalized mean square error (NMSE), given by

$$\text{NMSE} = 10 \log_{10} \left(\frac{\sum_{i \in \eta} \|\widehat{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_i\|_F^2}{\sum_{i \in \eta} \|\boldsymbol{\Sigma}_i\|_F^2} \right), \tag{4.13}$$

with $\|\cdot\|_F$ denoting a Frobenius norm. We also use a quadratic loss (QL) given by [67]

$$\text{QL} = \max_{i \in \eta} \ \|\widehat{\boldsymbol{\Sigma}}_i \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{I}_3\|_2, \tag{4.14}$$

where $\|\cdot\|_2$ denotes a spectral norm. Note that both the NMSE and quadratic loss depend on all elements of the polarimetric covariance matrix. Additionally, the structural similarity (SSIM) index [161] is computed on the images corresponding to the log of the diagonal covariance matrix elements. The average SSIM of the images corresponding to the three intensity elements is recorded.

Whitened Covariance Image Evaluation

It is common in the despeckling literature to analyze the image formed by the element-wise ratio between the speckled single-channel image and the filtered image. The analysis of such ratio images is usually limited to an assessment of visible structure, the presence of which indicates an undue erosion of the underlying image

structure. For a multichannel analogue to the ratio image, we adopt a whitened image $\boldsymbol{W}$, formed as

$$\boldsymbol{W}_i = \widehat{\boldsymbol{\Sigma}}_i^{-1/2} \boldsymbol{Z}_i \widehat{\boldsymbol{\Sigma}}_i^{-1/2} \quad \text{for } i \in \eta. \tag{4.15}$$

For a scene with fully-developed speckle, a faithfully recovered covariance image results in $\boldsymbol{W}_i$ that are independently and identically distributed as

$$\boldsymbol{W}_i \sim \mathrm{CW}(\boldsymbol{W}_i; L^{-1}\boldsymbol{I}_3, L), \tag{4.16}$$

for $L \geq 3$. We introduce indexes regarding the whitened image which are intended to detect erosion of structure in the underlying image, providing an enrichment to the qualitative analysis of the ratio image.

The first is a Kolmogorov-Smirnov test computed on $\boldsymbol{W}^{jj}$ for $j \in \{1, 2, 3\}$. For fully-developed speckle, each diagonal element of the whitened image ought to independently and identically follow a gamma distribution. The Kolmogorov-Smirnov test compares the empirical distribution function of a sample with a theorized cumulative distribution function, with a smaller Kolmogorov-Smirnov statistic (KSS) indicating greater correspondence.

Second, we use a modified Box-Pierce test [162], given by

$$\mathrm{BP2} = |\eta|\big(-1 + \sum_{d_1=0}^{K}\sum_{d_2=0}^{K} \rho_{\boldsymbol{v}}^2(d_1, d_2)\big), \tag{4.17}$$

to detect autocorrelations in the whitened image with $\rho_{\boldsymbol{v}}(d_1, d_2)$ denoting the normalized sample autocorrelation function for image $\boldsymbol{v}$ at lag (d_1, d_2). When the fully developed speckle hypothesis is satisfied, the whitened image should be spatially decorrelated. The purpose of this test is to detect degradations of the spatial structure of the image from the despeckling process. In experiments, we use a maximum lag of $K = 5$.

In simulated data experiments, adherence to the speckle model is ensured by construction, and these whitened image metrics are valid everywhere. In the measured data experiments, we limit the quantitative analysis to a region of the image which appears to lack man-made media and is more likely to produce fully-developed speckle. In quantitative assessments, we report the KSS and BP2 results corresponding to $\boldsymbol{W}^{11}$.

Incoherent Decomposition Parameters

The entropy (H), anisotropy (An), and mean alpha angle ($\overline{\alpha}$) are parameters related to the eigenvalues and eigenvectors of the coherency matrix [163]. These parameters are sensitive to the type of scattering mechanism present within the resolution cell and are commonly used to compare PolSAR despeckling performance [17,43,125]. In the simulated data experiments, we report the relative absolute error between the true values and the estimates of these parameters as in [17, 43], using the median as an indicator of aggregate performance across the image.

Polarization Signatures

PolSAR data admit the synthesis of scattering coefficients corresponding to an arbitrary pair of transmit and receive polarizations. This allows the plotting of polarization signatures, 3D visualizations of the normalized scattering power density as a function of the polarizations of the transmitting and receiving antennas [164]. Because a visualization of all possible receive polarizations for each transmit polarization is difficult, it is common to plot either a co-polarization signature, in which the receive polarization is identical to the transmit polarization, or a cross-polarization

signature, in which the two are orthogonal [16, 17, 43]. We use both such plots in comparisons using simulated PolSAR data.

Measured Data Assessment

Next we list the performance indexes used exclusively on the measured PolSAR data, for which ground truth reference is unavailable. The first is the equivalent number of looks (ENL), which assesses suppression of speckle in homogeneous regions. We use a trace moment-based ENL estimator designed for PolSAR images [165]:

$$\mathrm{ENL} = \frac{\left[\,\mathrm{tr}(\langle\widehat{\boldsymbol{\Sigma}}\rangle)\right]^2}{\left\langle\,\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}})\right\rangle - \mathrm{tr}(\langle\widehat{\boldsymbol{\Sigma}}\rangle\langle\widehat{\boldsymbol{\Sigma}}\rangle)}, \tag{4.18}$$

where $\langle\cdot\rangle$ denotes a spatial ensemble average.

We use the edge-preservation detection ratio of average (EPD-ROA) [76] to assess edge-preservation ability, defined for each j^{th} diagonal entry of the covariance matrix as

$$\text{EPD-ROA} = \frac{\sum_{i\in\eta}\left|\widehat{\boldsymbol{\Sigma}}_i^{jj}/\widehat{\boldsymbol{\Sigma}}_{i+d}^{jj}\right|}{\sum_{i\in\eta}\left|\boldsymbol{Z}_i^{jj}/\boldsymbol{Z}_{i+d}^{jj}\right|} \tag{4.19}$$

with lag $d = (d_1, d_2)$. EPD-ROA values closer to one indicate better edge preservation.

We use target-to-clutter ratio (TCR) to assess the preservation of point targets during despeckling [16]. The TCR is computed over a patch containing a point scatterer amongst homogeneous clutter, and it measures the absolute difference in decibels before and after despeckling of the intensity ratio between the strong scatterer and the surrounding area. Finally, we compute the mean of ratio (MOR), which is a ratio between the mean intensity in a homogeneous area before and after despeckling. A MOR close to one indicates good radiometric preservation. The TCR, EPD-ROA, and MOR results are reported as averages over the three intensity channels.

4.5 Results

In this section, we discuss experimental results computed using both simulated and measured polarimetric SAR data which are intended to compare the proposed method to existing despeckling methods.

4.5.1 Simulated Data Experiments

We first study the performance of the compared methods on a set of 18 ground truth covariance images, each formed by the same multilooking procedure used to generate training data. These data comprise a set of 256×256 images from collections which were not used for training. From the ground truth covariance images, we simulate speckle according to (4.4) and (4.5) using $L = 1$ and $L = 4$. For each number of looks, each clean image was used to generate 10 speckled images used in experimental trials. Experiments were computed in MATLAB R2020a with a 3.2 GHZ Intel Core i5-6500 CPU and a Nvidia GTX 1060 6GB GPU. Training and application of CNNs were carried out with the MatConvNet toolbox [166]. The MuLoG codes were run in parallel using four CPU cores.

In Fig. 4.2 the despeckled results computed from a simulated single-look ($L = 1$) image are displayed in an RGB representation based on the Pauli decomposition. The image depicting the ground truth is shown in the center panel along with the original data and the results from the compared methods. A visual analysis shows that the boxcar and IDAN filters reduce speckle in homogeneous regions yet fail to preserve fine structural detail, while the extended sigma filter appears to better preserve edges and point targets. The three MuLoG results are qualitatively similar to each other with the BM3D variant suffering from over-smoothing and poorer preservation of point

targets. Each of the CNN-based methods result in an apparently faithful restoration of the ground truth covariance image, with the proposed approach providing a better preservation of texture.

The proposed approach outperforms the compared methods on both the single-look and four-look data according to each of the loss index results reported in Table 4.1 and Table 4.2. Though operating with a running time similar to the extended sigma and IDAN filters, MS-CNN provides greater than 3 dB improvement in average NMSE over either in both the $L = 1$ and $L = 4$ simulations. The MuLoG variants provide comparable performance to MS-CNN in terms of the loss indexes but are more computationally intensive. The quantitative results regarding the incoherent decomposition parameters of entropy, anisotropy, and alpha angle are shown in Table 4.3. In these assessments of polarimetric information preservation, the proposed method is likewise shown to perform best among the compared methods.

Table 4.1: Quantitative despeckling assessment results from simulated single-look ($L = 1$) PolSAR data. The mean results over ten independent realizations of the speckle process from each of 18 ground truth covariance images of 256×256 pixels are shown with standard deviations in gray.

	NMSE (dB)	QL	SSIM
Original	0.92 ±0.7	33.5 ±17.6	0.574 ±0.10
Boxcar	-3.54 ±1.2	181.0 ±231.6	0.550 ±0.15
Ext. sigma	0.44 ±0.7	91.6 ±126.7	0.664 ±0.09
IDAN	-2.57 ±0.9	15.2 ±8.2	0.568 ±0.12
MuLoG-BM3D	-5.36 ±1.1	24.9 ±11.4	0.707 ±0.11
MuLoG-CNN	-5.23 ±1.1	27.2 ±12.2	0.722 ±0.10
MuLoG-CNN-P	-5.28 ±1.1	27.1 ±13.1	0.727 ±0.10
MS-CNN	-5.83 ±1.0	15.0 ±8.0	0.762 ±0.08

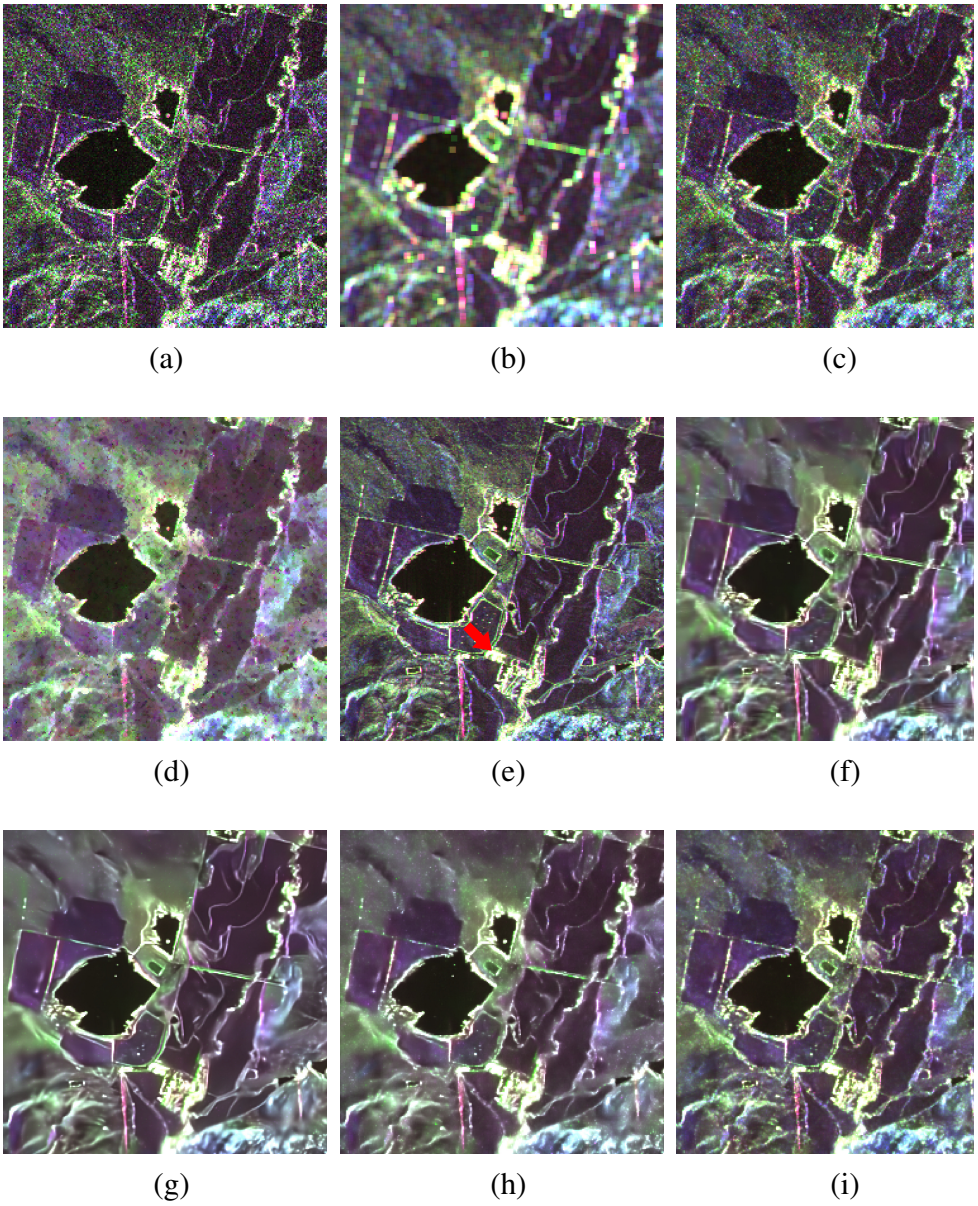


Figure 4.2: Comparison of despeckling algorithms using a 256 × 256 pixel simulated single-look PolSAR image [6]. The results are displayed in an RGB representation based on the Pauli decomposition. (a) Original data. (b) Boxcar. (c) Extended sigma. (d) IDAN. (e) Ground truth. (f) MuLoG-BM3D. (g) MuLoG-CNN. (h) MuLoG-CNN-P. (i) MS-CNN.

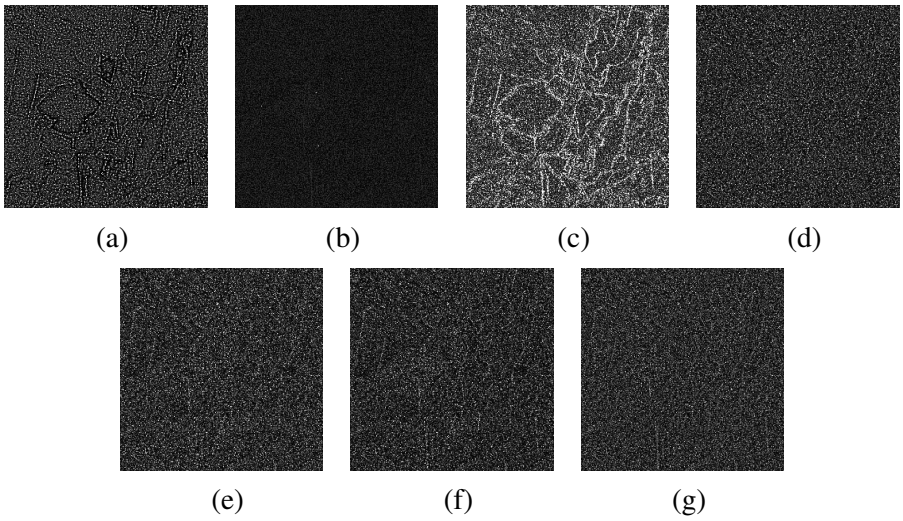


Figure 4.3: Whitened images from the simulated single-look PolSAR data. (a) Boxcar. (b) Extended sigma. (c) IDAN. (d) MuLoG-BM3D. (e) MuLoG-CNN. (f) MuLoG-CNN-P. (g) MS-CNN.

Table 4.2: Quantitative despeckling assessment results from simulated four-look ($L = 4$) PolSAR data. The mean results over ten independent realizations of the speckle process from each of 18 ground truth covariance images of 256×256 pixels are shown with standard deviations in gray.

	NMSE (dB)	QL	SSIM	Time (s)
Original	-4.85 ±0.6	14.3 ±7.6	0.820 ±0.06	-
Boxcar	-4.15 ±1.4	138.2 ±171.0	0.562 ±0.15	-
Ext. sigma	-4.54 ±0.7	29.9 ±29.9	0.844 ±0.05	0.4
IDAN	-4.39 ±1.1	12.9 ±4.7	0.701 ±0.09	0.4
MuLoG-BM3D	-7.39 ±1.0	12.2 ±6.9	0.864 ±0.05	24.4
MuLoG-CNN	-7.17 ±1.1	12.0 ±7.0	0.873 ±0.05	10.4
MuLoG-CNN-P	-7.44 ±0.9	12.9 ±9.7	0.870 ±0.05	10.9
MS-CNN	-8.28 ±0.7	11.7 ±8.1	0.876 ±0.05	0.2

Fig. 4.3 shows the $\boldsymbol{W}^{11}$ element of the whitened images produced via (4.15). Structure is visible in the boxcar, extended sigma, and IDAN whitened images, indicating a loss of spatial detail from the despeckling process. The whitened images from

Table 4.3: Quantitative polarimetric information preservation assessment results from simulated PolSAR data. The mean results over ten independent realizations of the speckle process from each of 18 ground truth covariance images of 256×256 pixels are shown.

	Single-look					Four-look				
	KSS	BP2	H	An	$\overline{\alpha}$	KSS	BP2	H	An	$\overline{\alpha}$
Boxcar	0.074	1581.0	0.097	0.855	0.113	0.186	3506.8	0.081	0.709	0.092
Ext. sigma	0.305	717.2	0.342	1.100	0.185	0.349	399.2	0.153	1.031	0.122
IDAN	0.163	1009.1	0.089	0.887	0.097	0.150	1500.3	0.073	0.757	0.082
MuLoG-BM3D	0.035	270.1	0.082	0.777	0.081	0.036	282.9	0.073	0.734	0.076
MULoG-CNN	0.026	148.2	0.085	0.775	0.082	0.067	1779.5	0.074	0.724	0.080
MuLoG-CNN-P	0.026	123.7	0.083	0.772	0.082	0.073	168.5	0.074	0.732	0.080
MS-CNN	0.035	268.7	0.073	0.676	0.077	0.074	625.6	0.061	0.602	0.065

MuLoG show very little visible structure, while apparent structure is visible within the road at the bottom of the MS-CNN image. The visual analysis is supported by the modified Box-Pierce test results in Table 4.3, which indicate significant autocorrelation in the whitened images from the boxcar, extended sigma, and IDAN methods in the $L = 1$ case. MuLoG-CNN-P exhibits the least autocorrelation according to the Box-Pierce test for any of the methods in both the $L = 1$ and $L = 4$ experiments. The Kolmogorov-Smirnov statistics in Table 4.3 show the MuLoG variants to adhere best to the theorized statistical properties of the whitened image, followed by MS-CNN.

In Fig. 4.4, co-polarization and cross-polarization signatures computed from the average covariance matrix within the 7×6 pixel region identified by the red arrow in Fig. 4.2 are shown for a subset of the compared despeckling methods. It is notable that the MS-CNN signatures, both co-polarization and cross-polarization, are in closer agreement with the respective ground truth signatures than those of the other methods shown, indicating preservation of polarimetric information. Of the methods

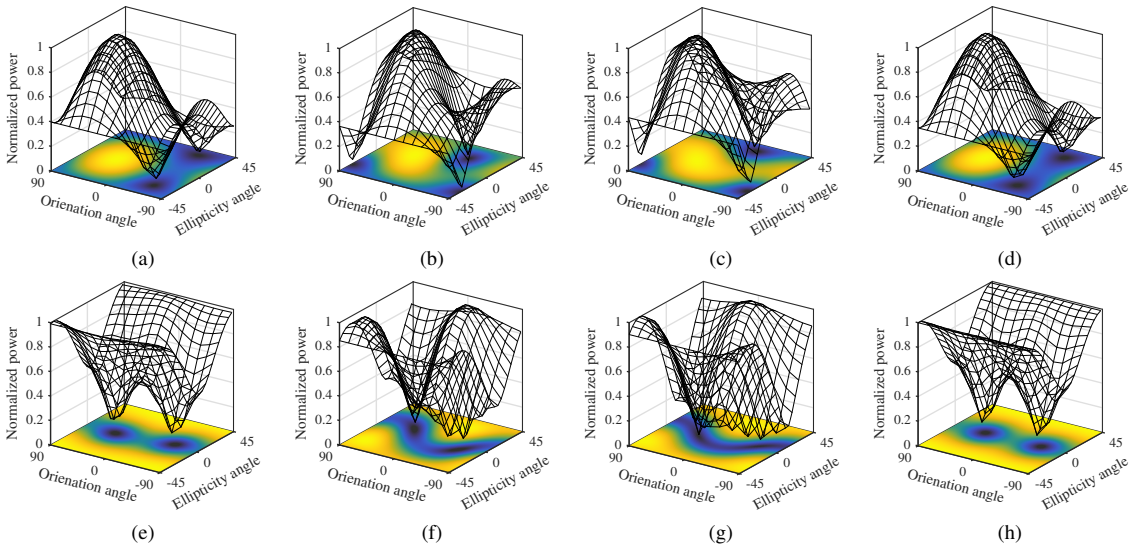


Figure 4.4: Co-polarization (top row) and cross-polarization (bottom row) signatures computed from a small region of a simulated single-look PolSAR image. The ellipticity and orientation angles of the transmitting antenna are given in degrees. (a,e) Ground truth. (b,f) Extended sigma. (c,g) MuLoG-CNN. (d,h) MS-CNN.

not displayed, IDAN produced signatures distinctly different from the ground truth, and the other MuLoG variants produced results similar to those from MuLoG-CNN.

4.5.2 Measured Data Experiments

We next compare the despeckling approaches using measured PolSAR images from the Pi-SAR sensor. Pi-SAR, an L-band airborne PolSAR system with a 3 m resolution in both azimuth and slant range, was developed by the former Communications Research Laboratory and NASDA (now JAXA) of Japan [80, 167]. A single-look 512×512 pixel PolSAR image from data collected by Pi-SAR over Tsukuba, Japan is shown in Fig. 4.5 with boxes indicating regions in which specific performance indexes are computed: white boxes for ENL and MOR, green for TCR, yellow for EPD-ROA, and blue for the whitened image assessments. To allow a detailed comparison, a

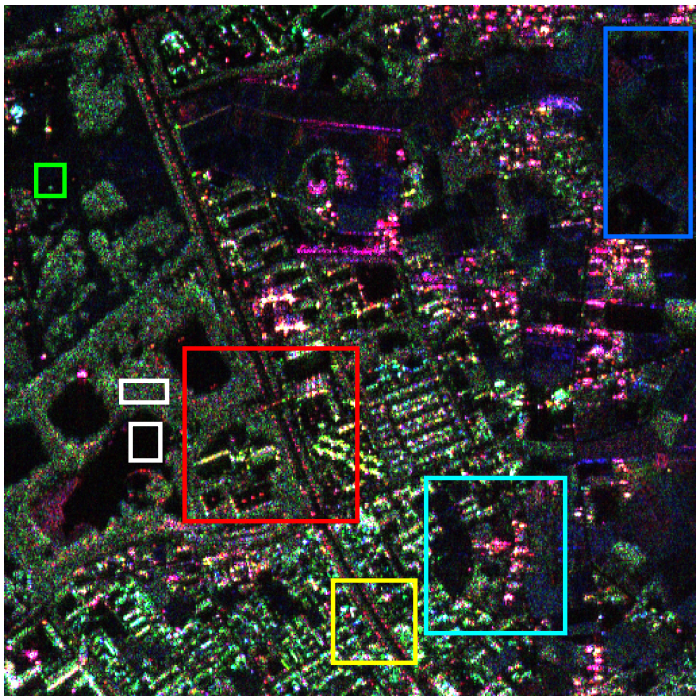


Figure 4.5: Single-look PolSAR image of 512 × 512 pixels from the Pi-SAR sensor used in despeckling comparison. The boxes indicate regions in which performance indexes are computed: white boxes for ENL and MOR, green for TCR, yellow for EPD-ROA, and blue for the whitened image assessments. The region within the red box is depicted in Fig. 4.6.

128×128 pixel region cropped from within the red box in Fig. 4.5 is shown in Fig. 4.6 along with the despeckled images from that region. Despeckling results from a four-look 200×200 pixel Pi-SAR image of Niigata City, Japan are shown in Fig. 4.7. The quantitative results computed from the two images are shown in Table 4.4 and Table 4.5.

From the cropped images in Fig. 4.6, it is clear that the MuLoG variants provide significant speckle suppression in homogeneous regions. MS-CNN provides a less smooth image with a greater preservation of texture within the vegetated regions. The ENL results show the CNN-based MuLoG variants to provide the greatest speckle

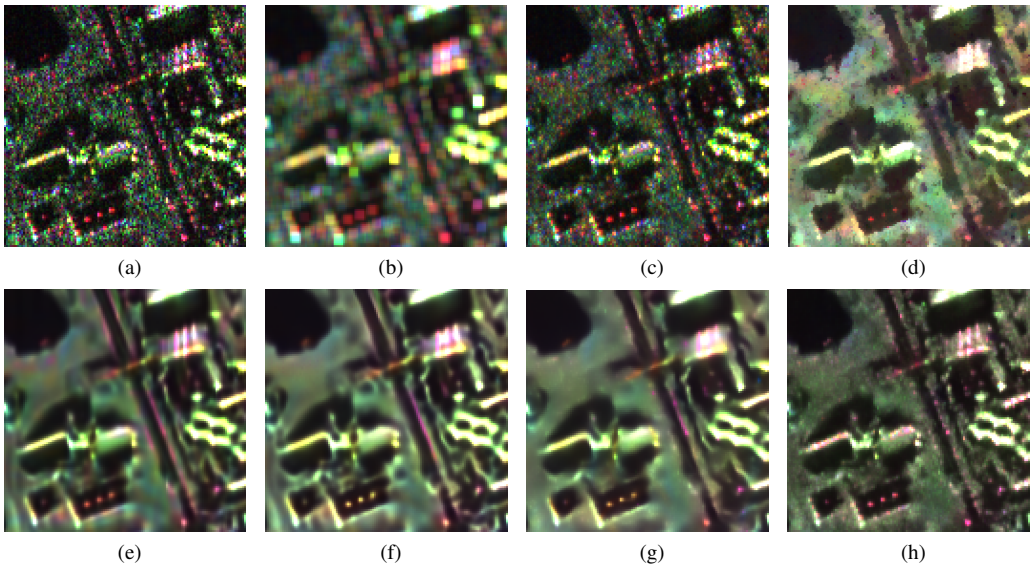


Figure 4.6: Detailed comparison of the despeckling algorithms using a 128 × 128 region cropped from Fig. 4.5. (a) Original data. (b) Boxcar. (c) Extended sigma. (d) IDAN. (e) MuLoG-BM3D. (f) MuLoG-CNN. (g) MuLoG-CNN-P. (h) MS-CNN.

suppression on the single-look data, followed by MS-CNN, which reduces speckle significantly more than the conventional methods. IDAN returns the highest ENL on the four-look data followed by the proposed method. Lacking a clean reference image, it is difficult to judge the correct balance between speckle suppression and textural preservation.

Notably, the extended sigma filter appears to preserve point targets more than the other methods, while the IDAN filter obscures point scatterers present in the original image. The MuLoG variants tend to smear point scatterers more than either the extended sigma filter or MS-CNN. The "EPD-HD" and "EPD-VD" columns in Table 4.4 and Table 4.5 show the EPD-ROA computed using immediate horizontally and vertically neighboring pixels, respectively. The EPD-ROA results show the extended sigma filter and MS-CNN to best preserve edges in both the single-look and

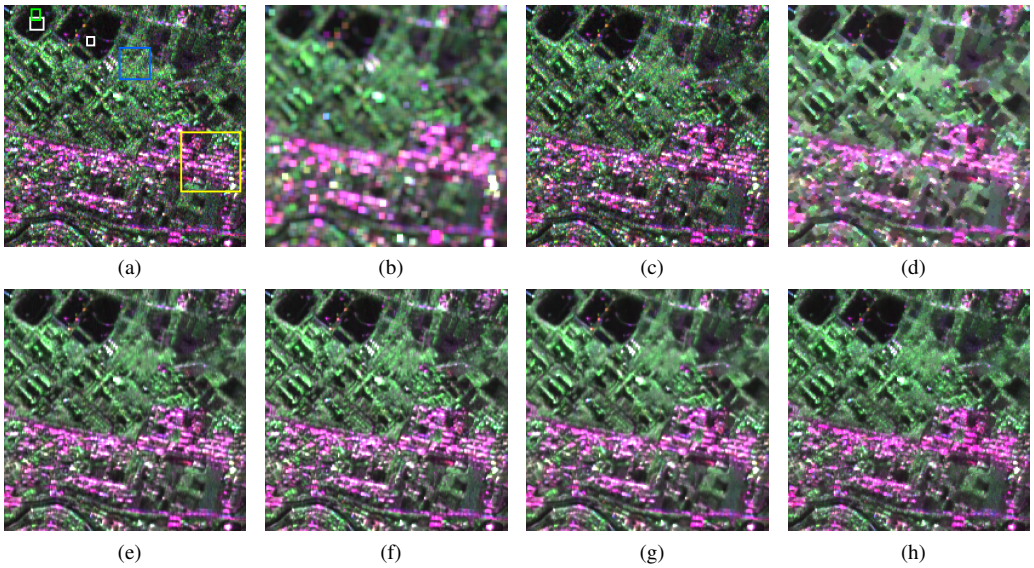


Figure 4.7: Comparison of despeckling methods using a four-look PolSAR image of 200×200 pixels. The boxes in (a) indicate regions in which performance indexes are computed: white boxes for ENL and MOR, green for TCR, yellow for EPD-ROA, and blue for the whitened image assessments. (a) Original data. (b) Boxcar. (c) Extended sigma. (d) IDAN. (e) MuLoG-BM3D. (f) MuLoG-CNN. (g) MuLoG-CNN-P. (h) MS-CNN.

four-look images, and the TCR and MOR indexes show largely mixed results. The Kolmogorov-Smirnov statistics in Table 4.4 and Table 4.5 show MS-CNN to underperform all methods but IDAN, indicating substandard fit to the theorized distribution. The extended sigma filter and MuLoG-BM3D both show a lower modified Box-Pierce statistic than the proposed method.

The $\boldsymbol{W}^{11}$ element of the whitened images formed from the region bound by the cyan box in the single-look image in Fig. 4.5 are shown in Fig. 4.8. Spatial structure is visible in the whitened images from each method, most notably from IDAN. The whitened images from the MuLoG variants appear similar to that of the proposed method, with each having small constant areas. Note that the simulated data

Table 4.4: Quantitative despeckling assessment results from the measured PolSAR data ($L = 1$).

	ENL	EPD-HD	EPD-VD	TCR	MOR	KSS	BP2
Original	2.3	-	-	-	-	-	-
Boxcar	15.0	0.33	0.30	6.32	1.01	0.067	554.7
IDAN	35.3	0.33	0.30	6.21	0.94	0.120	727.9
Ext. sigma	3.9	0.41	0.40	0.71	1.00	0.313	346.0
MuLoG-BM3D	61.6	0.34	0.30	2.84	1.03	0.068	422.8
MuLoG-CNN	110.1	0.34	0.29	1.99	1.03	0.071	621.2
MuLoG-CNN-P	170.5	0.33	0.29	2.02	1.00	0.096	707.5
MS-CNN	89.3	0.35	0.31	1.97	1.00	0.100	480.3

Table 4.5: Quantitative despeckling assessment results from the measured PolSAR data ($L = 4$).

	ENL	EPD-HD	EPD-VD	TCR	MOR	KSS	BP2
Original	3.9	-	-	-	-	-	-
Boxcar	9.1	0.64	0.54	7.62	1.12	0.084	217.3
IDAN	51.5	0.71	0.55	5.68	0.94	0.095	159.4
Ext. sigma	4.9	0.90	0.82	0.11	0.99	0.358	162.6
MuLoG-BM3D	22.3	0.77	0.64	1.33	0.94	0.061	122.4
MuLoG-CNN	23.6	0.78	0.64	0.41	0.94	0.090	107.1
MuLoG-CNN-P	33.8	0.75	0.61	0.98	0.94	0.057	163.8
MS-CNN	38.1	0.79	0.67	2.07	0.89	0.087	152.0

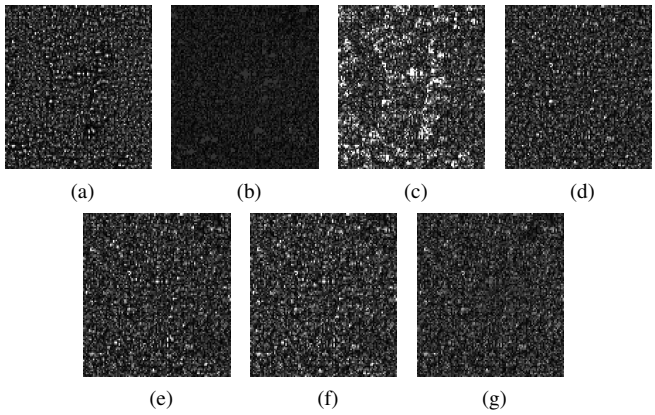


Figure 4.8: Whitened single-look PolSAR images from the region bound by the cyan box in Fig. 4.5. (a) Boxcar. (b) Extended sigma. (c) IDAN. (d) MuLoG-BM3D. (e) MuLoG-CNN. (f) MuLoG-CNN-P. (g) MS-CNN.

experiments contain only fully-developed speckle by construction, and thus spatial correlations in the whitened images indicated an undue erosion of polarimetric information by the despeckling process. For measured data, this is not necessarily the case, e.g., with the presence of man-made media.

Fig. 4.9 shows the estimated entropy of the four-look image from a subset of the considered despeckling methods. The entropy image from the extended sigma filter has a granular character similar to that of the original data. Compared to the original image and the extended sigma filter result, both MuLoG and MS-CNN yield high entropy values within the vegetated regions, indicating a more random scattering process. The MuLoG and MS-CNN results are quite similar with MuLoG appearing smoother and MS-CNN better preserving small low entropy regions. The estimated alpha angle images in Fig. 4.10 show an extended sigma filter result that again appears similar to the original data. The MuLoG image is much smoother with apparent structure over the man-made media. The result from the proposed

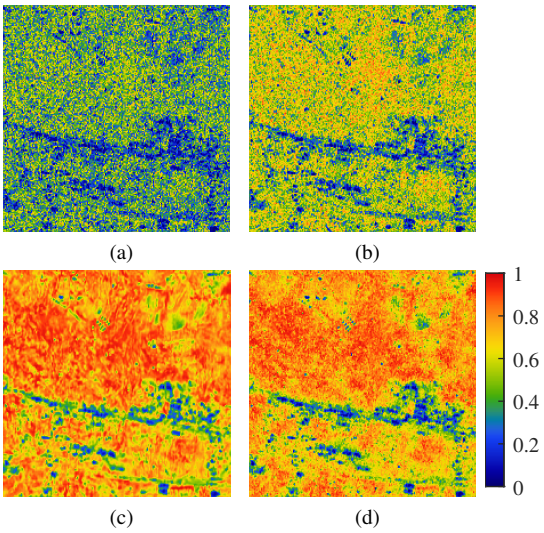


Figure 4.9: Estimated entropy from the four-look measured data. (a) Original data. (b) Extended sigma. (c) MuLoG-CNN. (d) MS-CNN.

method is also smooth compared to the original data with low-angle regions that are less contiguous than the MuLoG image. A satisfying interpretation of these results is difficult due to the lack of ground truth. Note that the MuLoG variants not shown were found to produce results similar to MuLoG-CNN in estimates of both entropy and mean alpha angle.

4.6 Conclusion

4.6.1 Summary

In this chapter, we have proposed a CNN-based polarimetric SAR despeckling approach that operates in the matrix logarithm domain and uses a residual learning architecture that has been successful in natural image denoising. We have implemented this approach with a training set of spatially multilooked PolSAR images that are corrupted by a simulated speckle process.

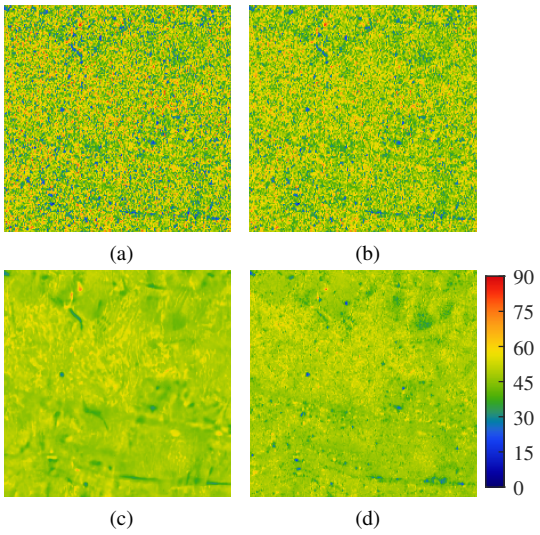


Figure 4.10: Estimated alpha angle (in degrees) from the four-look measured data. (a) Original data. (b) Extended sigma. (c) MuLoG-CNN. (d) MS-CNN.

In experimental results on both simulated and measured PolSAR data, the proposed method has demonstrated improved scene feature preservation and comparable despeckling performance compared to state-of-the-art methods. Further, the proposed approach provides improved despeckling over conventional filtering methods and operates with a comparably fast running time afforded by the use of CNNs.

4.6.2 Future Work

A continuation of the work in this chapter might entail application to other multichannel SAR imaging modalities. Alternative methods of procuring unspeckled SAR data for training purposes might also be considered, e.g., using temporally multilooked images. Another avenue for future work is the use of complex-valued CNNs

in a similar approach; accounting for the relationship between the real and imaginary components of the covariance matrix in this way could improve despeckling performance and allow for training with fewer samples.

Chapter 5: Maximum Likelihood Estimation in Mixed Integer Linear Models

5.1 Introduction

In many sensing and communication applications, an unknown quantity is encoded in the phase of a signal. Because phase is unambiguous only on $[0, 2\pi)$, a measurement may experience an integer number of phase wraps. Let $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{Q}$ denote the fields of real numbers, integers, and rational numbers, respectively. To estimate the quantity of interest, we consider a linear model with an observation vector $\boldsymbol{y} \in \mathbb{R}^m$ that is related to parameter vectors $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{k} \in \mathbb{Z}^m$ as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{M}\boldsymbol{k} + \boldsymbol{u}. \tag{5.1}$$

The encoding matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full column rank and no all-zero row, $\boldsymbol{M} \in \mathbb{R}^{m \times m}$ is nonsingular, and the zero-mean noise $\boldsymbol{u} \in \mathbb{R}^m$ is assumed to be Gaussian with covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$. The model in (5.1) arises in applications such as MIMO decoding [26–28], frequency estimation [22–24], ranging [18–21], interferometry [25], direction of arrival estimation [29], multichannel modulo sampling [30], and magnetic resonance imaging [9, 31, 168].

The maximum likelihood (ML) estimate of $\boldsymbol{x}$ and $\boldsymbol{k}$ in (5.1) is given by

$$\{\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{k}}\} = \underset{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{k}\in\mathbb{Z}^m}{\arg\min}\ \mathcal{L}_{x,k}(\boldsymbol{x}; \boldsymbol{y}), \tag{5.2}$$

where

$$\mathcal{L}_{x,k}\left(\boldsymbol{x}, \boldsymbol{k}\,; \boldsymbol{y}\right) = \left\| \boldsymbol{\Sigma}^{-1/2} \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \boldsymbol{M}\boldsymbol{k}\right) \right\|^2. \tag{5.3}$$

For general $\boldsymbol{A}$, $\boldsymbol{M}$, and covariance $\boldsymbol{\Sigma}$, (5.2) is an NP-hard mixed-integer quadratic program; for $\boldsymbol{M}^{-1}\boldsymbol{A} \in \mathbb{Q}^{m\times n}$, (5.2) is equivalent to a *closest lattice point problem* (CPP) [169, 170]. As such, it can be solved with Voronoi cell algorithms [171] or methods known as sphere decoders [172,173]. But these methods have non-polynomial worst-case computational complexity [174], and the computation time depends on the random noise realization. Polynomial-time, and even linear-time, algorithms for the CPP problem have long been known for certain lattices [175–177]. In [32], a special case of (5.1) is considered, for which (5.2) can be solved in polynomial time by leveraging the classical Chinese remainder theorem (CRT). The case in [32] has that $n = 1$, $\boldsymbol{\Sigma}$ and $\boldsymbol{M}$ are diagonal, and $\boldsymbol{M}^{-1}\boldsymbol{A}$ meets a certain coprime restriction. In some applications, this coprime restriction limits potential data acquisition strategies, and relaxation of the restriction has been shown to enable design of $\boldsymbol{A}$ with improved estimation error variance [21, 178, 179].

In this chapter[7], we first survey the existing literature, then provide background on the CRT and lattices. We present two algorithms for the problem in (5.2) that, for fixed n, have worst-case computational complexity that is polynomial in m. The

[7]This chapter is largely drawn from [7] (© 2023 IEEE) and [180], coauthored with Shen Zhao and Lee C. Potter. Manuscript [180] was supported in part by FA238523C0002 and has been granted public release, PA# AFRL-2024-1975. Distribution Statement A. Approved for public release: distribution is unlimited. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

two polynomial-time methods require that $\boldsymbol{\Sigma}$ and $\boldsymbol{M}$ be diagonal, as in [32]. First, we describe a solution method that allows for a multivariate unknown $\boldsymbol{x} \in \mathbb{R}^n$, but the method applies to a restricted class of rational $\boldsymbol{A}$ matrices. The solution, similar to its existing counterpart for $n = 1$ [32], is found in two parts: one part real-valued and the other integer-valued. The real part is found by constructing and evaluating a short list of candidate solutions. The integer part is found through a multivariate extension of the CRT [181, 182]. The second method lifts the restriction on $\boldsymbol{A}$ and allows arbitrary rational $\boldsymbol{A}$ with full column rank; this method works by simply evaluating a list of candidate solutions. We then provide a method that applies more generally to the problem in (5.2). This approach relies on sphere decoding and requires only that $\boldsymbol{M}^{-1}\boldsymbol{A}$ be rational. Unlike the other methods in this chapter, this more general method does not require that the covariance matrix $\boldsymbol{\Sigma}$ be diagonal, thus accommodating correlated noise. Application of the theory is illustrated in computed examples. Most proofs are relegated to appendices.

5.2 Prior Art

For $n = 1$, methods based on the CRT have been used to address special cases of (5.1) [32, 183–185]. Alternatively, parameter estimation involving (5.1) can be addressed by solving a closest lattice point problem [169, 170, 179]. Though, to use conventional algorithms such as sphere decoding, it is necessary to find a basis for a lattice generated by a rank-deficient matrix. Li et al. [20] study the $n = 1$ case of (5.1) with $\boldsymbol{M} = \boldsymbol{I}$, $\boldsymbol{\Sigma} = \boldsymbol{I}$, and $\boldsymbol{A} = [\,\Gamma_1^{-1}, \Gamma_2^{-1}, \ldots, \Gamma_m^{-1}\,]^{\mathsf{T}}$, where the elements of $\boldsymbol{\Gamma}$ are pairwise coprime integers (wavelengths, in ranging); explicit construction of a lattice basis is given. The case considered in [20] is broadened in [21] to allow for

arbitrary integer entries in $\boldsymbol{\Gamma}$. The lattice basis is explicitly derived in terms of the entries of $\boldsymbol{A}$, with a construction that relies on unimodular extension of $\boldsymbol{A}$. Though, for $n > 1$, the unimodular extension does not, in general, exist [186].

The CRT-based estimator proposed in [32] applies to the $n = 1$ case of (5.1) with diagonal $\boldsymbol{\Sigma}$ and $\boldsymbol{M}$ such that $\boldsymbol{M}^{-1}\boldsymbol{A} = \alpha[\,\Gamma_1^{-1}, \Gamma_2^{-1}, \ldots, \Gamma_m^{-1}\,]^{\mathsf{T}}$, for some $\alpha \in \mathbb{R}$, with $\boldsymbol{\Gamma} \in \mathbb{Z}^m$ having pairwise coprime entries. Thus, the method in [32] accommodates non-white, uncorrelated noise. In [187], the $n = 1$ case of (5.1) is considered with arbitrary covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{M}^{-1}\boldsymbol{A} = [\,\Gamma_1^{-1}, \Gamma_2^{-1}, \ldots, \Gamma_m^{-1}\,]^{\mathsf{T}}$, with $\boldsymbol{\Gamma} \in \mathbb{Z}^m$ having pairwise coprime entries. Hence, [187] accommodates correlated noise. In [188, 189], (5.1) is studied for modulo sampling in the noiseless case with $\boldsymbol{M} = \boldsymbol{I}$, sparse structure in $\boldsymbol{x}$, and a restriction on $\boldsymbol{A}$. In [190], a method from [177] is modified to solve the special case of (5.2) with $n = 1$, diagonal $\boldsymbol{\Sigma}$, and $\boldsymbol{M}^{-1}\boldsymbol{A}$ not subject to a coprime restriction.

The general problem in (5.1) for $n > 1$ is widely addressed for MIMO communications. There, the intrinsic rank deficiency is overcome by regularization or, equivalently, maximum *a posteriori* probability (MAP) estimation [19, 26–28]. The problem in (5.2) is equivalent to the CPP [7]

$$\arg\min_{\boldsymbol{k}\in\mathbb{Z}^m} \|\boldsymbol{B}\boldsymbol{k} - \boldsymbol{B}\boldsymbol{y}\|^2, \tag{5.4}$$

for $\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{M} - \boldsymbol{E}\boldsymbol{E}^{\mathsf{T}}$ and $\boldsymbol{E} = \boldsymbol{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1/2}$. The CPP solution method in [177] requires $\boldsymbol{B}$ to be nonsingular, and its complexity depends on the eigenstructure of $\boldsymbol{B}^T\boldsymbol{B}$. However, $\boldsymbol{B}$ is inherently singular, and thus, [177] does not apply to our problem.

5.3 Background

5.3.1 Lattices

For $\boldsymbol{Q} \in \mathbb{R}^{c \times d}$ with full column rank, the set

$$\Lambda_{\boldsymbol{Q}} = \{\boldsymbol{Q}\boldsymbol{k} \mid \boldsymbol{k} \in \mathbb{Z}^d\} \tag{5.5}$$

is the c-dimensional *lattice* generated by the *lattice basis* $\boldsymbol{Q}$ [191]. The integer d is the rank of the lattice. All lattice points $\boldsymbol{x} \in \Lambda_{\boldsymbol{Q}}$ admit a unique representation, $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{k}$, for $\boldsymbol{k} \in \mathbb{Z}^d$ [192]. If $\boldsymbol{Q}$ is invertible ($c = d$), $\Lambda_{\boldsymbol{Q}}$ is called a *full-rank* lattice. The set

$$\mathcal{P}_{\boldsymbol{Q}} = \{\boldsymbol{Q}\boldsymbol{\lambda} \mid \boldsymbol{\lambda} \in [0, 1)^n\} \tag{5.6}$$

is called the *fundamental parallelotope* [193] of $\Lambda_{\boldsymbol{Q}}$ for $\boldsymbol{Q}$. A fundamental parallelotope of $\Lambda_{\boldsymbol{Q}}$ contains no lattice points of $\Lambda_{\boldsymbol{Q}}$ except for its vertices.

A square matrix $\boldsymbol{U} \in \mathbb{Z}^{c \times c}$ is called *unimodular* if it has integer entries and $\det(\boldsymbol{U}) = \pm 1$, where $\det(\cdot)$ denotes the determinant of a matrix. The following are equivalent [194, Thm. 4.3].

(i) $\boldsymbol{U}$ is unimodular.

(ii) $\boldsymbol{U}^{-1}$ is unimodular.

(iii) The lattice generated by the columns of $\boldsymbol{U}$ is $\mathbb{Z}^c$.

(iv) The Hermite normal form of $\boldsymbol{U}$ is $\boldsymbol{I}_c$.

(v) $\boldsymbol{U}$ can be transformed into $\boldsymbol{I}_c$ by elementary column operations.

A lattice basis $\boldsymbol{Q}$ is not unique for a given lattice and, hence, neither is the corresponding fundamental parallelotope $\mathcal{P}_{\boldsymbol{Q}}$. We can express any lattice basis for $\Lambda_{\boldsymbol{Q}}$ as $\boldsymbol{Q}\boldsymbol{U}$, for some unimodular matrix $\boldsymbol{U}$, and $\Lambda_{\boldsymbol{Q}} = \Lambda_{\boldsymbol{Q}\boldsymbol{U}}$, for any unimodular $\boldsymbol{U}$.

The volume of a parallelotope, denoted by $\det \Lambda_{\boldsymbol{Q}} \triangleq \mathrm{vol}(\mathcal{P}_{\boldsymbol{Q}}) = \sqrt{\det(\boldsymbol{Q}^T\boldsymbol{Q})}$, is the same for all lattice bases. If $c = d$, the operation that translates $\boldsymbol{x} \in \mathbb{R}^d$ onto a fundamental parallelotope $\mathcal{P}_{\boldsymbol{Q}}$ is

$$\langle \boldsymbol{x} \rangle_{\boldsymbol{Q}} \triangleq \boldsymbol{x} - \boldsymbol{Q}\lfloor \boldsymbol{Q}^{-1}\boldsymbol{x} \rfloor, \tag{5.7}$$

where $\lfloor \cdot \rfloor$ returns the greatest integer less than or equal to its input, acting element-wise when applied to a vector.

5.3.2 Hermite Normal Form

Definition 2. *An integer matrix $\boldsymbol{H} \in \mathbb{Z}^{c\times d}$ is said to be in* Hermite normal form *(HNF) if it has entries that satisfy the following:*

$$\begin{aligned} h_{i,i} &> 0, \quad \forall i \\ h_{i,j} &= 0, \quad \forall i > j \\ 0 \le h_{i,j} &< h_{i,i}, \quad \forall i < j. \end{aligned} \tag{5.8}$$

A matrix $\boldsymbol{H}$ in HNF is upper triangular and has full column rank; it can be expressed as $\boldsymbol{H} = [\widetilde{\boldsymbol{H}}^{\mathsf{T}}, \boldsymbol{0}_{d,(c-d)}]^{\mathsf{T}}$, where $\widetilde{\boldsymbol{H}} \in \mathbb{Z}^{d\times d}$ is a nonsingular, upper triangular matrix.[8] Every full column rank integer matrix $\boldsymbol{T} \in \mathbb{Z}^{c\times d}$ has unique decomposition $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{T}$, with unimodular $\boldsymbol{U}$ and $\boldsymbol{H}$ in HNF [194, Cor. 4.2a]. The HNF and the corresponding unimodular transformation can be computed in polynomial time [191, p. 149].

5.3.3 Chinese Remainder Theorems

The classical CRT for a single unknown variable is as follows [195, 196].

[8] The HNF is sometimes defined to be lower triangular or according to various alternative restrictions on the matrix entries.

Theorem 3 (Chinese remainder theorem)**.** *Let* $\boldsymbol{b} \in \mathbb{Z}^m$*. Let* $\boldsymbol{c} \in \mathbb{Z}^m$ *have pairwise coprime entries. The system of linear congruences given by*

$$\mathbf{1}_m x \equiv \boldsymbol{b} \mod \boldsymbol{c} \tag{5.9}$$

has integer solutions, and any two solutions are congruent modulo $\prod_{i=1}^{m} c_i$*.*

There are solution methods to the univariate Chinese remaindering problem with complexity $O(m^2)$ [195, p. 23].

The CRT has been extended to accommodate multivariate unknowns [181, 182]. The following is a direct implication of the multivariate CRT in [182]:

Theorem 4 (Multivariate CRT)**.** *Let* $\boldsymbol{W} \in \mathbb{Z}^{m \times n}$*. Let* $\boldsymbol{c} \in \mathbb{Z}^m$ *have pairwise coprime entries. The system of linear congruences given by*

$$\boldsymbol{W}\boldsymbol{x} \equiv \boldsymbol{b} \mod \boldsymbol{c} \tag{5.10}$$

has integer solutions for all $\boldsymbol{b} \in \mathbb{Z}^m$ *iff*

$$\gcd([g_{i,1}, \ldots, g_{i,n}, c_i]^{\mathsf{T}}) = 1, \tag{5.11}$$

for each $i \in \{1, \ldots, m\}$*, where* $\gcd(\cdot)$ *is the positive column-wise greatest common divisor of an integer matrix input.*

Lemma 5. *Let* $\boldsymbol{W} \in \mathbb{Z}^{m \times n}$*;* $\boldsymbol{c} \in \mathbb{Z}^m$*;* $\boldsymbol{b} \in \mathbb{Z}^m$*; and* $\widetilde{\boldsymbol{W}} = [\boldsymbol{W}, \operatorname{diag}(\boldsymbol{c})]^{\mathsf{T}} \in \mathbb{Z}^{(m+n) \times m}$*. Let* $\boldsymbol{H} = [\widetilde{\boldsymbol{H}}^{\mathsf{T}}, \mathbf{0}_{m,n}]^{\mathsf{T}}$ *be the HNF of* $\widetilde{\boldsymbol{W}}$*. The system of congruences*

$$\boldsymbol{W}\boldsymbol{x} \equiv \boldsymbol{b} \mod \boldsymbol{c} \tag{5.12}$$

has integer solutions iff $\widetilde{\boldsymbol{H}}^{-\mathsf{T}}\boldsymbol{b} \in \mathbb{Z}^m$*.*

Proof. The system of linear congruences in (5.12) has integer solutions iff there are $\boldsymbol{k} \in \mathbb{Z}^m$ and $\boldsymbol{x} \in \mathbb{Z}^n$ such that $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{b} - \boldsymbol{c} \circ \boldsymbol{k}$ or, equivalently, if there is $\widetilde{\boldsymbol{k}} \in \mathbb{Z}^{m+n}$ such that $\widetilde{\boldsymbol{W}}^{\mathsf{T}}\widetilde{\boldsymbol{k}} = \boldsymbol{b}$. Here, $\circ$ denotes the element-wise product. From [194, Cor. 5.3b], $\widetilde{\boldsymbol{W}}^{\mathsf{T}}\widetilde{\boldsymbol{k}} = \boldsymbol{b}$ has an integer solution iff $\widetilde{\boldsymbol{H}}^{-\mathsf{T}}\boldsymbol{b} \in \mathbb{Z}^m$. □

The following multivariate CRT is more general than Theorem 4 as it does not assume that the moduli are pairwise coprime:

Theorem 6 (Generalized multivariate CRT)**.** *Let* $\boldsymbol{W} \in \mathbb{Z}^{m\times n}$*;* $\boldsymbol{c} \in \mathbb{Z}^m$*; and* $\widetilde{\boldsymbol{W}} = [\boldsymbol{W}, \operatorname{diag}(\boldsymbol{c})]^{\mathsf{T}} \in \mathbb{Z}^{(m+n)\times m}$*. Let* $\boldsymbol{H} = [\widetilde{\boldsymbol{H}}^{\mathsf{T}}, \boldsymbol{0}_{m,n}]^{\mathsf{T}}$ *be the HNF of* $\widetilde{\boldsymbol{W}}$*. The system of congruences*

$$\boldsymbol{W}\boldsymbol{x} \equiv \boldsymbol{b} \mod \boldsymbol{c} \tag{5.13}$$

has integer solutions for all $\boldsymbol{b} \in \mathbb{Z}^m$ *iff* $\widetilde{\boldsymbol{H}} = \boldsymbol{I}_m$*.*

Proof. From Lemma 5, (5.13) has integer solutions for all $\boldsymbol{b} \in \mathbb{Z}^m$ iff $\widetilde{\boldsymbol{H}}^{-\mathsf{T}}\boldsymbol{b} \in \mathbb{Z}^m$ for all $\boldsymbol{b} \in \mathbb{Z}^m$, which is true iff $\widetilde{\boldsymbol{H}}^{-\mathsf{T}} \in \mathbb{Z}^{m\times m}$. As $\widetilde{\boldsymbol{H}}^{\mathsf{T}} \in \mathbb{Z}^{m\times m}$, $\widetilde{\boldsymbol{H}}^{-\mathsf{T}} \in \mathbb{Z}^{m\times m}$ iff $\widetilde{\boldsymbol{H}}^{\mathsf{T}}$, and hence $\widetilde{\boldsymbol{H}}$, is unimodular. Because $\widetilde{\boldsymbol{H}}$ is upper-triangular, its determinant is the product of its diagonal entries. Thus, if $\widetilde{\boldsymbol{H}}$ is unimodular, its diagonal entries are ones. By the definition of the HNF, the off-diagonal entries of $\widetilde{\boldsymbol{H}}$ are zeros. Therefore, if $\widetilde{\boldsymbol{H}}$ is unimodular, then $\widetilde{\boldsymbol{H}} = \boldsymbol{I}_m$. The converse is also true. Therefore, $\widetilde{\boldsymbol{H}}$ is unimodular iff $\widetilde{\boldsymbol{H}} = \boldsymbol{I}_m$.

□

A method is given in Algorithm 7 for computing an integer solution to a multivariate system of congruences.

Algorithm 7 Solution to a system of linear congruences with a multivariate unknown.

Input: $\boldsymbol{b} \in \mathbb{Z}^m$, $\boldsymbol{c} \in \mathbb{Z}^m$, and $\boldsymbol{W} \in \mathbb{Z}^{m \times n}$, such that $\boldsymbol{W}\boldsymbol{x} \equiv \boldsymbol{b} \mod \boldsymbol{c}$ has integer solutions.

Output: Solution $\boldsymbol{x} \in \mathbb{Z}^n$ such that $\boldsymbol{W}\boldsymbol{x} \equiv \boldsymbol{b} \mod \boldsymbol{c}$.

1: Let $\widetilde{\boldsymbol{W}} = [\boldsymbol{W}, \operatorname{diag}(\boldsymbol{c})]^\mathsf{T}$

2: By Theorem 6, $\boldsymbol{H} = [\boldsymbol{I}_m, \boldsymbol{0}_{m,n}]^\mathsf{T}$ is the HNF of $\widetilde{\boldsymbol{W}}$, with $\boldsymbol{H} = \boldsymbol{U}\widetilde{\boldsymbol{W}}$ for unimodular $\boldsymbol{U} \in \mathbb{Z}^{(m+n)\times(m+n)}$. Let $\widetilde{\boldsymbol{U}} \in \mathbb{Z}^{m \times n}$ be the first m rows of the first n columns of $\boldsymbol{U}$.

3: $\boldsymbol{x} \leftarrow \widetilde{\boldsymbol{U}}^\mathsf{T}\boldsymbol{b}$

5.3.4 Negative Log-Likelihood Function

The negative log-likelihood function for $\boldsymbol{x}$ and $\boldsymbol{k}$ in (5.1) is, up to an additive constant, proportional to $\mathcal{L}_{x,k}$ in (5.3). Note that $\mathcal{L}_{x,k}$ is periodic in $\boldsymbol{x}$ and $\boldsymbol{k}$, with

$$\mathcal{L}_{x,k}\left(\boldsymbol{x}, \boldsymbol{k}\,; \boldsymbol{y}\right) = \mathcal{L}_{x,k}\left(\boldsymbol{x} + \boldsymbol{\delta}, \boldsymbol{k} - \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{\delta}\,; \boldsymbol{y}\right), \tag{5.14}$$

for any $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{k} \in \mathbb{Z}^m$, and $\boldsymbol{\delta} \in \{\boldsymbol{x} \mid \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{x} \in \mathbb{Z}^m\}$. If $\boldsymbol{M}^{-1}\boldsymbol{A}$ is rational, then

$$\{\boldsymbol{x} \mid \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{x} \in \mathbb{Z}^m\} = \Lambda_{\boldsymbol{V}}, \tag{5.15}$$

for some invertible lattice basis $\boldsymbol{V} \in \mathbb{R}^{n \times n}$. The basis $\boldsymbol{V}$ can be formed as in [7]. First, let $\boldsymbol{M}^{-1}\boldsymbol{A} = \boldsymbol{Q}_{\text{num}} \oslash \boldsymbol{Q}_{\text{den}}$, for $\boldsymbol{Q}_{\text{num}}, \boldsymbol{Q}_{\text{den}} \in \mathbb{Z}^{m \times n}$, and let $\boldsymbol{D} = \operatorname{diag}(\operatorname{lcm}(\boldsymbol{Q}_{\text{den}})) \in \mathbb{Z}^{n \times n}$, where $\oslash$ denotes element-wise division, and $\operatorname{lcm}(\cdot)$ gives the column-wise positive least common multiple of an integer-valued matrix input, with zeros ignored. Let

$$\boldsymbol{T} = \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{D} \in \mathbb{Z}^{m \times n}, \tag{5.16}$$

and let $\boldsymbol{H} = [\widetilde{\boldsymbol{H}}^\mathsf{T}, \boldsymbol{0}_{n,(m-n)}]^\mathsf{T}$ be the HNF of $\boldsymbol{T}$, with

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{T}, \tag{5.17}$$

for some unimodular $\boldsymbol{U} \in \mathbb{Z}^{m\times m}$. From (5.15) and (5.16),

$$\begin{aligned}\boldsymbol{x} \in \Lambda_{\boldsymbol{V}} &\Leftrightarrow \boldsymbol{A}\boldsymbol{x} \in \mathbb{Z}^m \Leftrightarrow \boldsymbol{T}\boldsymbol{D}^{-1}\boldsymbol{x} \in \mathbb{Z}^m \\ &\Leftrightarrow \boldsymbol{U}^{-1}\boldsymbol{H}\boldsymbol{D}^{-1}\boldsymbol{x} \in \mathbb{Z}^m \Leftrightarrow \boldsymbol{H}\boldsymbol{D}^{-1}\boldsymbol{x} \in \mathbb{Z}^m \\ &\Leftrightarrow \boldsymbol{x} \in \boldsymbol{D}\widetilde{\boldsymbol{H}}^{-1}\mathbb{Z}^n,\end{aligned}$$

where we have used that $\boldsymbol{U}$ is bijective on $\mathbb{Z}^m$. Thus,

$$\Lambda_{\boldsymbol{V}} = \{\boldsymbol{D}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{\lambda} \mid \boldsymbol{\lambda} \in \mathbb{Z}^n\} \tag{5.18}$$

is a full-rank lattice with a basis $\boldsymbol{V} = \boldsymbol{D}\widetilde{\boldsymbol{H}}^{-1}$. Therefore, for any $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{k} \in \mathbb{Z}^m$,

$$\mathcal{L}_{x,k}\left(\boldsymbol{x}, \boldsymbol{k}\,; \boldsymbol{y}\right) = \mathcal{L}_{x,k}\left(\boldsymbol{x} + \boldsymbol{V}\boldsymbol{\lambda}, \boldsymbol{k} - \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{V}\boldsymbol{\lambda}; \boldsymbol{y}\right), \tag{5.19}$$

for any $\boldsymbol{\lambda} \in \mathbb{Z}^n$. The parallelotope $\mathcal{P}_{\boldsymbol{V}}$ defines a region in which the solution to (5.2) is unique, and we can therefore limit our search to this region.

5.4 ML Estimator for Restricted Encoding Matrices

In this section we provide an ML estimation procedure for a special case of (5.2), in which $n \geq 1$, the covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{M}$ are diagonal, and $\boldsymbol{M}^{-1}\boldsymbol{A}$ is a rational matrix that meets a certain condition. For simplicity in this section, we will assume, without loss of generality, that $\boldsymbol{M} = \boldsymbol{I}$. The restriction on $\boldsymbol{A}$ is analogous to the coprime moduli restriction used in [32]. Hence in this section, we extend the algorithm in [32] to provide polynomial-time (for fixed n) ML estimation for multivariate unknowns.

5.4.1 Preliminaries

In [32], it is assumed that $\boldsymbol{A} = \alpha[\,\Gamma_1^{-1}, \Gamma_2^{-1}, \ldots, \Gamma_m^{-1}\,]^{\mathsf{T}}$, for some $\alpha \in \mathbb{R}$, with $\boldsymbol{\Gamma} \in \mathbb{Z}^m$ having pairwise coprime entries. The solution for x in (5.2) is constructed

as a sum of two terms: The first is an estimate of a "common remainder," that is, a real number within an interval of length $|C|$, where $C = \alpha^{-1} \gcd(\boldsymbol{\Gamma})$. The common remainder is found by a search over a list of only m candidate values. The second term is an integer multiple of C, which can be solved for efficiently by the use of the univariate CRT.

For $n > 1$, the interval in which the common remainder is found becomes the n-dimensional parallelotope $\mathcal{P}_{\boldsymbol{C}}$ corresponding to a lattice basis $\boldsymbol{C} \in \mathbb{R}^{n \times n}$. The basis is given by $\boldsymbol{C} = \boldsymbol{V} \operatorname{diag}^{-1}(\boldsymbol{z})$, with $\boldsymbol{z} = (\operatorname{lcm}(\boldsymbol{AV}))^{\mathsf{T}} \in \mathbb{Z}^n$, and $\boldsymbol{V}$ found from $\boldsymbol{A}$ as in (5.18). Thus, $\Lambda_{\boldsymbol{V}}$ is a sub-lattice of $\Lambda_{\boldsymbol{C}}$, and $\mathcal{P}_{\boldsymbol{C}}$ provides a tiling of $\mathcal{P}_{\boldsymbol{V}}$, i.e., translated copies of $\mathcal{P}_{\boldsymbol{C}}$ can be arranged within $\mathcal{P}_{\boldsymbol{V}}$ such that $\mathcal{P}_{\boldsymbol{V}}$ is covered with no overlap between the copies. This tiling is illustrated, for the encoding matrix

$$\boldsymbol{A} = \begin{bmatrix} 1/2 & 0 & 1 \\ -1/6 & 1/3 & 2/3 \end{bmatrix}^{\mathsf{T}}, \tag{5.20}$$

in Fig. 5.1; in this case $\boldsymbol{V} = [2, 1; 0, 3]$ and $\boldsymbol{C} = [1, 1/3; 0, 1]$. Thus, $\boldsymbol{z} = [2, 3]^{\mathsf{T}}$.

Let $\boldsymbol{\beta} \in \mathbb{Z}^{m \times n}$ with

$$\beta_{j,k} = \begin{cases} ((\boldsymbol{a}_j \boldsymbol{C})_k)^{-1}, & (\boldsymbol{a}_j \boldsymbol{C})_k \neq 0 \\ 0, & \text{otherwise} \end{cases}, \tag{5.21}$$

where $\boldsymbol{a}_j \in \mathbb{Q}^{1 \times n}$ denotes the jth row of $\boldsymbol{A}$. Also let $\boldsymbol{\omega} \in \mathbb{Z}^m$ denote the row-wise positive least common multiple of $\boldsymbol{\beta}$, i.e.,

$$\boldsymbol{\omega} = \left(\operatorname{lcm}\left(\boldsymbol{\beta}^{\mathsf{T}}\right)\right)^{\mathsf{T}}. \tag{5.22}$$

For example with $n = 1$, $m = 3$, and $\boldsymbol{A} = [\frac{1}{2}, \frac{1}{4}, \frac{1}{6}]^{\mathsf{T}}$, we have $\boldsymbol{V} = 12$, $\boldsymbol{C} = 2$, $z = 1$, and $\boldsymbol{\beta} = [1, 2, 3]^{\mathsf{T}}$. Note that for $n = 1$, $\omega_j = |\beta_j|$, $j \in \{1, \ldots, m\}$.

The following condition is, for $n > 1$, analogous to the assumption of pairwise coprime moduli, used for $n = 1$ in [32]:

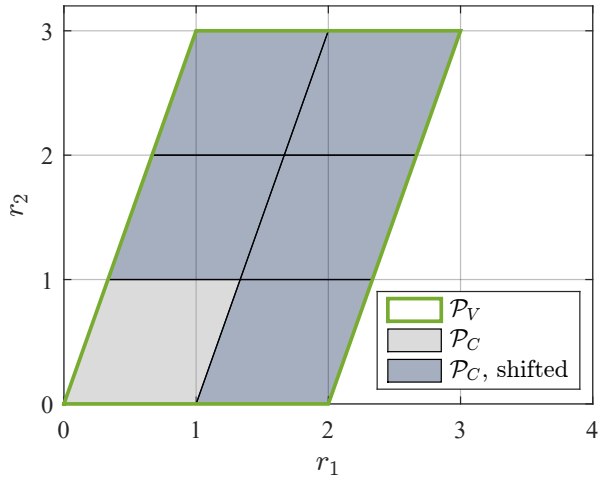


Figure 5.1: $\mathcal{P}_{\boldsymbol{V}}$ partitioned by six shifted copies of $\mathcal{P}_C$ for the encoding matrix in (5.20).

Condition 7. *For any $\boldsymbol{b} \in \mathbb{Z}^m$, there is $\boldsymbol{\gamma} \in \mathbb{Z}^n$ such that*

$$\boldsymbol{\omega} \circ (\boldsymbol{AC})\boldsymbol{\gamma} \equiv \boldsymbol{b} \mod \boldsymbol{\omega}. \tag{5.23}$$

The following proposition is directly implied by Theorem 6:

Proposition 8. *Let $\widetilde{\boldsymbol{W}} = [\boldsymbol{\omega} \circ (\boldsymbol{AC}), \operatorname{diag}(\boldsymbol{\omega})]^{\mathsf{T}} \in \mathbb{Z}^{(m+n)\times m}$ and $\boldsymbol{H} = [\widetilde{\boldsymbol{H}}^{\mathsf{T}}, \boldsymbol{0}_{m,n}]^{\mathsf{T}}$ be the HNF of $\widetilde{\boldsymbol{W}}$. Condition 7 is equivalent to $\widetilde{\boldsymbol{H}} = \boldsymbol{I}_m$.*

The solution method in this section requires that $\boldsymbol{A}$ meets Condition 7. We will show that this restriction leads to a fast ML estimation procedure.

A solution to (5.2) for $\boldsymbol{x}$ is constructed as

$$\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{r}} + \boldsymbol{C}\widehat{\boldsymbol{\gamma}} \in \mathcal{P}_{\boldsymbol{V}}, \tag{5.24}$$

with $\widehat{\boldsymbol{r}} \in \mathcal{P}_{\boldsymbol{C}}$ and $\widehat{\boldsymbol{\gamma}} \in \langle \mathbb{Z}^n \rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}$. Such an approach is appealing because the integer-valued unknown $\boldsymbol{\gamma}$ can be estimated efficiently thanks to the multivariable CRT; the

real-valued unknown $\boldsymbol{r}$ can, as in [32], be found by searching a relatively short list of candidate solutions.

5.4.2 Estimating the Integer Remainders

Recall that $\boldsymbol{C}^{-1}\boldsymbol{V} = \text{diag}(\boldsymbol{z})$, with $\boldsymbol{z} \in \mathbb{Z}^n$. According to the formulation in (5.24), we consider an equivalent problem to (5.2), given by:

$$\{\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{r}}, \widehat{\boldsymbol{k}}\} = \underset{\boldsymbol{\gamma}\in\langle\mathbb{Z}^n\rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}, \boldsymbol{r}\in\mathcal{P}_{\boldsymbol{C}}, \boldsymbol{k}\in\mathbb{Z}^m}{\arg\min} \mathcal{L}_{\gamma,r,k}(\boldsymbol{\gamma}, \boldsymbol{r}; \boldsymbol{y}), \tag{5.25}$$

where

$$\mathcal{L}_{\gamma,r,k}(\boldsymbol{\gamma}, \boldsymbol{r}, \boldsymbol{k}; \boldsymbol{y}) = \left\|\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y} - \boldsymbol{A}\left(\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\gamma}\right) - \boldsymbol{k}\right)\right\|^2.$$

For any $\boldsymbol{q}$, $\boldsymbol{k}$, and $\boldsymbol{\gamma}$ such that

$$\boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C}\boldsymbol{\gamma} + k) = \boldsymbol{q}, \tag{5.26}$$

we have that $\mathcal{L}_{q,r}(\boldsymbol{q}, \boldsymbol{r}; \boldsymbol{y}) = \mathcal{L}_{\gamma,r,k}(\boldsymbol{\gamma}, \boldsymbol{r}, \boldsymbol{k}; \boldsymbol{y})$, where

$$\mathcal{L}_{q,r}(\boldsymbol{q}, \boldsymbol{r}; \boldsymbol{y}) = \left\|\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r} - \boldsymbol{q} \oslash \boldsymbol{\omega}\right)\right\|^2. \tag{5.27}$$

The bijection between $\{\boldsymbol{q} \in \mathbb{Z}^m\}$ and $\{\boldsymbol{\gamma} \in \langle\mathbb{Z}^n\rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}, \boldsymbol{k} \in \mathbb{Z}^m\}$ shown by Theorem 27 suggests that, rather than finding $\widehat{\boldsymbol{\gamma}} \in \langle\mathbb{Z}^n\rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}$ and $\widehat{\boldsymbol{k}} \in \mathbb{Z}^m$ directly, we find $\widehat{\boldsymbol{q}} \in \mathbb{Z}^m$ such that (5.26) is satisfied. This leads to yet another equivalent problem, given by

$$\{\widehat{\boldsymbol{q}}, \widehat{\boldsymbol{r}}\} = \underset{\boldsymbol{q}\in\mathbb{Z}^m, \boldsymbol{r}\in\mathcal{P}_{\boldsymbol{C}}}{\arg\min} \mathcal{L}_{q,r}(\boldsymbol{q}, \boldsymbol{r}; \boldsymbol{y}). \tag{5.28}$$

For any $\boldsymbol{r} \in \mathcal{P}_{\boldsymbol{C}}$, $\widehat{\boldsymbol{q}}$ can be found by solving a CPP:

$$\widehat{\boldsymbol{q}} = \underset{\boldsymbol{q}\in\mathbb{Z}^m}{\arg\min} \left\|\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r} - \text{diag}^{-1}(\boldsymbol{\omega})\boldsymbol{q}\right)\right\|^2. \tag{5.29}$$

Because $\boldsymbol{\Sigma}^{-1/2}\operatorname{diag}^{-1}(\boldsymbol{\omega})$ has orthogonal columns (as it is diagonal), $\widehat{\boldsymbol{q}}$ can be found via Babai's rounding procedure [197]. That is [198, pg. 379],

$$\widehat{\boldsymbol{q}} = \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor, \tag{5.30}$$

where $\lceil \boldsymbol{x} \rfloor$ denotes element-wise rouding of $\boldsymbol{x}$ to the nearest integers, with $\lceil \frac{1}{2} \rfloor = 1$.

5.4.3 Estimating the Real-Valued Common Remainder

After inserting $\widehat{\boldsymbol{q}}$ into (5.28), we consider

$$\begin{aligned}
\widehat{\boldsymbol{r}} &= \underset{\boldsymbol{r}\in\mathcal{P}_C}{\arg\min}\ \mathcal{L}_{q,r}(\widehat{\boldsymbol{q}}, \boldsymbol{r}; \boldsymbol{y}) \\
&= \underset{\boldsymbol{r}\in\mathcal{P}_C}{\arg\min} \left\| \boldsymbol{\Sigma}^{-1/2}\, \mathrm{d}_{\mathbf{1}_m \oslash \boldsymbol{\omega}} \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r} \right) \right\|^2 \\
&= \underset{\boldsymbol{r}\in\mathcal{P}_C}{\arg\min}\, \mathcal{L}_r(\boldsymbol{r}; \boldsymbol{y}),
\end{aligned} \tag{5.31}$$

where $\mathbf{1}_m$ is a length-m column vector of ones, and

$$\mathrm{d}_{\boldsymbol{z}}(\boldsymbol{x}) \triangleq \boldsymbol{x} - \boldsymbol{z} \circ \lceil \boldsymbol{x} \oslash \boldsymbol{z} \rfloor, \tag{5.32}$$

for $\boldsymbol{z} \in \mathbb{R}^m$, with $\mathbf{0} \prec \boldsymbol{z}$. Here, $\prec$ and $\preceq$ denote element-wise inequalities. The operator $\mathrm{d}_{\boldsymbol{z}}$ has the property that

$$-\boldsymbol{z}/2 \preceq \mathrm{d}_{\boldsymbol{z}}(\boldsymbol{x}) \prec \boldsymbol{z}/2 \tag{5.33}$$

If $-\boldsymbol{z}/2 \preceq \boldsymbol{x} \prec \boldsymbol{z}/2$, then $\mathrm{d}_{\boldsymbol{z}}(\boldsymbol{x}) = \boldsymbol{x}$.

Define $\mathcal{H}$ as

$$\mathcal{H} = \left\{ \boldsymbol{r} \in \mathbb{R}^n \mid \exists j \in \{1, \ldots, m\} \text{ s.t. } \mathrm{d}_{\omega_j^{-1}}(y_j - \boldsymbol{a}_j \boldsymbol{r}) = -\tfrac{1}{2}\omega_j^{-1} \right\} \tag{5.34}$$

$$= \left\{ \boldsymbol{r} \in \mathbb{R}^n \mid \exists j \in \{1, \ldots, m\} \text{ s.t. } \omega_j \boldsymbol{a}_j \boldsymbol{r} - \langle \omega_j y_j + \tfrac{1}{2} \rangle_1 \in \mathbb{Z} \right\}. \tag{5.35}$$

Note that $\mathcal{H}$ is a union of countably infinitely many vector hyperplanes (subspaces with codimension one) in $\mathbb{R}^n$. Due to the rounding operation in d, $\mathcal{L}_r$ is non-differentiable for $\boldsymbol{r} \in \mathcal{H}$. The hyperplanes in $\mathcal{H}$ partition $\mathbb{R}^n$ into *chambers*, convex regions within which $\lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor$ is constant.[9] A chamber is thus uniquely identified by $\boldsymbol{\ell} \in \Phi \subset \mathbb{Z}^m$, according to

$$\Psi(\boldsymbol{\ell}) = \{\boldsymbol{r} \in \mathbb{R}^n \mid \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor = \boldsymbol{\ell}\}, \tag{5.36}$$

where $\Psi : \Phi \mapsto \mathbb{R}^n$, and Φ is the set of integer vectors that correspond to chambers in $\mathbb{R}^n$ formed by $\mathcal{H}$. The closure of any chamber is a polytope (the convex hull of finitely many points).

The function $\mathcal{L}_r$ in (5.31) is continuous for $\boldsymbol{r} \in \mathbb{R}^n$ and differentiable for $\boldsymbol{r} \in \mathbb{R}^n \backslash \mathcal{H}$. For any $\boldsymbol{\ell} \in \Phi$, the rounded portion of $\mathcal{L}_r$ is constant within $\boldsymbol{r} \in \Psi(\boldsymbol{\ell})$. Thus, $\mathcal{L}_r$ is quadratic over $\Psi(\boldsymbol{\ell})$, and $\mathcal{L}_r$ is a piecewise quadratic function. In Fig. 5.2, $\mathcal{L}_r$ for $\boldsymbol{r} \in \mathcal{P}_{\boldsymbol{C}}$ is shown, with value indicated by color, for a measurement from (5.1) with $\boldsymbol{A}$ from (5.20).

Let $\widetilde{\mathcal{H}} = \mathcal{P}_{\boldsymbol{C}} \cap \mathcal{H}$. The set $\widetilde{\mathcal{H}}$ is the intersection of $\mathcal{P}_{\boldsymbol{C}}$ with a union of finitely many hyperplanes, and it can be considered to partition $\mathcal{P}_{\boldsymbol{C}}$ into "wrapped chambers," i.e., chambers that are wrapped about the parallelotope $\mathcal{P}_{\boldsymbol{C}}$. The wrapped chamber corresponding to $\boldsymbol{\ell} \in \widetilde{\Phi}$ is

$$\begin{aligned} \widetilde{\Psi}(\boldsymbol{\ell}) &= \{\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} \mid \boldsymbol{r} \in \Psi(\boldsymbol{\ell})\} \\ &= \{\boldsymbol{r} \in \mathcal{P}_{\boldsymbol{C}} \mid \exists \boldsymbol{\lambda} \in \mathbb{Z}^n \text{ s.t. } \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor = \boldsymbol{\ell} + \boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda})\}. \end{aligned} \tag{5.37}$$

[9] The term "chamber" has elsewhere been used interchangeably with the more common term "cell" to refer to an intersection of half-planes in the context of hyperplane arrangements [199, 200]. Here, we use "chamber" to refer to a similar concept, with the distinction that the cells of hyperplane arrangements are open, whereas our chambers are neither open nor closed.

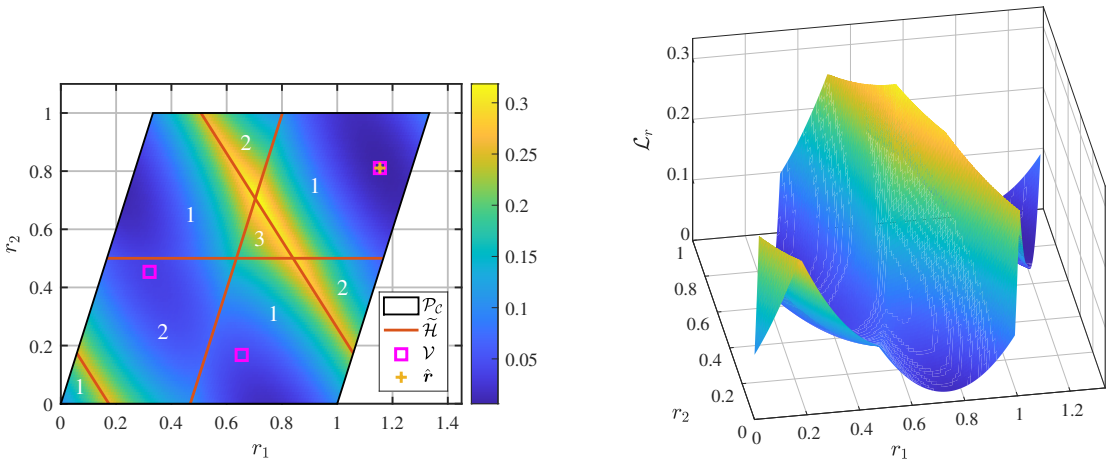


Figure 5.2: (a) The value of $\mathcal{L}_r$ is indicated by color over $\mathcal{P}_{\boldsymbol{C}}$. The wrapped hyperplanes in $\widetilde{\mathcal{H}}$ are shown in red. The candidate solutions for (5.31) are marked with magenta squares, with the solution $\widehat{\boldsymbol{r}}$ at the yellow cross. $\mathcal{P}_{\boldsymbol{C}}$ is partitioned into wrapped chambers, labeled 1–3. (b) $\mathcal{L}_r$ as a surface plot.

For any $\boldsymbol{\ell} \in \Phi$ and $\boldsymbol{\lambda} \in \mathbb{Z}^n$, $\widetilde{\Psi}(\boldsymbol{\ell}) = \widetilde{\Psi}(\boldsymbol{\ell} + \boldsymbol{\omega} \circ (\boldsymbol{AC\lambda}))$, as shown by Proposition 22 (Appendix B.1). In Fig. 5.2, $\widetilde{\mathcal{H}}$ is depicted by red line segments, and $\mathcal{P}_{\boldsymbol{C}}$ is shown to be divided by $\widetilde{\mathcal{H}}$ into three wrapped chambers, labeled with white text.

Theorem 9. *For some* $\boldsymbol{\ell} \in \Phi$, $\widehat{\boldsymbol{r}} \in \operatorname{int} \widetilde{\Psi}(\boldsymbol{\ell})$.

Proof. We have that $\widehat{\boldsymbol{r}} \in \mathcal{P}_{\boldsymbol{C}}$, and

$$\mathcal{P}_{\boldsymbol{C}} = \left(\cup_{\boldsymbol{\ell} \in \Phi} \operatorname{int} \widetilde{\Psi}(\boldsymbol{\ell}) \right) \cup \widetilde{\mathcal{H}}. \tag{5.38}$$

For no $\boldsymbol{r} \in \mathcal{H}$ does $\mathcal{L}_r$ have a local minimum at $\boldsymbol{r}$, as shown by Lemma 24 (Appendix B.1); the statement and proof follow those given for $n = 1$ in [32, Thm. 2]. As $\widetilde{\mathcal{H}} \subset \mathcal{H}$, we have that $\widehat{\boldsymbol{r}} \notin \widetilde{\mathcal{H}}$. Therefore, $\widehat{\boldsymbol{r}} \in \operatorname{int} \widetilde{\Psi}(\boldsymbol{\ell})$, for some $\boldsymbol{\ell} \in \Phi$. □

According to Theorem 9, $\widehat{\boldsymbol{r}} \in \operatorname{int} \widetilde{\Psi}(\boldsymbol{\ell})$, for some $\boldsymbol{\ell} \in \Phi$, where int denotes the interior of a set. Now assume, for any $\widehat{\boldsymbol{\ell}} \in \Phi$, that $\widehat{\boldsymbol{r}} \in \widetilde{\Psi}(\widehat{\boldsymbol{\ell}})$. From (5.37), for any

$\boldsymbol{r} \in \widetilde{\Psi}(\widehat{\boldsymbol{\ell}})$, that

$$\mathcal{L}_r(\boldsymbol{r};\boldsymbol{y}) = \left\| \boldsymbol{\Sigma}^{-1/2} \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r} - \widehat{\boldsymbol{\ell}} \oslash \boldsymbol{\omega} - \boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda} \right) \right\|^2, \tag{5.39}$$

for some $\boldsymbol{\lambda} \in \mathbb{Z}^n$. Therefore, $\widehat{\boldsymbol{r}}$ can be found by generalized least squares as

$$\begin{aligned} \widehat{\boldsymbol{r}} &= (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\ell}} \oslash \boldsymbol{\omega} - \boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda}) \\ &= \langle (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\ell}} \oslash \boldsymbol{\omega}) \rangle_{\boldsymbol{C}} \\ &= \langle (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor \oslash \boldsymbol{\omega}) \rangle_{\boldsymbol{C}}, \end{aligned} \tag{5.40}$$

for any $\boldsymbol{r} \in \widetilde{\Psi}(\widehat{\boldsymbol{\ell}})$. Because $\widehat{\boldsymbol{\ell}}$ is unknown (i.e., we do not know which wrapped cell contains $\widehat{\boldsymbol{r}}$), it is useful to find a set χ containing one point from each wrapped chamber (see Appendix B.3 for deteails). We then form $\mathcal{V}$, a set of candidate solutions to (5.31), as

$$\mathcal{V} = \left\{ \langle (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor \oslash \boldsymbol{\omega}) \rangle_{\boldsymbol{C}} \mid \boldsymbol{r} \in \chi \right\}. \tag{5.41}$$

For the example in Fig. 5.2, there are three wrapped chambers and, thus, three elements in $\mathcal{V}$, which are marked with magenta squares; the solution to (5.2) is shown with a yellow cross. The solution to (5.31) is

$$\widehat{\boldsymbol{r}} = \underset{\boldsymbol{r} \in \mathcal{V}}{\arg\min}\, \mathcal{L}_r(\boldsymbol{r};\boldsymbol{y}), \tag{5.42}$$

and it can be found by evaluating $\mathcal{L}_r$ for each element of the candidate set $\mathcal{V}$.

5.4.4 Constructing the ML Solution

By inserting $\widehat{\boldsymbol{r}}$ into (5.30), we get

$$\widehat{\boldsymbol{q}} = \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{r}}) \rfloor. \tag{5.43}$$

Algorithm 8 ML solution for (5.1) with Condition 7 satisfied.

Input: Measurements $\boldsymbol{y} \in \mathbb{R}^m$; diagonal covariance $\boldsymbol{\Sigma}$; and full column rank $\boldsymbol{A} \in \mathbb{Q}^{m \times n}$ satisfying Condition 7.

Output: ML solution $\{\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{k}}\}$.

1: Let χ be a set containing a point in $\widetilde{\Psi}(\boldsymbol{\ell})$ for each $\boldsymbol{\ell} \in \Phi$
2: Let $\mathcal{V}$ be the set of candidate solutions in (5.41)
3: $\widehat{\boldsymbol{r}} \leftarrow \arg\min_{\boldsymbol{r} \in \mathcal{V}} \mathcal{L}_r(\boldsymbol{r}; \boldsymbol{y})$
4: $\widehat{\boldsymbol{q}} \leftarrow \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{r}}) \rfloor$
5: Use Algorithm 7 to find $\widehat{\boldsymbol{\gamma}}$ as the solution for $\boldsymbol{\gamma}$ to $\boldsymbol{\omega} \circ (\boldsymbol{AC}\boldsymbol{\gamma}) \equiv \widehat{\boldsymbol{q}} \mod \boldsymbol{\omega}$
6: $\widehat{\boldsymbol{x}} \leftarrow \langle \boldsymbol{C}\widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{r}} \rangle_{\boldsymbol{V}}$
7: $\widehat{\boldsymbol{k}} \leftarrow \lceil \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{x}} \rfloor$

We then find $\widehat{\boldsymbol{\gamma}} \in \langle \mathbb{Z}^n \rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}$ as the solution to the following system of congruences:

$$\boldsymbol{\omega} \circ (\boldsymbol{AC})\boldsymbol{\gamma} \equiv \widehat{\boldsymbol{q}} \mod \boldsymbol{\omega}, \tag{5.44}$$

using Algorithm 7; the congruences are known to have a solution by Theorem 6. Finally, the solution to (5.2) is formed as

$$\widehat{\boldsymbol{x}} = \langle \boldsymbol{C}\widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{r}} \rangle_{\boldsymbol{V}} \tag{5.45}$$

$$\widehat{\boldsymbol{k}} = \lceil \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{x}} \rfloor . \tag{5.46}$$

The proposed method to solve (5.2) for this case is summarized in Algorithm 8.

5.4.5 Computational Complexity

The ML solution method in Algorithm 8 involves two primary tasks. The first is to solve a system of linear congruence equations in multiple integer variables, which can be done in polynomial time [194, Cor. 5.3b]. The solution method given in Algorithm 7 computes a HNF, which can be reused for subsequent measurements from the same $\boldsymbol{A}$. The second task is to build a set of candidate solutions for (5.31) and evaluate them. The number of candidate solutions is given by $|\mathcal{V}| = |\{\widetilde{\Psi}(\boldsymbol{\ell})\}_{\boldsymbol{\ell} \in \Phi}|$.

For fixed $\boldsymbol{V}$, $\boldsymbol{C}$, and n, we get that $|\mathcal{V}|$ is $O(m^n)$, as shown in Appendix B.4. Here again, much of the computation to form the set of candidate solutions can be reused for inference on subsequent measurements. Thus, for fixed $\boldsymbol{V}$, $\boldsymbol{C}$, and n, both primary tasks can be completed in time that is polynomial in m.

5.4.6 Comment on the $n = 1$ Case

The problem treated in this section is simpler when $n = 1$, which is the scenario considered in [32]. With $n = 1$, $\omega_j = |\beta_j| = |(Ca_j)^{-1}|$, for $j \in \{1, \ldots, m\}$, and $C^{-1}V = \operatorname{lcm}(\boldsymbol{\beta})$. The fundamental parallelotope $\mathcal{P}_V$ is simply an interval with length $|V|$. A univariate CRT method [195, p. 23] can be used rather than Algorithm 7 in step 5 of Algorithm 8. The construction of χ is particularly simple in this case, as discussed in Appendix B.3. The set χ, and hence $\mathcal{V}$, has cardinality m. Therefore, with $n = 1$, just m candidates need be considered to estimate the real-valued unknown, as in [32].

5.5 ML Estimator for General Encoding Matrices

In this section, we remove the restriction on $\boldsymbol{A}$ given by Condition 7 and provide an ML estimator for $n \geq 1$ that constructs a short list of candidate points in $\mathbb{R}^n$ provably containing the ML solution. That is, we consider the case of (5.2) in which $n \geq 1$, the covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{M}$ are diagonal, and $\boldsymbol{A} \in \mathbb{Q}^{m \times n}$ is an arbitrary matrix with full column rank. As in Section 5.4, we assume, without loss of generality that $\boldsymbol{M} = \boldsymbol{I}_m$. The development parallels Section 5.4; however, the multivariate CRT is no longer necessarily applicable, typically resulting in a larger list of candidate solutions.

According to Proposition 36, for any $\boldsymbol{x}$, the Babai rounding method [197] gives the optimal $\boldsymbol{k}$ in (5.2) as $\widehat{\boldsymbol{k}} = \lceil \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \rfloor$. Inserting this into $\mathcal{L}_{x,k}$ gives us

$$\mathcal{L}_{x,k}(\boldsymbol{x}, \widehat{\boldsymbol{k}}; \boldsymbol{y}) = \mathcal{L}_x(\boldsymbol{x}; \boldsymbol{y}) = \left\| \boldsymbol{\Sigma}^{-1/2} \, \mathrm{d}_{\mathbf{1}} \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right) \right\|^2, \tag{5.47}$$

with d defined in (5.32). We define $\mathcal{H}_{\mathrm{NC}}$ as the set of points at which $\mathcal{L}_x$ is non-differentiable:

$$\mathcal{H}_{\mathrm{NC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \exists j \in \{1, \ldots, m\} \text{ s.t. } \boldsymbol{a}_j \boldsymbol{x} - \langle y_j + \tfrac{1}{2} \rangle_1 \in \mathbb{Z} \right\}, \tag{5.48}$$

using the subscript $_{\mathrm{NC}}$ to distinguish from analogous quantities in Section 5.4. The hyperplanes in $\mathcal{H}_{\mathrm{NC}}$ partition $\mathbb{R}^n$ into chambers, within which $\lceil \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \rfloor$ is constant. A chamber is thus uniquely identified by $\boldsymbol{\ell} \in \Phi_{\mathrm{NC}}$, according to

$$\Psi_{\mathrm{NC}}(\boldsymbol{\ell}) = \{ \boldsymbol{x} \in \mathbb{R}^n \mid \lceil \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \rfloor = \boldsymbol{\ell} \}, \tag{5.49}$$

where $\Psi_{\mathrm{NC}} : \Phi_{\mathrm{NC}} \mapsto \mathbb{R}^n$, and $\Phi_{\mathrm{NC}} \subset \mathbb{Z}^m$ is the set of integer vectors that correspond to chambers in $\mathbb{R}^n$ formed by $\mathcal{H}_{\mathrm{NC}}$. Let $\widetilde{\mathcal{H}}_{\mathrm{NC}} = \mathcal{P}_{\boldsymbol{V}} \cap \mathcal{H}_{\mathrm{NC}}$; i.e., $\widetilde{\mathcal{H}}_{\mathrm{NC}}$ is the intersection of $\mathcal{P}_{\boldsymbol{V}}$ with a union of finitely many hyperplanes, and it can be considered to partition $\mathcal{P}_{\boldsymbol{V}}$ into wrapped chambers. Now, similarly to (5.37), we define $\widetilde{\Psi}_{\mathrm{NC}}(\boldsymbol{\ell})$ as

$$\begin{aligned} \widetilde{\Psi}_{\mathrm{NC}}(\boldsymbol{\ell}) =& \{ \langle \boldsymbol{x} \rangle_{\boldsymbol{V}} \mid \boldsymbol{x} \in \Psi_{\mathrm{NC}}(\boldsymbol{\ell}) \} \\ =& \{ \boldsymbol{x} \in \mathcal{P}_{\boldsymbol{V}} \mid \exists \boldsymbol{\lambda} \in \mathbb{Z}^n \text{ s.t. } \lceil \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \rfloor = \boldsymbol{\ell} + \boldsymbol{A}\boldsymbol{V}\boldsymbol{\lambda} \}. \end{aligned} \tag{5.50}$$

Note that (5.50) uses the lattice basis $\boldsymbol{V}$, rather than basis $\boldsymbol{C}$ used in (5.37). By an analogous statement to Theorem 9, we have that $\widehat{\boldsymbol{x}} \in \operatorname{int} \widetilde{\Psi}_{\mathrm{NC}}(\boldsymbol{\ell})$, for some $\boldsymbol{\ell} \in \Phi_{\mathrm{NC}}$.

Assume, for any $\widehat{\boldsymbol{\ell}} \in \Phi_{\mathrm{NC}}$, that $\widehat{\boldsymbol{x}} \in \widetilde{\Psi}_{\mathrm{NC}}(\widehat{\boldsymbol{\ell}})$. From (5.50) we have, for any $\boldsymbol{x} \in \widetilde{\Psi}_{\mathrm{NC}}(\widehat{\boldsymbol{\ell}})$, that

$$\mathcal{L}_x(\boldsymbol{x}; \boldsymbol{y}) = \left\| \boldsymbol{\Sigma}^{-1/2} \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \widehat{\boldsymbol{\ell}} - \boldsymbol{A}\boldsymbol{V}\boldsymbol{\lambda} \right) \right\|^2, \tag{5.51}$$

for some $\boldsymbol{\lambda} \in \mathbb{Z}^n$. Therefore, $\widehat{\boldsymbol{x}}$ can be found by generalized least squares as

$$\begin{aligned}\widehat{\boldsymbol{x}} &= (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\ell}} - \boldsymbol{A}\boldsymbol{V}\boldsymbol{\lambda}) \\ &= \langle (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\ell}})\rangle_{\boldsymbol{V}} \\ &= \langle (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \lceil \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \rfloor)\rangle_{\boldsymbol{V}}, \end{aligned} \tag{5.52}$$

for any $\boldsymbol{x} \in \widetilde{\Psi}_{\mathrm{NC}}(\widehat{\boldsymbol{\ell}})$. We again form a set of candidate solutions to (5.2), as

$$\mathcal{V}_{\mathrm{NC}} = \left\{ \langle (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A})^{\dagger}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \lceil \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \rfloor)\rangle_{\boldsymbol{V}} \mid \boldsymbol{x} \in \chi_{\mathrm{NC}} \right\}, \tag{5.53}$$

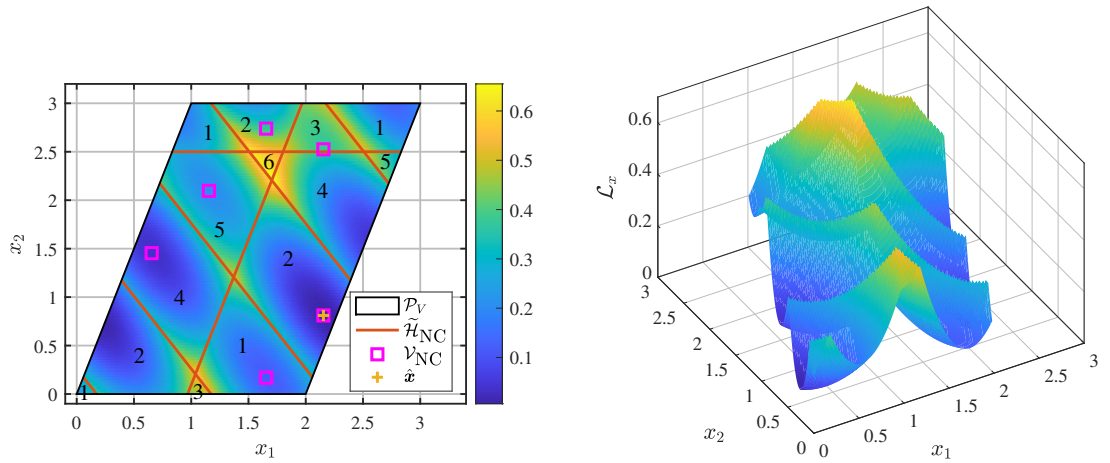
where the set χ_{NC} is constructed so as to contain a point in each wrapped chamber. The set χ_{NC} can be constructed analogously to χ in Section 5.4. The solution to (5.2) for $\boldsymbol{x}$ is

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathcal{V}_{\mathrm{NC}}}{\arg\min}\, \mathcal{L}_x(\boldsymbol{x}; \boldsymbol{y}), \tag{5.54}$$

and it can be found by evaluating $\mathcal{L}_x$ for each element of the candidate set $\mathcal{V}_{\mathrm{NC}}$. The solution for $\boldsymbol{k}$ is then given by $\widehat{\boldsymbol{k}} = \lceil \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{x}} \rfloor$.

In Fig. 5.3, $\mathcal{L}_x$ for $\boldsymbol{x} \in \mathcal{P}_{\boldsymbol{V}}$ is shown, with value indicated by color, with $\boldsymbol{A}$ from (5.20) for the measurement as Fig. 5.2. In this case, the more general method given in this section yields six wrapped chambers, and hence six candidate solutions, whereas the approach in Section 5.4 must evaluate only three candidate solutions.

The proposed method to solve (5.2) for this case is summarized in Algorithm 9. Note that this method is more straightforward than Algorithm 8, in the sense that it involves only a search over a candidate set and does not require solving a system of congruence equations. The cardinality of the candidate set is again polynomial in m, as shown in Appendix B.5. However, $|\mathcal{V}_{\mathrm{NC}}|$ is typically larger than $|\mathcal{V}|$, the cardinality of the candidate set constructed in Section 5.4.



(a) (b)

Figure 5.3: (a) The value of $\mathcal{L}_x$ is indicated by color over $\mathcal{P}_{\boldsymbol{V}}$. The wrapped hyperplanes in $\widetilde{\mathcal{H}}_{\mathrm{NC}}$ are shown in red. The candidate solutions for (5.2) are marked with magenta squares, with the solution $\widehat{\boldsymbol{x}}$ at the yellow cross. $\mathcal{P}_{\boldsymbol{V}}$ is partitioned into wrapped chambers, labeled 1–6. (b) $\mathcal{L}_x$ as a surface plot.

Algorithm 9 ML solution for (5.1) with unrestricted $\boldsymbol{A}$.

Input: Measurements $\boldsymbol{y} \in \mathbb{R}^m$; diagonal covariance $\boldsymbol{\Sigma}$; and $\boldsymbol{A} \in \mathbb{Q}^{m \times n}$.
Output: ML solution $\{\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{k}}\}$.

1: Let χ_{NC} be a set containing a point in $\widetilde{\Psi}_{\mathrm{NC}}(\boldsymbol{\ell})$ for each $\boldsymbol{\ell} \in \Phi_{\mathrm{NC}}$
2: Let $\mathcal{V}_{\mathrm{NC}}$ be the set of candidate solutions defined in (5.53)
3: $\widehat{\boldsymbol{x}} \leftarrow \arg\min_{\boldsymbol{x} \in \mathcal{V}_{\mathrm{NC}}} \mathcal{L}_x(\boldsymbol{x}; \boldsymbol{y})$
4: $\widehat{\boldsymbol{k}} \leftarrow \lceil \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{x}} \rfloor$

As in Section 5.4, the problem considered in this section is simpler when $n = 1$. In that case, $C^{-1}V = \mathrm{lcm}(\boldsymbol{\beta})$, and $|\chi_{\mathrm{NC}}| = |\mathcal{V}_{\mathrm{NC}}| \leq mV/C$, as shown in Appendix B.5. A benefit of both Algorithm 8 and Algorithm 9 is that much of the computation can be done ahead of time and reused on arbitrarily many sets of measurements when the same $\boldsymbol{A}$ is used repeatedly.

5.6 ML Estimator for General Mixed Integer Linear Models

The methods presented in Section 5.4 and Section 5.5 require that the covariance $\boldsymbol{\Sigma}$ be diagonal, and thus, that the noise in (5.2) be uncorrelated. Though, in certain applications the noise is correlated, such as when the measurements correspond to phase differences appearing in sufficient statistics for multi-snapshot, multi-channel data [29]. In this section we provide an ML solution method for (5.2) that applies more generally than the methods presented in Section 5.4 and Section 5.5.

Some applications considered in this chapter share three features: (1) no prior knowledge of the distribution of the parameters, rendering MAP estimation inappropriate or subject to extensive tuning of a regularization parameter; (2) correlated noise; (3) multivariate unknowns, $\boldsymbol{x} \in \mathbb{R}^n$. In this section, we provide an ML solution method for the problem in (5.1) that simultaneously accommodates these three characteristics.

For this method, we require only that $\boldsymbol{M}^{-1}\boldsymbol{A}$ be rational. We also assume in this section, without loss of generality, that the measurements have been whitened; thus, $\boldsymbol{\Sigma} = \boldsymbol{I}_m$. A MATLAB implementation of this approach is available.[10]

5.6.1 Estimation Task

For any $\boldsymbol{k}$, minimization of (5.3) over $\boldsymbol{x}$ yields

$$\boldsymbol{x}_k = \boldsymbol{A}^{\dagger}\left(\boldsymbol{y} - \boldsymbol{M}\boldsymbol{k}\right), \tag{5.55}$$

[10] https://github.com/tuckerda/Mixed-Integer-Linear-Models-MLE

where † denotes the Moore-Penrose pseudo-inverse. Minimizing $\mathcal{L}(\boldsymbol{x}_{\boldsymbol{k}}, \boldsymbol{k}; \boldsymbol{y})$ with respect to the integer vector $\boldsymbol{k}$ gives us

$$\boldsymbol{k}^\star = \underset{\boldsymbol{k}\in\mathbb{Z}^m}{\arg\min}\ \mathcal{L}(\boldsymbol{x}_{\boldsymbol{k}}, \boldsymbol{k}; \boldsymbol{y}) = \underset{\boldsymbol{k}\in\mathbb{Z}^m}{\arg\min}\ \|\widetilde{\boldsymbol{y}} - \boldsymbol{B}\boldsymbol{k}\|^2, \tag{5.56}$$

an integer least squares problem (ILS) with $\boldsymbol{B} = (\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^\dagger)\boldsymbol{M}$ and $\widetilde{\boldsymbol{y}} = \boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{y}$. Note that the range of $\boldsymbol{B}$ is the nullspace of $\boldsymbol{A}^\mathsf{T}$. For $m = n$, we get that $\boldsymbol{B} = \boldsymbol{0}$, and any $\boldsymbol{k} \in \mathbb{Z}^m$ minimizes (5.56).

The estimation task is then to find the point in the lattice $\Lambda_B = \{\boldsymbol{B}\boldsymbol{k} \mid \boldsymbol{k} \in \mathbb{Z}^m\}$ that is closest to the transformed noisy observation $\widetilde{\boldsymbol{y}}$. This is a closest lattice point problem, which is NP-hard [201], though the expected complexity may be a polynomial [174]. Note that the lattice-generating matrix $\boldsymbol{B}$ has nullity n and is therefore not a lattice basis. To use conventional algorithms such as sphere decoding [169,170] to solve (5.56), it is necessary to first find a basis for $\Lambda_{\boldsymbol{B}}$.

5.6.2 Lattice Basis Construction

To find a basis for Λ_B, we use that $\boldsymbol{B}\boldsymbol{T} = \boldsymbol{0}$. Partition $\boldsymbol{U}^{-1}$ from (5.17) into $\widetilde{\boldsymbol{G}} \in \mathbb{Z}^{m\times n}$ and $\boldsymbol{G}_\perp \in \mathbb{Z}^{m\times(m-n)}$ such that

$$\boldsymbol{U}^{-1} = [\widetilde{\boldsymbol{G}}, \boldsymbol{G}_\perp]. \tag{5.57}$$

Because only the first n rows of $\boldsymbol{H}$ are non-zero, we have from (5.17) that $\widetilde{\boldsymbol{G}} = \boldsymbol{T}\widetilde{\boldsymbol{H}}^{-1}$. Using that unimodular $\boldsymbol{U}^{-1}$ is bijective on $\mathbb{Z}^m$, we have

$$\begin{aligned}
\Lambda_B &= \boldsymbol{B}\mathbb{Z}^m = \boldsymbol{B}\boldsymbol{U}^{-1}\mathbb{Z}^m \\
&= \boldsymbol{B}\widetilde{\boldsymbol{G}}\mathbb{Z}^n + \boldsymbol{B}\boldsymbol{G}_\perp\mathbb{Z}^{m-n} \\
&= \boldsymbol{B}\boldsymbol{T}\boldsymbol{H}_n^{-1}\mathbb{Z}^n + \boldsymbol{B}\boldsymbol{G}_\perp\mathbb{Z}^{m-n}.
\end{aligned} \tag{5.58}$$

Since $\boldsymbol{BT} = \boldsymbol{0}$, this gives us

$$\Lambda_B = \boldsymbol{B}\boldsymbol{G}_{\perp}\mathbb{Z}^{m-n}, \tag{5.59}$$

and therefore, $\widetilde{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{G}_{\perp}$ is a basis for Λ_B.

An alternative to the basis construction in (5.59) is to apply the modified LLL algorithm [202, 203] to $\boldsymbol{B}$, as in [187]. Unlike the original LLL algorithm [204], the modified LLL algorithm accepts as input a linearly dependent set of vectors. The modified LLL algorithm operates by detecting linear dependencies among a set of vectors, and this process can be plagued by rounding errors [202]. We find that this method is prone to numerical difficulties even for small m [205].

5.6.3 Lattice Basis Reduction

In order to solve (5.56), we first compute the LLL-QRZ factorization of $\widetilde{\boldsymbol{B}}$ as

$$\widetilde{\boldsymbol{B}} = \boldsymbol{Q}[\boldsymbol{R}^{\mathsf{T}}, \boldsymbol{0}_{m-n,n}]^{\mathsf{T}}\boldsymbol{Z}^{-1} = \boldsymbol{Q_1}\boldsymbol{R}\boldsymbol{Z}^{-1} \tag{5.60}$$

where $\boldsymbol{Q} = [\boldsymbol{Q}_1, \boldsymbol{Q}_2] \in \mathbb{R}^{m\times m}$ is orthonormal, $\boldsymbol{Z}^{m\times m}$ is unimodular, and $\boldsymbol{R}^{(m-n)\times(m-n)}$ is upper-triangular [206]. The matrix $\boldsymbol{R}$ satisfies the size-reduction and Lovász conditions [204], given respectively by

$$|r_{i,j}| \leq |r_{i,i}|/2, \qquad 1 \leq i < j \leq m-n \tag{5.61}$$

$$\delta r_{i-1,i-1}^2 \leq r_{i-1,i}^2 + r_{i,i}^2, \qquad 1 < i \leq m-n, \tag{5.62}$$

with $\delta \in (1/4, 1]$. We use this factorization to pose a new closest lattice point problem that is equivalent to (5.56), given by

$$\boldsymbol{z}^{\star} = \underset{\boldsymbol{z}\in\mathbb{Z}^{(m-n)}}{\arg\min} \ \|\boldsymbol{Q}_1^{\mathsf{T}}\widetilde{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{z}\|^2. \tag{5.63}$$

The closest lattice point problem in (5.63) can be solved using a search algorithm such as sphere decoding. For our examples, we use a depth-first sphere decoding algorithm [169, 170, 207]. After a solution has been found for (5.63), we get that $\boldsymbol{k}^\star = \boldsymbol{G}_\perp \boldsymbol{Z}\boldsymbol{z}^\star$ and $\boldsymbol{x}_{k^\star} = \boldsymbol{A}^\dagger (\boldsymbol{y} - \boldsymbol{M}\boldsymbol{k}^\star)$. Thus, for any optimal $\{\boldsymbol{k}^\star, \boldsymbol{x}_{\boldsymbol{k}^\star}\}$ in (5.3), there exists an equivalent solution $\{\widehat{\boldsymbol{k}}, \widehat{\boldsymbol{x}}\}$ with $\widehat{\boldsymbol{x}} \in \mathcal{P}_{\boldsymbol{V}}$, given by

$$\widehat{\boldsymbol{x}} = \langle \boldsymbol{x}_{\boldsymbol{k}^\star} \rangle_{\mathcal{P}} \tag{5.64}$$

and $\widehat{\boldsymbol{k}} = \boldsymbol{k}^\star - \boldsymbol{M}^{-1}\boldsymbol{A}(\langle \widehat{\boldsymbol{x}} \rangle_{\mathcal{P}} - \boldsymbol{x}_{\boldsymbol{k}^\star})$. From (5.55), (5.63), and (5.64),

$$\widehat{\boldsymbol{x}} = \left\langle \boldsymbol{A}^\dagger (\boldsymbol{y} - \boldsymbol{M}\boldsymbol{G}_\perp \boldsymbol{Z}\boldsymbol{z}^\star) \right\rangle_{\mathcal{P}}. \tag{5.65}$$

5.6.4 Babai Point

Simple rounding of the non-integer least squares solution to (5.63) can, in some cases, provide the ML solution $\boldsymbol{\ell}^\star$, thereby foregoing the complexity of sphere decoding. Define

$$\boldsymbol{\ell}_{\mathrm{B}} = \lceil \widetilde{\boldsymbol{B}}^\dagger \widetilde{\boldsymbol{y}} \rfloor \tag{5.66}$$

where $\lceil \cdot \rfloor$ denotes element-wise rounding to the nearest integer. The solution $\boldsymbol{\ell}_{\mathrm{B}}$ is referred to as the *Babai* point [197] or the *zero-forcing equalization* point [174]. Define $d_{\min}$ as the length of the shortest non-zero lattice point in Λ_B [208]. As noted in [19, 178], for example, if

$$\|\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{B}}\boldsymbol{\ell}_{\mathrm{B}}\| \leq d_{\min}/2, \tag{5.67}$$

then $\boldsymbol{\ell}_{\mathrm{B}} = \boldsymbol{\ell}^\star$. For applications in which the same $\widetilde{\boldsymbol{B}}$ is used with many noisy measurements [209], then $\widetilde{\boldsymbol{B}}^\dagger$ may be pre-computed and (5.67) tested before resorting to sphere decoding for integer solution in (5.63), thereby providing accelerated computation, on average. Further, if Λ_B admits an orthogonal basis $\widetilde{\boldsymbol{B}}\boldsymbol{U}$, for some unimodular

$\boldsymbol{U}$, then the $(m-n)$-dimensional problem in (5.63) decouples for that basis. In that case, $\lceil (\widetilde{\boldsymbol{B}}\boldsymbol{U})^{\dagger}\widetilde{\boldsymbol{y}} \rfloor$ is likewise optimal, as shown by Proposition 36 (Appendix B.6).

Other instances of lattices that admit polynomial-time solution of (5.63) are given in [24, 176, 190, 210].

5.6.5 Estimation Procedure

An ML parameter estimation procedure for the model in (5.1) is given in Algorithm 10. Note that steps 1–4, 6, and 7 can be computed beforehand and reused to form estimates from different sets of noisy measurements that share the same $\boldsymbol{A}$ and $\boldsymbol{M}$. This reuse of precomputed values can greatly reduce computation times when many measurements have the same noise covariance and sensor configuration.

We also seek to improve the average computational cost by first computing the Babai point as in (5.66) and determining whether it is optimal as in (5.67). If the Babai point is optimal, then sphere decoding can be foregone. Otherwise we solve a closest lattice point problem using a sphere decoding algorithm [169, 170, 207]. In the case that $m = n$, we have that any $\boldsymbol{k} \in \mathbb{Z}^m$ is optimal in (5.56), and we can give $\boldsymbol{k}^{\star}$ an arbitrary value.

5.6.6 Estimator Distribution

We next consider the distribution of the estimator, which in turn enables optimized design of $\boldsymbol{A}$ [31, 178, 211]. The Voronoi cell of Λ_B, denoted $\mathrm{Vor}(\Lambda_B)$, is the closure of the subset in $\mathbb{R}^m$ containing all points closer to the origin than to any other lattice point. The lattice Λ_B lies in the nullspace of $\boldsymbol{A}^{\mathsf{T}}$, and $\mathrm{Vor}(\Lambda_B)$ is therefore unbounded in the range of $\boldsymbol{A}$. The integer parameter estimate $\widehat{\boldsymbol{k}}$ is considered *correct* if $\boldsymbol{B}\widehat{\boldsymbol{k}} = \boldsymbol{B}\boldsymbol{k}$ [178] or, equivalently, if $\boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{u} \in \mathrm{Vor}(\Lambda_B)$ [19]. From [178],

Algorithm 10 ML parameter estimation for linear model in (5.1).

Input: Measurements $\boldsymbol{y} \in \mathbb{R}^m$; full-column rank $\boldsymbol{A} \in \mathbb{R}^{m\times n}$; and invertible $\boldsymbol{M} \in \mathbb{R}^{m\times m}$, such that $\boldsymbol{M}^{-1}\boldsymbol{A} \in \mathbb{Q}^{m\times n}$.

Output: ML parameter estimates $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{k}}$

Compute $\boldsymbol{V}$, a lattice basis for Λ, as in (5.18), (5.16), (5.17)
Compute $\boldsymbol{G}_\perp$ as in (5.57)
$\boldsymbol{B} \leftarrow (\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^\dagger)\boldsymbol{M}$
$\widetilde{\boldsymbol{B}} \leftarrow \boldsymbol{B}\boldsymbol{G}_\perp$
if $m > n$ **then**
  Compute the LLL-QRZ factorization of $\widetilde{\boldsymbol{B}}$, with $\boldsymbol{Q}_1$ and $\boldsymbol{Z}$ as given in (5.60).
  $d_{\min} \leftarrow \min_{\boldsymbol{z}\in\mathbb{Z}^{m-n},\boldsymbol{z}\neq\boldsymbol{0}} \|\boldsymbol{R}\boldsymbol{z}\|$
  $\widetilde{\boldsymbol{y}} \leftarrow \boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{y}$
  $\boldsymbol{z}_{\mathrm{B}} \leftarrow \lceil \boldsymbol{R}^{-1}\boldsymbol{Q}_1^{\mathsf{T}}\widetilde{\boldsymbol{y}} \rfloor$
  if $\|\boldsymbol{Q}_1^{\mathsf{T}}\widetilde{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{z}_{\mathrm{B}}\| \leq \frac{1}{2}d_{\min}$ **then**
    $\boldsymbol{z}^\star \leftarrow \boldsymbol{z}_{\mathrm{B}}$
  else
    $\boldsymbol{z}^\star \leftarrow \arg\min_{\boldsymbol{z}\in\mathbb{Z}^{(m-n)}} \|\boldsymbol{Q}_1^{\mathsf{T}}\widetilde{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{z}\|^2$
  end if
  $\boldsymbol{k}^\star \leftarrow \boldsymbol{G}_\perp\boldsymbol{Z}\boldsymbol{z}^\star$
else
  $\boldsymbol{k}^\star \leftarrow \boldsymbol{0}$
end if
$\boldsymbol{x}_{\boldsymbol{k}^\star} \leftarrow \boldsymbol{A}^\dagger(\boldsymbol{y} - \boldsymbol{M}^{-1}\boldsymbol{k}^\star)$
$\widehat{\boldsymbol{x}} \leftarrow \langle \boldsymbol{x}_{\boldsymbol{k}^\star} \rangle_{\mathcal{P}}$
$\widehat{\boldsymbol{k}} \leftarrow \boldsymbol{k}^\star - \boldsymbol{M}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{x}} - \boldsymbol{x}_{\boldsymbol{k}^\star})$.

$\boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{u} \in \mathrm{Vor}(\Lambda_B)$ if and only if $\boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{s} \in \mathrm{Vor}(\Lambda_B)$ for all $\boldsymbol{s} \in \mathbb{R}^n$. Since $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{s}$ for some $\boldsymbol{s}$,

$$\boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{u} \in \mathrm{Vor}(\Lambda_B) \Leftrightarrow \boldsymbol{u} \in \mathrm{Vor}(\Lambda_B). \tag{5.68}$$

That is, the probability that $\widehat{\boldsymbol{k}}$ is correct is equal to the probability that $\boldsymbol{u}$ is in $\mathrm{Vor}(\Lambda_B)$. This probability, denoted P_c, is upper bounded by the probability that $\boldsymbol{u}$ is within a sphere of volume $\det\Lambda_B$ and lower bounded by the probability that $\boldsymbol{u}$ is in a sphere of radius $d_{\min}/2$ [19]. Thus,

$$F_{r,0}\left(d_{\min}^2/4\right) \leq P_c \leq F_{r,0}\left((\det\Lambda_B/\alpha)^{2/r}\right) \tag{5.69}$$

where $F_{r,0}$ is the central χ^2 cumulative distribution function (cdf) with $r = m - n$ degrees of freedom, $\alpha = \pi^{r/2}/\Gamma\left(1 + r/2\right)$, and Γ is the Gamma function.

Conditioned on $\boldsymbol{k}^\star = \boldsymbol{G}_\perp \boldsymbol{\ell}^\star$, the estimate $\boldsymbol{x}_{\boldsymbol{k}^\star}$ (without translation onto $\mathcal{P}$) in (5.55) is linear in $\boldsymbol{y}$ and is therefore Gaussian with variance $\boldsymbol{\Sigma}_{\mathrm{ML}} = \left(\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\right)^{-1}$. Let $\mathcal{N}_{\mathcal{P}}(\mu, \Sigma)$ denote a Gaussian distribution wrapped onto $\mathcal{P}$. Given the true $\boldsymbol{x}$ and $\boldsymbol{\ell}$, by total probability the distribution of the estimate $\widehat{\boldsymbol{x}}$ is then a mixture of wrapped Gaussians,

$$f(\widehat{\boldsymbol{x}} \,|\, \boldsymbol{x}, \boldsymbol{\ell}) = \sum_{\boldsymbol{\ell}' \in \mathbb{Z}^r} w_{\boldsymbol{\ell}'} \mathcal{N}_{\mathcal{P}}(\boldsymbol{x} + \boldsymbol{A}^\dagger \boldsymbol{M} \boldsymbol{G}_\perp (\boldsymbol{\ell} - \boldsymbol{\ell}'), \boldsymbol{\Sigma}_{\mathrm{ML}}). \tag{5.70}$$

Each weight $w_{\boldsymbol{\ell}'}$ in the mixture is the integration of the probability density of $\widetilde{\boldsymbol{y}}$ over a translated Voronoi cell, $\left\{\mathrm{Vor}(\Lambda_B) + \widetilde{\boldsymbol{B}}(\boldsymbol{\ell}' - \boldsymbol{\ell})\right\}$. We can numerically approximate this distribution by limiting the sum to the top several closest lattice points and approximating $w_{\boldsymbol{\ell}'}$, as in (5.69), using a non-central χ^2 cdf with non-centrality parameter $\|\widetilde{\boldsymbol{B}}(\boldsymbol{\ell}' - \boldsymbol{\ell})\|^2$. For the non-central case, the right-hand side term in (5.69) provides only an approximation when $\boldsymbol{\ell}' \neq \boldsymbol{\ell}$, while the left-hand side remains a valid lower-bound. To illustrate, Fig. 5.4 displays the distribution of $f(\widehat{\boldsymbol{x}} \,|\, \boldsymbol{x}, \boldsymbol{\ell})$ for whitened matrices $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1/2}$,

$$\boldsymbol{A} = \frac{\boldsymbol{\Sigma}^{-1/2}}{6} \begin{bmatrix} -2 & 7 \\ 2 & 2 \\ 4 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \frac{1}{100} \begin{bmatrix} 4 & 1 & -1 \\ 1 & 4 & 1 \\ -1 & 1 & 4 \end{bmatrix}.$$

Here, the probability of correctly detecting the wrapping integers is only $P_c \approx 0.6988$, with upper and lower bounds in (5.69) equal because $r = 1$. The most probable six components are visible via the colormap used to display the Gaussian mixture on the fundamental parallelotope.

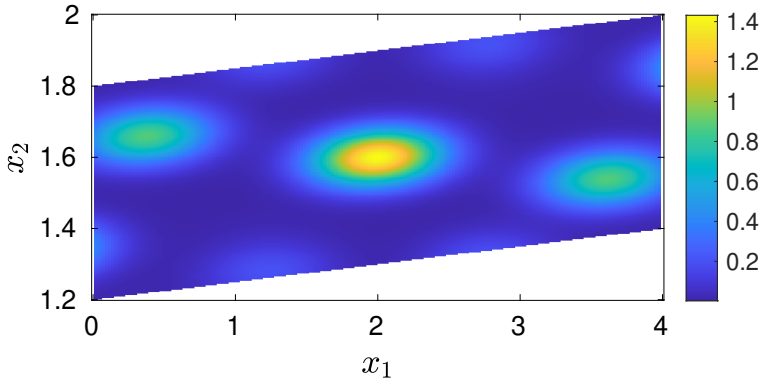


Figure 5.4: Example distribution, $f(\widehat{\boldsymbol{x}} \,|\, \boldsymbol{x}, \boldsymbol{\ell})$, for $m = 3$, $n = 2$.

5.6.7 Direction of Arrival Estimation

Consider a planar array of L isotropic sensors that is used to estimate the direction of arrival (DoA) of a narrowband signal $s(t) \in \mathbb{C}$ from a single emitter located in the far field of the array. The demodulated array output vector is given by

$$\boldsymbol{b}(t) = \boldsymbol{a}(\boldsymbol{x})s(t) + \boldsymbol{\varepsilon}(t), \qquad t = 1, 2, \ldots, T, \tag{5.71}$$

where $\boldsymbol{a}(\boldsymbol{x})$ is the array steering vector and $\boldsymbol{\varepsilon}(t)$ is zero-mean circular Gaussian noise that is uncorrelated across snapshots indexed by t. The array steering vector is given by

$$\boldsymbol{a}(\boldsymbol{x}) = \left[e^{i\pi \boldsymbol{p}_1^{\mathsf{T}} \boldsymbol{x}} \quad e^{i\pi \boldsymbol{p}_2^{\mathsf{T}} \boldsymbol{x}} \quad \cdots \quad e^{i\pi \boldsymbol{p}_L^{\mathsf{T}} \boldsymbol{x}}\right]^{\mathsf{T}}, \tag{5.72}$$

The position vectors $\boldsymbol{p}_j \in \mathbb{R}^n$, for $j \in \{1, 2, \ldots, L\}$, give the sensor positions in units of one-half wavelength. For linear arrays, $n = 1$ and $x = \sin\theta$, where θ is the azimuth angle of the source. For planar arrays, $n = 2$ and $\boldsymbol{x} = [\cos\theta\cos\phi,\, \sin\theta\cos\phi]^{\mathsf{T}}$, with ϕ denoting the elevation angle to the source. For general arrays, $n = 3$ and $\boldsymbol{x} = [\cos\theta\cos\phi,\, \sin\theta\cos\phi,\, \sin\phi]^{\mathsf{T}}$.

The sample covariance, given by $\widehat{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{b}(t)\boldsymbol{b}^{\mathsf{H}}(t)$, is a sufficient statistic of the data [29], and the MLE for $\boldsymbol{x}$ in (5.71) is given by [22, p. 288]

$$\arg\max_{\boldsymbol{x}\in\mathcal{B}} \ \boldsymbol{a}^{\mathsf{H}}(\boldsymbol{x})\widehat{\boldsymbol{R}}\boldsymbol{a}(\boldsymbol{x}), \tag{5.73}$$

where $\mathcal{B}$ is the set of feasible values for $\boldsymbol{x}$. For $n = 1$, the set $\mathcal{B}_1 = \left\{\boldsymbol{x} \mid \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}$ is the unit ball in $\mathbb{R}$, $\|x\| \leq 1$. For a planar array, $n = 2$ and the set $\mathcal{B}_2 = \left\{\boldsymbol{x} \mid \theta \in [0, 2\pi), \phi \in \left[0, \frac{\pi}{2}\right]\right\}$ of all DoAs is the unit ball in $\mathbb{R}^2$, $\|\boldsymbol{x}\| \leq 1$. These DoAs correspond to source directions on the upper hemisphere. If, for the planar array, the source is constrained to lie in the plane of the sensors, then $\mathcal{B}_2'$ is the unit circle in the plane. For a 3D array, $n = 3$ and the set $\mathcal{B}_3 = \left\{\boldsymbol{x} \mid \theta \in [0, 2\pi), \phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}$ of all DoAs is a sphere in $\mathbb{R}^3$, $\{\boldsymbol{x} \,|\, \|\boldsymbol{\omega}\| = 1\}$.

The evaluation of the MLE in (5.73) often involves a costly grid search over the non-convex cost surface. This grid search can be accelerated by the use of a multi-resolution grid search as in [212], though this involves some tuning of the succession of grid spacings over which the search will be conducted. The grid search will also necessarily introduce a bias.

Phase unwrapping estimators [23,24,31] offer a fast grid-free alternative. A phase unwrapping estimator uses only the phase information to estimate the direction of arrival. As such, this option is viable only when estimating the DoA of a single source [213–215].

Let $\boldsymbol{v}$ denote the vector of $\binom{L}{2}$ lower-triangular entries of $\widehat{\boldsymbol{R}}$ in column-major order. That is,

$$\boldsymbol{v} = [\,\widehat{r}_{2,1}, \widehat{r}_{3,1}, \ldots, \widehat{r}_{L,L-1}\,]^{\mathsf{T}} \tag{5.74}$$

where $\widehat{r}_{i,j}$ is the $(i,j)^{\text{th}}$ entry of $\widehat{\boldsymbol{R}}$. The whitened measurement vector $\boldsymbol{y}$ in (5.1) is then given by

$$\boldsymbol{y} = \widetilde{\boldsymbol{\Sigma}}^{-1/2}(2\pi)^{-1}\angle\boldsymbol{v}, \tag{5.75}$$

where $\angle$ returns the phase of a complex quantity, and $\widetilde{\boldsymbol{\Sigma}} = \operatorname{cov}\left((2\pi)^{-1}\angle\boldsymbol{v}\right)$. The covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ can be approximated as in [31]; see Appendix C for details. The encoding matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \tfrac{1}{2}\widetilde{\boldsymbol{\Sigma}}^{-1/2}\left[\,(\boldsymbol{p}_2 - \boldsymbol{p}_1), (\boldsymbol{p}_3 - \boldsymbol{p}_1), \ldots, (\boldsymbol{p}_L - \boldsymbol{p}_{L-1})\,\right]^{\mathsf{T}} \tag{5.76}$$

and $\boldsymbol{M} = \widetilde{\boldsymbol{\Sigma}}$. This construction allows the use of Algorithm 10 to estimate DoA from only the phase data of the sample covariance matrix. This approach offers a fast, grid-free alternative to the MLE in (5.73).

The phase unwrapping estimator conducts a depth-first search to find a solution to (5.63), though the solution may not correspond to a direction in the feasible set $\mathcal{B}$ [216]. A non-feasible $\widehat{\boldsymbol{\omega}}$ is projected to the feasible set as

$$\widehat{\boldsymbol{\omega}}_{\text{MLPUE}} = \underset{\boldsymbol{\omega}\in\mathcal{B}}{\arg\min} \min_{\boldsymbol{q}\in\mathbb{Z}^n} \|\widehat{\boldsymbol{x}} - \boldsymbol{\omega} - \boldsymbol{V}\boldsymbol{q}\|. \tag{5.77}$$

More generally, multiple solutions may be accepted from the sphere decoding step, only to keep the projected solution maximizing the likelihood.

5.7 Examples

In this section we provide computed examples for the three ML solution methods given in the chapter.

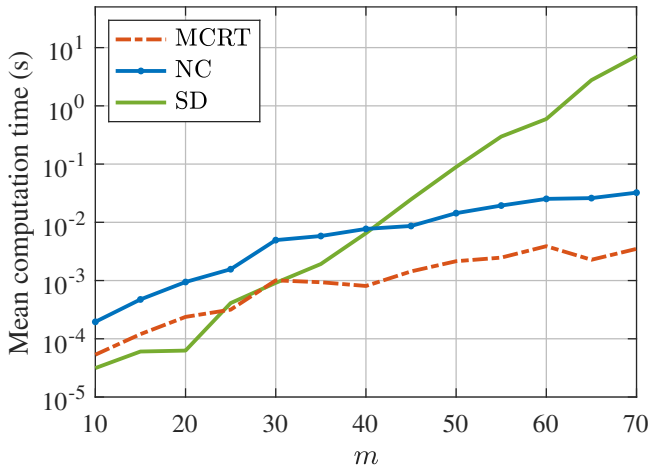


Figure 5.5: Mean computation time plotted against m for Algorithm 8 (MCRT), Algorithm 9 (NC), and the sphere decoding approach of [7] (SD).

5.7.1 Computation Time Comparison

First, we compare the computation times of the methods. We form a series of encoding matrices $\boldsymbol{A}$ with $n = 2$ and $m \in [10, 70]$. Each such $\boldsymbol{A}$ satisfies Condition 7 of Section 5.4 and has $\boldsymbol{V} = \boldsymbol{I}_2$ and $\boldsymbol{C} = \text{diag}([1/5, 1/3]^{\mathsf{T}})$. The noise has covariance $\boldsymbol{\Sigma} = 10\boldsymbol{I}_m$. In Fig. 5.5, the average times to compute the ML estimate via Algorithm 8, (labeled "MCRT"), Algorithm 9 (labeled "NC"), and the Algorithm 10 (labeled "SD") are plotted. The results are from 100 random draws of noise and parameters $\boldsymbol{x}$ and $\boldsymbol{k}$. The algorithms have identical estimation error; here, we compare only the computation time. Note that each algorithm uses a precomputation step that is not included in the timing results. This step need only be done once for a given $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$. For reasonably small m, the sphere decoding approach is faster than either polynomial-time method proposed here. As m becomes large, the mean computation time of the sphere decoding approach increases rapidly [174]; indeed,

for no fixed signal-to-noise ratio (SNR) is there a polynomial upper bound on the expected complexity of sphere decoding that holds for all m [217]; the rate of the exponential function dictating average complexity depends on SNR. Compared to the polynomial-time algorithms, the sphere decoding approach also exhibits greater variance in computing time, which can result in undesirable decoding and complications in hardware implementation [218]. Nevertheless, the sphere decoding approach remains an efficient option for many practical problems involving (5.1).

5.7.2 Encoding Matrices in Application

Table 5.1 refers to encoding matrices from the literature and provides context relevant to the two polynomial time algorithms given in this chapter. For sensor arrays, $\boldsymbol{A}$ corresponds to pairwise differences of the element positions. Table 5.1 shows that there are encoding matrices in use that meet Condition 7 and those that do not. For the $\boldsymbol{A}$ matrix in [219], $K = K_{\mathrm{NC}}$. Thus, despite meeting Condition 7, Algorithm 8 provides no computational benefit, and it is advisable to use the simpler Algorithm 9. Though, for the encoding matrix made by the wavelengths denoted "set E" in [178], the computational benefit is significant, with $K_{\mathrm{NC}} > 585K$.

5.7.3 Direction of Arrival Estimation With Planar Array

As an example for the estimation approach in Section 5.6, we simulate array outputs from the cross-shaped five-element planar array configuration from [8], with positions given by $[0, 0]^{\mathsf{T}}$, $[-4, 0]^{\mathsf{T}}$, $[0, 4]^{\mathsf{T}}$, $[0, -5]^{\mathsf{T}}$, and $[5, 0]^{\mathsf{T}}$, in units of one-half wavelength. The MLE in (5.73) is evaluated using a multi-resolution grid search over $\boldsymbol{x}$ [212]. We use an initial grid search with a grid spacing in $\boldsymbol{x}$ of 10^{-2}, followed

Table 5.1: Algorithms and problem sizes for examples of the linear model (5.1) appearing in application.

Reference	m	n	Condition 7	Alg.	K	K_{NC}
[216]	10	3	false	Alg. 9	-	264
[219]	10	3	true	Alg. 9	200	200
[220]	10	1	false	Alg. 9	-	34
[221]	10	2	false	Alg. 9	-	1143
[178, set E]	5	1	true	Alg. 8	5	2927
[178, set F]	5	1	false	Alg. 9	-	2311
Appendix B.5	16	3	true	Alg. 8	1087	8953

by a second search over the grid cells adjacent to the optimal point in the first search, with a spacing of 10^{-3}.

We compare the MLE in (5.73), which uses the full sample covariance data, amplitude and phase, to an approximate MLE that uses phase unwrapping estimation. The phase unwrapping estimator uses only the phase information of the sample covariance. Phase unwrapping estimation allows us to use Algorithm 10. To do so, we make the approximation that the phase noise follows a zero-mean Gaussian distribution with the covariance matrix determined as in [31]; see Appendix C for details.

The root mean square error (RMSE) results of the two estimators from $T = 10$ snapshots are plotted in Fig. 5.6 against signal-to-noise ratio (SNR) for a true angle of arrival of $\theta = \pi/2$, $\phi = \pi/4$. The results are computed from 10^6 trials at each SNR. The approximate MLE based on phase unwrapping (labeled MLPUE in Fig. 5.6, for "ML phase unwrapping estimator") matches the RMSE performance of the MLE. The square root of the trace (denoted $\mathrm{tr}(\cdot)$ in Fig. 5.6) of the Cramér-Rao bound (CRB) [222] of $\boldsymbol{x}$ in (5.71) is also shown. Both estimators approach this lower bound

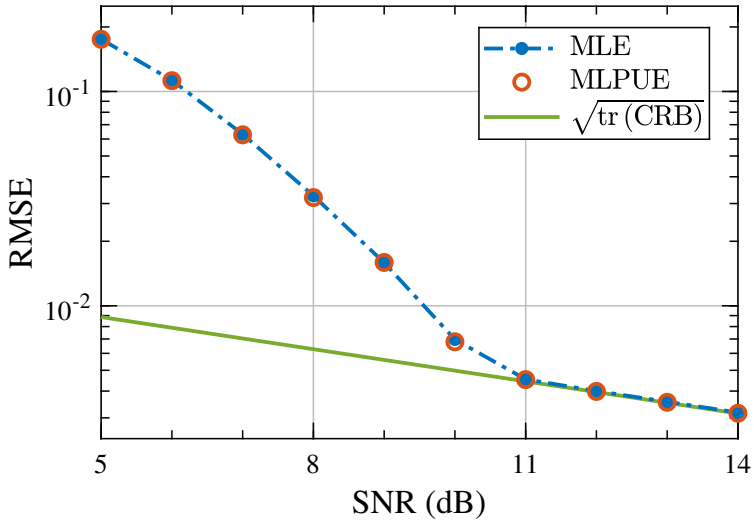


Figure 5.6: DoA estimation performance against SNR with $T = 10$ snapshots from five-element planar array [8]. RMSE results for the MLE of (5.73) and the phase uwrapping MLE are shown along with the square root of the trace of the CRB for $\boldsymbol{x}$. The results are averaged over 10^6 trials.

for SNR values exceeding a threshold [223] of roughly 11 dB. In this example, the MLPUE approach is on average over 85 times faster than the multi-resolution grid search.

5.7.4 Magnetic Resonance Flow Imaging

We next demonstrate the use of the estimation approach in Section 5.6 in phase-contrast magnetic resonance imaging (PC-MRI). PC-MRI encodes the velocity of spins into the image voxel phase through the first moment, $\boldsymbol{p} \in \mathbb{R}^3$, of the time varying gradient field:

$$b_k(t) = s(t)e^{i(\phi_0 + \gamma \boldsymbol{p}_k^{\mathsf{T}} \boldsymbol{x})} + \varepsilon_k(t), \tag{5.78}$$

where $k = 1, 2, \ldots, L$ indexes encodings, ϕ_0 is an unknown background phase, γ is the gyromagnetic ratio, $\boldsymbol{x} \in \mathbb{R}^3$ is velocity, and $\varepsilon_k(t)$ is zero-mean circular Gaussian

noise that is uncorrelated across T coils indexed by t. Proceeding with the sample covariance matrix as in Section 5.7.3, $\boldsymbol{y}$ is formed from the complex measurements as in (5.75). Each entry in $\boldsymbol{y}$ is a scaled phase difference from two encodings, averaged over coils, thereby canceling the unknown background phase.

Pelc et al. [9] introduced 3-directional flow encoding and a velocity estimator. Phase differences from balanced four-point encoding yield

$$\boldsymbol{A}^{\mathsf{T}} = \frac{\gamma\rho}{\pi}\begin{bmatrix} 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & 1 & -1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}, \tag{5.79}$$

where ρ is the absolute value of the first moment in each spatial direction. Phase differences are then pre-processed to conveniently yield a decoupled set of three equations in $n = 3$ unknown velocity components. Specifically, the pre-processing is given by a linear operator $\boldsymbol{P} \in \mathbb{R}^{3\times 6}$, given by

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}. \tag{5.80}$$

This preprocessing yields $\boldsymbol{PA} = \frac{2\gamma\rho}{\pi}\boldsymbol{I}$ and a least squares estimate, $\boldsymbol{x}_{\text{LS}} = \frac{\pi}{2\gamma\rho}\boldsymbol{Py}$.

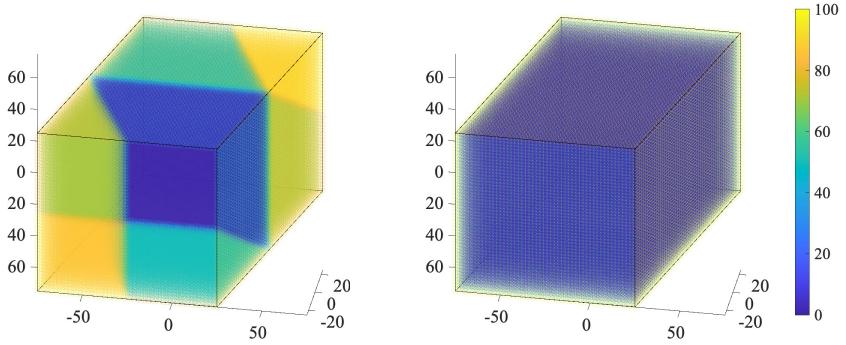


Figure 5.7: Velocity RMSE on $\mathcal{P}_{\boldsymbol{V}} \in \mathbb{R}^3$ for (a) the estimator in [9] and (b) the MLPUE. Velocity components and RMSE are shown in cm/s.

Fig. 5.7 shows RMSE versus true $\boldsymbol{x}$ across one choice of the fundamental parallelotope $\mathcal{P}_{\boldsymbol{V}}$. The example is computed for $\mathrm{var}(\varepsilon_k(t)) = 1$, $s(t) = 10$, $T = 8$ coils, and $\gamma\rho = \frac{\pi}{100}$ s/cm, with 10^3 trials at each $\boldsymbol{x}$. The estimator $\boldsymbol{x}_{\mathrm{LS}}$ is aliased on half of $\mathcal{P}$, as shown in Fig. 5.7a, whereas the MLPUE allows un-aliased estimation across most of $\mathcal{P}$, illustrated in Fig. 5.7b. With the MLPUE, wrap-around error is encountered only when the true speed is within approximately 2 cm/s of a boundary. With the four-point balanced encoding, the error covariance matrices for both $\boldsymbol{x}_{\mathrm{LS}}$ and the MLPUE are a scaled identity, signifying i.i.d. errors; but, the processing in [9] results in a 95% confidence sphere with 3.7% larger volume, because it ignores noise correlation in the phase differences. Finally, the sphere decoding for the MLPUE is, on average, over 500 times faster than grid-search estimation in (5.73) using 100 grid points per dimension [224].

5.8 Summary

In this chapter, we have proposed three methods for the ML estimation task in (5.2). Generally, the problem can be solved using the sphere decoding approach given in Section 5.6. The computation time of this approach is dependent on the noise realization, and the worst-case computational complexity is not polynomial in m, the length of the measurement vector. In [32], an ML solution method for (5.2) is given for the special case of a single unknown variable ($n = 1$), diagonal noise covariance, and a restriction on the encoding matrix $\boldsymbol{A}$. This method has worst-case polynomial-time complexity. In this chapter, we have presented estimation algorithms that extend polynomial-time computation to a wider range of problems with diagonal covariance. The extensions presented here account for both multivariate unknowns

($n > 1$) and general encoding matrices with full column rank. The method given in Section 5.4 use a multivariate CRT to further reduce algorithm complexity.

5.9 Future Work

In this chapter, we have described two ML estimation methods for the problem in (5.2) that have polynomial complexity in m, the number of measurements. Both of these methods require that the measurement noise be uncorrelated. We have also described a sphere decoding approach that accommodates correlated noise, though the worst-case computational complexity is non-polynomial in m. It remains to find a polynomial-time algorithm that accommodates correlated noise. A structure is present for the special case of (5.2) with measurements corresponding to phase differences and the covariance matrix $\mathbf{\Sigma}$ found according to the phase noise approximation described in Appendix C. This structure may admit an efficient estimation procedure.

Another avenue of continued work is the leveraging of the proposed techniques for MAP estimation. In this chapter, we have considered estimation from a single set of measurements. Though, often, as in phase-contrast MRI, a parameter is estimated for each voxel in an image. The per-voxel estimation proposed here could be a component in an estimation method that leverages spatial correlations to improve estimation performance, as in [31, 225].

It is often necessary to estimate the DoA of multiple targets simultaneously, and a notable limitation of the methods in this chapter is that they apply only to situations with a single target. Future research could extend these methods to apply to multiple targets. Such an extension would also allow for the simultaneous estimation

of multiple frequency components in frequency estimation problems and the heights of multiple scatterers within a resolution cell in interferometric SAR.

Chapter 6: Aliasing in Nonuniform Arrays

6.1 Introduction

A sensor array is said to be "alias-free" on some set $\mathcal{B}$ of directions of arrival (DoAs) if array steering vectors are distinct for all DoAs in $\mathcal{B}$. In this chapter[11], a necessary and sufficient condition is presented for an arbitrary sensor array to be alias-free. The results apply to linear, planar, and 3D arrays; they hold for any predefined subset of DoAs for which the array must be unambiguous. The results extend existing literature to provide simple geometrical interpretation for use in array design.

In [33], Godara and Cantoni give necessary and sufficient conditions for alias-free arrays in one, two, or three dimensions. For L sensors, the test entails a combinatorial search over L-tuples of integers. For planar and 3D arrays, [33] also provides easily tested, but restrictive, sufficient conditions, requiring a smallest inter-sensor spacing not exceeding $\lambda/2$, where λ is the wavelength of the signal. The sufficient conditions have been widely used in design [226–228]. A slightly less restrictive sufficient condition for planar arrays is given in [229, Thm. 2].

Motivated by array design, alternative sufficient conditions have been devised for linear arrays. To avoid aliasing over the full range of azimuth angles, uniform linear

[11]This chapter is largely drawn from [216], coauthored with Shen Zhao, Rizwan Ahmad, and Lee C. Potter ().

arrays require an inter-sensor spacing that is no more than $\lambda/2$ [230]. It may be prohibitive to have two sensors within $\lambda/2$, due either to physical size limitations or mutual coupling. Nonuniform linear arrays allow for alias-free operation despite a minimum inter-sensor spacing that exceeds $\lambda/2$ [8, 221, 231, 232]. A difference *co-array* is a virtual array with positions given by the pairwise differences of the physical sensor locations. Particular attention has been given to array geometries for which the co-array has uniform sampling and is fully filled, such as minimum redundancy arrays [233], nested arrays [234], and co-prime arrays [235]. Thinned arrays refer to the special case of a uniform array from which some elements have been removed. For a thinned linear array of L sensors with $\lambda/2$ spacing, a necessary and sufficient condition for no aliasing is given in [236]: for reference sensor at 0 and positions $n_i\lambda/2$, the integers $\{n_2, n_3, \ldots, n_L\}$ must have a greatest common divisor (gcd) of 1. Specialized to a nonuniform linear array of $L = 3$ sensors at positions $\{0, q_1, q_2\}$ with $0 < q_1 < q_2$, [237] posits the sufficient condition that $q_2/q_1 = m/\ell$ for co-prime integers m, ℓ, with $|\ell| > \lfloor 2q_1/\lambda \rfloor$ or $|m| > \lfloor 2q_2/\lambda \rfloor$. Likewise for $L = 3$, [213, 238] establish the sufficient condition that the co-array contains a pair of elements not more than $\lambda/2$ apart.

A task very similar to array design arises in phase-contrast magnetic resonance imaging [31, 168, 225, 239, 240]. There, the velocity of spins at a voxel is encoded into image phase differences. Encoded images are acquired by adopting a time-varying gradient field, and the first moments of these waveforms are analogous to sensor positions. For velocity in one or several dimensions, unambiguous performance is required up to some maximum speed, such as 300 cm/s.

6.2 Signal Model

Consider data collected from an array with L sensors receiving a narrow-band signal from a single emitter located in the far field. The demodulated complex-valued $L \times 1$ array output vector is given by

$$\boldsymbol{b}(t) = \boldsymbol{a}(\boldsymbol{x})s(t) + \boldsymbol{\varepsilon}(t), \qquad t = 1, 2, \ldots, T, \tag{6.1}$$

where $s(t) \in \mathbb{C}$ is the baseband complex amplitude of the impinging wavefront, T is the number of snapshots, $\boldsymbol{a}(\boldsymbol{x}) \in \mathbb{C}^{L\times 1}$ is the array steering vector, $\boldsymbol{\varepsilon}(t) \in \mathbb{C}^{L\times 1}$ is circularly symmetric Gaussian noise with zero mean and covariance $\sigma^2 \boldsymbol{I}_L$, and $\boldsymbol{\epsilon}(t)$ is uncorrelated across snapshots.

For an array of isotropic sensors in n dimensions, the array steering vector is given by

$$\boldsymbol{a}(\boldsymbol{x}) = \begin{bmatrix} e^{i\pi \boldsymbol{p}_1^\mathsf{T}\boldsymbol{x}} & e^{i\pi \boldsymbol{p}_2^\mathsf{T}\boldsymbol{x}} & \ldots & e^{i\pi \boldsymbol{p}_L^\mathsf{T}\boldsymbol{x}} \end{bmatrix}^\mathsf{T}, \tag{6.2}$$

where $i = \sqrt{-1}$ and the vector $\boldsymbol{x} \in \mathbb{R}^d$ corresponds to the DoA of the impinging signal. The position vectors $\boldsymbol{p}_j \in \mathbb{R}^n$, for $j \in \{1, 2, \ldots, L\}$, give the sensor positions in units of one-half wavelength. For linear arrays, $n = 1$ and $x = \sin\theta$, where θ is the azimuth angle of the source. For planar arrays, $n = 2$ and $\boldsymbol{x} = [\cos\theta\cos\phi, \sin\theta\cos\phi]^\mathsf{T}$, with ϕ denoting the elevation angle to the source. For general arrays, $n = 3$ and $\boldsymbol{x} = [\cos\theta\cos\phi, \sin\theta\cos\phi, \sin\phi]^\mathsf{T}$.

6.3 Alias-free

For an array to be alias-free, the steering vectors in (6.2) must be distinct for distinct DoAs. Using the first sensor as a phase reference, the $(L-1)$ phase differences

are $\pi\left[\boldsymbol{p}_k - \boldsymbol{p}_1\right]^{\mathsf{T}}\boldsymbol{x}$, for $k \in \{2, 3, \ldots, L\}$. We define the $(L-1)$-by-n encoding matrix

$$\boldsymbol{A} = \tfrac{1}{2}\left[\,(\boldsymbol{p}_2 - \boldsymbol{p}_1), (\boldsymbol{p}_3 - \boldsymbol{p}_1), \ldots, (\boldsymbol{p}_L - \boldsymbol{p}_1)\,\right]^{\mathsf{T}}, \tag{6.3}$$

and assume that $\operatorname{rank}(\boldsymbol{A}) = n$ [33, Thm. 4]. Note that the encoding matrix $\boldsymbol{A}$ is determined entirely by the geometry of the array. Next, define $\mathcal{B}$ as the set of arrival directions. For $n = 1$, the set $\mathcal{B}_1 = \left\{\boldsymbol{x} \mid \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}$ is the unit ball in $\mathbb{R}$, $\|\boldsymbol{x}\| \leq 1$. For a planar array, $n = 2$ and the set $\mathcal{B}_2 = \left\{\boldsymbol{x} \mid \theta \in [0, 2\pi), \phi \in \left[0, \frac{\pi}{2}\right]\right\}$ of all DoAs is the unit ball in $\mathbb{R}^2$, $\|\boldsymbol{x}\| \leq 1$. These DoAs correspond to source directions on the upper hemisphere. If, for the planar array, the source is constrained to lie in the plane of the sensors, then $\mathcal{B}_2'$ is the unit circle in the plane. For a 3D array, $n = 3$ and the set $\mathcal{B}_3 = \left\{\boldsymbol{x} \mid \theta \in [0, 2\pi), \phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}$ of all DoAs is a sphere in $\mathbb{R}^3$, $\{\boldsymbol{x} \mid \|\boldsymbol{x}\| = 1\}$. These DoAs correspond to source directions on the sphere. Note that $\mathcal{B}$ may alternatively be defined as some subset of $\mathcal{B}_n$ in n dimensions, for example, with prior knowledge restricting the range of possible DoAs [241].

Definition 10 (Alias-free on $\mathcal{B}$)**.** *An array with the encoding matrix $\boldsymbol{A}$ is alias-free on $\mathcal{B}$ if, for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{B}$,*

$$\boldsymbol{A}(\boldsymbol{x}_2 - \boldsymbol{x}_1) \in \mathbb{Z}^{L-1} \Rightarrow \boldsymbol{x}_2 = \boldsymbol{x}_1.$$

When considering the set of all directions of arrival, $\mathcal{B} = \mathcal{B}_n$ and the array is simply said to be "alias-free."

Thus, we consider the system of linear congruence equations

$$\boldsymbol{A}\boldsymbol{x} \equiv \boldsymbol{0}_{L-1} \mod \boldsymbol{1}_{L-1}, \tag{6.4}$$

which can be reformulated as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{k}, \tag{6.5}$$

where $\boldsymbol{k} \in \mathbb{Z}^{L-1}$ is called a vector of *wrapping integers*. The set of solutions to (6.4) for an array with encoding matrix $\boldsymbol{A}$ is

$$\Omega_{\boldsymbol{A}} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \in \mathbb{Z}^{L-1}\}. \tag{6.6}$$

If (6.4) admits any nonzero solution, then $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$; that is, the solutions form a lattice $\Lambda_{\boldsymbol{V}} \subset \mathbb{R}^n$ [181, 242, 243], for some lattice basis $\boldsymbol{V}$. The solution set $\Omega_{\boldsymbol{A}}$ is thus either $\{\boldsymbol{0}\}$ or a lattice; in either case the solution set is closed under addition. Note that if $\Omega_{\boldsymbol{A}} = \{\boldsymbol{0}\}$, then the array is said to be alias-free on all $\mathbb{R}^n$.

We have that $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ if and only if, for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{A}\boldsymbol{x} \in \mathbb{Z}^{L-1} \Leftrightarrow \boldsymbol{x} \in \Lambda_{\boldsymbol{V}}. \tag{6.7}$$

Definition 11. *A matrix $\boldsymbol{Z} \in \mathbb{Z}^{c\times d}$ with $c \geq d$ is called* primitive *if, for all $\boldsymbol{q} \in \mathbb{Q}^d$ [244],*

$$\boldsymbol{Z}\boldsymbol{q} \in \mathbb{Z}^c \Rightarrow \boldsymbol{q} \in \mathbb{Z}^d. \tag{6.8}$$

A column vector is primitive if its elements have a greatest common divisor of unity [245]. Any c-by-d primitive matrix can be extended to a c-by-c unimodular matrix [186, 244].

Lemma 12. *A matrix $\boldsymbol{Z} \in \mathbb{Z}^{c\times d}$ with $c \geq d$ is primitive if and only if $\boldsymbol{U}\boldsymbol{Z}$ is primitive for any unimodular $\boldsymbol{U} \in \mathbb{Z}^{c\times c}$.*

Proof. For any unimodular $\boldsymbol{U} \in \mathbb{Z}^{c\times c}$ and any $\boldsymbol{q} \in \mathbb{Q}^d$,

$$\boldsymbol{U}\boldsymbol{Z}\boldsymbol{q} \in \mathbb{Z}^c \Leftrightarrow \boldsymbol{Z}\boldsymbol{q} \in \boldsymbol{U}^{-1}\mathbb{Z}^c \Leftrightarrow \boldsymbol{Z}\boldsymbol{q} \in \mathbb{Z}^c.$$

From Definition 11, $\boldsymbol{Z}$ is primitive if and only if $\boldsymbol{U}\boldsymbol{Z}$ is primitive. □

Lemma 13. *For invertible $\boldsymbol{V} \in \mathbb{R}^{n\times n}$, if $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{V}$ is primitive, then $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$.*

Proof. If $\boldsymbol{Z} = \boldsymbol{AV}$ is primitive, then for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\begin{aligned} \boldsymbol{Ax} \in \mathbb{Z}^{L-1} &\Leftrightarrow \boldsymbol{ZV}^{-1}\boldsymbol{x} \in \mathbb{Z}^{L-1} \Leftrightarrow \boldsymbol{V}^{-1}\boldsymbol{x} \in \mathbb{Z}^n \\ &\Leftrightarrow \boldsymbol{x} \in \boldsymbol{V}\mathbb{Z}^n \Leftrightarrow \boldsymbol{x} \in \Lambda_{\boldsymbol{V}}. \end{aligned} \tag{6.9}$$

For the second equivalence, sufficiency is given by Definition 11, and necessity is due to $\boldsymbol{Z}$ being integer-valued. From (6.7), $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$. □

Proposition 14. *For invertible $\boldsymbol{V} \in \mathbb{R}^{n\times n}$, $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ if and only if $\boldsymbol{A} = \boldsymbol{ZV}^{-1}$, for some primitive matrix $\boldsymbol{Z} \in \mathbb{Z}^{(L-1)\times n}$.*

Proof. Necessity: This is given by Lemma 13.

Sufficiency: Assume that $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$. From (6.7), for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{Ax} \in \mathbb{Z}^{L-1} \Leftrightarrow \boldsymbol{x} \in \Lambda_{\boldsymbol{V}} \Leftrightarrow \boldsymbol{x} \in \boldsymbol{V}\mathbb{Z}^n. \tag{6.10}$$

Thus, $\boldsymbol{AV}\mathbb{Z}^n \subset \mathbb{Z}^{L-1}$, which is true if and only if $\boldsymbol{AV} \in \mathbb{Z}^{(L-1)\times n}$.

From (6.10), for any $\boldsymbol{q} \in \mathbb{Q}^n$,

$$\boldsymbol{AVq} \in \mathbb{Z}^{L-1} \Leftrightarrow \boldsymbol{Vq} \in \boldsymbol{V}\mathbb{Z}^n \Leftrightarrow \boldsymbol{q} \in \mathbb{Z}^n. \tag{6.11}$$

Therefore, from Definition 11, $\boldsymbol{Z} = \boldsymbol{AV}$ is primitive. □

Corollary 15. *For any invertible $\boldsymbol{V} \in \mathbb{R}^{n\times n}$, $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ if and only if $\Omega_{\boldsymbol{UA}} = \Lambda_{\boldsymbol{V}}$, for any unimodular $\boldsymbol{U} \in \mathbb{Z}^{(L-1)\times(L-1)}$.*

Proof. From Lemma 12, $\boldsymbol{AV}$ is primitive if and only if $\boldsymbol{UAV}$ is primitive. Thus, according to Lemma 13, $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ if and only if $\Omega_{\boldsymbol{UA}} = \Lambda_{\boldsymbol{V}}$. □

Corollary 16. *For any invertible $\boldsymbol{V} \in \mathbb{R}^{n\times n}$, $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ if and only if $\Omega_{\boldsymbol{AS}} = \Lambda_{\boldsymbol{S}^{-1}\boldsymbol{V}}$, for any invertible $\boldsymbol{S} \in \mathbb{R}^{n\times n}$.*

Proof. From Proposition 14, if $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$, then $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{V}^{-1}$, for some primitive $\boldsymbol{Z}^{(L-1)\times n}$. This is true if and only if $\boldsymbol{A}\boldsymbol{S} = \boldsymbol{Z}\boldsymbol{V}^{-1}\boldsymbol{S}$, for any invertible $\boldsymbol{S} \in \mathbb{R}^{n\times n}$, which is true if and only if $\Omega_{\boldsymbol{A}\boldsymbol{S}} = \Lambda_{\boldsymbol{S}^{-1}\boldsymbol{V}}$. □

For linear arrays, the lattice basis becomes a scalar $v \in \mathbb{R}$. From Proposition 14, we have that a 1D array has $\Omega_{\boldsymbol{A}} = \Lambda_v$ if and only if the sensor displacements in (6.3) are setwise coprime integer multiples of a common real-valued factor $1/v$.

If $\boldsymbol{A}$ is rational, $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ is a full-rank lattice with invertible basis $\boldsymbol{V} \in \mathbb{R}^{n\times n}$; the basis $\boldsymbol{V}$ can be formed as in (5.16), (5.17), (5.18), (5.18) in Section 5.3.4 [7], with $\boldsymbol{M} = \boldsymbol{I}_{L-1}$. Therefore, a sufficient condition for $\Omega_{\boldsymbol{A}}$ to be a full-rank lattice is for $\boldsymbol{A}$ to be rational, i.e., for the sensors to have been placed on a Cartesian grid, such as might naturally happen in fabrication. The parallelotope $\mathcal{P}_{\boldsymbol{V}}$ defines a region in which the solution to (6.4) is unique.

For a given DoA $\boldsymbol{x}_0$, let $\boldsymbol{\phi}_0 = 2\pi(\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{k}_0) = \angle\boldsymbol{a}(\boldsymbol{x}_0)$ be the phases of the steering vector; $\boldsymbol{k}_0$ is the associated vector of wrapping integers. The least-squares solution $\widehat{\boldsymbol{x}} = \left\langle \boldsymbol{A}^{\dagger}\left(\frac{\boldsymbol{\phi}_0}{2\pi} + \widehat{\boldsymbol{k}}\right)\right\rangle_{\boldsymbol{V}}$ for $\widehat{\boldsymbol{k}} \neq \boldsymbol{k}_0$ corresponds to a beampattern side-lobe. Here, $\langle\cdot\rangle_{\boldsymbol{V}}$ refers to the operation defined in (5.7), by which a vector is translated to the parallelotope $\mathcal{P}_{\boldsymbol{V}}$.

6.4 Necessary and Sufficient Condition

For the remainder of this chapter, we assume that $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$, for some invertible $\boldsymbol{V} \in \mathbb{R}^{n\times n}$. From Definition 10, we immediately have:

Proposition 17. *An array is alias-free on $\mathcal{B}$ if and only if translations $\mathcal{B} + \{\boldsymbol{u}_i\}$ and $\mathcal{B} + \{\boldsymbol{u}_j\}$ are disjoint for all $\boldsymbol{u}_i \neq \boldsymbol{u}_j \in \Lambda_{\boldsymbol{V}}$.*

Specializing to the case of the full range of DoAs, defined as $\mathcal{B}_n$ above, yields the following corollary.

Corollary 18. *An array is alias-free if and only if*

$$\min_{\boldsymbol{u}\in\Lambda_{\boldsymbol{V}}\setminus\{\boldsymbol{0}\}} \|\boldsymbol{u}\| > 2, \tag{6.12}$$

i.e., the smallest norm of a non-zero lattice point in $\Lambda_{\boldsymbol{V}}$ *is greater than* 2.

Proof. Recall $\mathcal{B}_n$ as defined in 6.3: unit balls in $\mathbb{R}$ and $\mathbb{R}^2$ or the sphere in $\mathbb{R}^3$. Then, $\mathcal{B}+\{\boldsymbol{u}_i\}$ and $\mathcal{B}+\{\boldsymbol{u}_j\}$ have an empty intersection for all $\boldsymbol{u}_i \neq \boldsymbol{u}_j \in \Lambda_{\boldsymbol{V}}$ if and only if $\|\boldsymbol{u}_j - \boldsymbol{u}_i\| > 2$ for all $\boldsymbol{u}_i \neq \boldsymbol{u}_j \in \Lambda_{\boldsymbol{V}}$. As lattices are closed under addition, the array is alias-free if and only if $\|\boldsymbol{u}\| > 2$, for all $\boldsymbol{u} \in \Lambda_{\boldsymbol{V}} \setminus \boldsymbol{0}$. □

Godara and Cantoni [33, Thm. 6] provide an alternative necessary and sufficient test for alias-free on the full set of DoAs: the image of the unit ball under $\boldsymbol{A}$ contains no integer point other than the origin. The test is applied by combinatorial search over $\boldsymbol{k} \in \mathbb{Z}^{L-1}$, with $\|\boldsymbol{k}\| < \|\boldsymbol{A}\|_2$. In contrast, construction of the lattice basis $\boldsymbol{V}$ provides for an easily tested necessary and sufficient condition for arbitrary $\mathcal{B}$. Specialization to the set of all possible directions of arrival in one or several dimensions is considered below.

6.4.1 Linear Arrays

For $n = 1$, the fundamental parallelotope $\mathcal{P}_{\boldsymbol{v}} \subset \mathbb{R}$ reduces to an interval and is unique, up to a sign. By the Chinese remainder theorem and Corollary 18, a linear array is alias-free if and only if

$$\mathrm{LCM}\left(\mathbf{1}_{(L-1)} \oslash \boldsymbol{A}\right) > 2, \tag{6.13}$$

where LCM$(\cdot)$ denotes the least common multiple, which is the smallest positive real number that is an integer multiple of all input numbers. For example,

$$\text{LCM}\left(\tfrac{\pi}{3}, \tfrac{\pi}{2}, \tfrac{2\pi}{3}\right) = 2\pi.$$

Equation (6.13) is an equivalent and alternative test to [33, Thm. 11]: there exists sensor separation less than or equal to $\lambda/2$, or for minimum spacing $q\alpha > \lambda/2$ with integer q, there exists a third element such that its distance from any one of the two elements is not an integer multiple of α. The simplicity of (6.13) aids in design.

6.4.2 Nonlinear Arrays

For planar or 3D arrays, the lattice basis, and hence the corresponding fundamental parallelotope, are not unique. Equation (6.12) provides a simple necessary and sufficient test, in contrast to the widely used sufficient condition [33, Thm. 8]. Note that evaluating (6.12) is equivalent to solving a shortest lattice vector problem in n dimensions.

6.4.3 Invariances and Scaling

From (6.6), we see that $\Lambda_{\boldsymbol{V}}$ is invariant to translations of the point set $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L\}$. According to Corollary 15, $\Lambda_{\boldsymbol{V}}$ is invariant to unimodular transformation of $\boldsymbol{A}$. Corollary 16 describes the effect of an invertible transformation of $\boldsymbol{A}$ on the lattice of solutions to (6.4); this property can be used to examine the effects of dilation or contraction of a point set in array design. We also have the following property showing that an alias-free array remains alias-free following orthonormal transformation (including rotation or reflection) of its point set.

Proposition 19. *If the array with encoding matrix* $\boldsymbol{A}$ *is alias-free, then the array with encoding matrix* $\boldsymbol{AB}^{\mathsf{T}}$ *is alias-free, for any* n*-by-*n *matrix* $\boldsymbol{B}$ *with orthonormal columns.*

Proof. Assume that $\Omega_{\boldsymbol{A}} = \Lambda_{\boldsymbol{V}}$ and that the array with encoding matrix $\boldsymbol{A}$ is alias-free. By Corollary 16, $\Omega_{\boldsymbol{AB}^{\mathsf{T}}} = \Lambda_{\boldsymbol{BV}}$ (note that $\boldsymbol{B}^{-1} = \boldsymbol{B}^{\mathsf{T}}$). From Corollary 18,

$$\begin{aligned}\min_{\boldsymbol{u}\in\Lambda_{\boldsymbol{V}}\setminus\{\boldsymbol{0}\}} \|\boldsymbol{u}\| > 2 &\Leftrightarrow \min_{\boldsymbol{\lambda}\in\mathbb{Z}^n\setminus\{\boldsymbol{0}\}} \|\boldsymbol{V\lambda}\| > 2 \\ &\Leftrightarrow \min_{\boldsymbol{\lambda}\in\mathbb{Z}^n\setminus\{\boldsymbol{0}\}} \|\boldsymbol{BV\lambda}\| > 2 \Leftrightarrow \min_{\boldsymbol{u}\in\Lambda_{\boldsymbol{BV}}\setminus\{\boldsymbol{0}\}} \|\boldsymbol{u}\| > 2.\end{aligned}$$

Therefore, the array with encoding matrix $\boldsymbol{AB}^{\mathsf{T}}$ is alias-free. □

6.5 Examples

6.5.1 Linear Array

In [221] a four-element linear array is proposed for meteor trajectory estimation with element positions given, in units of one-half wavelength, by $\{0, 4, 9, 40\}$. The array, an extension of an interferometric system in [8], is alias-free for the full range of azimuth angles. The application requires a minimum inter-sensor spacing much larger than $\lambda/2$ to limit the effects of mutual coupling. As an example, we design an alternative array configuration to minimize the Cramér-Rao lower bound (CRB) on estimation error variance subject to the constraints that the minimum inter-sensor spacing and peak side-lobe level meet those of the array in [221].

The source signal $\{s(t)\}$ follows a zero-mean circularly symmetric Gaussian distribution with power ρ^2 and is independently and identically distributed across snapshots, indexed by t [22, 230]. Under this model with T snapshots and L sensors, the

CRB is given by (e.g., [222, 223])

$$\mathrm{CRB} = \frac{L \cdot \mathrm{snr} + 1}{2\pi^2 T L^2 M \mathrm{snr}^2} \tag{6.14}$$

where snr denotes the signal-to-noise ratio ρ^2/σ^2 and

$$M = \frac{1}{L}\sum_{k=1}^{L}\left(p_k - \frac{1}{L}\sum_{n=1}^{L} p_n\right)^2 \tag{6.15}$$

is the moment of inertia.

For a 20 dB signal-to-noise ratio as in [8], the array in [221] has a CRB of 5.09×10^{-7} and a peak side-lobe level of 0.369 dB. Assuming that $|\theta| < 80°$ [221], we follow a design procedure similar to that in [31] to obtain the element locations $0.986 \cdot \{0, 5, 41, 78\}$. This new array configuration has a peak side-lobe level of 0.369 dB, a reduced CRB of 1.25×10^{-7}, and from (6.13), an unambiguous interval for $\sin\theta$ of 1.97.

Next, we interpret (6.13) for an arbitrary three-element linear array with rational positions. Without loss of generality, let $p_1 = 0$ and $(p_2 - p_1) \leq (p_3 - p_2)$. Select co-prime integers q and ℓ such that $\frac{p_3-p_1}{p_3-p_2} = \frac{q}{\ell} \in (1, 2]$. Then, from (6.13) with $\mathcal{B} = (-1, 1)$, the array is alias-free on $\mathcal{B}$ if and only if

$$\frac{q-\ell}{(p_2-p_1)} \geq 1. \tag{6.16}$$

In [213, 214, 238], the co-array is constrained to have a pair of positions not more than $\lambda/2$ apart; from (6.16), this sufficient condition is not necessary, allowing additional design flexibility. If $[-1, 1]$ is a strict subset of the unambiguous interval $\mathcal{P}_{\boldsymbol{V}}$, then the two end-fire DoAs, $\pm\sin(\pi/2)$, are disambiguated.

6.5.2 Planar Arrays

Consider a uniform circular array with $L = 6$ and one-wavelength spacing [246–248]. The lattice has basis $\boldsymbol{V} = \mathrm{diag}(\{2, 2\sqrt{3}/3\})$. Corollary 18 is violated, because a vertex in $\mathcal{P}_{\boldsymbol{V}}$, given by $\left[0, \frac{2}{3}\sqrt{3}\right]^{\mathsf{T}}$, has norm $\frac{2}{3}\sqrt{3} \approx 1.1547 < 2$. This is evident in Fig. 6.1a, where the boundary of $\mathcal{B}_2$ is shown translated to each of the several lattice points displayed, and the translated unit balls overlap. For a regular hexagonal array, the largest spacing at which the unit balls become tangent is $\frac{\sqrt{3}}{3}\lambda \approx 0.5774$ wavelength. The minimum spacing $\frac{\sqrt{3}}{3}\lambda > \frac{1}{2}\lambda$ violates the widely used sufficient condition [33, Thm. 9].

Fig. 6.1b depicts the lattice and shifted unit balls for a four-element array with positions $[0, 0]^{\mathsf{T}}$, $[\pm 1.71, 0]^{\mathsf{T}}$, and $[0.26, 2.11]^{\mathsf{T}}$. The array is alias-free on $\mathcal{B}_2$, although no fundamental parallelogram contains a unit ball and the smallest inter-sensor spacing is 0.855λ.

Fig. 6.1c shows the lattice and shifted $\mathcal{B}$ for a four-element array with a reduced set of DoAs given by $\mathcal{B} = \{\boldsymbol{x} \mid \theta \in [-\pi/3, \pi/3],\, \phi \in [0, \pi/2]\}$. Note that the array is not alias-free for the full set of DoAs $\mathcal{B}_2$; the minimum separation between the parallelogram vertices is 1.335, and the condition of Corollary 18 is not satisfied. However, for the subset $\mathcal{B}$, the condition in Proposition 17 is satisfied, and the array is alias-free on $\mathcal{B}$. The lack of aliasing is illustrated by the disjointedness of the hourglass-shaped translations of $\mathcal{B}$ in Fig. 6.1c. The array is designed from the desired parallelogram via $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{V}^{-1}$ with primitive matrix $\boldsymbol{Z}$ yielding positions $[0, 0]^{\mathsf{T}}$, $[2.67, 0.42]^{\mathsf{T}}$, $[1.78, 2.48]^{\mathsf{T}}$, and $[0.89, 4.54]^{\mathsf{T}}$; the minimum separation between sensors is 1.122λ. This example shows how Proposition 17 may be used to design arrays that avoid aliasing over a subset of DoAs, as is sought in [221].

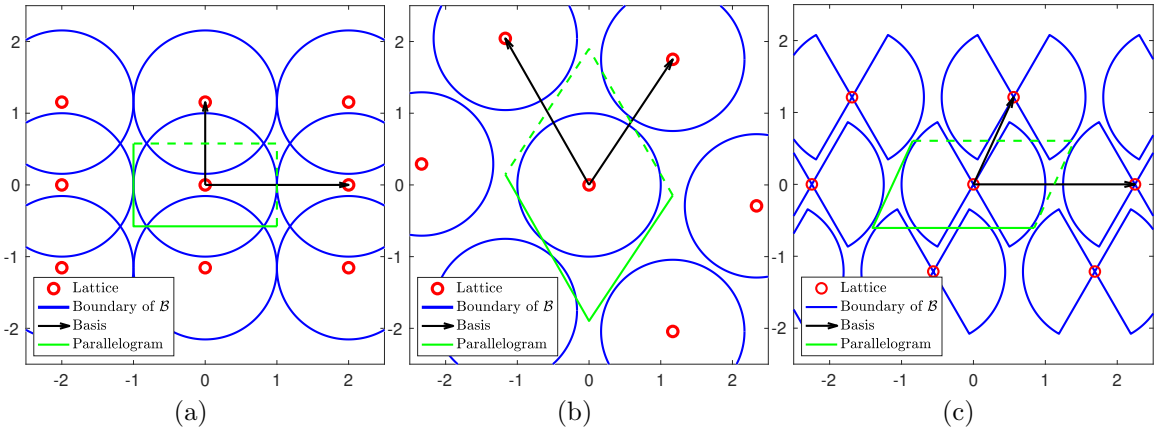
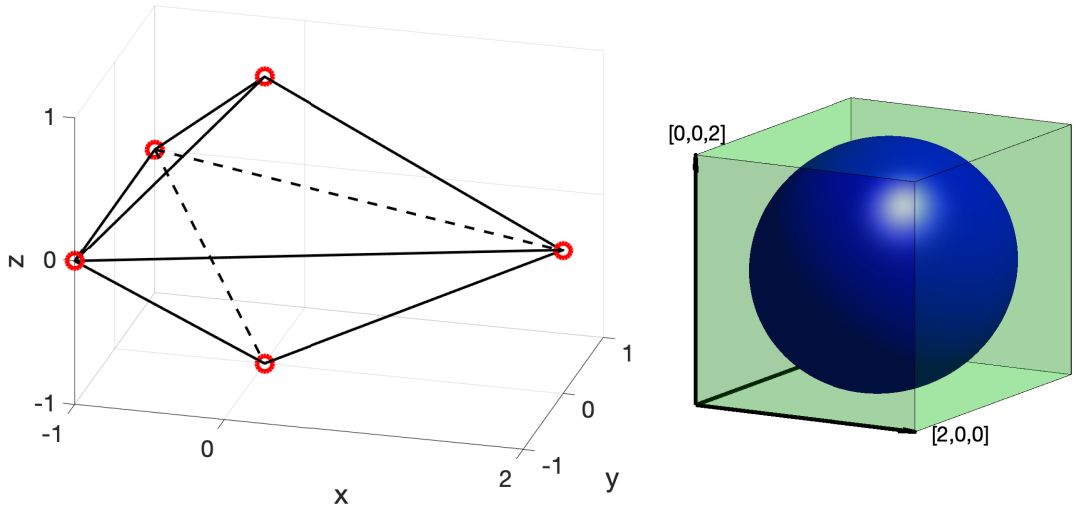


Figure 6.1: (a) For a uniform circular array of six sensors with λ spacing, illustration of the lattice Λ, a basis $\boldsymbol{V}$, a fundamental parallelogram, and the disc of all DoAs, $\mathcal{B}_2$, translated periodically per the lattice. In this case, the translated discs intersect, and the array is aliased. (b) The lattice and shifted $\mathcal{B}_2$ for a four-sensor planar array that is alias-free, despite a minimum sensor separation of 0.855λ. (c) A second four-sensor planar array in which $\mathcal{B}$ is the set of DoAs with $|\theta| \leq 60°$; the array is alias-free on $\mathcal{B}$.

6.5.3 3D Array

A bi-tetrahedron of five sensors [219, 249, 250] is shown in Fig. 6.2a with positions $[0, 0, \pm 1]^{\mathsf{T}}$, $[2, 0, 0,]^{\mathsf{T}}$, and $[-1, \pm 1, 0,]^{\mathsf{T}}$ in units of one-half wavelength. The positions correspond to the vertices of an irregular polyhedron and are chosen so that there exists a cubic fundamental parallelotope circumscribing a sphere of diameter 2. The minimum distance between sensors is $\sqrt{3}\lambda/2$. This is 73 % too large to satisfy the commonly used sufficient condition [33, Thm. 8], which requires that the minimum distance between sensors not exceed $\lambda/2$. The tangent points on each face correspond to the three pairs $(\mathrm{az}, \mathrm{el})$ of ambiguous DoAs, $(0, \pm 90°)$, $(\pm 90°, 0)$, and $(90° \pm 90°, 0)$; any small contraction of the array geometry removes these points of tangency. The



(a) (b)

Figure 6.2: For a bi-tetrahedral (a) array of five sensors, (b) illustration of a fundamental parallelotope and the sphere of all DoAs, $\mathcal{B}_3$, translated to the center of the parallelotope. Spheres translated to lattice points are merely tangent; the array is alias-free other than three pairs of ambiguous DoAs.

array is not isotropic: the CRB for x_1 is three times smaller than for x_2 and x_3. The unambiguous set for 3D arrays is underestimated in some prior literature, e.g., [9,219].

6.6 Summary

In this chapter, the results in [33] have been extended to provide a simply tested necessary and sufficient condition characterizing array geometries that provide unambiguous steering vectors. First, Proposition 17 considers an arbitrary set, $\mathcal{B}$, of candidate DoAs, with application illustrated in Fig. 6.1c. In contrast, [33] only considers the case in which $\mathcal{B}$ corresponds to all possible DoAs. Second, the test in (6.12) provides a simple necessary and sufficient condition for alias-free array geometries in

2D and 3D, offering a less restrictive alternative to a widely used sufficient condition [33, Thms. 8, 9]. The test in (6.12) is easily applied, yet it offers increased flexibility for array geometry design over the sufficient condition in [33]. Third, the direct test for 1D arrays in (6.13) is simpler than [33, Thm. 11] and therefore advantageous in a design procedure. The new results facilitate the design of nonuniform arrays for widely separated sensors or increased aperture without an increase in the number of sensors. The unambiguous set of spatial frequencies was shown to be a parallelotope, admitting both intuitive geometric interpretation and application of the test for any subset of DoAs.

6.7 Future Work

The properties given in this chapter may allow for efficient 2D and 3D sensor array design and analysis techniques similar to those offered for 1D arrays in [31, 178]. In particular, Lemma 13 allows for an array design process that begins with the specification of a desired lattice basis $\boldsymbol{V}$. Next, a set of candidate primitive matrices can be formed. For each $\boldsymbol{Z}$ in the set, the encoding matrix $\boldsymbol{A}$ is given by $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{V}^{-1}$, and from $\boldsymbol{A}$, the sensor locations can be determined. The result is a set of sensor configurations that can be analyzed with respect to the CRB and the probability of an outlier [223]. This approach may lead to an efficient design process, in which only the sensor configurations with a desirable unambiguous region will be considered. Corollary 16 can also be used to scale candidate array configurations.

Note that the signal model in (6.2) uses the traditional assumption of isotropic sensor elements. In [251], ambiguity analysis with arbitrary complex-valued gain

functions per sensor was considered. Future research might address these alternative scenarios.

Appendix A: Variational Inference Details for Chapter 3

The gamma distribution's probability density is denoted by

$$\mathcal{G}(v;\alpha,\beta) = \frac{\beta^{\alpha} v^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta v\}, \tag{A.1}$$

with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$. The inverse gamma distribution's probability density is denoted by

$$\mathcal{IG}(v;\alpha,\beta) = \frac{\beta^{\alpha} v^{-(\alpha+1)}}{\Gamma(\alpha)} \exp\left\{-\beta v^{-1}\right\} \tag{A.2}$$

with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. From (3.7) and (3.14), the variational factor associated with $r_i^{(1)}$ is updated as

$$\begin{aligned}
\log q(r_i^{(1)}) =& \log p(y_i^{(1)}|r_i^{(1)}) + \langle \log p(r_i^{(1)} \mid \boldsymbol{z}^{(1)}) \rangle_{q(\boldsymbol{z}^{(1)})} \\
&+ \text{const.} \\
=& -\log r_i^{(1)} - \frac{|y_i^{(1)}|^2}{r_i^{(1)}} - (1 + a|S_i^r|) \log r_i^{(1)} \\
&- \sum_{j \in S_i^r} \frac{a\langle z_j^{(1)} \rangle}{r_i^{(1)}} + \text{const.},
\end{aligned} \tag{A.3}$$

giving $q(r_i^{(1)}) = \mathcal{IG}(r_i^{(1)}; \alpha_{r,i}^{(1)}, \beta_{r,i}^{(1)})$, with

$$\alpha_{r,i}^{(1)} = 1 + a|S_i^r| \tag{A.4a}$$

$$\beta_{r,i}^{(1)} = |y_i^{(1)}|^2 + \sum_{j \in S_i^r} a\langle z_j^{(1)} \rangle. \tag{A.4b}$$

The factor associated with $r_i^{(2)}$ is updated as

$$\begin{aligned}\log q(r_i^{(2)}) =& \big\langle \log p(y_i^{(2)} \mid y_i^{(1)}, r_i^{(2)}, c_i) \big\rangle_{q(\boldsymbol{c})} \\ &+ \langle \log p(r_i^{(2)} \mid \boldsymbol{z}^{(2)}) \rangle_{q(\boldsymbol{z}^{(2)})} + \text{const.} \qquad \text{(A.5)} \\ =& -\langle \delta(c_i+1) \rangle \left[\frac{|y_i^{(2)} - \gamma e^{-j\phi_i} y_i^{(1)}|^2}{r_i^{(2)}(1-\gamma^2)} \right] \\ &- \langle \delta(c_i-1) \rangle \left[\frac{|y_i^{(2)}|^2}{r_i^{(2)}} \right] - \big(1 + a|S_i^r|\big) \log r_i^{(2)} \\ &- \log r_i^{(2)} - \sum_{j \in S_i^r} \frac{a \langle z_j^{(2)} \rangle}{r_i^{(2)}} + \text{const.}, \qquad \text{(A.6)}\end{aligned}$$

giving $q(r_i^{(2)}) = \mathcal{IG}(r_i^{(2)}; \alpha_{r,i}^{(2)}, \beta_{r,i}^{(2)})$, with

$$\begin{aligned}\alpha_{r,i}^{(2)} &= 1 + a|S_i^r| \qquad \text{(A.7a)} \\ \beta_{r,i}^{(2)} &= \frac{\langle \delta(c_i+1) \rangle |y_i^{(2)} - \gamma e^{-j\phi_i} y_i^{(1)}|^2}{(1-\gamma^2)} \\ &\quad + \langle \delta(c_i-1) \rangle |y_i^{(2)}|^2 + \sum_{j \in S_i^r} a \langle z_j^{(2)} \rangle. \qquad \text{(A.7b)}\end{aligned}$$

The time-k auxiliary variable factors for $k \in \{1, 2\}$ are updated according to

$$\begin{aligned}\log q(z_j^{(k)}) =& \langle \log p(z_j^{(k)} \mid \boldsymbol{r}^{(k)}) \rangle_{q(\boldsymbol{r}^{(k)})} + \text{const.} \\ =& (-1 + a|S_j^z|) \log z_j^{(k)} \\ &- \sum_{i \in S_j^z} \left\langle \frac{a}{r_i^{(k)}} \right\rangle z_j^{(k)} + \text{const.} \qquad \text{(A.8)}\end{aligned}$$

This implies that $q(z_j^{(k)}) = \mathcal{G}(z_j^{(k)}; \alpha_{z,j}^{(k)}, \beta_{z,j}^{(k)})$, with

$$\begin{aligned}\alpha_{z,j}^{(k)} &= a|S_j^z| \qquad \text{(A.9a)} \\ \beta_{z,j}^{(k)} &= \sum_{i \in S_j^z} \left\langle \frac{a}{r_i^{(k)}} \right\rangle. \qquad \text{(A.9b)}\end{aligned}$$

Appendix B: Proofs for Chapter 5

B.1 Proofs for Section 5.4

This appendix has statements that are relevant to the ML solution method in Section 5.4, in which Condition 7 is assumed to be met.

Proposition 20. *For any* $\boldsymbol{r} \in \mathcal{H}$ *and* $\boldsymbol{\lambda} \in \mathbb{Z}^n$, $\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\lambda} \in \mathcal{H}$.

Proof. From (5.35), $\omega_j \boldsymbol{a}_j \boldsymbol{r} - \langle \omega_j y_j + \frac{1}{2} \rangle_1 \in \mathbb{Z}$, for some $j \in \{1, \ldots, m\}$. Then for any $\boldsymbol{\lambda} \in \mathbb{Z}^n$,

$$\omega_j \boldsymbol{a}_j (\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\lambda}) - \langle \omega_j y_j + \tfrac{1}{2} \rangle_1 \in \mathbb{Z}, \tag{B.1}$$

as $\omega_j \boldsymbol{a}_j \boldsymbol{C}\boldsymbol{\lambda} \in \mathbb{Z}$. Therefore, $\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\lambda} \in \mathcal{H}$. □

Proposition 21. *If* $\boldsymbol{r} \in \mathcal{H}$, *then* $\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} \in \widetilde{\mathcal{H}}$.

Proof. From (5.7), $\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} = \boldsymbol{r} - \boldsymbol{C} \lfloor \boldsymbol{C}^{-1} \boldsymbol{r} \rfloor$. Thus from Proposition 20, $\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} \in \mathcal{H}$, and since $\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} \in \mathcal{P}_{\boldsymbol{C}}$, $\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} \in \widetilde{\mathcal{H}}$. □

Proposition 22. *If* $\widetilde{\boldsymbol{r}} \in \widetilde{\Psi}(\boldsymbol{\ell})$, *then* $\widetilde{\boldsymbol{r}} \in \widetilde{\Psi}(\boldsymbol{\ell} + \boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda}))$, *for any* $\boldsymbol{\lambda} \in \mathbb{Z}^n$.

Proof. From (5.37), if $\widetilde{\boldsymbol{r}} \in \widetilde{\Psi}(\boldsymbol{\ell})$, then there is $\boldsymbol{r} \in \Psi(\boldsymbol{\ell})$ such that $\langle \boldsymbol{r} \rangle_{\boldsymbol{C}} = \widetilde{\boldsymbol{r}}$. From (5.36), $\lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r}) \rfloor = \boldsymbol{\ell}$, and therefore, for any $\boldsymbol{\lambda} \in \mathbb{Z}^n$,

$$\begin{aligned} \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{r} - \boldsymbol{C}\boldsymbol{\lambda} + \boldsymbol{C}\boldsymbol{\lambda})) \rfloor &= \boldsymbol{\ell} \\ \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{r} - \boldsymbol{C}\boldsymbol{\lambda})) \rfloor &= \boldsymbol{\ell} + \boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda}) \end{aligned} \tag{B.2}$$

Thus, $\boldsymbol{r} - \boldsymbol{C}\boldsymbol{\lambda} \in \Psi(\boldsymbol{\ell} + \boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda}))$. Since $\langle \boldsymbol{r} - \boldsymbol{C}\boldsymbol{\lambda} \rangle_{\boldsymbol{C}} = \widetilde{\boldsymbol{r}}$, $\widetilde{\boldsymbol{r}} \in \widetilde{\Psi}(\boldsymbol{\ell} + \boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda}))$, according to (5.37). □

Proposition 23. *For any* $\boldsymbol{\lambda} \in \mathbb{Z}^n$, $\mathcal{L}_r(\boldsymbol{r}; \boldsymbol{y}) = \mathcal{L}_r(\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\lambda}; \boldsymbol{y})$

Proof.

$$\begin{aligned} \mathcal{L}_r(\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\lambda}; \boldsymbol{y}) =& \left\| \boldsymbol{\Sigma}^{-1/2} \, \mathrm{d}_{\mathbf{1}_m \oslash \boldsymbol{\omega}} \left(\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{r} + \boldsymbol{C}\boldsymbol{\lambda}) \right) \right\|^2 \\ =& \Big\| \boldsymbol{\Sigma}^{-1/2} \big(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r} - \boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda} \\ & - \lceil \boldsymbol{\omega} \circ (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{r} - \boldsymbol{A}\boldsymbol{C}\boldsymbol{\lambda}) \rfloor \oslash \boldsymbol{\omega} \big) \Big\|^2 \\ =& \mathcal{L}_r(\boldsymbol{r}; \boldsymbol{y}) \end{aligned}$$

□

Lemma 24. *At no element of* $\mathcal{H}$ *does* $\mathcal{L}_r$ *have a local minimum.*

Proof. From (5.31),

$$\mathcal{L}_r(\boldsymbol{r}; \boldsymbol{y}) = \sum_{i=1}^{m} \frac{1}{\sigma_{ii}^2} \, \mathrm{d}^2_{\omega_i^{-1}} \left(y_i - \boldsymbol{a}_i \boldsymbol{r} \right).$$

For any $\boldsymbol{r}_0 \in \mathcal{H}$, $\mathrm{d}_{\omega_j^{-1}}(y_j - \boldsymbol{a}_j \boldsymbol{r}_0) = -\frac{1}{2}\omega_j^{-1}$, for some $j \in \{1, \ldots, m\}$, and therefore,

$$y_j = \boldsymbol{a}_j \boldsymbol{r}_0 + \omega_j^{-1} \lceil \omega_j (y_j - \boldsymbol{a}_j \boldsymbol{r}_0) \rfloor - \frac{1}{2} \omega_j^{-1}. \tag{B.3}$$

Let $\boldsymbol{u} \in \mathbb{R}^n$ be an arbitrary unit vector. The right directional derivative of $\mathcal{L}_r$ with respect to $\boldsymbol{u}$ is

$$\begin{aligned}
\partial_{\boldsymbol{u}^+}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) &= \lim_{h\to 0^+} \frac{\mathcal{L}_r(\boldsymbol{r}_0 + h\boldsymbol{u}) - \mathcal{L}_r(\boldsymbol{r}_0)}{h} \\
&= \lim_{h\to 0^+} \sum_{i=1}^{m} \frac{\sigma_{ii}^{-2}}{h}\big(\mathrm{d}^2_{\omega_i^{-1}}(y_i - \boldsymbol{a}_i(\boldsymbol{r}_0 + h\boldsymbol{u})) \\
&\quad - \mathrm{d}^2_{\omega_i^{-1}}(y_i - \boldsymbol{a}_i\boldsymbol{r}_0)\big) \\
&= \lim_{h\to 0^+} \frac{\sigma_{jj}^{-2}}{h}\big(\mathrm{d}^2_{\omega_j^{-1}}(y_j - \boldsymbol{a}_j(\boldsymbol{r}_0 + h\boldsymbol{u})) \\
&\quad - \mathrm{d}^2_{\omega_j^{-1}}(y_j - \boldsymbol{a}_j\boldsymbol{r}_0)\big) \\
&= \lim_{h\to 0^+} \frac{\sigma_{jj}^{-2}}{h}\Bigg(\big(\omega_j^{-1}\lceil \omega_j(y_j - \boldsymbol{a}_j\boldsymbol{r}_0)\rfloor \\
&\quad - \omega_j^{-1}\lceil \omega_j(y_j - \boldsymbol{a}_j(\boldsymbol{r}_0 + h\boldsymbol{u}))\rfloor \\
&\quad - \frac{1}{2}\omega_j^{-1} - h\boldsymbol{a}_j\boldsymbol{u}\big)^2 - \frac{1}{4}\omega_j^{-2}\Bigg).
\end{aligned} \tag{B.4}$$

If $\boldsymbol{a}_j\boldsymbol{u} > 0$, we have

$$\lceil \omega_j(y_j - \boldsymbol{a}_j(\boldsymbol{r}_0 + h\boldsymbol{u}))\rfloor - \lceil \omega_j(y_j - \boldsymbol{a}_j\boldsymbol{r}_0)\rfloor = -1, \tag{B.5}$$

in which we have used that $h > 0$. Using (B.5) in (B.4), we get that

$$\partial_{\boldsymbol{u}^+}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) = -(\sigma_{jj}^2\omega_j)^{-1}\boldsymbol{a}_j\boldsymbol{u}, \tag{B.6}$$

and thus $\partial_{\boldsymbol{u}^+}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) < 0$, when $\boldsymbol{a}_j\boldsymbol{u} > 0$. (Note that $\sigma_{jj}^2\omega_j > 0$.) Alternatively, if $\boldsymbol{a}_j\boldsymbol{u} \leq 0$,

$$\lceil \omega_j(y_j - \boldsymbol{a}_j(\boldsymbol{r}_0 + h\boldsymbol{u}))\rfloor - \lceil \omega_j(y_j - \boldsymbol{a}_j(\boldsymbol{r}_0))\rfloor = 0, \tag{B.7}$$

and from (B.4), we have

$$\partial_{\boldsymbol{u}^+}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) = (\sigma_{jj}^2\omega_j)^{-1}\boldsymbol{a}_j\boldsymbol{u}. \tag{B.8}$$

With $\boldsymbol{a}_j\boldsymbol{u} < 0$, we also have that $\partial_{\boldsymbol{u}^+}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) < 0$. Thus, $\partial_{\boldsymbol{u}^+}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) < 0$. Following a similar procedure for the left derivative, we find that $\partial_{\boldsymbol{u}^-}\mathcal{L}_r(\boldsymbol{r}_0;\boldsymbol{y}) < 0$. Thus, $\mathcal{L}_r$

does not have a local minimum at $\boldsymbol{r}_0$, and we conclude that $\mathcal{L}_r$ has a local minimum at no point in $\mathcal{H}$. □

Lemma 25. *If $\boldsymbol{\omega}$ has pairwise coprime entries, then Condition 7 is satisfied.*

Proof. From (5.22), $\omega_i = \operatorname{lcm}([\beta_{i,1}, \ldots, \beta_{i,n}]^\mathsf{T})$, for each $i = 1, \ldots, m$. So, there is $h_{i,j} \in \mathbb{Z}$, for each $j \in \{1, \ldots, n\}$, such that

$$h_{i,j} = \begin{cases} \frac{\omega_i}{\beta_{i,j}}, & \beta_{i,j} \neq 0 \\ 0, & \text{else,} \end{cases}$$

with $\gcd([h_{i,1}, \ldots, h_{i,n}]^\mathsf{T}) = 1$. Thus, from (5.21),

$$\gcd([\omega_i \boldsymbol{a}_i \boldsymbol{C}, \omega_i]^\mathsf{T}) = \gcd([h_{i,1}, \ldots, h_{i,n}, \omega_i]^\mathsf{T}) = 1.$$

According to Theorem 4, there are integer solutions to

$$\boldsymbol{\omega} \circ (\boldsymbol{A}\boldsymbol{C})\boldsymbol{\lambda} \equiv \boldsymbol{q} \mod \boldsymbol{\omega}, \tag{B.9}$$

for any $\boldsymbol{q} \in \mathbb{Z}^m$. Hence, Condition 7 is satisfied. □

Lemma 25 shows that $\boldsymbol{\omega}$ having pairwise coprime entries is sufficient to satisfy Condition 7, but it is not necessary. As an example, consider the encoding matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix}^\mathsf{T}, \tag{B.10}$$

which has $\boldsymbol{\omega} = [1, 2, 4]^\mathsf{T}$The entries of $\boldsymbol{\omega}$ are not pairwise coprime, yet $\boldsymbol{A}$ satisfies Condition 7.

Proposition 26. *With $n = 1$, Condition 7 is satisfied iff $\boldsymbol{\omega}$ has pairwise coprime entries.*

Proof. ($\Leftarrow$): Given by Lemma 25.

($\Rightarrow$): Let $\widetilde{\boldsymbol{W}} = [\boldsymbol{\omega} \circ (\boldsymbol{AC}), \mathrm{diag}(\boldsymbol{\omega})]^{\mathsf{T}} \in \mathbb{Z}^{(m+n)\times m}$, and let $\boldsymbol{H} = [\widetilde{\boldsymbol{H}}^{\mathsf{T}}, \boldsymbol{0}_{m,n}]^{\mathsf{T}}$ be the HNF of $\widetilde{\boldsymbol{W}}$. From Proposition 8, Condition 7 is equivalent to $\widetilde{\boldsymbol{H}} = \boldsymbol{I}_m$. As $\widetilde{\boldsymbol{W}} = \boldsymbol{U}^{-1}\boldsymbol{H}$, Condition 7 is equivalent to the possibility of extending $\widetilde{\boldsymbol{W}}$ to an $(m+n)\times(m+n)$ unimodular matrix. According to [186, Lemma 2], this is possible iff the full-rank minors of $\widetilde{\boldsymbol{W}}$ are setwise coprime. For $n = 1$, we have that $\omega_j a_j C = \pm 1$, for $j \in \{1, 2, \ldots, m\}$. Therefore, each element of the first row of $\widetilde{\boldsymbol{W}}$ is ± 1, and the absolute values of the full-rank minors of $\widetilde{\boldsymbol{W}}$ are given by

$$\mathcal{M} = \Big\{ \prod_{i=1}^{m} \omega_i \Big\} \cup \Big\{ \prod_{i\neq j} \omega_i \Big\}_{j=1}^{m}. \tag{B.11}$$

For example, if $\boldsymbol{A} = [\,1/2, 1/3, 1/5\,]^{\mathsf{T}}$, then $C = 1$ and $\boldsymbol{\omega} = [2, 3, 5]^{\mathsf{T}}$. Thus,

$$\widetilde{\boldsymbol{W}} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \tag{B.12}$$

and $\mathcal{M} = \{30, 15, 10, 6\}$.

Suppose that $\mathcal{M}$ is setwise coprime and that $\boldsymbol{\omega}$ does not have pairwise coprime entries. Then $\omega_k = \alpha\omega_\ell$ for some $k \neq \ell$ and $\alpha \in \mathbb{Z}$. Thus, the factor ω_ℓ is present in each element in $\mathcal{M}$. This is a contradiction, as we have assumed that $\mathcal{M}$ is setwise coprime. Thus, if the full-rank minors of $\widetilde{\boldsymbol{W}}$ are setwise coprime, then $\boldsymbol{\omega}$ has pairwise coprime entries. □

Theorem 27. *For encoding matrix $\boldsymbol{A}$ satisfying Condition 7, the function $f : \{\langle\mathbb{Z}^n\rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}} \times \mathbb{Z}^m\} \mapsto \mathbb{Z}^m$ defined by $f(\boldsymbol{\gamma}, \boldsymbol{k}) = \boldsymbol{\omega} \circ (\boldsymbol{AC\gamma} + \boldsymbol{k})$ is a bijection.*

Proof. Onto: As $\boldsymbol{A}$ meets Condition 7, for any $\boldsymbol{q} \in \mathbb{Z}^m$, there is $\boldsymbol{\lambda} \in \mathbb{Z}^n$ such that

$$\boldsymbol{\omega} \circ (\boldsymbol{AC})\boldsymbol{\lambda} = \boldsymbol{q} \mod \boldsymbol{\omega}. \tag{B.13}$$

Let $\boldsymbol{\gamma} = \langle \boldsymbol{\lambda} \rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}$. It follows from (B.13) and (5.7) that

$$\begin{aligned} &\boldsymbol{\omega} \circ (\boldsymbol{AC})(\boldsymbol{\gamma} + \boldsymbol{C}^{-1}\boldsymbol{V}\lfloor \boldsymbol{V}^{-1}\boldsymbol{C}\boldsymbol{\lambda} \rfloor) \equiv \boldsymbol{q} \mod \boldsymbol{\omega} \\ \Leftrightarrow\ &\boldsymbol{\omega} \circ (\boldsymbol{AC}\boldsymbol{\gamma}) \equiv \boldsymbol{q} \mod \boldsymbol{\omega}, \end{aligned}$$

where we have used that $\boldsymbol{AV} \in \mathbb{Z}^{m \times n}$. Therefore, $f(\boldsymbol{\gamma}, \boldsymbol{k}) = \boldsymbol{\omega} \circ (\boldsymbol{AC}\boldsymbol{\gamma} + \boldsymbol{k}) = \boldsymbol{q}$ for some $\boldsymbol{k} \in \mathbb{Z}^m$.

One-to-one: For any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \langle \mathbb{Z}^n \rangle_{\boldsymbol{C}^{-1}\boldsymbol{V}}$ and $\boldsymbol{k}_1, \boldsymbol{k}_2 \in \mathbb{Z}^m$, assume that $f(\boldsymbol{\gamma}_1, \boldsymbol{k}_1) = f(\boldsymbol{\gamma}_2, \boldsymbol{k}_2)$. Then,

$$\begin{aligned} &\boldsymbol{\omega} \circ (\boldsymbol{AC}\boldsymbol{\gamma}_1 + \boldsymbol{k}_1) = \boldsymbol{\omega} \circ (\boldsymbol{AC}\boldsymbol{\gamma}_2 + \boldsymbol{k}_2) \\ \Rightarrow\ &\boldsymbol{AC}(\boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_1)) \in \mathbb{Z}^m \Leftrightarrow \boldsymbol{C}(\boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_1) \in \Lambda_{\boldsymbol{V}} \\ \Leftrightarrow\ &\boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_1 \in \boldsymbol{C}^{-1}\boldsymbol{V}\mathbb{Z}^n \Leftrightarrow \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 \Leftrightarrow \boldsymbol{k}_1 = \boldsymbol{k}_2. \end{aligned} \tag{B.14}$$

□

Proposition 28. *If Condition 7 is satisfied for an encoding matrix $\boldsymbol{A}$, then it is satisfied for the encoding matrix $\widetilde{\boldsymbol{A}} = \boldsymbol{AS}$, where $\boldsymbol{S}$ is any n-by-n invertible matrix.*

Proof. If $\boldsymbol{V}$ is the basis for $\{\boldsymbol{x} \mid \boldsymbol{Ax} \in \mathbb{Z}^m\}$, then according to Corollary 16, $\boldsymbol{S}^{-1}\boldsymbol{V}$ is the basis for $\{\boldsymbol{x} \mid \widetilde{\boldsymbol{A}}\boldsymbol{x} \in \mathbb{Z}^m\}$. If $\boldsymbol{C}$ is a basis for a super-lattice of $\Lambda_{\boldsymbol{V}}$ (as defined in Section 5.4.1), then $\widetilde{\boldsymbol{C}} = \boldsymbol{S}^{-1}\boldsymbol{C}$ is a basis for a super-lattice of $\Lambda_{\widetilde{\boldsymbol{V}}}$. Continuing with this notation, we get that $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and that $\widetilde{\boldsymbol{\omega}} = \boldsymbol{\omega}$. Thus, for any $\boldsymbol{b} \in \mathbb{Z}^m$,

$$\widetilde{\boldsymbol{\omega}} \circ (\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{C}})\boldsymbol{\gamma} \equiv \boldsymbol{b} \mod \widetilde{\boldsymbol{\omega}} \Leftrightarrow \boldsymbol{\omega} \circ (\boldsymbol{AC})\boldsymbol{\gamma} \equiv \boldsymbol{b} \mod \boldsymbol{\omega}.$$

□

B.2 Correspondence Between Hyperplane Intersections and Chambers

Let $\mathcal{U}$ denote the set of points at which n hyperplanes in $\mathcal{H}$ intersect, and let $\widetilde{\mathcal{U}} = \mathcal{P}_{\boldsymbol{C}} \cap \mathcal{U} = \langle \mathcal{U} \rangle_{\boldsymbol{C}}$. Recall that $\mathcal{H}$, defined in (5.35), comprises hyperplanes (subspaces of codimension one) in $\mathbb{R}^n$ defined by the normal vectors $\{\boldsymbol{a}_j^{\mathsf{T}}\}_{j \in \{1,\ldots,m\}}$, where $\boldsymbol{a}_j \in \mathbb{Q}^{1 \times n}$ is the jth row of $\boldsymbol{A}$. Because $\boldsymbol{A}$ is assumed to have full column rank, we have that $\mathcal{U}$ is nonempty. Further, we assume that there is no intersection of more than n hyperplanes in $\mathcal{H}$, which is true almost surely.

Let $\operatorname{cl} \Psi(\boldsymbol{\ell})$ denote the polytope given by the closure of $\Psi(\boldsymbol{\ell})$. Define a directed graph G, in which each node is a point in $\mathcal{U}$. Let any pair of nodes in G be connected by an edge if both points occupy the same edge in $\operatorname{cl} \Psi(\boldsymbol{\ell})$, for any $\boldsymbol{\ell} \in \Phi$. The direction of the edges will be determined by a linear objective function on the vertices.

Let $G(\Psi(\boldsymbol{\ell}))$ be the sub-graph of G having only the nodes that are vertices of $\operatorname{cl} \Psi(\boldsymbol{\ell})$. The graph $G(\Psi(\boldsymbol{\ell}))$ is thus a *polytopal digraph*. Given a polytope and a linear objective function on the vertices of the polytope, a polytopal digraph is a directed graph with nodes and edges corresponding to the polytope vertices and extremal edges, with each edge directed toward the vertex with the higher objective function value [252, 253]. The linear objective function we use is $p(\boldsymbol{v}) = \boldsymbol{d}^{\mathsf{T}} \boldsymbol{v}$, where $\boldsymbol{d} \in \mathbb{R}^n$ is an arbitrary unit vector such that $\boldsymbol{d}^{\mathsf{T}} \boldsymbol{a}_j \neq 0$, for all $j \in \{1, \ldots, m\}$. This ensures that $\boldsymbol{d}$ is orthogonal to no hyperplane in $\mathcal{H}$. Thus, the edge between any pair of connected nodes $\boldsymbol{v}$ and $\boldsymbol{\eta}$ is directed towards $\boldsymbol{v}$ if $\boldsymbol{d}^{\mathsf{T}} \boldsymbol{v} > \boldsymbol{d}^{\mathsf{T}} \boldsymbol{\eta}$, and towards $\boldsymbol{\eta}$ otherwise. Also, let $G(\boldsymbol{v})$ denote a sub-graph of G containing only the node $\boldsymbol{v}$ and the nodes connected to $\boldsymbol{v}$.

Proposition 29. *Let $\boldsymbol{v} \in \mathcal{U}$. For any node $\boldsymbol{\eta}$ connected to $\boldsymbol{v}$, $\boldsymbol{d}^{\mathsf{T}} \boldsymbol{v} \neq \boldsymbol{d}^{\mathsf{T}} \boldsymbol{\eta}$.*

Proof. Both $\boldsymbol{v}$ and $\boldsymbol{\eta}$ are in one of the hyperplanes that constitute $\mathcal{H}$. So, $\boldsymbol{v} - \boldsymbol{\eta}$ is also in that hyperplane. We assume that $\boldsymbol{d}$ is orthogonal to no hyperplane in $\mathcal{H}$, and thus, $\boldsymbol{d}^\mathsf{T}(\boldsymbol{v} - \boldsymbol{\eta}) \neq 0$. □

Lemma 30. *For any $\boldsymbol{v} \in \mathcal{U}$, the remaining nodes in $G(\boldsymbol{v})$ can be partitioned into pairs such that, within each pair, the edge between $\boldsymbol{v}$ and one neighboring node directs towards $\boldsymbol{v}$, and the edge between $\boldsymbol{v}$ and the other neighboring node directs away from $\boldsymbol{v}$.*

Proof. Any vertex $\boldsymbol{v} \in \mathcal{U}$ is an intersection of $\boldsymbol{n}$ hyperplanes, and there are $2n$ nodes in G that are connected to $\boldsymbol{v}$. The points in $\mathcal{U}$ appear along lines formed by the intersection of $n - 1$ hyperplanes. Hence, for any point connected to $\boldsymbol{v}$ in G, there will be a corresponding point (also connected to $\boldsymbol{v}$) in the opposing direction. Each pair of opposing points is collinear with $\boldsymbol{v}$, and $\boldsymbol{v}$ is located between the pair. Let $\{\eta_1, \eta_2\}$ be any such pair. We have that $\boldsymbol{\eta}_2 = \boldsymbol{v} - \alpha(\boldsymbol{\eta}_1 - \boldsymbol{v})$ for some $\alpha > 0$. (This expresses that $\boldsymbol{\eta}_1$ is in the opposite direction from $\boldsymbol{v}$ as $\boldsymbol{\eta}_2$.) The edge between $\boldsymbol{v}$ and $\boldsymbol{\eta}_1$ directs to $\boldsymbol{\eta}_1$ iff

$$\begin{aligned} \boldsymbol{d}^\mathsf{T}(\boldsymbol{\eta}_1 - \boldsymbol{v}) > 0 &\Leftrightarrow \boldsymbol{d}^\mathsf{T}(-\alpha(\boldsymbol{\eta}_1 - \boldsymbol{v})) < 0 \\ &\Leftrightarrow \boldsymbol{d}^\mathsf{T}(\boldsymbol{\eta}_2 - \boldsymbol{v}) < 0, \end{aligned}$$

which is true iff the edge between $\boldsymbol{v}$ and $\boldsymbol{\eta}_2$ directs to $\boldsymbol{v}$. Hence, one of $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2\}$ has an edge directed towards $\boldsymbol{v}$, and one has an edge directed away from $\boldsymbol{v}$. □

Lemma 31. *For any $\boldsymbol{v} \in \mathcal{U}$, there is only one chamber, denoted $\Psi(\boldsymbol{\ell})$ for some $\boldsymbol{\ell} \in \Phi$, for which $\boldsymbol{v}$ is a sink (source) in $G(\Psi(\boldsymbol{\ell}))$. (Note that $\boldsymbol{v} \in \operatorname{cl} \Psi(\boldsymbol{\ell})$.)*

Proof. Any of the 2^n chambers that have $\boldsymbol{v}$ in its closure can be uniquely identified with a set consisting of one node from each of the n pairs described by Lemma 30.

One of each such pair has an edge that directs towards (away from) $\boldsymbol{v}$. Therefore, for only one of the 2^n chambers are all edges between $\boldsymbol{v}$ and vertices in the chamber's closure directed towards (away from) $\boldsymbol{v}$. Thus, for this chamber, denoted $\Psi(\boldsymbol{\ell})$, $\boldsymbol{v}$ is a sink (source) in $G(\Psi(\boldsymbol{\ell}))$. □

Theorem 32 ([253, Theorem 19.4.1])**.** *A polytopal digraph has one source and one sink.*

Theorem 33. *There is a bijection between* $\mathcal{U}$ *and* $\{\Psi(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\Phi}$.

Proof. Let $\pi : \mathcal{U} \mapsto \{\Psi(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\Phi}$ be a function that maps $\boldsymbol{v} \in \mathcal{U}$ to the (unique, from Lemma 31) chamber $\Psi(\boldsymbol{\ell})$ for which $\boldsymbol{v}$ is a source in $G(\Psi(\boldsymbol{\ell}))$.

Onto: According to Theorem 32, one vertex of $\operatorname{cl}\Psi(\boldsymbol{\ell})$ is a source in $G(\Psi(\boldsymbol{\ell}))$. For any $\boldsymbol{\ell} \in \Phi$, the source in $G(\Psi(\boldsymbol{\ell}))$ is given by $\pi(\Psi(\boldsymbol{\ell}))$. Therefore, π is onto.

One-to-one: Assume that $\pi(\boldsymbol{v}) = \pi(\boldsymbol{\eta}) = \Psi(\boldsymbol{\ell})$, for some $\boldsymbol{\ell} \in \Phi$. Thus, both $\boldsymbol{v}$ and $\boldsymbol{\eta}$ are sources in $G(\Psi(\boldsymbol{\ell}))$. Because $G(\Psi(\boldsymbol{\ell}))$ is a polytopal digraph with a unique source, $\boldsymbol{v} = \boldsymbol{\eta}$. Therefore, π is one-to-one.

□

Lemma 34. *For any* $\boldsymbol{v} \in \mathcal{U}$ *and any* $\boldsymbol{\lambda} \in \mathbb{Z}^n$, $\pi(\boldsymbol{v}+\boldsymbol{C\lambda}) = \pi(\boldsymbol{v})+\{\boldsymbol{C\lambda}\}$, *a Minkowski sum.*

Proof. The chamber $\pi(\boldsymbol{v})$ is formed by hyperplanes in $\mathcal{H}$. Due to the periodicity in $\mathcal{H}$, $\pi(\boldsymbol{v} + \boldsymbol{C\lambda})$ is a translation of $\pi(\boldsymbol{v})$ by $\boldsymbol{C\lambda}$. □

Theorem 35. *There is a bijection between* $\widetilde{\mathcal{U}}$ *and* $\{\widetilde{\Psi}(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\Phi}$.

Proof. Define a mapping $\widetilde{\pi} : \widetilde{\mathcal{U}} \mapsto \{\widetilde{\Psi}(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\Phi}$, which is given by $\widetilde{\pi}(\widetilde{\boldsymbol{v}}) = \langle\pi(\widetilde{\boldsymbol{v}})\rangle_{\boldsymbol{C}}$.

Onto: According to Theorem 33: for any $\boldsymbol{\ell} \in \Phi$ there is $\boldsymbol{v} \in \mathcal{U}$ such that $\pi(\boldsymbol{v}) = \Psi(\boldsymbol{\ell})$. Then $\pi(\langle \boldsymbol{v} \rangle_{\boldsymbol{C}} + \boldsymbol{C}\lfloor \boldsymbol{C}^{-1}\boldsymbol{v} \rfloor) = \Psi(\boldsymbol{\ell})$, as $\boldsymbol{v} = \langle \boldsymbol{v} \rangle_{\boldsymbol{C}} + \boldsymbol{C}\lfloor \boldsymbol{C}^{-1}\boldsymbol{v} \rfloor$. From Lemma 34, we have that $\pi(\langle \boldsymbol{v} \rangle_{\boldsymbol{C}}) + \{\boldsymbol{C}\lfloor \boldsymbol{C}^{-1}\boldsymbol{v} \rfloor\} = \Psi(\boldsymbol{\ell})$. Thus,

$$\begin{aligned} \langle \pi(\langle \boldsymbol{v} \rangle_{\boldsymbol{C}}) \rangle_{\boldsymbol{C}} &= \langle \Psi(\boldsymbol{\ell}) \rangle_{\boldsymbol{C}} \\ \widetilde{\pi}(\langle \boldsymbol{v} \rangle_{\boldsymbol{C}}) &= \widetilde{\Psi}(\boldsymbol{\ell}). \end{aligned} \tag{B.15}$$

Therefore, $\widetilde{\pi}$ is onto.

One-to-one: Assume, for any $\boldsymbol{v}, \boldsymbol{\eta} \in \mathcal{U}$ that $\widetilde{\pi}(\boldsymbol{v}) = \widetilde{\pi}(\boldsymbol{\eta})$, and therefore that $\langle \pi(\boldsymbol{v}) \rangle_{\boldsymbol{C}} = \langle \pi(\boldsymbol{\eta}) \rangle_{\boldsymbol{C}}$. Then, $\pi(\boldsymbol{v}) = \pi(\boldsymbol{\eta}) + \boldsymbol{C}\boldsymbol{\lambda}$, for some $\boldsymbol{\lambda} \in \mathbb{Z}^n$. From Lemma 34, $\pi(\boldsymbol{v}) = \pi(\boldsymbol{\eta} + \boldsymbol{C}\boldsymbol{\lambda})$, and from Theorem 33, $\boldsymbol{v} = \boldsymbol{\eta} + \boldsymbol{C}\boldsymbol{\lambda}$. Since $\boldsymbol{v}, \boldsymbol{\eta} \in \mathcal{P}_{\boldsymbol{C}}$, this gives us $\boldsymbol{v} = \boldsymbol{\eta}$. Thus, π is one-to-one.

Therefore, $\widetilde{\pi}$ is a bijection between $\widetilde{\mathcal{U}}$ and $\{\widetilde{\Psi}(\boldsymbol{\ell})\}_{\boldsymbol{\ell} \in \Phi}$.

□

The correspondence between $\widetilde{\mathcal{U}}$ and $\{\widetilde{\Psi}(\boldsymbol{\ell})\}_{\boldsymbol{\ell} \in \Phi}$ is shown in Fig. B.1. For each of the three elements in $\widetilde{\mathcal{U}}$ (marked with blue circles), the wrapped chamber given by $\widetilde{\pi}$ is indicated by an arrow.

B.3 Constructing χ

B.3.1 Constructing χ with $n \geq 1$

Algorithm 8 in Section 5.4 requires a set χ that contains a point in $\widetilde{\Psi}(\boldsymbol{\ell})$ for each $\boldsymbol{\ell} \in \Phi$. According to Theorem 35 in Appendix B.2, we can construct χ by forming a point in $\widetilde{\pi}(\boldsymbol{v})$, for each $\boldsymbol{v} \in \widetilde{\mathcal{U}}$.

Any $\boldsymbol{v} \in \widetilde{\mathcal{U}}$ is the intersection of n hyperplanes defined by $\{\boldsymbol{a}_{j_i}^{\mathsf{T}}\}_{i \in \{1,\ldots,n\}}$, for some $\boldsymbol{j} \in \{1, \ldots, m\}^n$. Let $\{\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n\}$ be the n vertices in $\operatorname{cl} \pi(\boldsymbol{v})$ that are connected to $\boldsymbol{v}$

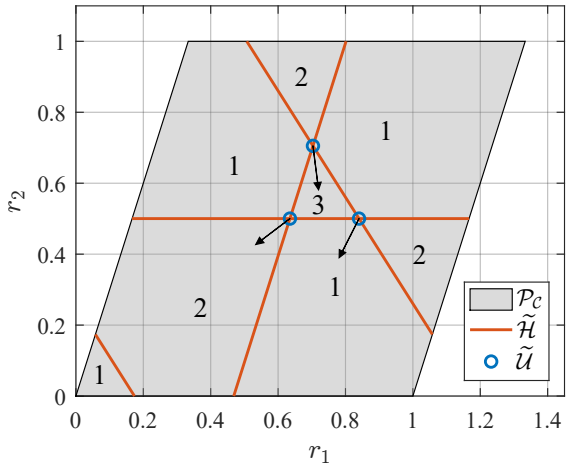


Figure B.1: $\mathcal{P}_{\boldsymbol{C}}$ is partitioned into wrapped chambers, labeled 1–3. The points at which two wrapped hyperplanes in $\widetilde{\mathcal{H}}$ (shown in red) intersect are marked with blue circles (denoted $\widetilde{\mathcal{U}}$). The correspondence of each such point to a wrapped chamber (as described in Appendix B.2) is shown by an arrow.

in G. (Recall that the closure of $\pi(\boldsymbol{v})$ is a polytope.) Consider an n-simplex defined by the $n+1$ extremal points $\{\boldsymbol{v}, \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n\}$. This simplex is a subset of $\operatorname{cl} \pi(\boldsymbol{v})$.

We find a point in the interior of this simplex as the convex combination

$$\phi(\boldsymbol{v}) = \alpha_0 \boldsymbol{v} + \sum_{i=1}^{n} \alpha_i \boldsymbol{\eta}_i, \tag{B.16}$$

where $\alpha_0 + \sum_{i=1}^{n} \alpha_i = 1$ and $\alpha_i \in (0,1)$, for $i \in \{0, \ldots, n\}$. First note that

$$\boldsymbol{\eta}_i = \boldsymbol{v} + (\boldsymbol{I}_n - \boldsymbol{P}_i)\boldsymbol{d}\gamma_i \tag{B.17}$$

for some $\gamma_i > 0$, where $\boldsymbol{P}_i$ is the orthogonal projection onto $\operatorname{span}(\{\boldsymbol{a}_{j_q}^{\mathsf{T}}\}_{q \in \{1,\ldots,n\} \setminus i})$. That is, $(\boldsymbol{I}_n - \boldsymbol{P}_i)$ is a projection to the line that contains both $\boldsymbol{v}$ and $\boldsymbol{\eta}_i$.

From (B.16) and (B.17), we have

$$\phi(\boldsymbol{v}) = \alpha_0 \boldsymbol{v} + \sum_{i=1}^{n} \alpha_i (\boldsymbol{v} + (\boldsymbol{I}_n - \boldsymbol{P}_i)\boldsymbol{d}\gamma_i) \tag{B.18}$$

$$= \boldsymbol{v} + \sum_{i=1}^{n} \alpha_i (\boldsymbol{I}_n - \boldsymbol{P}_i)\boldsymbol{d}\gamma_i. \tag{B.19}$$

We choose the weights such that $\alpha_i = \frac{\epsilon}{\gamma_i}$, for each $i \in \{1, \ldots, n\}$, for some $\epsilon > 0$. Letting $\alpha_0 = 1 - \sum_{i=1}^{n} \frac{\epsilon}{\gamma_i}$, we have that $\sum_{i=0}^{n} \alpha_i = 1$. For a convex combination, it is necessary that $\sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \frac{\epsilon}{\gamma_i} < 1$, and thus, that $\epsilon < (\sum_{i=1}^{n} \gamma_i^{-1})^{-1}$. Recalling that $\boldsymbol{d}$ is a unit vector, we have from (B.17) that

$$\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{\eta}_i - \boldsymbol{v}) = \boldsymbol{d}^{\mathsf{T}}(\boldsymbol{I}_n - \boldsymbol{P}_i)\boldsymbol{d}\gamma_i$$

$$\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{\eta}_i - \boldsymbol{v}) = \|(\boldsymbol{I}_n - \boldsymbol{P}_i)\boldsymbol{d}\|^2 \gamma_i$$

$$\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{\eta}_i - \boldsymbol{v}) \leq \gamma_i.$$

Therefore, it is sufficient that

$$\epsilon < \left(\sum_{i=1}^{n} \left(\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{\eta}_i - \boldsymbol{v}) \right)^{-1} \right)^{-1} \tag{B.20}$$

to ensure that $\sum_{i=1}^{n} \frac{\epsilon}{\gamma_i} < 1$. We have a point,

$$\phi(\boldsymbol{v}) = \boldsymbol{v} + \epsilon \sum_{i=1}^{n} (\boldsymbol{I}_n - \boldsymbol{P}_i)\boldsymbol{d}, \tag{B.21}$$

in the interior of the simplex and, thus, in $\pi(\boldsymbol{v})$. Then $\langle \phi(\boldsymbol{v}) \rangle_{\boldsymbol{C}} \in \widetilde{\pi}(\boldsymbol{v})$. The set χ is formed as $\chi = \{ \langle \phi(\boldsymbol{v}) \rangle_{\boldsymbol{C}} \mid \boldsymbol{v} \in \widetilde{\mathcal{U}} \}$.

B.3.2 Comment on the $n = 1$ Case

For $n = 1$, the construction of χ is greatly simplified. In this case the unit vector is merely $d = \pm 1$ (we assume $d = 1$, without loss of generality), and $(\boldsymbol{I}_n - \boldsymbol{P}_i) = 1$. We also have that $\widetilde{\mathcal{U}} = \widetilde{\mathcal{H}}$. Thus, each point in χ is given by $\phi(v) = v + \epsilon$, for some

$v \in \widetilde{\mathcal{H}}$, where ϵ is chosen such that $\epsilon < |\eta - v|$, for any $v \neq \eta \in \widetilde{\mathcal{H}}$. Therefore, $\chi = \{\langle v + \epsilon \rangle_{\boldsymbol{C}} \mid v \in \widetilde{\mathcal{H}}\}$.

Note that, for $n = 1$, we have that $\mathcal{H} = \{a_j^{-1} y_j + C(\lambda + 1/2) \mid \lambda \in \mathbb{Z}\}$, and

$$\widetilde{\mathcal{H}} = \{\langle a_j^{-1} y_j + C/2 \rangle_C \mid j \in \{1, \ldots, m\}\}. \tag{B.22}$$

Therefore, χ has a cardinality of m.

B.4 Bounds on Problem Size for Restricted Encoding Matrices

This appendix provides bounds on quantities related to the size of the problem considered in Section 5.4, in which we assume that Condition 7 is met by $\boldsymbol{A}$.

B.4.1 Upper Bound on N

From (5.6), for any $\boldsymbol{r} \in \mathcal{P}_{\boldsymbol{C}}$ and some $\boldsymbol{\lambda} \in [0, 1)^n$,

$$\begin{aligned}
\omega_j \boldsymbol{a}_j \boldsymbol{r} &= \omega_j \boldsymbol{a}_j \boldsymbol{C} \boldsymbol{\lambda} \\
&\in \omega_j \boldsymbol{a}_j \boldsymbol{C} [0, 1)^n \\
&= \left[\min(\mathbf{0}_n^\mathsf{T}, \omega_j \boldsymbol{a}_j \boldsymbol{C}) \mathbf{1}_n,\ \max(\mathbf{0}_n^\mathsf{T}, \omega_j \boldsymbol{a}_j \boldsymbol{C}) \mathbf{1}_n \right) \\
&= \left[t_j^{\min},\ t_j^{\max} \right),
\end{aligned} \tag{B.23}$$

with $t_j^{\min}, t_j^{\max} \in \mathbb{Z}$, for each $j \in \{1, \ldots, m\}$ and min and max denoting element-wise minimum and maximum operations, respectively.

From (5.35) and (B.23), any $\boldsymbol{r} \in \widetilde{\mathcal{H}}$ satisfies the following linear equality:

$$\omega_j \boldsymbol{a}_j \boldsymbol{r} = \langle \omega_j y_j + \tfrac{1}{2} \rangle_1 + \ell, \tag{B.24}$$

for some $j \in \{1, \ldots, m\}$ and $\ell \in \{t_j^{\min}, \ldots, t_j^{\max} - 1\}$. Note that $\{t_j^{\min}, \ldots, t_j^{\max} - 1\}$ is a set of integers with cardinality $N^j = \sum_{k=1}^n |\omega_j (\boldsymbol{a}_j \boldsymbol{C})_k|$. Thus, (B.24) describes a

total of $N = \sum_{j=1}^{m} N^j$ hyperplanes, and $\widetilde{\mathcal{H}} \subset \mathcal{H}_{\text{sub}}$, where

$$\begin{aligned}\mathcal{H}^{\text{sub}} =& \big\{ \boldsymbol{r} \in \mathbb{R}^n \mid \exists j \in \{1, \dots, m\} \text{ s.t.} \\ & \omega_j \boldsymbol{a}_j \boldsymbol{r} - \langle \omega_j y_j + \tfrac{1}{2} \rangle_1 \in \{ t_j^{\min}, \dots, t_j^{\max} - 1 \} \big\}, \end{aligned} \tag{B.25}$$

i.e., $\mathcal{H}^{\text{sub}}$ is a union of the N hyperplanes in $\mathcal{H}$ that intersect with $\mathcal{P}_{\boldsymbol{C}}$. Thus, $\widetilde{\mathcal{H}} = \mathcal{P}_{\boldsymbol{C}} \cap \mathcal{H} = \mathcal{P}_{\boldsymbol{C}} \cap \mathcal{H}^{\text{sub}}$.

Recall that $\boldsymbol{\beta} \in \mathbb{Z}^{m \times n}$, with

$$\beta_{j,k} = \begin{cases} 1/(\boldsymbol{a}_j \boldsymbol{C})_k, & (\boldsymbol{a}_j \boldsymbol{C})_k \neq 0 \\ 0, & \text{otherwise} \end{cases},$$

and that $\boldsymbol{\beta}_j \in \mathbb{Z}^{1 \times n}$ is the jth row of $\boldsymbol{\beta}$. Thus, if $\beta_{j,i} \neq 0$, $\beta_{j,i} = \frac{z_i}{(\boldsymbol{a}_j \boldsymbol{V})_i}$. For each $j \in \{1, \dots, m\}$,

$$\begin{aligned} N^j &= \sum_{k=1}^{n} |\operatorname{lcm}(\boldsymbol{\beta}_j^{\mathsf{T}})(\boldsymbol{a}_j \boldsymbol{C})_k| \\ &\leq \sum_{k=1}^{n} \left| \left(\prod_{\{i | \beta_{j,i} \neq 0\}} \beta_{j,i} \right) (\boldsymbol{a}_j \boldsymbol{C})_k \right| \\ &= \sum_{k=1}^{n} \prod_{\{i \neq k | \beta_{j,i} \neq 0\}} |\beta_{j,i}| \leq \prod_{\{i | \beta_{j,i} \neq 0\}} n \, |\beta_{j,i}| \\ &\leq \prod_{\{i | \beta_{j,i} \neq 0\}} n \left| \frac{z_i}{(\boldsymbol{a}_j \boldsymbol{V})_i} \right| \leq n \prod_{i=1}^{n} z_i, \end{aligned} \tag{B.26}$$

in which we we have used that $\boldsymbol{a}_j \boldsymbol{V} \in \mathbb{Z}^{1 \times n}$. From (B.26), an upper bound on N is given by

$$N \leq mn \left(\prod_{i=1}^{n} z_i \right) = mn \det(\boldsymbol{C}^{-1} \boldsymbol{V}). \tag{B.27}$$

Thus, for fixed $\boldsymbol{C}$, $\boldsymbol{V}$, and n, N is $O(m)$.

B.4.2 Bound on Number of Wrapped Chambers

Let $\mathcal{U} \subset \mathcal{H}$ denote the set of points at which n hyperplanes in $\mathcal{H}$ intersect, and let $\mathcal{U}^{\text{sub}} \subset \mathcal{U}$ denote the set of points at which n hyperplanes in $\mathcal{H}^{\text{sub}}$ intersect. $|\mathcal{U}^{\text{sub}}|$

is maximized when none of the N hyperplanes that it comprises are parallel to one another, in which case, $|\mathcal{U}^{\text{sub}}| = \binom{N}{n}$. From (B.27), an upper bound on $|\mathcal{U}^{\text{sub}}|$ is given by

$$|\mathcal{U}^{\text{sub}}| \leq \binom{N}{n} \leq (m \det(\boldsymbol{C}^{-1}\boldsymbol{V}))^n n^{n-1}. \tag{B.28}$$

Thus, for fixed $\boldsymbol{C}$, $\boldsymbol{V}$, and n, $|\mathcal{U}^{\text{sub}}|$ is $O(m^n)$.

In general, some hyperplanes in $\mathcal{H}^{\text{sub}}$ will be parallel to one another, and this admits a refinement of the upper bound on $|\mathcal{U}^{\text{sub}}|$. Letting τ denote the set of n-combinations of $\{1, \ldots, m\}$, an upper bound on $|\mathcal{U}^{\text{sub}}|$ is given by

$$|\mathcal{U}^{\text{sub}}| \leq \sum_{\mathcal{I} \in \tau} \prod_{j \in \mathcal{I}} N^j. \tag{B.29}$$

We have that $\widetilde{\mathcal{U}} = \mathcal{P}_{\boldsymbol{C}} \cap \mathcal{U} = \mathcal{P}_{\boldsymbol{C}} \cap \mathcal{U}^{\text{sub}}$ and, therefore, that $|\widetilde{\mathcal{U}}| \leq |\mathcal{U}^{\text{sub}}|$.

Let $K = |\{\widetilde{\Psi}(\boldsymbol{\ell}) \mid \boldsymbol{\ell} \in \Phi\}|$ denote the number of wrapped chambers. We have from Theorem 35 that there is a bijection between the set of wrapped chambers and $\widetilde{\mathcal{U}}$. Thus,

$$K = |\widetilde{\mathcal{U}}| \leq |\mathcal{U}^{\text{sub}}|, \tag{B.30}$$

and the upper bounds on $|\mathcal{U}^{\text{sub}}|$ in (B.28) and (B.29) apply to the number of wrapped chambers as well. Thus, K is $O(m^n)$.

B.5 Bounds on Problem Size for General Encoding Matrices

This appendix provides bounds on quantities related to the size of the problem considered in Section 5.5, in which the restriction on $\boldsymbol{A}$ used in Section 5.4 is lifted.

B.5.1 Bound on N_{NC}

For $\boldsymbol{x} \in \mathcal{P}_{\boldsymbol{V}}$, we have, for some $\boldsymbol{\lambda} \in [0,1)^n$, that

$$\begin{aligned}\boldsymbol{a}_j\boldsymbol{x} &= \boldsymbol{a}_j\boldsymbol{V}\boldsymbol{\lambda} \\ &\in \boldsymbol{a}_j\boldsymbol{V}[0,1)^n \\ &= \big[\min(\boldsymbol{0}_n^{\mathsf{T}}, \boldsymbol{a}_j\boldsymbol{V})\boldsymbol{1}_n,\, \max(\boldsymbol{0}_n^{\mathsf{T}}, \boldsymbol{a}_j\boldsymbol{V})\boldsymbol{1}_n\big) \\ &= \big[t_j^{\min},\, t_j^{\max}\big) \end{aligned} \tag{B.31}$$

with $t_j^{\min}, t_j^{\max} \in \mathbb{Z}$, for each $j \in \{1,\dots,m\}$.

From (5.48) and (B.31), any $\boldsymbol{x} \in \widetilde{\mathcal{H}}_{\mathrm{NC}}$ satisfies

$$\boldsymbol{a}_j\boldsymbol{x} = \langle y_j + \tfrac{1}{2}\rangle_1 + \ell, \tag{B.32}$$

for some $j \in \{1,\dots,m\}$ and $\ell \in \{t_j^{\min},\dots,t_j^{\max}-1\}$. Note that $\{t_j^{\min},\dots,t_j^{\max}-1\}$ is a set of integers with cardinality $N_{\mathrm{NC}}^j = \sum_{k=1}^n |(\boldsymbol{a}_j\boldsymbol{V})_k|$. Thus (B.32) describes a total of $N_{\mathrm{NC}} = \sum_{j=1}^m N_{\mathrm{NC}}^j$ hyperplanes. We therefore have that $\widetilde{\mathcal{H}}_{\mathrm{NC}} \subset \mathcal{H}_{\mathrm{NC}}^{\mathrm{sub}}$, where

$$\begin{aligned}\mathcal{H}_{\mathrm{NC}}^{\mathrm{sub}} = \big\{&\boldsymbol{x} \in \mathbb{R}^n \mid \exists j \in \{1,\dots,m\} \text{ s.t.} \\ &\boldsymbol{a}_j\boldsymbol{x} - \langle y_j + \tfrac{1}{2}\rangle_1 \in \{t_j^{\min},\dots,t_j^{\max}-1\}\big\},\end{aligned} \tag{B.33}$$

i.e., $\mathcal{H}_{\mathrm{NC}}^{\mathrm{sub}}$ is a union of the N_{NC} hyperplanes in $\mathcal{H}_{\mathrm{NC}}$ that intersect with $\mathcal{P}_{\boldsymbol{V}}$. Therefore, $\widetilde{\mathcal{H}}_{\mathrm{NC}} = \mathcal{P}_{\boldsymbol{V}} \cap \mathcal{H}_{\mathrm{NC}} = \mathcal{P}_{\boldsymbol{V}} \cap \mathcal{H}_{\mathrm{NC}}^{\mathrm{sub}}$.

Recalling that $\boldsymbol{z} = (\mathrm{lcm}(\boldsymbol{A}\boldsymbol{V}))^{\mathsf{T}}$, we have an upper bound on N_{NC}^j given by $N_{\mathrm{NC}}^j \le \sum_{k=1}^n z_k$. Therefore,

$$N_{\mathrm{NC}} \le m\boldsymbol{z}^{\mathsf{T}}\boldsymbol{1}_n = m\,\mathrm{tr}(\boldsymbol{C}^{-1}\boldsymbol{V}). \tag{B.34}$$

B.5.2 Comparison Between N and N_{NC}

Like N_{NC}, N in Appendix B.4 determines the complexity of an ML solution method. Both quantities give the number of hyperplanes that intersect a parallelotope. The method in Section 5.4, to which N is relevant, has a restrictive condition on the encoding matrix $\boldsymbol{A}$. It may be surprising that, even when Condition 7 is met, it is possible that $N > N_{\mathrm{NC}}$. This is the case, for example, when

$$\boldsymbol{A} = \begin{bmatrix} 3 & 0 & 1 & 3 \\ 4 & 4 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{\mathsf{T}}, \tag{B.35}$$

with $\boldsymbol{V} = \boldsymbol{I}_3$ and $\boldsymbol{C} = \mathrm{diag}([\frac{1}{3}, \frac{1}{4}, 1]^{\mathsf{T}})$. In this case, $N = 26$ and $N_{\mathrm{NC}} = 20$. Note that for $n = 1$, $N = m$, and thus, $N \leq N_{\mathrm{NC}}$.

B.5.3 Bound on the Number of Wrapped Chambers

Following Appendix B.4, upper bounds on $|\mathcal{U}^{\mathrm{sub}}|$ are given by

$$|\mathcal{U}^{\mathrm{sub}}| \leq \binom{N_{\mathrm{NC}}}{n} \leq (m \,\mathrm{tr}(\boldsymbol{C}^{-1}\boldsymbol{V}))^n n^{-1}. \tag{B.36}$$

$$|\mathcal{U}^{\mathrm{sub}}| \leq \sum_{\mathcal{I} \in \tau} \prod_{j \in \mathcal{I}} N_{\mathrm{NC}}^j. \tag{B.37}$$

We have that $\widetilde{\mathcal{U}} = \mathcal{P}_{\boldsymbol{V}} \cap \mathcal{U} = \mathcal{P}_{\boldsymbol{V}} \cap \mathcal{U}^{\mathrm{sub}}$ and, therefore, that $|\widetilde{\mathcal{U}}| \leq |\mathcal{U}^{\mathrm{sub}}|$. Let $K_{\mathrm{NC}} = |\{\widetilde{\Psi}_{\mathrm{NC}}(\boldsymbol{\ell}) \mid \boldsymbol{\ell} \in \Phi\}|$ denote the number of wrapped chambers. By a statement analogous to Theorem 35 in Appendix B.2, it can be shown that there is a bijection between the set of wrapped chambers and $\widetilde{\mathcal{U}}$. Thus,

$$K_{\mathrm{NC}} = |\widetilde{\mathcal{U}}| \leq |\mathcal{U}^{\mathrm{sub}}|, \tag{B.38}$$

and the upper bounds on $|\mathcal{U}^{\mathrm{sub}}|$ in (B.36) and (B.37) apply to the number of wrapped chambers as well. Therefore, K is of order $O(m^n)$ (for fixed n, $\boldsymbol{C}$, and $\boldsymbol{V}$).

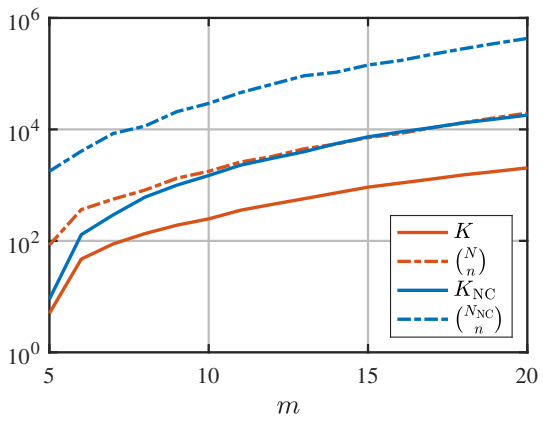


Figure B.2: The number of wrapped chambers for Algorithm 8 is plotted in red against m; the red dash-dotted line is the upper bound on K from (B.28). The number of wrapped chambers for Algorithm (9) is in blue; the blue dash-dotted line is the upper bound on K_{NC} from (B.36).

B.5.4 Comparison Between K and K_{NC}

Here we compare the number of wrapped chambers for the algorithms in Section 5.4 (denoted K) and Section 5.5 (denoted K_{NC}). The number of wrapped chambers for the two methods are shown against m in Fig. B.2. The encoding matrices are such that, as m increases, $\boldsymbol{V}$ and $\boldsymbol{C}$ are constant. The encoding matrices are made to satisfy Condition 7 from Section 5.4. In Fig. B.2, K and K_{NC} are plotted against $m = 5, \ldots, 20$. The upper bound on K from (B.28) is shown as well as the upper bound on K_{NC} from (B.36). The plot shows the computational benefit that Algorithm 8 may have over Algorithm 9 when Condition 7 is met. For example, with $m = 20$, $K_{\mathrm{NC}} > 8K$.

B.6 Proofs for Section 5.6

Proposition 36. *For any $\boldsymbol{t} \in \mathbb{R}^m$ and $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ having orthogonal (though, not necessarily orthonormal) columns,*

$$\widehat{\boldsymbol{k}} = \operatorname*{arg\,min}_{\boldsymbol{k} \in \mathbb{Z}^n} \|\boldsymbol{t} - \boldsymbol{B}\boldsymbol{k}\| = \lceil \boldsymbol{B}^\dagger \boldsymbol{t} \rfloor. \tag{B.39}$$

That is, the Babai rounding method [197] solves a closest lattice point problem when the lattice basis has orthogonal columns.

Proof.

$$\begin{aligned}
\|\boldsymbol{t} - \boldsymbol{B}\boldsymbol{k}\|^2 =& \|(\boldsymbol{B}\boldsymbol{B}^\dagger + \boldsymbol{P}_{\boldsymbol{B}}^{\perp})\boldsymbol{t} - \boldsymbol{B}\boldsymbol{k}\|^2 \\
=& \|\boldsymbol{P}_{\boldsymbol{B}}^{\perp}\boldsymbol{t}\|^2 + \|\boldsymbol{B}\boldsymbol{B}^\dagger\boldsymbol{t}\|_2^2 + \|\boldsymbol{B}\boldsymbol{k}\|^2 - 2\boldsymbol{t}^\mathsf{T}\boldsymbol{B}\boldsymbol{k} \\
=& \|\boldsymbol{P}_{\boldsymbol{B}}^{\perp}\boldsymbol{t}\|^2 + \sum_{i=1}^{n} \|\boldsymbol{b}_i\|^2((\boldsymbol{B}^\dagger\boldsymbol{t})_i - k_i)^2,
\end{aligned}$$

where $\boldsymbol{B}\boldsymbol{B}^\dagger$ is the orthogonal projection onto the range of $\boldsymbol{B}$, $\boldsymbol{P}_B^{\perp}$ is the orthogonal projection onto the kernel of $\boldsymbol{B}$, and $\boldsymbol{b}_i$ is the ith column of $\boldsymbol{B}$. Since the objective is separable in $\boldsymbol{k}$, it is clear that $\widehat{k}_i = \lceil (\boldsymbol{B}^\dagger\boldsymbol{t})_i \rfloor$, and $\widehat{\boldsymbol{k}} = \lceil \boldsymbol{B}^\dagger\boldsymbol{t} \rfloor$. □

Appendix C: Phase Noise Model

In [31], the noise in the phase entries of a sample covariance matrix from multi-snapshot data is modeled by jointly Gaussian random variables. In this appendix, we rederive this approximation and express the resulting approximate covariance matrix. Following the signal model in Section 5.6.7 and Section 6.2, consider an array of L isotropic sensors that is impinged upon by a narrowband signal $s(t) \in \mathbb{C}$ from a single emitter located in the far field of the array. The demodulated array output vector is given by

$$\boldsymbol{b}(t) = \boldsymbol{a}(\boldsymbol{x})s(t) + \boldsymbol{\varepsilon}(t), \qquad t = 1, ..., T, \tag{C.1}$$

where $\boldsymbol{a}(\boldsymbol{x})$ is the array steering vector for the parameter $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{\varepsilon}(t)$ is zero-mean circular Gaussian noise that is uncorrelated across snapshots indexed by t. The array steering vector is given by

$$\boldsymbol{a}(\boldsymbol{x}) = \begin{bmatrix} e^{i\pi \boldsymbol{p}_1^\mathsf{T}\boldsymbol{x}} & e^{i\pi \boldsymbol{p}_2^\mathsf{T}\boldsymbol{x}} & \cdots & e^{i\pi \boldsymbol{p}_L^\mathsf{T}\boldsymbol{x}} \end{bmatrix}^\mathsf{T}, \tag{C.2}$$

where $i = \sqrt{-1}$. The position vectors $\boldsymbol{p}_j \in \mathbb{R}^n$, for $j \in \{1, 2, \ldots, L\}$, give the sensor positions in units of one-half wavelength.

The sample correlation matrix of $\boldsymbol{b}$ is $\widehat{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{b}(t)\boldsymbol{b}^\mathsf{H}(t)$. Let $\boldsymbol{v}$ denote the vector of $\binom{L}{2}$ lower-triangular entries of $\widehat{\boldsymbol{R}}$ in column-major order. That is,

$$\boldsymbol{v} = [\, \widehat{r}_{2,1}, \widehat{r}_{3,1}, \ldots, \widehat{r}_{L,L-1} \,]^\mathsf{T} \tag{C.3}$$

where $\widehat{r}_{i,j}$ is the (i,j)th entry of $\widehat{\boldsymbol{R}}$.

Under the conditional (deterministic) signal model [222], the distribution of $\widehat{r}_{i,j}$ depends on the fixed, unknown signal $\{s(t)\}$. Because we do not observe this signal directly, we involve the noisy array outputs in the following approximation:

$$|s(t)|^2 \approx \tfrac{1}{L}\boldsymbol{b}^{\mathsf{H}}(t)\boldsymbol{b}(t). \tag{C.4}$$

Using (C.4), we have

$$\mathrm{E}\left[\widehat{r}_{i,j}\right] \approx \widehat{\rho}a_i(\boldsymbol{x})\overline{a}_j(\boldsymbol{x}), \qquad i \neq j \tag{C.5}$$

$$\mathrm{var}\left(\widehat{r}_{i,j}\right) \approx T^{-1}\left(2\widehat{\rho}\sigma^2 + \sigma^4\right), \qquad i \neq j \tag{C.6}$$

$$\mathrm{cov}\left(\widehat{r}_{i,j}, \widehat{r}_{k,j}\right) \approx T^{-1}a_i(\boldsymbol{x})\overline{a}_k(\boldsymbol{x})\widehat{\rho}\sigma^2, \qquad i \neq j,\, k \neq j,\, i \neq k \tag{C.7}$$

$$\mathrm{cov}\left(\widehat{r}_{i,j}, \widehat{r}_{i,k}\right) \approx T^{-1}\overline{a}_j(\boldsymbol{x})a_k(\boldsymbol{x})\widehat{\rho}\sigma^2, \qquad i \neq j,\, i \neq k,\, j \neq k \tag{C.8}$$

$$\mathrm{cov}\left(\widehat{r}_{i,j}, \widehat{r}_{k,\ell}\right) = 0, \qquad i \neq j,\, k \neq \ell,\, i \neq k,\, j \neq \ell, \tag{C.9}$$

where $^-$ denotes complex conjugation, and

$$\widehat{\rho} \triangleq \frac{1}{LT}\sum_{t=1}^{T}\boldsymbol{b}^{\mathsf{H}}(t)\boldsymbol{b}(t). \tag{C.10}$$

In the above, we have used the conditional signal model for $\{s(t)\}$. If the signal follows the unconditional model, then $\widehat{\rho} = T^{-1}\sum_{t=1}^{T}\mathrm{E}\left[|s(t)|^2\right]$ may be used, rather than estimating the power of $\{s(t)\}$ from the data.

The complex logarithm of $z \in \mathbb{C}\setminus\{0\}$ is given by $\mathrm{Log}\, z \triangleq \log|z| + i\angle z$. Let

$$\begin{aligned} \boldsymbol{f}\colon \{\mathbb{C}\setminus 0\}^m &\to [-\tfrac{1}{2},\tfrac{1}{2})^m \\ \boldsymbol{z} &\mapsto \frac{i}{4\pi}\left(\mathrm{Log}\,\overline{\boldsymbol{z}} - \mathrm{Log}\,\boldsymbol{z}\right). \end{aligned} \tag{C.11}$$

Thus, $\boldsymbol{f}(\boldsymbol{z}) = (2\pi)^{-1}\angle\boldsymbol{z} \in [-\frac{1}{2},\frac{1}{2})^m$ for any $\boldsymbol{z} \in \{\mathbb{C}\setminus 0\}^m$.

Now, let $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{v}) \in [-\frac{1}{2},\frac{1}{2})^{\binom{L}{2}}$. We use nonlinear error propagation to approximate the covariance matrix of $\boldsymbol{y}$, i.e., we find the covariance of a first-order Taylor

approximation of $\boldsymbol{y}$. The linear approximation of $\boldsymbol{y}$ around its expected value is

$$\boldsymbol{y} \approx \widetilde{\boldsymbol{y}} = \boldsymbol{f}(\mathrm{E}[\boldsymbol{v}]) + \boldsymbol{J}_{\boldsymbol{f}}(\mathrm{E}[\boldsymbol{v}])(\boldsymbol{v} - \mathrm{E}[\boldsymbol{v}]) + \boldsymbol{J}^c_{\boldsymbol{f}}(\mathrm{E}[\boldsymbol{v}])(\overline{\boldsymbol{v}} - \mathrm{E}[\overline{\boldsymbol{v}}]) \tag{C.12}$$

$$= \boldsymbol{f}(\mathrm{E}[\boldsymbol{v}]) + 2\,\mathrm{Re}\,\{\boldsymbol{J}_{\boldsymbol{f}}(\mathrm{E}[\boldsymbol{v}])(\boldsymbol{v} - \mathrm{E}[\boldsymbol{v}])\}, \tag{C.13}$$

in which $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{z}) \in \mathbb{R}^{m\times m}$ is the Jacobian of $\boldsymbol{f}$ evaluated at $\boldsymbol{z}$ and $\boldsymbol{J}^c_{\boldsymbol{f}}(\boldsymbol{z}) \in \mathbb{R}^{m\times m}$ is the conjugate Jacobian of $\boldsymbol{f}$ [254]. That is, $(\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{z}))_{j,k} = \frac{\partial f_j(\boldsymbol{z})}{\partial v_k}$ and $(\boldsymbol{J}^c_{\boldsymbol{f}}(\boldsymbol{z}))_{j,k} = \frac{\partial f_j(\boldsymbol{z})}{\partial \overline{v}_k}$, for $j, k \in \{1, \ldots, m\}$. Note that $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{z}) = \overline{\boldsymbol{J}^c}_{\boldsymbol{f}}(\boldsymbol{z})$, as $\boldsymbol{f}$ is real-valued [254]. The Jacobian of $\boldsymbol{f}$ evaluated at $\boldsymbol{z} \in \{\mathbb{C} \setminus 0\}^m$ is

$$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{z}) = \mathrm{diag}\left(-\frac{i}{4\pi}\left[\, z_1^{-1}, z_2^{-1}, \ldots, z_m^{-1} \,\right]^{\mathsf{T}}\right). \tag{C.14}$$

The covariance of $\widetilde{\boldsymbol{y}}$ is

$$\mathrm{cov}(\widetilde{\boldsymbol{y}}) = \mathrm{cov}(2\,\mathrm{Re}\,\{\boldsymbol{J}_{\boldsymbol{f}}(\mathrm{E}[\boldsymbol{v}])(\boldsymbol{v} - \mathrm{E}[\boldsymbol{v}])\}) \tag{C.15}$$

$$= 2\boldsymbol{J}_{\boldsymbol{f}}(\mathrm{E}[\boldsymbol{v}])\,\mathrm{cov}(\boldsymbol{v})(\boldsymbol{J}_{\boldsymbol{f}}(\mathrm{E}[\boldsymbol{v}]))^{\mathsf{H}}. \tag{C.16}$$

Thus,

$$\begin{aligned}
\mathrm{var}(\widetilde{y}_{i,j}) &= (2\pi)^{-2}\frac{\mathrm{var}(\widehat{r}_{i,j})}{2|\mathrm{E}\,[\widehat{r}_{i,j}]|^2} = (2\pi)^{-2}\frac{1 + 2\frac{\widehat{\rho}}{\sigma^2}}{2T\left(\frac{\widehat{\rho}}{\sigma^2}\right)^2}, \qquad i > j \\
\mathrm{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{k,j}) &= (2\pi)^{-2}\frac{\mathrm{cov}(\widehat{r}_{i,j}, \widehat{r}_{k,j})}{2\,\mathrm{E}\,[\widehat{r}_{i,j}]\,\mathrm{E}\left[\overline{\widehat{r}}_{k,j}\right]} = (2\pi)^{-2}\frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, \qquad i > j,\ k > j,\ i \neq k \\
\mathrm{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{i,k}) &= (2\pi)^{-2}\frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, \qquad i > j,\ i > k,\ j \neq k \\
\mathrm{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{j,k}) &= -\,\mathrm{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{k,j}) = -(2\pi)^{-2}\frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, \qquad i > j > k \\
\mathrm{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{k,i}) &= -(2\pi)^{-2}\frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, \qquad k > i > j \\
\mathrm{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{k,\ell}) &= 0, \qquad i > j,\ k > \ell,\ \{i,j\} \cap \{k,\ell\} = \varnothing.
\end{aligned} \tag{C.17}$$

Equivalently, for any $i > j$, $k > \ell$, we have that

$$\operatorname{cov}(\widetilde{y}_{i,j}, \widetilde{y}_{k,\ell}) = \begin{cases} (2\pi)^{-2} \frac{1+2\frac{\widehat{\rho}}{\sigma^2}}{2T\left(\frac{\widehat{\rho}}{\sigma^2}\right)^2}, & i = k,\ j = \ell \\ (2\pi)^{-2} \frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, & i \neq k,\ j = \ell, \\ (2\pi)^{-2} \frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, & i = k,\ j \neq \ell, \\ -(2\pi)^{-2} \frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, & j = k \\ -(2\pi)^{-2} \frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}}, & i = \ell \\ 0, & \{i,j\} \cap \{k,\ell\} = \varnothing \end{cases} \tag{C.18}$$

Hence, the covariance matrix approximation can be expressed as

$$\operatorname{cov}(\widetilde{\boldsymbol{y}}) = (2\pi)^{-2} \frac{1}{2T\frac{\widehat{\rho}}{\sigma^2}} \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathsf{T}} + \left(\frac{\widehat{\rho}}{\sigma^2}\right)^{-1} \boldsymbol{I} \right), \tag{C.19}$$

where $\boldsymbol{\Phi} \in \{-1, 0, 1\}^{\binom{L}{2} \times L}$ is a matrix with rank $L-1$ that corresponds to the ordering of elements in $\boldsymbol{v}$. When $L = 4$, for example, $\boldsymbol{\Phi}$ is given by

$$\boldsymbol{\Phi} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}. \tag{C.20}$$

Bibliography

[1] Google, "Wright-Patterson Air Force Base Area B," 2021, Accessed Sept. 16, 2021. [Online]. Available: https://www.google.com/maps

[2] S. M. Scarborough, L. Gorham, M. J. Minardi, U. K. Majumder, M. G. Judge, L. Moore, L. Novak, S. Jaroszewksi, L. Spoldi, and A. Pieramico, "A challenge problem for SAR change detection and data compression," in *Proc. SPIE*, vol. 7699, 2010, pp. 287–291.

[3] Directorate Space, Security and Migration, European Commission Joint Research Centre, "Copernicus Emergency Management Service, EMSR515 Melini: Grading product, version 1, release 1," Jul. 2021, Accessed May 12, 2022. [Online]. Available: https://emergency.copernicus.eu

[4] Google Earth version 9.163.0.0, "Troodos Mountains, Cyprus," May 2020, Accessed June 2, 2022. [Online]. Available: https://earth.google.com

[5] Copernicus, "Sentinel-1 data, processed by ESA," Jul. 2021, Retrieved from ASF DAAC May 12, 2022. [Online]. Available: https://asf.alaska.edu/

[6] NASA, "Dataset: UAVSAR," Retrieved from ASF DAAC August 18, 2020, 2020. [Online]. Available: https://asf.alaska.edu/

[7] D. Tucker, S. Zhao, and L. C. Potter, "Maximum likelihood estimation in mixed integer linear models," *IEEE Signal Process. Lett.*, vol. 30, pp. 1557–1561, 2023.

[8] J. Jones, A. R. Webster, and W. K. Hocking, "An improved interferometer design for use with meteor radars," *Radio Sci.*, vol. 33, no. 1, pp. 55–65, 1998.

[9] N. J. Pelc, M. A. Bernstein, A. Shimakawa, and G. H. Glover, "Encoding strategies for three-direction phase-contrast MR imaging of flow," *J. Magn. Reson. Imaging*, vol. 1, no. 4, pp. 405–413, 1991.

[10] L. M. Novak, S. D. Halversen, G. Owirka, and M. Hiett, "Effects of polarization and resolution on SAR ATR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 1, pp. 102–116, Jan. 1997.

[11] E. Ertin and L. C. Potter, "Polarimetric classification of scattering centers using M-ary Bayesian decision rules," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 3, pp. 738–749, Jul. 2000.

[12] C. López-Martínez and X. Fabregas, "Polarimetric SAR speckle noise model," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2232–2242, Oct. 2003.

[13] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Am.*, vol. 66, no. 11, pp. 1145–1150, Nov. 1976.

[14] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1017–1028, 1994.

[15] R. G. Gallagher, *Circularly-Symmetric Gaussian Random Vectors.* Unpublished, Jan. 2008. [Online]. Available: https://pdfs.semanticscholar.org/815e/637fafa233b44e2067f6ce345734012bd844.pdf

[16] X. Ma, P. Wu, Y. Wu, and H. Shen, "A review on recent developments in fully polarimetric SAR image despeckling," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 743–758, Mar. 2018.

[17] S. Foucher and C. López-Martínez, "Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1751–1764, Apr. 2014.

[18] P. Teunissen, "The least-squares ambiguity decorrelation adjustment: A method for fast GPS integer ambiguity estimation," *J. Geod.*, vol. 70, pp. 65–82, 11 1995.

[19] A. Hassibi and S. Boyd, "Integer parameter estimation in linear models with applications to GPS," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2938–2952, 1998.

[20] W. Li, X. Wang, X. Wang, and B. Moran, "Distance estimation using wrapped phase measurements in noise," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1676–1688, 2013.

[21] A. Akhlaq, R. G. McKilliam, and R. Subramanian, "Basis construction for range estimation by phase unwrapping," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2152–2156, 2015.

[22] P. Stoica and R. Moses, *Spectral Analysis of Signals.* Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[23] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 832–835, 1985.

[24] R. G. McKilliam, B. G. Quinn, I. V. L. Clarkson, and B. Moran, "Frequency estimation by phase unwrapping," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 2953–2963, 2010.

[25] W. Xu, E.-C. Chang, L. K. Kwoh, H. Lim, W. Cheng, and A. Heng, "Phase-unwrapping of SAR interferogram with multi-frequency or multi-baseline," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 1994, pp. 730–732.

[26] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.

[27] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, 2004.

[28] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933–953, 2006.

[29] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

[30] L. GAN and H. Liu, "High dynamic range sensing using multi-channel modulo samplers," in *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2020, pp. 1–5.

[31] S. Zhao, R. Ahmad, and L. C. Potter, "Venc design and velocity estimation for phase contrast MRI," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3712–3724, 2022.

[32] W. Wang, X. Li, W. Wang, and X.-G. Xia, "Maximum likelihood estimation based robust Chinese remainder theorem for real numbers and its fast algorithm," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3317–3331, 2015.

[33] L. C. Godara and A. Cantoni, "Uniqueness and linear independence of steering vectors in array space," *J. Acoust. Soc. Am.*, vol. 70, no. 2, pp. 467–475, Aug. 1981.

[34] D. Tucker and L. C. Potter, "Speckle suppression in multi-channel coherent imaging: A tractable Bayesian approach," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1429–1439, 2020.

[35] J.-S. Lee, "A simple speckle smoothing algorithm for synthetic aperture radar images," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, no. 1, pp. 85–89, 1983.

[36] J.-S. Lee, M. R. Grunes, and G. de Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, 1999.

[37] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, vol. 2, 2005, pp. 60–65.

[38] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In)SAR denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2021–2038, 2015.

[39] H. Zhong, J. Zhang, and G. Liu, "Robust polarimetric SAR despeckling based on nonlocal means and distributed Lee filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4198–4210, Jul. 2014.

[40] G. Liu and H. Zhong, "Nonlocal means filter for polarimetric SAR data despeckling based on discriminative similarity measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 514–518, Feb. 2014.

[41] H. Li and Z. Qin, "PolSAR image speckle reduction algorithm based on TV-PDE," in *Proc. SPIE*, vol. 8002, 2011, pp. 178 – 184.

[42] C.-A. Deledalle, L. Denis, S. Tabti, and F. Tupin, "MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction?" *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4389–4403, 2017.

[43] X. Nie, H. Qiao, and B. Zhang, "A variational model for PolSAR data speckle reduction based on the Wishart distribution," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1209–1222, Apr. 2015.

[44] X. Nie, H. Qiao, B. Zhang, and X. Huang, "A nonlocal TV-based variational method for PolSAR data speckle reduction," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2620–2634, Jun. 2016.

[45] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[46] H. E. Güven, A. Güngör, and M. Çetin, "An augmented Lagrangian method for complex-valued compressed SAR imaging," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 235–250, Sep. 2016.

[47] M. Çetin, I. Stojanović, N. O. Önhon, K. Varshney, S. Samadi, W. C. Karl, and A. S. Willsky, "Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 27–40, Jul. 2014.

[48] L. C. Potter, E. Ertin, J. T. Parker, and M. Çetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, Jun. 2010.

[49] N. A. Goodman and L. C. Potter, "Pitfalls and possibilities of radar compressive sensing," *Appl. Opt.*, vol. 54, no. 8, pp. C1–C13, Mar. 2015.

[50] S. Zhang, Y. Liu, and X. Li, "Sparse aperture InISAR imaging via sequential multiple sparse Bayesian learning," in *Sensors*, 2017.

[51] G. Xu, M. Xing, L. Zhang, Y. Liu, and Y. Li, "Bayesian inverse synthetic aperture radar imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1150–1154, Nov. 2011.

[52] H. Liu, B. Jiu, H. Liu, and Z. Bao, "Superresolution ISAR imaging based on sparse Bayesian learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5005–5013, Aug. 2014.

[53] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001. [Online]. Available: https://doi.org/10.1162/15324430152748236

[54] C. J. Pellizzari, R. Trahan, H. Zhou, S. Williams, S. E. Williams, B. Nemati, M. Shao, and C. A. Bouman, "Synthetic aperature LADAR: A model-based approach," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 901–916, Dec. 2017.

[55] ——, "Optically-coherent image formation and denoising using plug and play inversion framework," *Applied Optics*, vol. 56, no. 16, pp. 4735–4744, Jun. 2017.

[56] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

[57] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, Mar. 2017.

[58] X. Dong and Y. Zhang, "SAR image reconstruction from undersampled raw data using maximum a posteriori estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1651–1664, Apr. 2015.

[59] D. C. Munson Jr, J. Dennis O'Brien, and K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proceedings of the IEEE*, vol. 17, pp. 917–925, Sep. 1983.

[60] T. J. Kragh and A. A. Kharbouch, "Monotonic iterative algorithms for SAR image restoration," in *2006 IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 645–648.

[61] L. M. Novak, M. C. Burl, R. D. Chaney, and G. J. Owirka, "Optimal processing of polarimetric synthetic-aperture radar imagery," *The Lincoln Laboratory Journal*, pp. 273–290, 1990.

[62] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimensional Systems and Signal Processing*, vol. 5, no. 3, pp. 263–279, Jul. 1994.

[63] J. N. Ash, "Joint imaging and change detection for robust exploitation in interrupted SAR environments," in *Proc. SPIE*, vol. 8746, 2013, pp. 144–152.

[64] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.

[65] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[66] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[67] P. Zwiernik, C. Uhler, and D. Richards, "Maximum likelihood estimation for linear Gaussian covariance models," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 79, no. 4, pp. 1269–1292, 2016.

[68] J. M. Tang and Y. Saad, "A probing method for computing the diagonal of a matrix inverse," *Numerical Linear Algebra with Applications*, vol. 19, no. 3, pp. 485–501, Apr. 2010.

[69] M. Kubale, *Graph Colorings*. Providence, RI, USA: American Mathematical Society, 2004.

[70] N. Nguyen and Q. H. Liu, "The regular Fourier matrices and nonuniform fast Fourier transforms," *SIAM Journal on Scientific Computing*, vol. 21, no. 1, pp. 283–293, 1999.

[71] R. Ahmad, C. D. Austin, and L. C. Potter, "Toeplitz embedding for fast iterative regularized imaging," in *SPIE Algorithms for Synthetic Aperture Radar Imagery*, vol. 8051, 2011.

[72] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Carnegie Mellon University, Pittsburgh, PA, USA, Tech. Rep., 1994.

[73] G. Farage, S. Foucher, and G. B. Bénié, "Comparison of PolSAR speckle filtering techniques," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2006, pp. 1760–1763.

[74] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[75] Hua Xie, L. E. Pierce, and F. T. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2196–2212, Oct. 2002.

[76] H. Feng, B. Hou, and M. Gong, "SAR image despeckling based on local homogeneous-region segmentation by using pixel-relativity measurement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2724–2737, Jul. 2011.

[77] L. Torres, S. J. S. Sant'Anna, C. da Costa Freitas, and A. C. Frery, "Speckle reduction in polarimetric sar imagery with stochastic distances and nonlocal means," *Pattern Recognit.*, vol. 47, no. 1, pp. 141–157, Jan. 2014.

[78] Y. Yamaguchi, T. Nishikawa, M. Sengoku, and W. M. Boerner, "2-D polarimetric imaging by an FM-CW radar," in *Proc. Proc. IEEE Antennas Propag. Soc. Int. Symp. URSI Nat. Radio Sci. Meeting*, vol. 3, Jun. 1994, pp. 1998–2001.

[79] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "ESA's sentinel missions in support of earth system science," *Remote Sens. Environ.*, vol. 120, pp. 84–90, 2012, the Sentinel Missions - New Opportunities for Science.

[80] Y. Yamaguchi, K. Kimura, H. Yamada, S. Uratsuka, and W.-M. Boerner, "L-band polarimetric AIR/Pi-SAR images around Niigata city," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Jun. 2002, pp. 423–425.

[81] C. Bouman and K. Sauer, "A generalized gaussian image model for edge-preserving map estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, July 1993.

[82] L. Denis, C. Deledalle, and F. Tupin, "From patches to deep learning: Combining self-similarity and neural networks for sar image despeckling," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, July 2019, pp. 5113–5116.

[83] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[84] C. J. Pellizzari, M. F. Spencer, and C. A. Bouman, "Coherent-image reconstruction using convolutional neural networks," in *Imaging and Applied Optics*, 2019, p. MTu4D.4.

[85] D. Tucker, J. N. Ash, and L. C. Potter, "SAR coherent change detection with variational expectation maximization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2163–2175, 2023.

[86] M. Preiss and N. J. S. Stacy, "Coherent change detection: Theoretical description and experimental results," Defence Sci. Technol. Org., Tech. Rep. DSTO-TR-1851, 2006.

[87] D. A. Yocky and B. F. Johnson, "Repeat-pass dual-antenna synthetic aperture radar interferometric change-detection post-processing," *Photogrammetric Eng. and Remote Sens.*, pp. 425–429, 1998.

[88] T. Strozzi, U. Wegmuller, C. L. Werner, A. Wiesmann, and V. Spreckels, "JERS SAR interferometry for land subsidence monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 7, pp. 1702–1708, 2003.

[89] R. Touzi, A. Lopes, J. Bruniquel, and P. W. Vachon, "Coherence estimation for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 135–149, 1999.

[90] F. Gatelli, A. Monti Guarnieri, F. Parizzi, P. Pasquali, C. Prati, and F. Rocca, "The wavenumber shift in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 855–865, 1994.

[91] C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. New York, NY, USA: Springer, 1996.

[92] M. Cha, R. D. Phillips, P. J. Wolfe, and C. D. Richmond, "Two-stage change detection for synthetic aperture radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6547–6560, 2015.

[93] M. Preiss, D. A. Gray, and N. J. S. Stacy, "Detecting scene changes using synthetic aperture radar interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2041–2054, 2006.

[94] D. E. Wahl, D. A. Yocky, C. V. Jakowatz, and K. M. Simonson, "A new maximum-likelihood change estimator for two-pass SAR coherent change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2460–2469, 2016.

[95] M. Newey, G. Benitz, and S. Kogon, "A generalized likelihood ratio test for SAR CCD," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst., Comput.*, 2012, pp. 1727–1730.

[96] R. Sabry, "A new coherency formalism for change detection and phenomenology in SAR imagery: A field approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 3, pp. 458–462, 2009.

[97] J.-S. Lee, S. R. Cloude, K. P. Papathanassiou, M. R. Grunes, and I. H. Woodhouse, "Speckle filtering and coherence estimation of polarimetric SAR interferometry data for forest applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2254–2263, 2003.

[98] C. López-Martínez, X. Fabregas, and E. Pottier, "Wavelet transform-based interferometric SAR coherence estimator," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 831–834, 2005.

[99] C.-A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1441–1452, 2011.

[100] G. Ferraioli, C.-A. Deledalle, L. Denis, and F. Tupin, "PARISAR: Patch-based estimation and regularized inversion for multibaseline SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1626–1636, 2018.

[101] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, "InSAR-BM3D: A nonlocal filter for SAR interferometric phase restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3456–3467, 2018.

[102] Y. Yang, X. Cong, K. Long, Y. Luo, W. Xie, and Q. Wan, "MRF model-based joint interrupted SAR imaging and coherent change detection via variational Bayesian inference," *Signal Process.*, vol. 151, pp. 144–154, 2018.

[103] O. Dikmen and A. T. Cemgil, "Gamma Markov random fields for audio source modeling," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 589–601, 2010.

[104] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Am. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, 2017.

[105] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, 2008.

[106] A. T. Cemgil and O. Dikmen, "Conjugate gamma Markov random fields for modelling nonstationary sources," in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*, 2007, pp. 697–705.

[107] Y. Altmann, M. Pereyra, and S. McLaughlin, "Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 174–185, 2015.

[108] Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. S. Buller, S. McLaughlin, and A. Hero, "Robust spectral unmixing of sparse multispectral lidar waveforms using gamma Markov random fields," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 658–670, 2017.

[109] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Raleigh, NC, USA: SciTech, 2004.

[110] V. Cevher, M. Duarte, C. Hegde, and R. Baraniuk, "Sparse signal recovery using Markov random fields," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 257–264.

[111] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[112] O. Dikmen and A. T. Cemgil, "Inference and parameter estimation in gamma chains," Univ. of Cambridge, Tech. Rep. CUED/F-IN-FENG/TR.596, Feb. 2008.

[113] Z.-P. Liang, "A model-based method for phase unwrapping," *IEEE Trans. Med. Imag.*, vol. 15, no. 6, pp. 893–897, 1996.

[114] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2570–2583, 1992.

[115] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 48, no. 3, pp. 259–302, 1986.

[116] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2115–2118.

[117] S. R. Cloude, *Polarisation: Applications in Remote Sensing*. New York, NY, USA: Oxford Univ. Press, 2009.

[118] E. Ertin, C. D. Austin, S. Sharma, R. L. Moses, and L. C. Potter, "GOTCHA experience report: three-dimensional SAR imaging with complete circular apertures," in *Proc. SPIE*, vol. 6568, 2007, pp. 9 – 20.

[119] D. Tucker and L. C. Potter, "Polarimetric SAR despeckling with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

[120] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, 1997.

[121] I. Hajnsek, E. Pottier, and S. R. Cloude, "Inversion of surface parameters from polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 727–744, 2003.

[122] X. Huang, J. Wang, and J. Shang, "An integrated surface parameter inversion scheme over agricultural fields at early growing stages by means of C-band polarimetric RADARSAT-2 imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2510–2528, 2016.

[123] J.-S. Lee, T. L. Ainsworth, Y. Wang, and K.-S. Chen, "Polarimetric SAR speckle filtering and the extended sigma filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1150–1160, 2015.

[124] J. Chen, Y. Chen, W. An, Y. Cui, and J. Yang, "Nonlocal filtering for polarimetric SAR data: A pretest approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1744–1754, May 2011.

[125] G. Vasile, E. Trouv/'e, J.-S. Lee, and V. Buzuloiu, "Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric SAR parameters estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1609–1621, 2006.

[126] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.

[127] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshop*, 2017, pp. 1132–1140.

[128] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw.*, vol. 30, no. 11, pp. 3212–3232, 2019.

[129] M. Dahmane, S. Foucher, M. Beaulieu, F. Riendeau, Y. Bouroubi, and M. Benoit, "Object detection in Pleiades images using deep features," in *2016 IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1552–1555.

[130] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Netw.*, vol. 131, pp. 251–275, 2020.

[131] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN based image denoising," 2017, *arXiv:1710.04026*.

[132] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "SAR image despeckling through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5438–5441.

[133] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, 2017.

[134] Q. Zhang, Q. Yuan, J. Li, Z. Yang, X. Ma, H. Shen, and L. Zhang, "Learning a dilated residual network for SAR image despeckling," *Remote Sens.*, vol. 10, pp. 1–18, Feb. 2018.

[135] F. Lattari, B. Gonzalez Leon, F. Asaro, A. Rucci, C. Prati, and M. Matteucci, "Deep Learning for SAR Image Despeckling," *Remote Sens.*, vol. 11, no. 13, p. 1532, Jun. 2019.

[136] X. X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler, "Deep learning meets SAR," Jun. 2020, *arXiv:2006.10027*.

[137] L. Lin, J. Li, Q. Yuan, and H. Shen, "Polarimetric SAR image super-resolution via deep convolutional neural network," in *2019 IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3205–3208.

[138] H. Shen, L. Lin, J. Li, Q. Yuan, and L. Zhao, "A residual convolutional neural network for polarimetric SAR image super-resolution," *J. Photogrammetry Remote Sens.*, vol. 161, pp. 90–108, 2020.

[139] Z. Zhang, H. Wang, F. Xu, and Y. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, 2017.

[140] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land Use Classification in Remote Sensing Images by Convolutional Neural Networks," 2015, *arXiv:1508.00092*.

[141] F. Xu, H. Wang, and Y. Jin, "Deep learning as applied in SAR target recognition and terrain classification," *J. Radars*, vol. 6, pp. 136–148, Apr. 2017.

[142] R. Haensch and O. Hellwich, "Complex-valued convolutional neural networks for object detection in PolSAR data," in *8th Eur. Conf. Synthetic Aperture Radar*, 2010, pp. 1–4.

[143] S. Foucher, M. Beaulieu, M. Dahmane, and F. Cavayas, "Deep speckle noise filtering," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5311–5314.

[144] S. Foucher, M. Beaulieu, F. Cavayas, and M. Dahmane, "Revisiting an iterative speckle filtering technique," in *2019 IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5213–5216.

[145] X. Yang, T. Pan, W. Yang, and H. Li, "PolSAR image despeckling using trained models on single channel SAR images," in *Proc. 6th Asia-Pacific Conf. Synthetic Aperture Radar*, 2019, pp. 1–4.

[146] T. Pan, D. Peng, W. Yang, and H.-C. Li, "A filter for SAR image despeckling using pre-trained convolutional neural network model," *Remote Sens.*, vol. 11, Oct. 2019.

[147] E. Dalsasso, X. Yang, L. Denis, F. Tupin, and W. Yang, "SAR image despeckling by deep neural networks: from a pre-trained model to an end-to-end training strategy," *Remote Sens.*, vol. 12, no. 16, 2020.

[148] D. Cozzolino, L. Verdoliva, G. Scarpa, and G. Poggi, "Nonlocal CNN SAR image despeckling," *Remote Sens.*, vol. 12, no. 6, p. 1006, Mar. 2020.

[149] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–17, 2021.

[150] E. Dalsasso, L. Denis, and F. Tupin, "SAR2SAR: a self-supervised despeckling algorithm for SAR images," 2020, *arXiv:2006.15037*.

[151] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, "Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives," *IEEE Geosci. Remote Sens. Magazine*, vol. 9, no. 2, pp. 29–51, 2021.

[152] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.

[153] D. Nagar and A. Gupta, "Expectations of functions of complex Wishart matrix," *Acta Applicandae Math.*, vol. 113, pp. 265–288, Mar. 2011.

[154] J.-S. Lee, J.-H. Wen, T. L. Ainsworth, K.-S. Chen, and A. J. Chen, "Improved sigma filter for speckle filtering of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 202–213, 2009.

[155] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Int. Conf. Mach. Learn.*, ser. ICML'15, 2015, p. 448–456.

[156] X. Yang, L. Denis, F. Tupin, and W. Yang, "SAR image despeckling using pretrained convolutional neural network models," in *Proc. Joint Urban Remote Sens. Event*, May 2019, pp. 1–4.

[157] P. A. Rosen, S. Hensley, K. Wheeler, G. Sadowy, T. Miller, S. Shaffer, R. Muellerschoen, C. Jones, S. Madsen, and H. Zebker, "UAVSAR: New NASA airborne SAR system for research," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 22, no. 11, pp. 21–28, 2007.

[158] A. Lapini, T. Bianchi, F. Argenti, and L. Alparone, "Blind speckle decorrelation for SAR image despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1044–1058, 2014.

[159] A. Arienzo, F. Argenti, L. Alparone, and M. Gherardelli, "Accurate despeckling and estimation of polarimetric features by means of a spatial decorrelation of the noise in complex PolSAR data," *Remote Sens.*, vol. 12, no. 2, p. 331, 2020.

[160] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *2015 IEEE Int. Conf. Comput. Vision (ICCV)*, 2015, pp. 1026–1034.

[161] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[162] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *J. Am. Stat. Assoc.*, vol. 65, no. 332, pp. 1509–1526, 1970.

[163] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, 1996.

[164] J. J. van Zyl, H. A. Zebker, and C. Elachi, "Imaging radar polarization signatures: Theory and observation," *Radio Sci.*, vol. 22, no. 4, pp. 529–543, 1987.

[165] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3795–3809, 2009.

[166] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, ser. MM '15, 2015, pp. 689–692.

[167] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1699–1706, 2005.

[168] M. Markl, A. Frydrychowicz, S. Kozerke, M. Hope, and O. Wieben, "4D flow MRI," *J. Magn. Reson. Imaging*, vol. 36, no. 5, pp. 1015–1036, 2012.

[169] A. Ghasemmehdi and E. Agrell, "Faster recursions in sphere decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3530–3536, 2011.

[170] X.-W. Chang and T. Zhou, "MILES: MATLAB package for solving mixed integer least squares problems," *GPS Solutions*, vol. 11, pp. 289–294, 11 2007.

[171] D. Micciancio and P. Voulgaris, "A deterministic single exponential time algorithm for most lattice problems based on voronoi cell computations," *SIAM Journal on Computing*, vol. 42, no. 3, pp. 1364–1391, 2013.

[172] M. Pohst, "On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications," *SIGSAM Bull.*, vol. 15, no. 1, p. 37–44, Feb. 1981.

[173] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, no. 170, pp. 463–471, 1985.

[174] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, 2005.

[175] J. Conway and N. Sloane, "Fast quantizing and decoding and algorithms for lattice quantizers and codes," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 227–232, 1982.

[176] R. G. McKilliam, W. D. Smith, and I. V. L. Clarkson, "Linear-time nearest point algorithms for Coxeter lattices," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1015–1022, 2010.

[177] S. Sahraei and M. Gastpar, "Polynomially solvable instances of the shortest and closest vector problems with applications to compute-and-forward," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7780–7792, 2017.

[178] A. Akhlaq, R. McKilliam, R. Subramanian, and A. Pollok, "Selecting wavelengths for least squares range estimation," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5205–5216, 2016.

[179] W. Li, X. Wang, and B. Moran, "Wireless signal travel distance estimation using non-coprime wavelengths," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 27–31, 2017.

[180] D. Tucker, S. Zhao, and L. C. Potter, "Multivariate maximum likelihood estimation with the Chinese remainder theorem," *IEEE Trans. Signal Process.*, 2024, to be submitted.

[181] O. Knill, "A multivariable Chinese remainder theorem," *arXiv preprint arXiv:1206.5114*, 2012.

[182] B. Sury, "Multivariable Chinese remainder theorem," *Resonance*, vol. 20, pp. 206–216, Mar. 2015.

[183] O. Goldreich, D. Ron, and M. Sudan, "Chinese remaindering with errors," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1330–1338, 2000.

[184] X.-G. Xia and G. Wang, "Phase unwrapping and a robust Chinese remainder theorem," *IEEE Signal Process. Lett.*, vol. 14, no. 4, pp. 247–250, 2007.

[185] L. Xiao, X.-G. Xia, and W. Wang, "Multi-stage robust Chinese remainder theorem," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4772–4785, 2014.

[186] G. Maze, J. Rosenthal, and U. Wagner, "Natural density of rectangular unimodular integer matrices," *Linear Algebra Appl.*, vol. 434, no. 5, pp. 1319–1324, 2011.

[187] W. Li, X. Wang, X. Wang, and B. Moran, "Resolving phase measurement ambiguity in presence of coloured noise," *Electron. Lett.*, vol. 49, pp. 1188–1190, 2013.

[188] D. Prasanna, C. Sriram, and C. R. Murthy, "On the identifiability of sparse vectors from modulo compressed sensing measurements," *IEEE Signal Process. Lett.*, vol. 28, pp. 131–134, 2021.

[189] D. Prasanna, C. R. Murthy, and C. Sriram, "On the application of modulo-ADCs for compressed sensing," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 852–856.

[190] A. Benedetti, "Phase unwrapping with multiple wavelengths on the flat torus," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, 2020, pp. 1301–1304.

[191] D. Micciancio and S. Goldwasser, *Complexity of Lattice Problems: A Cryptographic Perspective*. Boston, MA, USA: Kluwer, 2002.

[192] M. R. Bremner, *Lattice Basis Reduction: An Introduction to the LLL Algorithm and Its Applications*. Boca Raton, USA: CRC Press, 2011.

[193] J. Martinet, *Perfect Lattices in Euclidean Spaces*. Berlin, Germany: Springer-Verlag, 2003.

[194] A. Schrijver, *Theory of Linear and Integer Programming*. Chichester, U.K.: Wiley, 1998.

[195] C. Ding, D. Pei, and A. Salomaa, *Chinese Remainder Theorem*. Singapore: World Scientific, 1996.

[196] N. Koblitz, *A Course in Number Theory and Cryptography*, 2nd ed. New York, NY, USA: Springer-Verlag, 1994.

[197] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, pp. 1–13, 1986.

[198] J. H. Silverman, J. Pipher, and J. Hoffstein, *An Introduction to Mathematical Cryptography*. New York, NY, USA: Springer, 2008.

[199] H. Edelsbrunner, J. O'Rourke, and R. Seidel, "Constructing arrangements of lines and hyperplanes with applications," *SIAM Journal on Computing*, vol. 15, no. 2, pp. 341–363, 1986.

[200] N. Sleumer, "Output-sensitive cell enumeration in hyperplane arrangements," in *Proc. 6th Scand. Workshop Algorithm Theory*, vol. 1432, 1998, pp. 300–309.

[201] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1212–1215, 2001.

[202] M. Pohst, "A modification of the LLL reduction algorithm." *J. Symbolic Comput.*, vol. 4, pp. 123–127, 1987.

[203] C. C. Sims, *Computation with Finitely Presented Groups*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[204] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, p. 515–534, 1982.

[205] H. Cohen, *A Course in Computational Algebraic Number Theory.* Berlin, Germany: Springer-Verlag, 1993.

[206] X.-W. Chang and G. H. Golub, "Solving ellipsoid-constrained integer least squares problems," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1071–1089, 2010.

[207] X.-W. Chang, "MILES: MATLAB package for solving mixed integer least squares problems," http://www.cs.mcgill.ca/~chang/software/MILES.php, School of Computer Science, McGill University, 2006, last updated: Dec., 2022. [Online]. Available: http://www.cs.mcgill.ca/~chang/software/MILES.php

[208] C. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Prog.*, vol. 66, p. 181–199, 1994.

[209] A. Pruitt, A. Rich, Y. Liu, N. Jin, L. Potter, M. Tong, S. Rajpal, O. Simonetti, and R. Ahmad, "Fully self-gated whole-heart 4D flow imaging from a 5-minute scan," *Magn. Reson. Med.*, vol. 85, no. 3, pp. 1222–1236, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28491

[210] S. Sahraei and M. Gastpar, "Polynomially solvable instances of the shortest and closest vector problems with applications to compute-and-forward," *IEEE Trans. Inf. Theory*, vol. 63, no. 12, pp. 7780–7792, 2017.

[211] F. Athley, "Optimization of element positions for direction finding with sparse arrays," in *Proc. 11th IEEE Signal Process. Workshop Statistical Signal Process.*, 2001, pp. 516–519.

[212] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[213] T. Ballal and C. J. Bleakley, "Phase-difference ambiguity resolution for a single-frequency signal," *IEEE Signal Process. Lett.*, vol. 15, pp. 853–856, 2008.

[214] H. Chen, T. Ballal, and T. Y. Al-Naffouri, "DOA estimation with non-uniform linear arrays: A phase-difference projection approach," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2435–2439, 2021.

[215] Y. Hu and S. Gannot, "Closed-form single source direction-of-arrival estimator using first-order relative harmonic coefficients," in *IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 726–730.

[216] D. Tucker, S. Zhao, R. Ahmad, and L. C. Potter, "Alias-free arrays," *IEEE Signal Process. Lett.*, vol. 29, pp. 2457–2461, 2022.

[217] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, 2005.

[218] Y. Nasser, S. Aubert, and F. Nouvel, "A simplified hard output sphere decoder for large MIMO systems with the use of efficient search center and reduced domain neighborhood study," *EURASIP Journal on Wireless Communications and Networking*, 10 2015.

[219] K. M. Johnson and M. Markl, "Improved SNR in phase contrast velocimetry with five-point balanced flow encoding," *Magn. Reson. Med.*, vol. 63, no. 2, pp. 349–355, 2010.

[220] C. Binter, V. Knobloch, R. Manka, A. Sigfridsson, and S. Kozerke, "Bayesian multipoint velocity encoding for concurrent flow and turbulence mapping," *Magn. Reson. Medicine*, vol. 69, no. 5, pp. 1337–1345, 2013.

[221] D. A. Holdsworth, "Angle of arrival estimation for all-sky interferometric meteor radar systems," *Radio Sci.*, vol. 40, no. 6, pp. 1–8, 2005.

[222] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, 1990.

[223] F. Athley, "Threshold region performance of maximum likelihood direction of arrival estimators," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1359–1373, 2005.

[224] H. Carrillo, A. Osses, S. Uribe, and C. Bertoglio, "Optimal dual-VENC unwrapping in phase-contrast MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1263–1270, 2019.

[225] M. Loecher and D. B. Ennis, "Velocity reconstruction with nonconvex optimization for low-velocity-encoding phase-contrast MRI," *Magn. Reson. Med.*, vol. 80, no. 1, pp. 42–52, 2018.

[226] H. Gazzah and S. Marcos, "Cramer-Rao bounds for antenna array design," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 336–345, 2006.

[227] H. Gazzah and K. Abed-Meraim, "Optimum ambiguity-free directional and omnidirectional planar antenna arrays for DOA estimation," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3942–3953, 2009.

[228] A. Ghani, F. Keyvani, and S. H. Sedighy, "Antenna array placement on limited bound for isotropic and optimal direction-of-arrival estimation," *IET Signal Process.*, vol. 12, no. 3, pp. 277–283, 2018.

[229] K.-C. Tan, S. S. Goh, and E.-C. Tan, "A study of the rank-ambiguity issues in direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 880–887, 1996.

[230] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2002.

[231] B. Friedlander, "On the Cramér-Rao bound for sparse linear arrays," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2020, pp. 1255–1259.

[232] K.-C. Tan, G.-L. Oh, and M. H. Er, "A study of the uniqueness of steering vectors in array processing," *Signal Process.*, vol. 34, no. 3, pp. 245–256, 1993.

[233] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, 1968.

[234] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, 2010.

[235] P. P. Vaidyanathan and P. Pal, "Sparse sensing with coprime arrays," in *44th Asilomar Conf. Signals, Systems and Computers*, 2010, pp. 1405–1409.

[236] H. Chen, Y. Wang, and S. Wan, "Performance improvement of estimation direction-of-arrival via array geometry arrangement," in *Proc. IEEE Int. Symp. Antennas and Propag. Soc.*, vol. 3, 1999, pp. 1600–1603.

[237] M. D. Zoltowski and C. P. Mathews, "Real-time frequency and 2-D angle estimation with sub-Nyquist spatio-temporal sampling," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2781–2794, 1994.

[238] T. Ballal and C. J. Bleakley, "DOA estimation of multiple sparse sources using three widely-spaced sensors," in *17th Eur. Signal Process. Conf.*, 2009, pp. 1978–1982.

[239] N. J. Pelc, R. J. Herfkens, A. Shimakawa, and D. R. Enzmann, "Phase contrast cine magnetic resonance imaging," *Magn. Reson. Q.*, vol. 7, no. 4, pp. 229–254, 1991.

[240] S. Zhao, R. Ahmad, and L. C. Potter, "Maximizing unambiguous velocity range in phase-contrast MRI with multipoint encoding," in *IEEE Int. Symp. Biomed. Imag.*, 2022, pp. 1–5.

[241] O. Lange and B. Yang, "Optimization of array geometry for direction-of-arrival estimation using a priori information," *Adv. Radio Sci.*, vol. 8, pp. 87–94, 2010.

[242] C. Gauss, *Disquisitiones Arithmeticae*, ser. Translated by A.A. Clarke. New Haven, CT: Yale University Press, 1966.

[243] A. Barvinok, "Math 669: Combinatorics, Geometry, and Complexity of Integer Points," math.lsa.umich.edu/~barvinok/c20.html, accessed 2021-10-11.

[244] J. Chen, Y. Feng, Y. Liu, and W. Wu, "On the probability of generating a primitive matrix," *J. Syst. Sci. Complex.*, 2024.

[245] I. Reiner, "Unimodular complements," *Amer. Math. Monthly*, vol. 63, no. 4, pp. 246–247, 1956.

[246] A. Kintz and I. J. Gupta, "Airborne antenna array calibration with signals of opportunity," in *IEEE Antennas Propag. Soc. Int. Symp. (APSURSI)*, 2013, pp. 1264–1265.

[247] J. L. Volakis, A. J. O'Brien, and C.-C. Chen, "Small and adaptive antennas and arrays for GNSS applications," *Proc. IEEE*, vol. 104, no. 6, pp. 1221–1232, 2016.

[248] H. Liao and A. Shamim, "A low-profile and high-aperture-efficiency hexagonal circularly polarized microstrip antenna array," *IEEE Antennas Wireless Propag. Lett.*, vol. 21, no. 3, pp. 615–619, 2022.

[249] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1003–1010, 2007.

[250] J. Gebbie, M. Siderius, P. L. Nielsen, and J. Miller, "Passive localization of noise-producing targets using a compact volumetric array," *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 80–89, 2014.

[251] D. Kastinen and J. Kero, "Probabilistic analysis of ambiguities in radar echo direction of arrival from meteors," *Atmos. Meas. Tech.*, vol. 13, no. 12, pp. 6813–6835, 2020.

[252] D. Avis, H. Miyata, and S. Moriyama, "Families of polytopal digraphs that do not satisfy the shelling property," *Computational Geometry*, vol. 46, no. 3, pp. 382–393, 2013.

[253] J. E. Goodman, J. O'Rourke, and C. D. Tóth, *Handbook of Discrete and Computational Geometry*, 3rd ed. Boca Raton, USA: CRC Press, 2017.

[254] K. Kreutz-Delgado, "The Complex Gradient Operator and the CR-Calculus," Jun. 2009, *arXiv:0906.4835*.