A STUDY ON TRANSMIT BEAMFORMING AND NOISY AUTOREGRESSIVE MODELING PROBLEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÖMER ÇAYIR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JANUARY 2023

Approval of the thesis:

A STUDY ON TRANSMIT BEAMFORMING AND NOISY AUTOREGRESSIVE MODELING PROBLEMS

submitted by **ÖMER ÇAYIR** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** _____

Prof. Dr. Çağatay Candan
Supervisor, **Electrical and Electronics Eng. Dept., METU** _____

Examining Committee Members:

Prof. Dr. Temel Engin Tuncer
Electrical and Electronics Eng. Dept., METU _____

Prof. Dr. Çağatay Candan
Electrical and Electronics Eng. Dept., METU _____

Prof. Dr. Orhan Arıkan
Electrical and Electronics Eng. Dept., Bilkent Univ. _____

Prof. Dr. Umut Orguner
Electrical and Electronics Eng. Dept., METU _____

Prof. Dr. Süleyman Serdar Kozat
Electrical and Electronics Eng. Dept., Bilkent Univ. _____

Date:06.01.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Ömer Çayır

Signature :

ABSTRACT

A STUDY ON TRANSMIT BEAMFORMING AND NOISY AUTOREGRESSIVE MODELING PROBLEMS

Çayır, Ömer
Ph.D., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Çağatay Candan

January 2023, 122 pages

This thesis presents algorithms for two problems in statistical signal processing. The first problem is the transmit beamformer design under the peak-to-average power ratio (PAPR) constraint. With the aim of establishing a trade-off between the power efficiency (maximizing the average transmitted power in the main lobe) and other metrics, such as the fluctuation of power in the main lobe and the peak-sidelobe level, how the PAPR constraint affects the design problem is examined. In general, unimodular weights, which have a constant magnitude and therefore ensure the lowest possible PAPR value, are used to maximize the average transmitted power at the expense of other performance metrics. It is empirically shown that even a minor relaxation of the design problem from the lowest PAPR condition leads to a significant improvement in performance metrics at a negligible loss in power efficiency. A solution based on the alternating direction method of multipliers (ADMM) is provided to achieve the trade-off between the performance metrics. Moreover, a consensus ADMM-based solution is presented for the equivalent problem in consensus form. The proposed solutions can be used for both narrowband and wideband beamformers. The second problem is the maximum likelihood autoregressive (AR) model parameter estimation

from the independent snapshots observed under additive white Gaussian noise. In addition to the AR model parameters, the measurement noise variance is included among the unknowns of the problem to develop a general solution covering several special cases, such as the case of known noise variance, noise-free snapshots, and the single snapshot operation. The presented solution is based on the expectation-maximization method, which is formulated by assigning the noise-free snapshots as the missing data. An approximate version of the suggested method, at a significantly reduced computational load with virtually no loss of performance, is also provided.

Keywords: Transmit Beamforming, Peak-to-Average Power Ratio, PAPR, Alternating Direction Method of Multipliers, ADMM, Autoregressive Process, Autoregressive Model Parameter Estimation, Multiple Snapshots, Expectation-Maximization

ÖZ

GÖNDERME HÜZMESİ OLUŞTURMA VE GÜRÜLTÜLÜ ÖZBAĞLANIMLI MODELLEME PROBLEMLERİ ÜZERİNE BİR ÇALIŞMA

Çayır, Ömer
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Çağatay Candan

Ocak 2023 , 122 sayfa

Bu tez, istatistiksel sinyal işlemede iki problem için algoritmalar sunmaktadır. İlk problem, tepe gücü/ortalama güç oranı (PAPR) kısıtı altında gönderme hüzmesi oluşturucu tasarımıdır. PAPR kısıtının tasarım problemini nasıl etkilediği, güç verimliliği (ana kulakta gönderilen ortalama gücü enbüyütme) ve, ana kulaktaki güç dalgalanması ve tepe-yan kulak düzeyi gibi, diğer ölçevler arasında bir ödünleşim oluşturmak amacıyla incelenmektedir. Genelde, sabit bir büyüklüğe sahip olan ve bu nedenle mümkün olan en düşük PAPR değerini sağlayan özdeş büyüklükteki ağırlıklar diğer başarım ölçevleri pahasına gönderilen ortalama gücü enbüyütmek için kullanılmaktadır. Tasarım probleminin en düşük PAPR koşulundan küçük bir gevşemesinin bile güç verimliliğinde göz ardı edilebilir bir kayıpla başarım ölçevlerinde önemli bir iyileşmeye yol açtığı deneysel olarak gösterilmektedir. Almaşık yön çarpanları yöntemine (ADMM) dayanan çözüm, ölçevler arasındaki ödünleşimi elde etmek için sağlanmaktadır. Ayrıca, mutabakat ADMM tabanlı çözüm mutabakat formundaki eşdeğer problem için sunulmaktadır. Önerilen çözümler hem dar bant hem de geniş bant hüzme oluşturucular için kullanılabilmektedir. İkinci problem, toplanır beyaz

Gauss gürültüsü altında gözlemlenen bağımsız anlık gözlemlerden en büyük olabilirlik özbağlanımlı (AR) model parametre kestirimidir. AR model parametrelerine ek olarak, ölçüm gürültüsü varyansı, bilinen gürültü varyansı durumu, gürültüsüz anlık gözlemler ve tek anlık gözlem çalışması gibi, birkaç özel durumu kapsayan genel bir çözüm geliştirmek için problemin bilinmeyenleri arasına dahil edilmektedir. Sunulan çözüm, gürültüsüz anlık gözlemlerin eksik veri olarak atanmasıyla formüle edilen beklenti-enbüyütme yöntemine dayanmaktadır. Önerilen yöntemin neredeyse hiç başarım kaybı olmadan önemli ölçüde azaltılmış bir hesaplama yükündeki bir yaklaşık sürümü de sağlanmaktadır.

Anahtar Kelimeler: Gönderme Hüzmesi Oluşturma, Tepe Gücü/Ortalama Güç Oranı, PAPR, Almaşık Yön Çarpanları Yöntemi, ADMM, Özbağlanımlı Süreç, Özbağlanımlı Model Parametresi Kestirimi, Çoklu Anlık Gözlemler, Beklenti-Enbüyütme

To my loving family

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Prof. Dr. Çağatay Candan, for his excellent guidance, encouragement, and inspiration granted during the course of this dissertation.

I would like to extend my gratitude to my committee members, Prof. Dr. T. Engin Tuncer, Prof. Dr. Orhan Arıkan, Prof. Dr. Umut Orguner, and Prof. Dr. S. Serdar Kozat, for being on my committee and giving me valuable feedback concerning this dissertation. Special thanks to the last two of them for being on my thesis monitoring committee and for their great support and enlightening discussions throughout my dissertation process.

I am thankful for the excellent collaboration with Şafak Bilgi Akdemir on the frequency estimation problem.

I am sincerely grateful to all my teachers, especially my primary school teacher Ayşe İnci, and to all faculty whom I met, specifically Assist. Prof. Dr. A. Semih Bingöl, Prof. Dr. Adnan Köksal, Dr. Arzu Koç, Assist. Prof. Dr. Mehmet Demirer, Prof. Dr. Mehmet Şafak, Assoc. Prof. Dr. Melek Diker Yücel, Prof. Dr. Mübeccel Demirekler, Prof. Dr. Yalçın Tanık, and Prof. Dr. Zafer Ünver.

I would like to thank the Department of EEE, Hacettepe University, for providing me with a decent engineering education. Special thanks to my graduation project supervisor, Prof. Dr. Cenk Toker. My deepest gratitude to Dr. Işıltan Sayın, Dr. Sevda Özdemir, and Assist. Prof. Dr. Şölen Kumbay Yıldız, who became my family members, much more than the research assistants during my undergraduate years. I would like to thank all staff, especially Bilal Ünver, Bünyamin Hastürk, and Mehmet Gürbüz, for their help in any way.

I would like to thank all members of the Department of EEE, METU, for giving me an outstanding experience over a decade. My deepest thanks to my friends, Ali Bulut

Üçüncü, Erdal Epçaçan, Oktay Sipahigil, Ömer Melih Gül, and Özlem Tuğfe Demir, for their joyful talks and all the help during my research assistantship.

I would like to acknowledge TÜBİTAK for the financial support via 2211/E National Ph.D. Scholarship Program.

I am deeply grateful to my parents, Elmas and Hüseyin, my brother Kadir and sister-in-law Sebila, and my nephews, Miray and Simay, for supporting and motivating me through my academic life.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xii

LIST OF TABLES xv

LIST OF FIGURES xvii

LIST OF ALGORITHMS xix

LIST OF ABBREVIATIONS xx

NOTATION xxi

CHAPTERS

1 INTRODUCTION 1

2 TRANSMIT BEAMFORMING PROBLEM WITH PAPR CONSTRAINT . 5

2.1 Related Works 5

2.2 PAPR Constraint 6

2.3 A Problem With PAPR Constraint 7

2.4 System Model and Problem Formulation 10

3 TRANSMIT BEAMFORMING PROBLEM WITH PAPR CONSTRAINT: ADMM-BASED SOLUTION 13

3.1 Related Works and Contributions 13
3.2 Proposed Method 15
3.3 Wideband Beamforming 22
3.4 Algorithm Convergence 24
3.5 Computational Complexity Considerations 24
3.6 Numerical Results 25
3.6.1 Effect of the Fixed Total Power 25
3.6.2 Main Lobe Average Power Versus PAPR 26
3.6.3 Radar Waveform Design With Unimodular Sequence 30
3.6.4 Wideband Transmit Beamforming 32
3.7 Conclusion 36
4 TRANSMIT BEAMFORMING PROBLEM WITH PAPR CONSTRAINT: CONSENSUS ADMM-BASED SOLUTION 37
4.1 Consensus ADMM 37
4.2 Proposed Method 39
4.3 An Extension of the Proposed Method to Wideband Beamforming . . 45
4.4 Computational Complexity Considerations 50
4.5 Numerical Results 50
4.5.1 Effect of the Fixed Total Power 51
4.5.2 Narrowband Transmit Beamforming 52
4.5.3 Wideband Transmit Beamforming 55
4.6 Conclusion 59
5 MAXIMUM LIKELIHOOD ESTIMATION OF NOISY AUTOREGRESSIVE MODEL PARAMETER WITH INDEPENDENT SNAPSHOTS . . . 61

5.1 Related Works and Contributions 61

5.2 Noisy AR Parameter Estimation Problem 64

5.2.1 Expectation-Maximization Formulation for the Noisy AR Parameter Estimation With Multiple Snapshots 65

5.2.2 AR Parameter Estimation Problem With Multiple Noise-free Snapshots 69

5.2.2.1 The First Stage: Weighted Forward-Backward Prediction With the Increased Number of Snapshots 69

5.2.2.2 The Second Stage: Maximizing Likelihood Around the First Stage Estimate 70

5.2.3 An Efficient Implementation for the Suggested Solution 73

5.2.3.1 Efficient Calculation of the Mean Vector 75

5.2.3.2 Approximating the Error Covariance Matrix 75

5.2.3.3 Efficient Calculation of Trace Term 77

5.2.3.4 Efficient Calculation of Optimal Perturbation Around the First Stage Estimate 79

5.2.4 Computational Complexity Considerations 84

5.3 Initialization of the Suggested Algorithm 85

5.4 Numerical Results 88

5.5 Conclusion 103

6 CONCLUSIONS 105

REFERENCES 107

APPENDICES

A PERFORMANCE METRICS FOR TRANSMIT BEAMFORMING 117

CURRICULUM VITAE 121

LIST OF TABLES

TABLES

Table 3.1 Performance metrics of 10 realizations for the narrowband transmit beamforming with 32 elements ULA and PAPR $\leq \sigma$ by using (3.8) with $P = 32$ and (3.9). 27

Table 3.2 Performance metrics for the narrowband transmit beamforming with 32 elements ULA and PAPR $\leq \sigma$. 29

Table 3.3 Comparison of the peak-to-peak power swing in the passband and PSL values for radar waveform design. 31

Table 3.4 Performance metrics for the wideband transmit beamforming with 144 elements URA and PAPR $\leq \sigma$. 36

Table 4.1 Performance metrics of 10 realizations for the narrowband transmit beamforming with 32 elements ULA and PAPR $\leq \sigma$ by using (4.15) with $P = 32$ and (4.16). 52

Table 4.2 Comparison of the mean and variance of PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs for the narrowband transmit beamforming experiment with 100 MC runs. 53

Table 4.3 Comparison of the mean and variance of PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs for the wideband transmit beamforming experiment with 100 MC runs. 55

Table 5.1 Hellinger distance comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the real-valued AR(4) process given in Experiment 2. 93

Table 5.2 Total MSE comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the real-valued AR(4) process given in Experiment 2. 94

Table 5.3 Hellinger distance comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the complex-valued AR(4) process given in Experiment 2. 95

Table 5.4 Total MSE comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the complex-valued AR(4) process given in Experiment 2. 96

Table 5.5 Hellinger distance comparison for the single snapshot real-valued AR(4) process given in Experiment 2. 98

Table 5.6 Total MSE comparison for the single snapshot real-valued AR(4) process given in Experiment 2. 99

Table 5.7 Comparison of the number of unstable system estimates out of 100 runs for the single snapshot real-valued AR(4) process given in Experiment 2. 100

Table 5.8 The average computational time (in seconds) comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots for the real-valued AR(4) process in Experiment 2. 101

Table 5.9 The average computational time (in seconds) comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots for the complex-valued AR(4) process in Experiment 2. 102

LIST OF FIGURES

FIGURES

Figure 3.1 Augmented Lagrangian function (3.4) versus iteration number for the narrowband transmit beamforming with 32 elements ULA. 26

Figure 3.2 Main lobe average power versus PAPR for the narrowband transmit beamforming with 32 elements ULA. 28

Figure 3.3 Beampattern versus PAPR for the narrowband transmit beamforming with 32 elements ULA. 30

Figure 3.4 Spectrum comparison for radar waveform design. 31

Figure 3.5 Autocorrelation function comparison for radar waveform design. . 32

Figure 3.6 Spectrum comparison with ANSLM for radar waveform design. . . 32

Figure 3.7 Autocorrelation function comparison with ANSLM for radar waveform design. 33

Figure 3.8 Beampatterns on cutting-planes for the wideband transmit beamforming with 144 elements URA and different PAPR values. 34

Figure 3.9 Two-dimensional viewed beampatterns on cutting-planes for the wideband transmit beamforming with 144 elements URA and different PAPR values. 35

Figure 4.1 Augmented Lagrangian function (4.11) versus iteration number for the narrowband transmit beamforming with 32 elements ULA. 51

Figure 4.2 The PSL-$P^{\mathrm{m}}_{\text{p-p}}$ pairs for the narrowband transmit beamforming experiment with 100 MC runs. 53

Figure 4.3 Histogram of optimal values of (2.10) for the narrowband transmit beamforming experiment with 100 MC runs. 54

Figure 4.4 Beampattern comparison for the one where PAPR-cADMM gives its largest PSL and smallest $P^{\mathrm{m}}_{\text{p-p}}$ values out of 100 MC runs. 54

Figure 4.5 The PSL-$P^{\mathrm{m}}_{\text{p-p}}$ pairs for the wideband transmit beamforming experiment with 100 MC runs. 56

Figure 4.6 Histogram of optimal values of (3.24) for the wideband transmit beamforming experiment with 100 MC runs. 56

Figure 4.7 Beampattern comparison for the one where PAPR-cADMM-wb gives its largest PSL value out of 100 MC runs: (a) PAPR-ADMM and (b) PAPR-cADMM-wb. 57

Figure 4.8 Beampattern comparison for the one where PAPR-cADMM-wb gives its smallest $P^{\mathrm{m}}_{\text{p-p}}$ value out of 100 MC runs: (a) PAPR-ADMM and (b) PAPR-cADMM-wb. 58

Figure 5.1 Signal model block diagram for the noisy AR parameter estimation problem. 64

Figure 5.2 Experiment 1 - Multiple noise-free snapshots: Likelihood value comparison. 90

Figure 5.3 Experiment 2 - Hellinger distance and total MSE comparisons of the proposed method at different numbers of snapshots for the real-valued AR(4). 91

Figure 5.4 Experiment 2 - Hellinger distance and total MSE comparisons of the proposed method at different numbers of snapshots for the complex-valued AR(4). 92

LIST OF ALGORITHMS

ALGORITHMS

Algorithm 2.1 Computation of the nearest vector having low PAPR. 8

Algorithm 3.1 Proposed method PAPR-ADMM for the transmit beamformer design under a PAPR constraint. 22

Algorithm 4.1 Proposed method PAPR-cADMM for the narrowband transmit beamformer design under a PAPR constraint. 44

Algorithm 4.2 Proposed method PAPR-cADMM-wb for the wideband transmit beamformer design under a PAPR constraint. 49

Algorithm 5.1 Suggested method for the AR parameter estimation from multiple noise-free snapshots. 74

Algorithm 5.2 The disturbance smoother. 76

Algorithm 5.3 Efficient calculation of the optimal perturbation around $\mathbf{a}_{\mathrm{FS}}$. . . 82

Algorithm 5.4 Proposed AR parameter estimation method. 83

Algorithm 5.5 Expectation step of EM for the proposed AR parameter estimation method given in Algorithm 5.4. 83

Algorithm 5.6 Maximization step of EM for the proposed AR parameter estimation method given in Algorithm 5.4. 84

Algorithm 5.7 Initialization of the proposed AR parameter estimation method. 88

LIST OF ABBREVIATIONS

ADMM	alternating direction method of multipliers
ACRB	asymptotic Cramér-Rao bound
ALF	augmented Lagrangian function
AR	autoregressive
ARMA	autoregressive moving average
CRB	Cramér-Rao bound
DFT	discrete Fourier transform
EM	expectation-maximization
FB	forward-backward
GS	Gohberg-Semencul
i.i.d.	independent and identically distributed
IIR	infinite impulse response
ISL	integrated sidelobe level
LPNN	Lagrange programming neural network
LS	least-squares
MSE	mean square error
MC	Monte Carlo
PSL	peak sidelobe level
PAPR	peak-to-average power ratio
SNR	signal-to-noise ratio
ULA	uniform linear array
URA	uniform rectangular array
wFB	weighted forward-backward
YW	Yule-Walker

NOTATION

a	a scalar (italic lowercase letter)
$\mathbf{a}$	a column vector (boldface lowercase letter)
$\mathbf{A}$	a matrix (boldface uppercase letter)
$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{R}_+$	set of nonnegative real numbers, *i.e.*, $\mathbb{R}_+ = \{a \vert a \in \mathbb{R}, a \geq 0\}$
$\mathbb{R}_{++}$	set of positive real numbers, *i.e.*, $\mathbb{R}_{++} = \{a \vert a \in \mathbb{R}, a > 0\}$
$\mathbb{R}^N$	set of the real-valued N-dimensional column vectors
$\mathbb{C}^N$	set of the complex-valued N-dimensional column vectors
$\mathbb{C}^{N \times L}$	set of the complex-valued $N \times L$ dimensional matrices
$\mathsf{Re}\{\cdot\}$	real part of a complex-valued argument
$\mathsf{Im}\{\cdot\}$	imaginary part of a complex-valued argument
$(\cdot)^*$	conjugate operator
$(\cdot)^{\mathsf{T}}$	transpose operator
$(\cdot)^{\mathsf{H}}$	Hermitian (conjugate transpose) operator
$(\cdot)^{-1}$	inverse operator
$\odot$	Hadamard (element-wise) product
$\Vert \cdot \Vert_2$	ℓ_2 norm (Euclidean norm)
$\mathrm{tr}\,(\cdot)$	trace operator
$\vert \cdot \vert$	absolute value operator
$\det(\cdot)$	determinant operator
$\angle(\cdot)$	phase angle operator (in radians)
$\mathrm{card}\,(\cdot)$	cardinality operator

v_n	the nth entry of vector $\mathbf{v}$
$v_{\ell,n}$	the nth entry of vector $\mathbf{v}_\ell$
$[\mathbf{M}]_{ij}$	the ith row and jth column entry of matrix $\mathbf{M}$
$\mathbf{v}_{\ell,n:n+P}$	vector $[\, v_{\ell,n} \;\; v_{\ell,n+1} \;\; \ldots \;\; v_{\ell,n+P} \,]^{\mathsf{T}}$ for a positive integer P
$\mathbf{v}_{\ell,n:-1:n-P}$	vector $[\, v_{\ell,n} \;\; v_{\ell,n-1} \;\; \ldots \;\; v_{\ell,n-P} \,]^{\mathsf{T}}$ for a positive integer P
$\mathbf{0}_N$	N-dimensional column vector with all entries being zero
$\mathbf{0}_{N\times N}$	$N \times N$ dimensional matrix with all entries being zero
$\mathbf{I}_N$	$N \times N$ identity matrix
j	$\sqrt{-1}$
$\mathbf{v}^{(k)}$	$\mathbf{v}$ at the iteration number k
$\mathcal{CN}(0,\sigma^2)$	density of the zero-mean complex-valued (circular symmetric) white Gaussian noise with variance σ^2
$\mathsf{f}(\cdot)$	probability density function accounts for the continuous random variables
$\mathsf{E}\{\cdot\}$	expectation operator

CHAPTER 1

INTRODUCTION

This thesis studies two problems of interest in statistical signal processing. The first problem is an array signal processing problem, and the second one is a spectrum estimation problem.

In a conventional narrowband transmit beamformer, each antenna element transmits the same low-pass equivalent signal scaled by a chosen complex-valued weight. The problem of transmit beampattern design is to determine the weights such that several requirements, such as flat-top beampattern, low sidelobes, power efficiency, etc., are satisfied. Typically, unimodular weights (weights with a constant magnitude) are used in transmit beamforming to maximize the transmission power. Power maximization is important to extend the instrumented range of the sensor, which is proportional to $(\text{average power} \times \text{aperture})^{1/4}$ [1] and also to enhance the signal-to-noise ratio (SNR) affecting the accuracy of estimation operations conducted by the sensor. In addition to the transmit power maximization, a beampattern with a flat main lobe and low sidelobes is highly desirable for the reliability of the sensing system, [1]. Considerations on the sidelobe level, such as the integrated sidelobe level and the peak sidelobe level, are main considerations for *receive* beamforming systems for which several efficient methods exist for their optimization [2, 3]. Unfortunately, the power maximization requirement, which is unique to the *transmit* beamforming application, conflicts with other requirements (sidelobe suppression, flat main lobe), and an engineering trade-off has to be made in the construction of transmit beamformers. In Chapters 2 to 4, we examine the transmit beamforming problem under a peak-to-average power ratio (PAPR) constraint to enable such a trade-off for the design of flat-top beampatterns with low sidelobes at high power efficiency.

In the second part of the thesis, a problem in autoregressive (AR) modeling is studied. The AR modeling of random signals is being used in many areas associated with statistical signal processing [4], such as radar signal processing, speech processing, and biomedical signal processing [5–8]. The richness of application venues for AR models can be attributed to their success in representation and also to the availability of efficient methods for model parameter estimation. In Chapter 5, we consider the parameter estimation of AR processes observed under white noise. Our main goal is to extend the maximum likelihood like AR model parameter estimator developed for a single noise-free snapshot in [9] to the operation with multiple snapshots corrupted by white noise. A computationally efficient version of the suggested method is also presented.

This thesis concludes the discussion and the contributions of the developed methods. The thesis contents are associated with the following papers either published or submitted.

- **Chapters 2 and 3:**

  Ö. Çayır and Ç. Candan, "Transmit beamformer design with a PAPR constraint to trade-off between beampattern shape and power efficiency," *Digit. Signal Process.*, vol. 99, p. 102674, Apr. 2020, doi: 10.1016/j.dsp.2020.102674.

  Software:

  - C++ code, 2022, doi: 10.24433/CO.4042158. v1.
  - MATLAB code, 2020, doi: 10.24433/CO.7886035.v1.
  - Python code, 2022, doi: 10.24433/CO.2603813.v1.
  - R code, 2022, doi: 10.24433/CO.1947438.v1.

- **Chapter 5:**

  Ö. Çayır and Ç. Candan, "Maximum likelihood autoregressive model parameter estimation with noise corrupted independent snapshots," *Signal Process.*, vol. 186, p. 108118, Sep. 2021, doi: 10.1016/j.sigpro.2021.108118.

  Software:

  - MATLAB code, 2021, doi: 10.24433/CO.7257794.v1.

- **Other studies not included in the thesis document:**

  Ö. Çayır, Ş. Bilgi Akdemir, and Ç. Candan, "Frequency estimation of a complex exponential under signal intermittency and noise," submitted to *Signal Process.*, Jul. 2022.

  Software:

  - MATLAB code, 2022, doi: 10.24433/CO.8142587.v1.

CHAPTER 2

TRANSMIT BEAMFORMING PROBLEM WITH PAPR CONSTRAINT

2.1 Related Works

The problem of transmit beamforming under a peak-to-average power ratio (PAPR) constraint has been studied for the MIMO (multiple-input and multiple-output) systems in the radar signal processing literature [10–15]. For instance, in [11], the waveform diversity feature allowing each antenna element to transmit independent waveforms is taken into account for the solution of transmit beamforming problem under PAPR constraint. Several methods in the literature, such as [10, 11, 13, 16], ignore the PAPR constraint at the initial design stage and project the solution obtained without the PAPR constraint onto the set of vectors satisfying the constraint via the operation given in [17]. These methods have been shown to provide a good performance in spite of the decoupling of the problem into two stages. Different from these methods, the sequence design problem under a PAPR constraint is transformed to an unconstrained problem which can be solved via a gradient-based numerical search in [18].

The mathematical formulation of the waveform design for active sensing (transmit code design) is very similar to the transmit beamforming problem. As in transmit beamforming, the waveform energy is known to be maximized with the choice of unimodular sequences for the code design problem. Unimodular sequences also maximize the achievable sensitivity by maximizing the energy incident upon and reflected from the targets [19]. Therefore, the design of periodic or aperiodic unimodular sequences with low autocorrelation sidelobes is a major goal for active sensing systems, see [10, 20–25]. Since the problem of low autocorrelation sidelobes is directly related to the flatness of the spectral shape of the signal, the spectral shaping problem with

unimodular sequences has been examined in several works including [21, 23, 26, 27]. In [26], the SHAPE algorithm is presented to design the sequences satisfying the temporal envelope and spectral constraints by introducing auxiliary variables. In [21], the Lagrange programming neural network (LPNN) is used to design unimodular sequences satisfying the spectral constraints. The ANSLM described in [27] minimizes the ratio of the highest stopband level to the lowest passband level at the cost of a higher ripple in the passband of the spectrum associated with a unimodular sequence.

2.2 PAPR Constraint

In this chapter, we describe the transmit beamforming problem with a constraint on the antenna array weight $\mathbf{w} = [\, w_1 \, \ldots \, w_N \,]^{\mathsf{T}} \in \mathbb{C}^N$ that is induced by

$$\mathrm{PAPR}\left(\mathbf{w}\right) = \frac{\max\limits_{n\in\{1,\ldots,N\}} |w_n|^2}{(1/N)\sum\limits_{n=1}^{N} |w_n|^2}, \tag{2.1}$$

where the numerator and denominator are the maximum and average power of $\mathbf{w}$ entries, respectively. We have $\mathrm{PAPR}\left(\cdot\right) : \mathbb{C}^N \rightarrow [1, N]$. The lower endpoint (1), which implies equal maximum and average powers, is reached when all entries have the same magnitude. The upper endpoint (N) is reached when $\mathbf{w}$ has a single nonzero entry. Moreover, the ratio (2.1) is independent of the phase of its argument, *i.e.*, $\angle\mathbf{w}$, and is scale-invariant, that is, $\mathrm{PAPR}\left(\mathbf{w}\right) = \mathrm{PAPR}\left(s\mathbf{w}\right)$ for any $s \in \mathbb{C}$ and $|s| \neq 0$.

The PAPR adjustment of the weight as a constraint in a problem is not straightforward, especially when the problem without this constraint has a solution whose PAPR is higher than that of the solution associated with the constrained problem. For instance, in beampattern design, the weight with a higher dynamic range, namely the weight that has a higher difference between the maximum and minimum magnitudes of its entries, is generally required to produce narrower transition widths or a main lobe with less power swing. These fine tunes are due to the entries whose magnitudes are close to the minimum. However, the dynamic range of weight should be set as low as possible to increase the power efficiency, a performance metric discussed later.

Before delving into the beamforming problem, we would like to give some notions on PAPR constraint through a fundamental problem having PAPR constraint.

2.3 A Problem With PAPR Constraint

For a given vector $\mathbf{s} \in \mathbb{C}^N$, we consider finding the nearest vector to $\mathbf{s}$ among the vectors with the total power $P \in \mathbb{R}_{++}$ ($P > 0$) and PAPR not exceeding $\sigma \in [1, N]$. The total power constraint is given to disambiguate the scale-invariant PAPR constraint by fixing the denominator of (2.1) as P/N. The problem can be given as

$$\begin{aligned} \mathbf{w}^\star = \underset{\mathbf{w}\in\mathbb{C}^N}{\operatorname{argmin}} \quad & \|\mathbf{w}-\mathbf{s}\|_2 \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{w}) \le \sigma \\ & \|\mathbf{w}\|_2^2 = P \end{aligned} \tag{2.2}$$

and can be solved by an alternating projection method given in [17, Algorithm 2]. The method therein is also outlined in Algorithm 2.1.

In the initialization step, δ^2 and P/N are equal to the desired peak and average power, respectively. In Line 6, the magnitude of solution entries corresponding to the zero entry indices of $\mathbf{s}$ is adjusted to make the remaining solution entries have the same magnitude δ. In Line 8, the scaling factor γ is computed to share the remaining power $P - k\delta^2$ among the smallest (in magnitude) entries. Line 10 determines whether there is any entry having power higher than the desired peak power or not.

We can express the closed-form solutions when $\sigma = 1$ and $\operatorname{PAPR}(\mathbf{s}) \le \sigma \le N$ in (2.2). For the lower bound of PAPR, *i.e.*, $\sigma = 1$, the solution is

$$\mathbf{w}^\star = \sqrt{P/N}\, e^{j\angle\mathbf{s}}, \tag{2.3}$$

which is unimodular and has the same phase as $\mathbf{s}$. The reader should notice that the nearest vectors $\mathbf{w}^\star$ and $\mathbf{s}$ always have the same phase independently of σ since $\operatorname{PAPR}(\cdot)$ does not depend on the phase of its argument.

When $\sigma = N$, the solution becomes

$$\mathbf{w}^\star = \frac{\sqrt{P}}{\|\mathbf{s}\|_2}\,\mathbf{s}, \tag{2.4}$$

which is also valid for $\sigma \ge \operatorname{PAPR}(\mathbf{s})$.

Both (2.3) and (2.4) are given as the function of P and $\mathbf{s}$. Hence, the relation between P and $\mathbf{s}$ should be examined to have prior knowledge about the value of P, which

Algorithm 2.1: Computation of the nearest vector having low PAPR [17, Algorithm 2].

Input : $\mathbf{s} \in \mathbb{C}^N$, $\sigma \geq 1$, $P \in \mathbb{R}_{++}$

Output : $\mathbf{w}^\star \in \mathbb{C}^N$

Initialization: $\mathbf{s} \leftarrow \mathbf{s}/\|\mathbf{s}\|_2$, $\delta = \sqrt{\sigma P/N}$, $k = 0$

1 Form $\mathcal{T}$ as the index set of the $N - k$ smallest (in magnitude) entries of $\mathbf{s}$.

2 **while** $\mathcal{T}$ is not unique **do** // Some entries of $\mathbf{s}$ may have the same magnitude

3 $\quad k = k + 1$

4 $\quad$ Go to Line 1

5 **end**

6 **if** $s_t = 0$ for all $t \in \mathcal{T}$ **then** // Many solutions

7 $$w_t^\star = \begin{cases} \sqrt{\dfrac{P - k\delta^2}{N - k}} & \text{if } t \in \mathcal{T} \\ \delta\, e^{j\angle s_t} & \text{otherwise.} \end{cases}$$

8 **else**

9 $\quad \gamma = \sqrt{\frac{P - k\delta^2}{\sum_{t \in \mathcal{T}} |s_t|^2}}$

10 $\quad$ **if** $\gamma\,|s_t| > \delta$ for any $t \in \mathcal{T}$ **then**

11 $\quad\quad k = k + 1$

12 $\quad\quad$ Go to Line 1

13 $\quad$ **else** // Unique solution

14 $$w_t^\star = \begin{cases} \gamma\, s_t & \text{if } t \in \mathcal{T} \\ \delta\, e^{j\angle s_t} & \text{otherwise.} \end{cases}$$

15 $\quad$ **end**

16 **end**

17 **return** $\mathbf{w}^\star$

changes the objective function value of (2.2). The effect of P on the objective function value is discussed via an example. We let the problem be

$$\begin{array}{ll} \text{minimize} & \|\mathbf{w} - \mathbf{s}\|_2^2 \\ \text{subject to} & \mathrm{PAPR}\,(\mathbf{w}) \leq \sigma \\ & \|\mathbf{w}\|_2^2 = P \end{array} \tag{2.5}$$

with variables $\mathbf{w} \in \mathbb{C}^N$ and $P \in \mathbb{R}_{++}$. For the lower bound of PAPR, *i.e.*, $\sigma = 1$, we have $\mathbf{w}^\star = \sqrt{P/N}\, e^{j\angle \mathbf{s}}$, as given in (2.3). After substituting this solution, which is a

function of P, for $\mathbf{w}$ in the objective function of (2.5), we get

$$\begin{aligned}\|\mathbf{w}-\mathbf{s}\|_2^2 &= \|\mathbf{w}\|_2^2 - \mathbf{w}^{\mathsf{H}}\mathbf{s} - \mathbf{s}^{\mathsf{H}}\mathbf{w} + \|\mathbf{s}\|_2^2 \\ &= P - 2\sqrt{\frac{P}{N}}\sum_{n=1}^{N}|s_n| + \|\mathbf{s}\|_2^2. \end{aligned} \tag{2.6}$$

Setting the partial derivative of (2.6) with respect to P equal to zero, we find the minimizer as

$$P = \frac{1}{N}\left(\sum_{n=1}^{N}|s_n|\right)^2. \tag{2.7}$$

For the upper bound of PAPR, *i.e.*, $\sigma = N$, we have $\mathbf{w}^\star = \sqrt{P}\,\mathbf{s}/\|\mathbf{s}\|_2$, as given in (2.4). After replacing this solution with $\mathbf{w}$, the objective function becomes

$$\|\mathbf{w}-\mathbf{s}\|_2^2 = P - 2\sqrt{P}\,\|\mathbf{s}\|_2 + \|\mathbf{s}\|_2^2, \tag{2.8}$$

and the minimizer of (2.8) is

$$P = \|\mathbf{s}\|_2^2. \tag{2.9}$$

Since the PAPR constraint in (2.5) is an inequality constraint, the minimizer given in (2.9) is also valid for $\sigma \geq \mathrm{PAPR}\,(\mathbf{s})$.

For the remaining values of PAPR, *i.e.*, $1 < \sigma < \mathrm{PAPR}\,(\mathbf{s})$, there is not a closed-form expression for $\mathbf{w}^\star$ due to the solution method of (2.2). Hence, only the interval of the minimizer can be given as

$$\frac{1}{N}\left(\sum_{n=1}^{N}|s_n|\right)^2 < P < \|\mathbf{s}\|_2^2$$

for $1 < \sigma < \mathrm{PAPR}\,(\mathbf{s})$. Furthermore, assuming that P_1 and P_2 are the minimizers associated with σ_1 and σ_2, respectively, if $\sigma_1 > \sigma_2$, then $P_1 > P_2$.

Note that (2.7) and (2.9) are equal only when $\mathrm{PAPR}\,(\mathbf{s}) = 1$. Moreover, the objective function of (2.5) is zero only for $\sigma \geq \mathrm{PAPR}\,(\mathbf{s})$.

Lemma 1. *The achievable objective function value of* (2.2) *is zero if and only if* $\sigma \geq \mathrm{PAPR}\,(\mathbf{s})$ *and* $P = \|\mathbf{s}\|_2^2$.

Proof 1. *Let* $\mathbf{w}^\star$ *be the solution of* (2.2)*, then* $\|\mathbf{w}^\star - \mathbf{s}\|_2 \geq 0$ *and equality holds, i.e., the achievable objective function value of* (2.2) *is zero, only when* $\mathbf{w}^\star = \mathbf{s}$*. Two vectors* $\mathbf{w}^\star$ *and* $\mathbf{s}$ *are linearly dependent if and only if* $\sigma \geq \mathrm{PAPR}\,(\mathbf{s})$*, as given in* (2.4)*. From* (2.4)*,* $\mathbf{w}^\star = \mathbf{s}$ *if and only if* $P = \|\mathbf{s}\|_2^2$*. The proof is completed.*

When we describe our solution for the beamforming problem of interest later, we will refer to Lemma 1 to make an assumption essential for improving the convergence property of the suggested solution.

Next, we describe the transmit beamforming problem under the PAPR constraint.

2.4 System Model and Problem Formulation

We consider a phased array system with isotropic elements. The parameters are given as follows:

- f_c is the center frequency of the wave transmitted.
- c is the propagation speed of the wave in the medium of interest.
- λ is the wavelength given as
$$\lambda = \frac{c}{f_c}.$$
- N is the number of sensors at fixed positions.
- L is the number of direction samples $\phi_\ell, \theta_\ell \in [-90^\circ, 90^\circ]$, $\ell = 1, \ldots, L$, where ϕ_ℓ and θ_ℓ correspond to azimuth and elevation angles, respectively. Typically, $L \gg N$, *e.g.*, $L = 20N$.
- $\mathbf{w} = [\, w_1 \ \ldots \ w_N \,]^{\mathsf{T}} \in \mathbb{C}^N$ is the transmitter weight vector.
- $P \in \mathbb{R}_{++}$ is the total power of $\mathbf{w}$ entries, *i.e.*, $\|\mathbf{w}\|_2^2 = P$.
- $\sigma \in [1, N]$ is the PAPR upper bound for $\mathbf{w}$, *i.e.*, $\mathrm{PAPR}\,(\mathbf{w}) \leq \sigma$.
- $\mathbf{p}_n = [\, p_{n,1} \ \ p_{n,2} \ \ p_{n,3} \,]^{\mathsf{T}} \in \mathbb{R}^3$ is the position of the nth sensor that is
$$\mathbf{p}_n = p_{n,1}\mathbf{u}_x + p_{n,2}\mathbf{u}_y + p_{n,3}\mathbf{u}_z, \quad n = 1, \ldots, N,$$
in the Cartesian coordinate system with the three-dimensional unit vectors $\mathbf{u}_x$, $\mathbf{u}_y$, and $\mathbf{u}_z$ corresponding to the directions x, y, and z, respectively.
- $\mathbf{k}_\ell \in \mathbb{R}^3$ is the wavenumber vector for the ℓth direction pair (ϕ_ℓ, θ_ℓ) that is
$$\mathbf{k}_\ell = (2\pi/\lambda)\,(\cos\phi_\ell \cos\theta_\ell \mathbf{u}_x + \sin\phi_\ell \cos\theta_\ell \mathbf{u}_y + \sin\theta_\ell \mathbf{u}_z), \quad \ell = 1, \ldots, L.$$

- $\mathbf{a}_\ell \in \mathbb{C}^N$ is the array steering vector for the ℓth direction pair (ϕ_ℓ, θ_ℓ) that is

$$\mathbf{a}_\ell = \left[\, e^{j\mathbf{k}_\ell^\mathsf{T}\mathbf{p}_1} \;\ldots\; e^{j\mathbf{k}_\ell^\mathsf{T}\mathbf{p}_N} \,\right]^\mathsf{T}, \quad \ell = 1, \ldots, L.$$

- $\mathbf{b} = [\, b_1 \;\ldots\; b_L \,]^\mathsf{T} \in \mathbb{R}_+^L$ is the magnitude of the desired beampattern, where b_ℓ corresponds to the ℓth direction pair (ϕ_ℓ, θ_ℓ) for $\ell = 1, \ldots, L$.

- $\mathbf{h} = [\, h_1 \;\ldots\; h_L \,]^\mathsf{T} \in \mathbb{R}_+^L$ contains the weighting coefficients for the squared magnitude errors, where h_ℓ corresponds to the ℓth direction pair (ϕ_ℓ, θ_ℓ) for $\ell = 1, \ldots, L$.

The goal is to approximate the magnitude of the desired transmit beampattern $\mathbf{b}$ by using the transmitter weight vector $\mathbf{w}$ that satisfies the PAPR constraint $\mathrm{PAPR}\,(\mathbf{w}) \leqslant \sigma$. We also introduce an equality constraint $\|\mathbf{w}\|_2^2 = P$ for total power to disambiguate the PAPR constraint as used in (2.2). The transmit beamforming problem of interest is to optimize $\mathbf{w} \in \mathbb{C}^N$ by minimizing the weighted least squares error under the PAPR constraint:

$$\begin{aligned} \text{minimize} \quad & \sum_{\ell=1}^{L} h_\ell \left(b_\ell - \left|\mathbf{a}_\ell^\mathsf{H}\mathbf{w}\right|\right)^2 \\ \text{subject to} \quad & \mathrm{PAPR}\,(\mathbf{w}) \leq \sigma \\ & \|\mathbf{w}\|_2^2 = P. \end{aligned} \tag{2.10}$$

This formulation has several difficulties as follows:

- The objective function is nondifferentiable and nonconvex due to the absolute value operator in the term $\left|\mathbf{a}_\ell^\mathsf{H}\mathbf{w}\right|$.

- The total power constraint $\|\mathbf{w}\|_2^2 = P$ induces a nonconvex feasible set.

- The PAPR constraint is highly nonlinear and difficult to characterize, see Section 2.3.

By transforming (2.10) into an equivalent problem having a convex objective function, we describe an ADMM-based method in the next chapter.

CHAPTER 3

TRANSMIT BEAMFORMING PROBLEM WITH PAPR CONSTRAINT: ADMM-BASED SOLUTION

In this chapter, we use the alternating direction method of multipliers (ADMM) to solve the transmit beamforming problem under a peak-to-average power ratio (PAPR) constraint. The ADMM is known to converge to the unique global optima for convex problems, [28]. Yet, for nonconvex problems, it is not possible to ensure the optimality of the solution or even the convergence of the ADMM iterations. In spite of this setback, ADMM has been successfully used in many problems.

3.1 Related Works and Contributions

In [23], unimodular sequences with low autocorrelation sidelobe are designed by ADMM after introducing auxiliary variables to separate the linear and quadratic terms of the objective function and to impose the nonconvex unimodularity constraint only on the linear term. Furthermore, the spectrally shaped waveform design problem is solved by using ADMM via controlling the passband ripple and peak stopband level. In [25], an ADMM-based approach is proposed to solve optimization problems subject to nonconvex magnitude constraints for frequency/angular domains, and several application examples are given to demonstrate the effectiveness of the proposed method. In [24], an ADMM-based solution PhareADMM is presented for the phase retrieval problem by introducing auxiliary magnitude and phase variables to circumvent the absolute value operator in the objective function. By a simple modification exploiting the unitary property of discrete Fourier transform (DFT), PhareADMM is used to design unimodular periodic sequences having low autocorrelation sidelobe,

[24]. The successful application of ADMM in many optimization problems has also initiated the generalized studies on its convergence, [29].

Our formulation is similar to that of PhareADMM given in [24]. The main difference is the inclusion of the PAPR constraint in the problem. The inclusion of the PAPR constraint requires the introduction of distinct penalty parameters for the beampattern shape and PAPR constraints. More specifically, the augmented Lagrangian function, which is the objective function minimized in primal and auxiliary variable update steps of ADMM, is defined by using distinct penalty parameters for the beampattern shape and PAPR constraints. We also suggest a simple relation to set the penalty parameters specifically for this problem. The suggested multiple penalty parameters are to control the constraint violations individually, as in [30, 31], [32, p. 292].

This chapter presents the following main contributions.

- We examine the transmit beamformer design problem under the PAPR constraint to design a flat-top beampattern with low sidelobes at a small sacrifice from the power efficiency. We use the phrase *power efficiency* to denote the deviation of the average transmitted power from its maximum value. The transmit power is maximized with unimodular weights ($\sigma = 1$). The ratio of the average transmitted power of an arbitrary weight vector to the maximum transmitted power is the power efficiency, see $P_{\text{eff}}(\mathbf{w}_\sigma)$ in Table 3.2.

- We formulate an ADMM-based solution with multiple penalty parameters for individually controlling the mismatches in the beampattern shape and PAPR constraints. Different from earlier approaches (finding a solution without the PAPR constraint and then projecting the solution onto a set satisfying the PAPR constraint), the suggested method generates a solution via "mixing" the optimization outputs for the spectral processing (*i.e.*, beampattern shape) and temporal processing (*i.e.*, PAPR constraint), where the mixture is controlled via the penalty parameters. The presented results are valid for both narrowband and wideband beamforming problems and can be extended to the waveform design problem.

3.2 Proposed Method

The ADMM is a distributed convex optimization method, [28]. According to [28], a well-defined projection onto nonconvex sets can also assist the convergence when solving nonconvex problems, which have a convex objective function and nonconvex constraints. Inspired by this observation, we re-express the objective function of (2.10) so as to avoid the nonconvex absolute value operator ($|\cdot|$).

In [24], the auxiliary magnitude and phase variables are used to circumvent the absolute value operator in the objective function. We define the auxiliary variables $\alpha_\ell \in \mathbb{R}_+$ and $\beta_\ell \in \mathbb{R}$ as

$$\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} = \alpha_\ell\, e^{j\beta_\ell}, \tag{3.1}$$

where

$$\alpha_\ell = \left|\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}\right| \text{ and } \beta_\ell = \angle(\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}),$$

for $\ell = 1, \ldots, L$. After replacing $\left|\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}\right|$ with α_ℓ, (2.10) becomes

$$\begin{aligned} \text{minimize} \quad & \sum_{\ell=1}^{L} h_\ell\,(b_\ell - \alpha_\ell)^2 \\ \text{subject to} \quad & \mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} = \alpha_\ell\, e^{j\beta_\ell} \\ & \mathrm{PAPR}\,(\mathbf{w}) \le \sigma \\ & \|\mathbf{w}\|_2^2 = P \end{aligned} \tag{3.2}$$

with variables $\mathbf{w} \in \mathbb{C}^N$, $\alpha_1, \ldots, \alpha_L \in \mathbb{R}_+$ and $\beta_1, \ldots, \beta_L \in \mathbb{R}$. Thus, in (3.2), the problem formulation in (2.10) has been converted to one with a convex objective function and nonconvex constraints.

We suggest introducing an auxiliary vector $\mathbf{v}$ to prevent the appearance of the weight vector $\mathbf{w}$ in the PAPR constraint. With this suggestion, a solution of (3.2) can be generated by imposing the beampattern and PAPR constraints as the individual subproblems of ADMM:

$$\begin{aligned} \text{minimize} \quad & \sum_{\ell=1}^{L} h_\ell\,(b_\ell - \alpha_\ell)^2 \\ \text{subject to} \quad & \mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} = \alpha_\ell\, e^{j\beta_\ell} \\ & \mathbf{w} = \mathbf{v} \\ & \mathrm{PAPR}\,(\mathbf{v}) \le \sigma \\ & \|\mathbf{w}\|_2^2 = P \end{aligned} \tag{3.3}$$

with variables $\mathbf{w}, \mathbf{v} \in \mathbb{C}^N$, $\alpha_1, \ldots, \alpha_L \in \mathbb{R}_+$ and $\beta_1, \ldots, \beta_L \in \mathbb{R}$. In (3.3), the beampattern shape α_ℓ, $\ell = 1, \ldots, L$, and $\mathrm{PAPR}\,(\mathbf{v})$ constraints are coupled via the equality constraint $\mathbf{w} = \mathbf{v}$.

The augmented Lagrangian for the application of ADMM on (3.3) requires the definition of penalty parameters as follows:

$$
\begin{aligned}
&\mathcal{L}_{\rho_L,\rho_N}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}, \mathbf{y}, \boldsymbol{\lambda}) \\
&\quad = \sum_{\ell=1}^{L} h_\ell \,(b_\ell - \alpha_\ell)^2 \\
&\qquad + \sum_{\ell=1}^{L} \mathsf{Re}\,\{y_\ell\}\, \mathsf{Re}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} - \alpha_\ell\, e^{j\beta_\ell}\right\} + \sum_{\ell=1}^{L} \mathsf{Im}\,\{y_\ell\}\, \mathsf{Im}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} - \alpha_\ell\, e^{j\beta_\ell}\right\} \\
&\qquad + \frac{\rho_L}{2} \sum_{\ell=1}^{L} \big(\mathsf{Re}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} - \alpha_\ell\, e^{j\beta_\ell}\right\}\big)^2 + \frac{\rho_L}{2} \sum_{\ell=1}^{L} \big(\mathsf{Im}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w} - \alpha_\ell\, e^{j\beta_\ell}\right\}\big)^2 \\
&\qquad + \sum_{n=1}^{N} \mathsf{Re}\,\{\lambda_n\}\, \mathsf{Re}\,\{w_n - v_n\} + \sum_{n=1}^{N} \mathsf{Im}\,\{\lambda_n\}\, \mathsf{Im}\,\{w_n - v_n\} \\
&\qquad + \frac{\rho_N}{2} \sum_{n=1}^{N} (\mathsf{Re}\,\{w_n - v_n\})^2 + \frac{\rho_N}{2} \sum_{n=1}^{N} (\mathsf{Im}\,\{w_n - v_n\})^2, \qquad (3.4)
\end{aligned}
$$

where $\rho_L > 0$ and $\rho_N > 0$ are distinct penalty parameters for beampattern shape and PAPR constraints, respectively, $\mathbf{y} = [\, y_1 \ \ldots \ y_L \,]^{\mathsf{T}} \in \mathbb{C}^L$ and $\boldsymbol{\lambda} = [\, \lambda_1 \ \ldots \ \lambda_N \,]^{\mathsf{T}} \in \mathbb{C}^N$ are the dual variables, $\boldsymbol{\alpha} = [\, \alpha_1 \ \ldots \ \alpha_L \,]^{\mathsf{T}} \in \mathbb{R}_+^L$, $\boldsymbol{\beta} = [\, \beta_1 \ \ldots \ \beta_L \,]^{\mathsf{T}} \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{C}^N$ are the auxiliary variables. In (3.4), the real and imaginary parts of the complex-valued terms from (3.3) are separately considered due to the nature of ADMM, which is built on the real-valued and convex functions.

In many ADMM applications, the augmented Lagrangian functions are defined with a single penalty parameter used for all constraints [30, 33, 34]. However, we utilize two different penalty parameters ρ_L and ρ_N in this application. The use of a different penalty parameter for every equality constraint provides a means of individually scaling the constraint violations, as suggested in [31]. Further details on this approach can be found in [35, Ch. 9].

We apply ADMM to minimize (3.4) under the PAPR constraint with respect to variable sets $\{\mathbf{w}\}$ and $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}\}$ separately. We describe the ADMM steps involved in each iteration as follows:

Step-1: For given variables at the iteration number k, $\mathbf{w}^{(k)}$, $\mathbf{y}^{(k)}$ and $\boldsymbol{\lambda}^{(k)}$, the auxiliary variables at $k+1$ can be determined by solving the problem

$$\begin{aligned}\left\{\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{v}^{(k+1)}\right\} = \operatorname{argmin} \quad & \mathcal{L}_{\rho_L,\rho_N}\left(\mathbf{w}^{(k)}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}, \mathbf{y}^{(k)}, \boldsymbol{\lambda}^{(k)}\right) \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{v}) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P\end{aligned} \tag{3.5}$$

with variables $\boldsymbol{\alpha} \in \mathbb{R}_+^L$, $\boldsymbol{\beta} \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{C}^N$.

Let $\mathbf{z} = [\, z_1 \, \ldots \, z_L \,]^{\mathsf{T}} = \mathbf{y}/\rho_L \in \mathbb{C}^L$ and $\boldsymbol{\tau} = [\, \tau_1 \, \ldots \, \tau_N \,]^{\mathsf{T}} = \boldsymbol{\lambda}/\rho_N \in \mathbb{C}^N$ for using the scaled form of ADMM. In (3.5), the objection function terms depending on $\mathsf{Re}\{\mathbf{v}\}$ can be manipulated algebraically as follows:

$$\begin{aligned}&\sum_{n=1}^{N} \mathsf{Re}\left\{\lambda_n^{(k)}\right\} \mathsf{Re}\left\{w_n^{(k)} - v_n\right\} + \frac{\rho_N}{2} \sum_{n=1}^{N} \left(\mathsf{Re}\left\{w_n^{(k)} - v_n\right\}\right)^2 \\ &\quad = -\frac{1}{2\rho_N} \sum_{n=1}^{N} \left(\mathsf{Re}\left\{\lambda_n^{(k)}\right\}\right)^2 + \frac{\rho_N}{2} \sum_{n=1}^{N} \left(\mathsf{Re}\left\{w_n^{(k)} - v_n\right\} + \left(\mathsf{Re}\left\{\lambda_n^{(k)}\right\}/\rho_N\right)\right)^2 \\ &\quad = -\frac{1}{2\rho_N} \sum_{n=1}^{N} \left(\mathsf{Re}\left\{\lambda_n^{(k)}\right\}\right)^2 + \frac{\rho_N}{2} \sum_{n=1}^{N} \left(\mathsf{Re}\left\{w_n^{(k)} - v_n + \tau_n^{(k)}\right\}\right)^2,\end{aligned}$$

where the first term is fixed since $\boldsymbol{\lambda}^{(k)}$ $\left(\text{and } \boldsymbol{\tau}^{(k)}\right)$ is given for (3.5) in Step-1, and does not affect the arguments minimizing (3.5). After using this approach, namely, completing square and ignoring the constant terms, for the imaginary counterpart in (3.5), we can combine the expressions depending on real and imaginary parts of $\mathbf{v}$ and get a quadratic term depending on $\mathbf{v}$ as

$$\begin{aligned}\frac{\rho_N}{2} \sum_{n=1}^{N} \left(\mathsf{Re}\left\{w_n^{(k)} - v_n + \tau_n^{(k)}\right\}\right)^2 + \frac{\rho_N}{2} \sum_{n=1}^{N} \left(\mathsf{Im}\left\{w_n^{(k)} - v_n + \tau_n^{(k)}\right\}\right)^2 & \\ = \frac{\rho_N}{2} \left(\mathbf{w}^{(k)} + \boldsymbol{\tau}^{(k)} - \mathbf{v}\right)^{\mathsf{H}} \left(\mathbf{w}^{(k)} + \boldsymbol{\tau}^{(k)} - \mathbf{v}\right) & \\ = \frac{\rho_N}{2} \left\|\mathbf{w}^{(k)} + \boldsymbol{\tau}^{(k)} - \mathbf{v}\right\|_2^2. &\end{aligned}$$

By applying the same procedure to the remaining terms corresponding to $\mathbf{y}^{(k)}$ $\left(\text{and } \mathbf{z}^{(k)}\right)$, we introduce a new objective function

$$\begin{aligned}g(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}) = \sum_{\ell=1}^{L} h_\ell \left(b_\ell - \alpha_\ell\right)^2 + \frac{\rho_L}{2} \sum_{\ell=1}^{L} \left|\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}^{(k)} + z_\ell^{(k)} - \alpha_\ell \, e^{j\beta_\ell}\right|^2 & \\ + \frac{\rho_N}{2} \left\|\mathbf{w}^{(k)} + \boldsymbol{\tau}^{(k)} - \mathbf{v}\right\|_2^2 &\end{aligned} \tag{3.6}$$

such that

$$g(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{v}) \stackrel{c}{=} \mathcal{L}_{\rho_L,\rho_N}\left(\mathbf{w}^{(k)},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{v},\mathbf{y}^{(k)},\boldsymbol{\lambda}^{(k)}\right),$$

where the equality holds for a constant c, and write a problem equivalent to (3.5) as

$$\begin{aligned}\left\{\boldsymbol{\alpha}^{(k+1)},\boldsymbol{\beta}^{(k+1)},\mathbf{v}^{(k+1)}\right\} = \operatorname{argmin} \quad & g(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{v}) \\ \text{subject to} \quad & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P\end{aligned} \tag{3.7}$$

with variables $\boldsymbol{\alpha} \in \mathbb{R}_+^L$, $\boldsymbol{\beta} \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{C}^N$.

In (3.7), $\mathbf{v}$ appears only in the last term of objective function. Hence, $\mathbf{v}^{(k+1)}$ can be determined by solving the equivalent problem

$$\begin{aligned}\mathbf{v}^{(k+1)} = \operatorname*{argmin}_{\mathbf{v}\in\mathbb{C}^N} \quad & \left\|\mathbf{w}^{(k)}+\boldsymbol{\tau}^{(k)}-\mathbf{v}\right\|_2^2 \\ \text{subject to} \quad & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P.\end{aligned} \tag{3.8}$$

The problem (3.8) has the form of (2.2), which can be solved by following the steps given in [17, Algorithm 2], see also Algorithm 2.1. The main role of $\mathbf{v}$ in this step is to control the PAPR of $\mathbf{w}$. By the residual convergence feature of ADMM, we should observe that the primal residual converges to zero for PAPR constraint, *i.e.*, $\mathbf{w}^{(k)} - \mathbf{v}^{(k)} \to \mathbf{0}_N$ as $k \to \infty$, and hence, the PAPR and norm of $\mathbf{w}$ will be too close to those of $\mathbf{v}$ after some iterations. However, this residual convergence depends on P. Because P affects the objective function value of (3.8), as discussed in Section 2.3. According to Lemma 1, the objective function value of (3.8) can take zero only when $\sigma \geq \mathrm{PAPR}\left(\mathbf{w}^{(k)}+\boldsymbol{\tau}^{(k)}\right)$ and $\|\mathbf{v}\|_2^2 = \|\mathbf{w}^{(k)}+\boldsymbol{\tau}^{(k)}\|_2^2$. As a consequence, to eliminate the possibly undesired effect of P on the residual convergence, we suggest replacing P in (3.8) by $\|\mathbf{w}^{(k)}+\boldsymbol{\tau}^{(k)}\|_2^2$. Thus, we use

$$\begin{aligned}\mathbf{v}^{(k+1)} = \operatorname*{argmin}_{\mathbf{v}\in\mathbb{C}^N} \quad & \left\|\mathbf{w}^{(k)}+\boldsymbol{\tau}^{(k)}-\mathbf{v}\right\|_2^2 \\ \text{subject to} \quad & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = \left\|\mathbf{w}^{(k)}+\boldsymbol{\tau}^{(k)}\right\|_2^2.\end{aligned} \tag{3.9}$$

We note that the feasible set of (3.9) varies with k unlike that of (3.8). Hence, $\mathbf{w}^{(k)}$ can converge to a point which is not feasible for (2.10) due to the violation of the total power constraint, *i.e.*, $\left\|\mathbf{w}^{(k)}\right\|_2^2 \neq P$, when $\mathbf{w}^{(k)} - \mathbf{v}^{(k)} \to \mathbf{0}_N$ as $k \to \infty$. By scaling

$\mathbf{w}^{(k)}$, we can obtain a point $\sqrt{P}\,\mathbf{w}^{(k)}/\left\|\mathbf{w}^{(k)}\right\|_2$ satisfying the total power constraint of (2.10). As seen from the PAPR definition in (2.1), the scaling of a vector does not change its PAPR, and we have

$$\mathrm{PAPR}\left(\sqrt{P}\,\mathbf{w}^{(k)}/\left\|\mathbf{w}^{(k)}\right\|_2\right) = \mathrm{PAPR}\left(\mathbf{w}^{(k)}\right).$$

Moreover, after the scaling, only the transmit beampattern shape is shifted up or down in dB scale, and the performance metrics, such as the peak sidelobe level (PSL) and the power swing in the main lobe, are not affected.

Similarly, $\boldsymbol{\beta}$ appears only in the middle term of (3.6), that is the objective function of (3.7). To determine $\boldsymbol{\beta}^{(k+1)}$, an equivalent problem is

$$\boldsymbol{\beta}^{(k+1)} = \operatorname{argmin} \sum_{\ell=1}^{L}\left|\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}^{(k)} + z_\ell^{(k)} - \alpha_\ell e^{j\beta_\ell}\right|^2 \tag{3.10}$$

with variables $\boldsymbol{\alpha} \in \mathbb{R}_+^L$ and $\boldsymbol{\beta} \in \mathbb{R}^L$. Let $\boldsymbol{\gamma} = [\,\gamma_1\ \ldots\ \gamma_L\,]^{\mathsf{T}} \in \mathbb{C}^L$, where $\gamma_\ell = \alpha_\ell\, e^{j\beta_\ell}$, $\ell = 1, \ldots, L$. Then, the objective function of (3.10) can be expressed as

$$\sum_{\ell=1}^{L}\left|\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}^{(k)} + z_\ell^{(k)} - \alpha_\ell e^{j\beta_\ell}\right|^2 = \left\|\mathbf{A}^{\mathsf{H}}\mathbf{w}^{(k)} + \mathbf{z}^{(k)} - \boldsymbol{\gamma}\right\|_2^2, \tag{3.11}$$

where $\mathbf{A} = [\,\mathbf{a}_1\ \ldots\ \mathbf{a}_L\,]$ is the $N \times L$ matrix consisting of array steering vectors.

As $\boldsymbol{\beta} = \angle\boldsymbol{\gamma}$ in (3.11), the solution of (3.10) becomes

$$\boldsymbol{\beta}^{(k+1)} = \angle\big(\mathbf{A}^{\mathsf{H}}\mathbf{w}^{(k)} + \mathbf{z}^{(k)}\big). \tag{3.12}$$

After replacing β_ℓ in the objective function of (3.7) by $\beta_\ell^{(k+1)}$, we obtain an equivalent problem of finding $\boldsymbol{\alpha}^{(k+1)}$ as follows:

$$\boldsymbol{\alpha}^{(k+1)} = \operatorname*{argmin}_{\boldsymbol{\alpha}\in\mathbb{R}_+^L} \left(\sum_{\ell=1}^{L} h_\ell\,(b_\ell - \alpha_\ell)^2 + (\rho_L/2)\sum_{\ell=1}^{L}(r_\ell - \alpha_\ell)^2\right), \tag{3.13}$$

where $r_\ell = \left|\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}^{(k)} + z_\ell^{(k)}\right|$, $\ell = 1, \ldots, L$.

Taking the derivative of the objective function in (3.13) with respect to α_ℓ and zeroing at $\alpha_\ell^{(k+1)}$, we get

$$0 = 2h_\ell\, b_\ell - 2h_\ell\, \alpha_\ell^{(k+1)} + \rho_L\, r_\ell - \rho_L \alpha_\ell^{(k+1)}, \quad \ell = 1, \ldots, L.$$

Since α_ℓ denotes the magnitude, *i.e.*, $\alpha_\ell \geq 0$, the solution is given as follows:

$$\alpha_\ell^{(k+1)} = \begin{cases} \dfrac{\rho_L\, r_\ell + 2h_\ell\, b_\ell}{\rho_L + 2h_\ell} & \text{if } (\rho_L\, r_\ell + 2h_\ell\, b_\ell)/(\rho_L + 2h_\ell) \geq 0 \text{ and } \rho_L + 2h_\ell \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.14}$$

for $\ell = 1, \ldots, L$.

For our application, $\rho_L > 0$, $r_\ell \geq 0$, $h_\ell \geq 0$ and $b_\ell \geq 0$ satisfy that $\rho_L\, r_\ell + 2h_\ell\, b_\ell \geq 0$ and $\rho_L + 2h_\ell > 0$, and hence, (3.14) can be simplified as

$$\alpha_\ell^{(k+1)} = \frac{\rho_L\, r_\ell + 2h_\ell\, b_\ell}{\rho_L + 2h_\ell}, \quad \ell = 1, \ldots, L. \tag{3.15}$$

Step-2: We update $\mathbf{w}$ by solving

$$\mathbf{w}^{(k+1)} = \underset{\mathbf{w}\in\mathbb{C}^N}{\operatorname{argmin}} \;\; \mathcal{L}_{\rho_L,\rho_N}\big(\mathbf{w}, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{v}^{(k+1)}, \mathbf{z}^{(k)}, \boldsymbol{\tau}^{(k)}\big). \tag{3.16}$$

According to (3.16), all variables except $\mathbf{w}$ are fixed in (3.4). Thus, the objective function of (3.7) that is modified from (3.4) due to the fixed $\mathbf{z}^{(k)}$ and $\boldsymbol{\tau}^{(k)}$ can be manipulated further by using the fixed

$$\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\alpha}^{(k+1)} \odot e^{j\boldsymbol{\beta}^{(k+1)}},$$

as in (3.11). Then, (3.16) is equivalent to

$$\mathbf{w}^{(k+1)} = \underset{\mathbf{w}\in\mathbb{C}^N}{\operatorname{argmin}} \;\; \left(\big\| \mathbf{A}^{\mathsf{H}}\mathbf{w} + \mathbf{z}^{(k)} - \boldsymbol{\gamma}^{(k+1)} \big\|_2^2 + (\rho_N/\rho_L) \big\| \mathbf{w} + \boldsymbol{\tau}^{(k)} - \mathbf{v}^{(k+1)} \big\|_2^2 \right), \tag{3.17}$$

which is quadratic in $\mathbf{w}$. Setting the gradient of the objective function in (3.17) with respect to $\mathbf{w}^*$ (conjugate of $\mathbf{w}$) equal to zero for $\mathbf{w}^{(k+1)}$ [36], we get

$$0 = \mathbf{A}\big(\mathbf{A}^{\mathsf{H}}\mathbf{w}^{(k+1)} + \mathbf{z}^{(k)} - \boldsymbol{\gamma}^{(k+1)}\big) + (\rho_N/\rho_L)\big(\mathbf{w}^{(k+1)} + \boldsymbol{\tau}^{(k)} - \mathbf{v}^{(k+1)}\big),$$

and find

$$\mathbf{w}^{(k+1)} = \left(\mathbf{A}\mathbf{A}^{\mathsf{H}} + \frac{\rho_N}{\rho_L}\mathbf{I}_N \right)^{-1} \left(\mathbf{A}(\boldsymbol{\gamma}^{(k+1)} - \mathbf{z}^{(k)}) + \frac{\rho_N}{\rho_L}\big(\mathbf{v}^{(k+1)} - \boldsymbol{\tau}^{(k)}\big) \right). \tag{3.18}$$

Step-3: The dual updates are as follows:

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \mathbf{A}^{\mathsf{H}}\mathbf{w}^{(k+1)} - \boldsymbol{\gamma}^{(k+1)}, \tag{3.19}$$

$$\boldsymbol{\tau}^{(k+1)} = \boldsymbol{\tau}^{(k)} + \mathbf{w}^{(k+1)} - \mathbf{v}^{(k+1)}. \tag{3.20}$$

We have described the steps to compute the transmitter weights satisfying the PAPR constraint. Then, we have a suggestion for penalty parameters selection. The suggested method has two distinct penalty parameters ρ_L and ρ_N, and they are jointly used only in (3.18). Here, we suggest that the ratio ρ_N/ρ_L should be equal to $\|\mathbf{A}\|_2$, which is also equal to the largest singular value of matrix $\mathbf{A}$, for keeping a balance between $\mathbf{A}\big(\boldsymbol{\gamma}^{(k+1)} - \mathbf{z}^{(k)}\big)$ and $(\rho_N/\rho_L)\big(\mathbf{v}^{(k+1)} - \boldsymbol{\tau}^{(k)}\big)$. Hence, ρ_N can be set as

$$\rho_N = \rho_L \|\mathbf{A}\|_2. \tag{3.21}$$

The suggested method is referred to as PAPR-ADMM outlined in Algorithm 3.1. The PAPR-ADMM requires an initial condition vector $\mathbf{w}^{(1)} \in \mathbb{C}^N$, which can be a unimodular vector for the transmit beamforming application. After initializing dual variables as $\mathbf{y}^{(1)} = \mathbf{0}_L$, $\boldsymbol{\lambda}^{(1)} = \mathbf{0}_N$, the steps described in this section are iteratively applied until the condition

$$\frac{\big\|\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\big\|_2}{\big\|\mathbf{w}^{(k-1)}\big\|_2} \leq 10^{-6}$$

or the maximum iteration number $k = k_{\max}$ is satisfied. Due to the update equation (3.18), the final transmitter weight vector $\mathbf{w}^{\text{final}}$ after termination of iterations, *i.e.*, $\mathbf{w}^{\text{final}} = \mathbf{w}^{(k)}$, may require an adjustment to meet the PAPR constraint of (2.10). In other words, $\mathrm{PAPR}\left(\mathbf{w}^{\text{final}}\right) \leq \sigma$ may not hold. We suggest implementing the PAPR update again on $\mathbf{w}^{\text{final}}$:

$$\begin{aligned} \mathbf{w}^{\star} = \underset{\mathbf{w}\in\mathbb{C}^N}{\operatorname{argmin}} \quad & \big\|\mathbf{w}^{\text{final}} - \mathbf{w}\big\|_2^2 \\ \text{subject to} \quad & \mathrm{PAPR}\left(\mathbf{w}\right) \leq \sigma \\ & \|\mathbf{w}\|_2^2 = P. \end{aligned} \tag{3.22}$$

Note that $\mathbf{w}^{\star}$ is simply the Euclidean projection of $\mathbf{w}^{\text{final}}$ onto the feasible set of the transmit beamforming problem (2.10) that is given by [17], see also Algorithm 2.1.

Conventionally ADMM is built on the real-valued and convex functions. In this chapter, we have given the formulation by separating the real and imaginary parts of complex variables, see (3.4). Then, we have utilized the complex-valued vector-matrix calculations to get an equivalent form, see the primal and dual updates. Depending on the application, *e.g.*, when $\mathbf{A}^{\mathsf{H}}\mathbf{w}$ is the DFT of $\mathbf{w}$, we can apply DFT and inverse DFT and simplify the matrix inverse calculation in (3.18).

Algorithm 3.1: Proposed method PAPR-ADMM for the transmit beamformer design under a PAPR constraint, also see [37].

Input : $\mathbf{w}^{(1)} \in \mathbb{C}^N$, $\mathbf{A} \in \mathbb{C}^{N\times L}$, $\mathbf{b} \in \mathbb{R}_+^L$, $\mathbf{h} \in \mathbb{R}_+^L$, $\rho_L > 0$, $\sigma \geq 1$, $P \in \mathbb{R}_{++}$, $k_{\max} \in \{1, 2, \ldots\}$

Output : $\mathbf{w}^\star \in \mathbb{C}^N$

Initialization: $\mathbf{w}^{(0)} = 2\mathbf{w}^{(1)}$, $\mathbf{y}^{(1)} = \mathbf{0}_L$, $\boldsymbol{\lambda}^{(1)} = \mathbf{0}_N$, $k = 1$

Compute ρ_N from (3.21)

Assign $\mathbf{z}^{(1)} = \mathbf{y}^{(1)}/\rho_L$ and $\boldsymbol{\tau}^{(1)} = \boldsymbol{\lambda}^{(1)}/\rho_N$

while $k \leq k_{\max}$ and $\left(\left\|\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\right\|_2 / \left\|\mathbf{w}^{(k-1)}\right\|_2\right) > 10^{-6}$ **do**

  // Step-1

  Determine $\mathbf{v}^{(k+1)}$ from (3.9) by applying Algorithm 2.1

  Compute $\boldsymbol{\beta}^{(k+1)}$ from (3.12)

  Compute $\boldsymbol{\alpha}^{(k+1)}$ from (3.15)

  // Step-2

  Compute $\mathbf{w}^{(k+1)}$ from (3.18) by using $\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\alpha}^{(k+1)} \odot e^{j\boldsymbol{\beta}^{(k+1)}}$

  // Step-3

  Compute $\mathbf{z}^{(k+1)}$ from (3.19)

  Compute $\boldsymbol{\tau}^{(k+1)}$ from (3.20)

  $k = k + 1$

end

Determine $\mathbf{w}^\star$ from (3.22) by using $\mathbf{w}^{\text{final}} = \mathbf{w}^{(k)}$ and applying Algorithm 2.1

return $\mathbf{w}^\star$

Next, we present an extension of the formulation to the wideband transmit beamformer design.

3.3 Wideband Beamforming

In wideband beamforming, the goal is to set the sensor weights that give the desired beampattern within the negligible difference for different operating frequencies/wavelengths. The wavelengths of interest correspond a limited bandwidth around the center frequency of the transmitter. Considering the well-known quality factor

definition,

$$Q_f = \frac{f_c}{\Delta_f}, \tag{3.23}$$

where f_c is the center frequency and Δ_f is the bandwidth of the transmitting elements for the definition of frequency samples $f_i \in [f_c - \Delta_f/2, f_c + \Delta_f/2]$, we compute the wavelength corresponding to f_i as

$$\lambda_i = \frac{c}{f_i}, \quad i = 1, \ldots, F.$$

For $i = 1, \ldots, F$, and $\ell = 1, \ldots, L$, we express the array steering vector corresponding to the ith operating frequency f_i and ℓth direction pair (ϕ_ℓ, θ_ℓ) as follows:

$$\mathbf{a}_{i\ell} = \left[\, e^{j\mathbf{k}_{i\ell}^{\mathsf{T}}\mathbf{p}_1} \;\ldots\; e^{j\mathbf{k}_{i\ell}^{\mathsf{T}}\mathbf{p}_N} \,\right]^{\mathsf{T}} \in \mathbb{C}^N,$$

where $\mathbf{k}_{i\ell} \in \mathbb{R}^3$ is the wavenumber given as

$$\mathbf{k}_{i\ell} = (2\pi/\lambda_i)[\, \cos\phi_\ell \cos\theta_\ell \mathbf{u}_x \;\; \sin\phi_\ell \cos\theta_\ell \mathbf{u}_y \;\; \sin\theta_\ell \mathbf{u}_z \,]^{\mathsf{T}}.$$

By concatenating the $N \times L$ dimensional matrices

$$\mathbf{A}_i = [\, \mathbf{a}_{i1} \;\ldots\; \mathbf{a}_{iL} \,], \quad i = 1, \ldots, F,$$

we define

$$\mathbf{A}_{\text{wb}} = [\, \mathbf{A}_1 \;\ldots\; \mathbf{A}_F \,],$$

which is an $N \times FL$ matrix of array steering vectors. Similarly, we can obtain the desired beampattern vector $\mathbf{b}_{\text{wb}} \in \mathbb{R}_+^{FL}$ by concatenating the desired beampatterns for operating frequency samples as

$$\mathbf{b}_{\text{wb}} = \left[\, \mathbf{b}_1^{\mathsf{T}} \;\ldots\; \mathbf{b}_F^{\mathsf{T}} \,\right]^{\mathsf{T}}.$$

Thus, the wideband beamforming problem with PAPR constraint is given as

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{FL} h_\ell \left(b_\ell - \left|\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}\right|\right)^2 \\ \text{subject to} & \text{PAPR}\,(\mathbf{w}) \le \sigma \\ & \|\mathbf{w}\|_2^2 = P \end{array} \tag{3.24}$$

with variable $\mathbf{w} \in \mathbb{C}^N$, where $\mathbf{a}_\ell$ is the ℓth column of $\mathbf{A}_{\text{wb}}$, b_ℓ is the ℓth entry of $\mathbf{b}_{\text{wb}}$, and $h_\ell \in \mathbb{R}_+$ is the weighting coefficient for the squared magnitude error corresponding to the ℓth direction pair, $\ell = 1, \ldots, FL$. This problem has the form of (2.10), and it can be solved by replacing $\mathbf{A}$ and $\mathbf{b}$ in Algorithm 3.1 with $\mathbf{A}_{\text{wb}}$ and $\mathbf{b}_{\text{wb}}$, respectively.

3.4 Algorithm Convergence

The convergence of ADMM for nonconvex problems is an important theoretic problem of the optimization literature, [29]. Algorithm convergence depends on the initial primal and dual points, the penalty parameter and also on the formulation of update steps, [28]. Unfortunately, the PAPR constraint (due to its nonlinearity) does not lend itself to further analysis. In our experience, PAPR-ADMM with the suggested settings has converged at all runs that we have implemented. The main difficulty that we face is not the convergence, the need of running PAPR-ADMM repeatedly with different initial conditions to avoid local optima. The sensitivity to the initial conditions is expected for given the multimodal nature of the beampattern matching problem, as discussed in [11]. To improve the performance, we suggest using the formulation (3.9) instead of (3.8) and examine this suggestion with some common parameter settings in Section 3.6.1. The reader is also invited to conduct Monte-Carlo trials of PAPR-ADMM with randomized initial conditions by running the ready-to-use MATLAB [37], Python [38], R [39] and C++ [40] codes.

3.5 Computational Complexity Considerations

We are able to present Algorithm 3.1, whose each step either has a closed-form expression or applies Algorithm 2.1. Algorithm 3.1 uses Algorithm 2.1 at Step-1 of each iteration and once after the termination of iterations. Algorithm 2.1 requires a sorting operation with the complexity of $\mathcal{O}\left(N \log N\right)$ operations and a processing of a vector with the worst case complexity of $\mathcal{O}\left(N^2\right)$.

In Step-1 of Algorithm 3.1, the necessary updates are implemented through (3.12) and (3.15) both requiring $\mathcal{O}\left(LN\right)$ multiplications. In Step-2, (3.18) requires $\mathcal{O}\left(LN^2\right)$. In Step-3, (3.19) and (3.20) require $\mathcal{O}\left(LN\right)$ and $\mathcal{O}\left(N\right)$, respectively. Thus, the computational complexity of each iteration is $\mathcal{O}\left(LN^2\right)$, which can be reduced further by computing the matrix inverse in (3.18) offline and storing the result. With the storage option, the computational complexity of Step-2 is reduced to $\mathcal{O}\left(N^2\right)$, and the overall complexity of each iteration becomes $\mathcal{O}\left(N^2 + LN\right)$ and $\mathcal{O}\left(N^2 + FLN\right)$ multiplications for narrowband and wideband beamformers, respectively.

3.6 Numerical Results

Unless otherwise is stated, we set $\rho_L = 25$, $k_{\max} = 10000$. For $\ell = 1, \ldots, L$, the weighting coefficient h_ℓ is set to δ_{m} if the ℓth direction sample is associated with the main lobe (or passband), and to δ_{s} otherwise (sidelobe or stopband), where $\delta_{\mathrm{s}}/\delta_{\mathrm{m}} = 100$. The initial condition weight vector $\mathbf{w}^{(1)} = \left[w_1^{(1)} \ \ldots \ w_N^{(1)} \right]^{\mathsf{T}}$ has the entries $w_n^{(1)} = e^{j2\pi\mu_n}$, $n = 1, \ldots, N$, where $\{\mu_n\}_{n=1}^N$ are the independent and identically distributed random variables, which are uniformly distributed on $(0, 1)$.

3.6.1 Effect of the Fixed Total Power

We assume a sonar system having $N = 32$ elements uniform linear array (ULA). The distance between neighboring sensors is $\lambda/2$ for the center frequency 150 kHz, where λ is determined by using the speed of sound in water, which is taken as 1500 m/s.

The design of a one-dimensional flat-top beampattern is considered in this experiment. The PSL of the desired beampattern is 30 dB. The main lobe is centered at $(0^\circ, 0^\circ)$ in (azimuth, elevation). The desired half-beamwidth and transition width are 45° and 8° in azimuth, respectively. The direction samples are taken by using a uniform grid spacing of 0.1°.

Our goal is to examine the algorithm performance when (3.8) is utilized instead of (3.9) in Step-1. We have 10 realizations for the initial weight vector. For each realization, we apply PAPR-ADMM by setting PAPR $= \sigma \in \{1, 2\}$, and $P = N$ with the formulation given by (3.8) (fixed total power), or (3.9) (varying total power).

In Figure 3.1, the augmented Lagrangian function (ALF) value (3.4) versus iteration number k is shown. Comparing Figures 3.1a and 3.1b, we notice that the fixed total power formulation prevents the monotonicity of ALF when PAPR $= 1$. For PAPR $= 2$, both formulations provide the monotonous ALF values after some iterations, as shown in Figures 3.1c and 3.1d. Moreover, Table 3.1 shows that the formulation (3.9) usually leads to a solution giving better performance metrics (less power swing in the main lobe, higher PSL) than those of (3.8). Hence, we suggest the usage of formulation (3.9) in Step-1 of Algorithm 3.1 for the performance improvement.

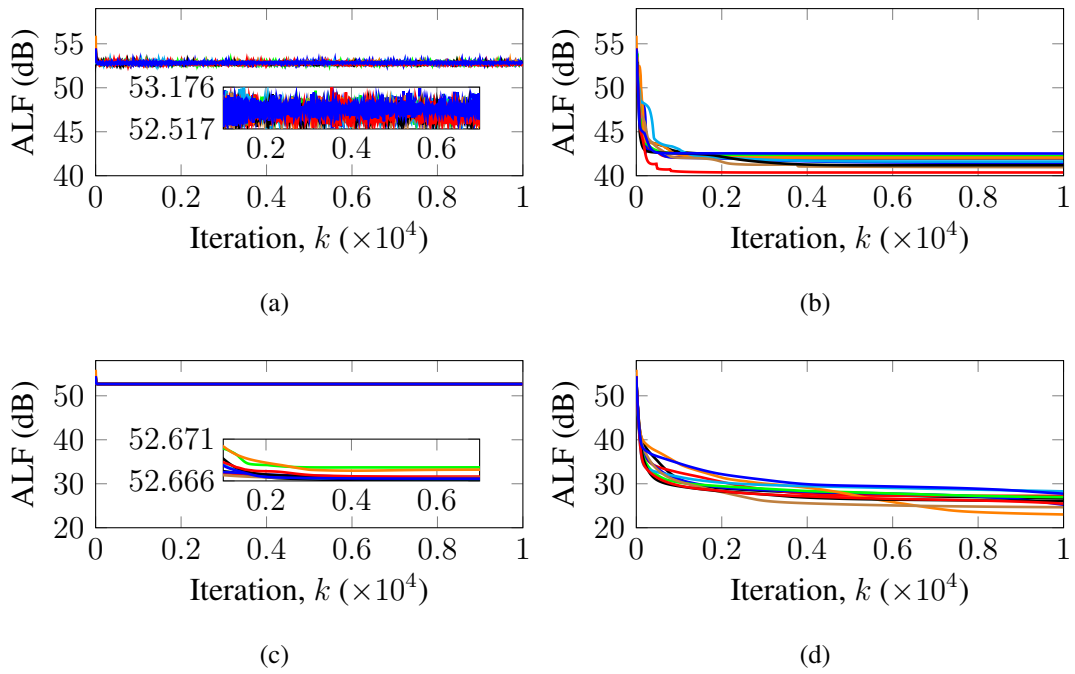


Figure 3.1: Augmented Lagrangian function (3.4) versus iteration number, $\mathcal{L}_{\rho_L,\rho_N}\big(\mathbf{w}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \mathbf{v}^{(k)}, \mathbf{y}^{(k)}, \boldsymbol{\lambda}^{(k)}\big)$, for the narrowband transmit beamforming with 32 elements ULA: (a) PAPR $= 1$ using (3.8), (b) PAPR $= 1$ using (3.9), (c) PAPR $= 2$ using (3.8), and (d) PAPR $= 2$ using (3.9).

3.6.2 Main Lobe Average Power Versus PAPR

We consider the sonar system given in Section 3.6.1 by setting PAPR as $\sigma \in \{1, 1.01, 1.02, 1.03, 1.04, 1.05, 1.1, 1.2, 1.25, 1.3, 1.4, 1.5, 1.6, 1.7, 1.75, 1.8, 1.9, 2, 3, 4, 5, 6\}$ and using PAPR-ADMM initialized with the same initial vector $\mathbf{w}^{(1)}$.

The goal is to observe the main lobe average power versus PAPR for narrowband transmit beamforming. As a strategy for maximizing the total transmitted power, transmitter weights are normalized so that the peak (or maximum) power is always fixed to the maximum allowed power that can be transmitted from an antenna. Thus, each weight $\mathbf{w}_\sigma$ for PAPR $= \sigma$ is scaled to have entries with the maximum magnitude of 1 before computing the main lobe average power. Consequently, the numerator of (2.1) is fixed to 1 for each $\mathbf{w}_\sigma$, and only $\|\mathbf{w}_\sigma\|_2^2$ is affected from the changed σ. Owing to the power conservation, we increase the average power of the designed beampattern by increasing $\|\mathbf{w}_\sigma\|_2^2$. Our method actually solves a constrained least-squares

Table 3.1: Performance metrics of 10 realizations for the narrowband transmit beamforming with 32 elements ULA and PAPR $\leq \sigma$ by using (3.8) with $P = 32$ and (3.9).

	$\sigma = 1$				$\sigma = 2$			
	$P^{\text{m}}_{\text{p-p}}$ (dB)		PSL (dB)		$P^{\text{m}}_{\text{p-p}}$ (dB)		PSL (dB)	
#	(3.8)	(3.9)	(3.8)	(3.9)	(3.8)	(3.9)	(3.8)	(3.9)
1	11.3612	15.6160	3.1925	8.2945	5.6236	1.6129	12.1844	28.0961
2	30.8255	16.8515	−16.9287	10.6568	6.1817	1.6942	12.0562	26.9719
3	23.6257	9.7819	−8.2082	21.0342	5.8507	1.3171	12.0346	27.4168
4	40.9117	14.5736	−27.3392	12.4576	5.2061	1.9593	12.3528	26.5214
5	11.8548	18.2355	5.3457	5.7737	5.3079	0.8077	12.3051	26.8273
6	8.7353	9.9073	6.8774	18.1448	6.0472	1.9623	12.0675	24.8963
7	8.7565	9.1254	5.6044	18.6577	5.7129	1.3863	12.1685	28.2452
8	7.8473	13.6166	8.6415	10.9221	6.0557	1.3971	11.9522	28.2863
9	9.0881	8.1983	6.4345	17.6731	5.9224	1.5101	12.0335	27.5341
10	25.4156	10.3492	−12.6823	14.1019	6.1426	1.9500	12.0646	26.7344

problem by trying to keep the ratio of average powers in the main lobe and sidelobe at some level. Hence, we observe that the main lobe average power is decreased when $\text{PAPR}\,(\mathbf{w}_\sigma)$ is increased, as shown in Figure 3.2. Owing to the same initial point, we can observe the achievable PAPR to which $\text{PAPR}\,(\mathbf{w}_\sigma)$ converges. This value is less than 5 for our parameters and initial point, and the results are not affected any more if we increase σ from 5 to 6.

For some PAPR values, the performance metrics of the designed beampatterns are given in Table 3.2. The performance metrics are as follows: $P^{\text{m}}_{\text{avg}}(\mathbf{w}_\sigma)$ is the main lobe average power (A.11), $P^{\text{m}}_{\text{min}}$ is the minimum power in the main lobe (A.12), $P^{\text{m}}_{\text{max}}$ is the maximum power in the main lobe (A.13), $P^{\text{m}}_{\text{p-p}}$ is the peak-to-peak power swing in the main lobe (A.14), $P_{\text{eff}}(\mathbf{w}_\sigma)$ is the power efficiency (A.15), $P^{\text{s}}_{\text{avg}}(\mathbf{w}_\sigma)$

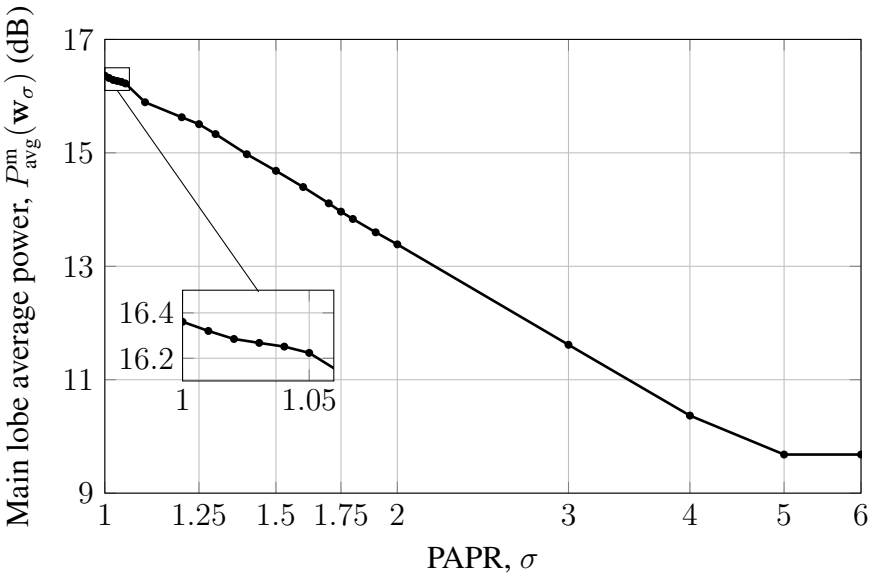


Figure 3.2: Main lobe average power versus PAPR for the narrowband transmit beamforming with 32 elements ULA (base-2 log scale is used for the PAPR axis).

is the sidelobe average power (A.16), $P^{\text{m/s}}_{\text{avg}}(\mathbf{w}_\sigma)$ is the ratio of average powers in the main lobe and sidelobe (A.19), and $R_{\max}(\mathbf{w}_\sigma)$ is the maximum range ratio (A.20). Further details on the performance metrics are given in Appendix A.

The beampatterns corresponding to $\sigma \in \{1, 1.6, 2, 5\}$ are shown in Figure 3.3. As a comparison, $\sigma = 1$ design (PAPR $= 1$ case) presents maximum average power over the illuminated sector ($P^{\text{m}}_{\text{avg}}(\mathbf{w}_\sigma) = 16.3609$ dB) at the expense of a large peak-to-peak power swing ($P^{\text{m}}_{\text{p-p}} = 9.7083$ dB). While $\sigma = 1.6$ design has about 2 dB less average power in the main lobe ($P^{\text{m}}_{\text{avg}}(\mathbf{w}_\sigma) = 14.3975$ dB), yet has a peak-to-peak power swing of 1.9813 dB in the main lobe. Furthermore, $\sigma = 1.6$ design has a better PSL than $\sigma = 1$ design by 6.5 dB. On the other extreme, when the PAPR constraint is ignored, which is the case of $\sigma = 5$ or $\sigma = 6$; the power swing in the main lobe is minimized to 0.5470 dB and PSL is improved to 28.7207 dB, yet the average power over the illuminated region is only 9.6810 dB which is about a quarter of $\sigma = 1$ design. The maximum range is decreased to 89.31% (about 10% range loss) and 68.08 (about 30% range loss) for $\sigma = 1.6$ and $\sigma = 5$ (or $\sigma = 6$) designs, respectively.

As noted previously, our aim is to let the designer trade-off between several metrics, as shown in Table 3.2, with a proper choice of PAPR value.

Table 3.2: Performance metrics for the narrowband transmit beamforming with 32 elements ULA and PAPR $\leq \sigma$.

σ	1	1.01	1.02	1.05	1.4	1.6	1.75	1.9	2	3	4	5, 6
$P^{\text{m}}_{\text{avg}}(\mathbf{w}_\sigma)$ (dB)	16.3609	16.3205	16.2850	16.2237	14.9758	14.3975	13.9638	13.5994	13.3865	11.6160	10.3677	9.6810
$P^{\text{m}}_{\text{p-p}}$ (dB)	9.7083	8.8528	8.5463	7.6724	3.6882	1.9813	1.0921	0.6956	0.5755	0.5489	0.5364	0.5470
PSL (dB)	19.9537	21.6018	21.7920	22.4198	24.1667	26.5034	27.3493	27.6373	28.7691	28.8939	28.6532	28.7207
PAPR $(\mathbf{w}_\sigma)$	1	1.01	1.02	1.05	1.4	1.6	1.75	1.9	2	3	4	4.6884
$P_{\text{eff}}(\mathbf{w}_\sigma)$ (%)	100.00	99.01	98.04	95.24	71.43	62.50	57.14	52.63	50.00	33.33	25.00	21.33
$P^{\text{m}}_{\text{min}}$ (dB)	9.9636	10.8448	11.1819	11.8113	12.5919	13.3165	13.3591	13.2474	13.0861	11.2791	9.9710	9.2830
$P^{\text{m}}_{\text{max}}$ (dB)	19.6719	19.6976	19.7282	19.4837	16.2801	15.2978	14.4512	13.9430	13.6616	11.8280	10.5074	9.8300
$P^{\text{s}}_{\text{avg}}(\mathbf{w}_\sigma)$ (dB)	-13.164	-13.295	-13.535	-13.620	-14.898	-15.525	-16.010	-16.385	-16.608	-18.382	-19.631	-20.318
$P^{\text{m/s}}_{\text{avg}}(\mathbf{w}_\sigma)$ (dB)	29.5250	29.6155	29.8201	29.8441	29.8736	29.9225	29.9742	29.9845	29.9943	29.9984	29.9992	29.9993
R_{max} (%)	100.00	99.77	99.56	99.21	92.34	89.31	87.11	85.30	84.26	76.10	70.82	68.08

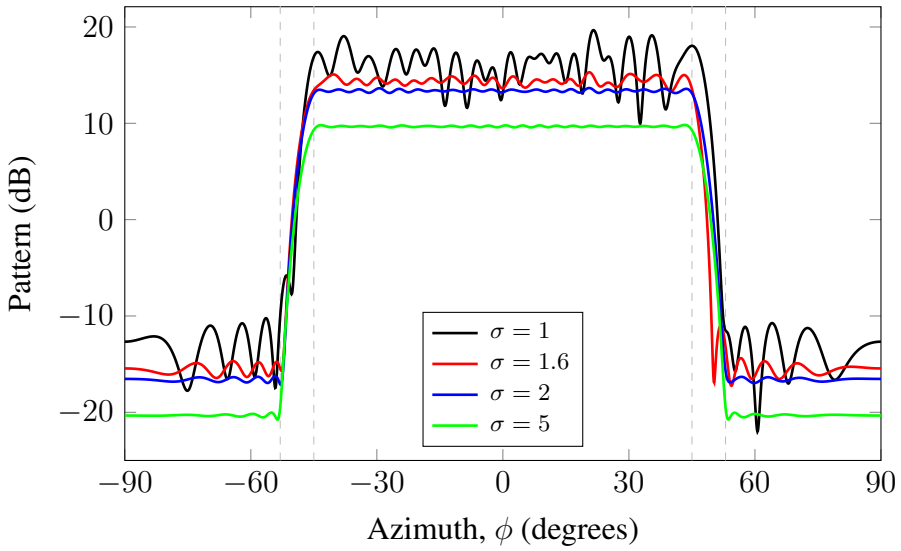


Figure 3.3: Beampattern versus PAPR for the narrowband transmit beamforming with 32 elements ULA.

3.6.3 Radar Waveform Design With Unimodular Sequence

The second experiment is related to a radar application, where a unimodular periodic sequence is required to communicate through a crowded channel. While the beamforming applications are our main concern, we also give this example to put forward the feasibility of the method in other applications.

We compare the proposed method PAPR-ADMM with the unimodular sequence design methods SHAPE [26], LPNN [21], ADMM [23] for the radar waveform design experiment described in [23]. The radar has the sample rate 810 kHz and the pulse duration 200 μs, and therefore, the sequence length is $N = 162$. With respect to the normalized frequency, the stopband intervals are $[0, 0.0617]$, $[0.0988, 0.2469]$, $[0.2593, 0.2840]$, $[0.3086, 0.3827]$, $[0.4074, 0.4938]$, $[0.5185, 0.5556]$, and $[0.9383, 1]$, and the rest of intervals correspond to the passband. The SHAPE[1] and ADMM require upper and lower masks for passband specified by the ripple constraint 0.2, whereas LPNN and PAPR-ADMM require the unity passband mask. The peak stopband level is $\eta = 0.01$ or $\eta = -20$ dB. The spectrum of each unimodular sequence is shown in Figure 3.4, and the corresponding peak-to-peak power swing in the passband, $P^{\text{m}}_{\text{p-p}}$,

[1] MATLAB code is available on `www.sal.ufl.edu/shape`

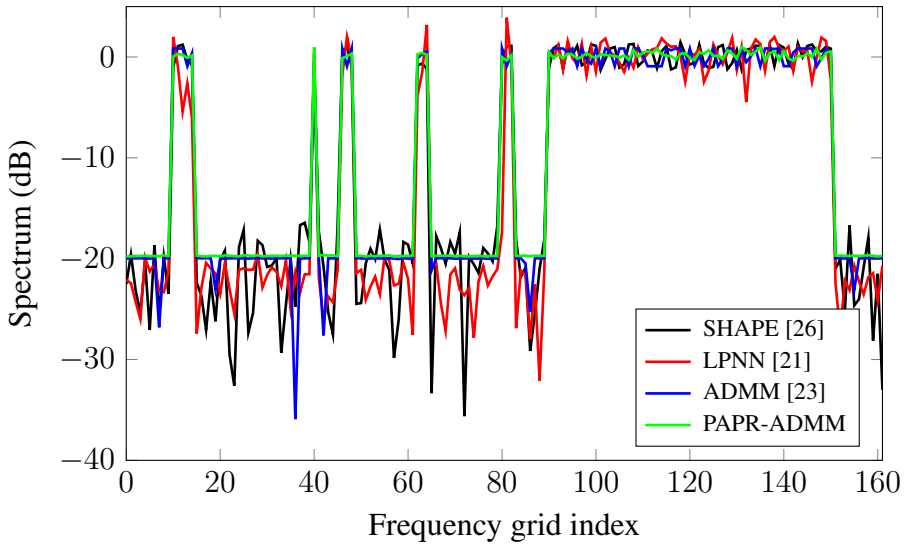


Figure 3.4: Spectrum comparison for radar waveform design.

Table 3.3: Comparison of the peak-to-peak power swing in the passband and PSL values for radar waveform design.

Method	SHAPE [26]	LPNN [21]	ADMM [23]	PAPR-ADMM
$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ (dB)	2.7351	9.9479	1.7607	1.3923
PSL (dB)	15.1515	13.5066	19.0311	19.2418

and PSL values are given in Table 3.3. Each sequence has comparable autocorrelation sidelobe levels, as shown in Figure 3.5.

Next, we compare our method with ANSLM [27]. For PAPR-ADMM, $\eta = -29$ dB, and the ratio of stopband and passband sample weighting coefficients is $\delta_{\mathrm{s}}/\delta_{\mathrm{m}} = 1000$ to increase the power fluctuations in the passband. The spectra of sequences designed by using ANSLM and PAPR-ADMM are shown in Figure 3.6. For ANSLM, PSL and $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ values are 24.1653 dB and 9.2059 dB, respectively. With PAPR-ADMM, we observe PSL as 25.7239 dB and $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ as 6.7645 dB. Both sequences have comparable autocorrelation sidelobe levels, as shown in Figure 3.7.

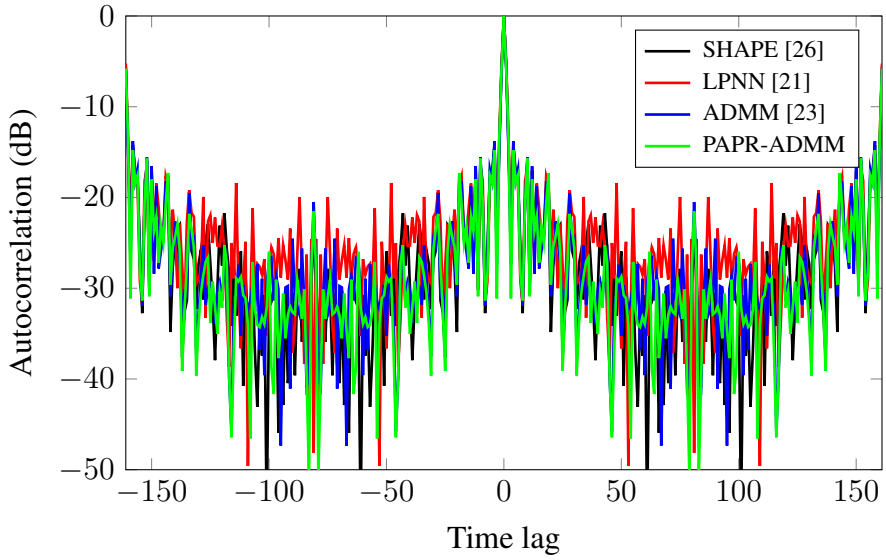


Figure 3.5: Autocorrelation function comparison for radar waveform design.

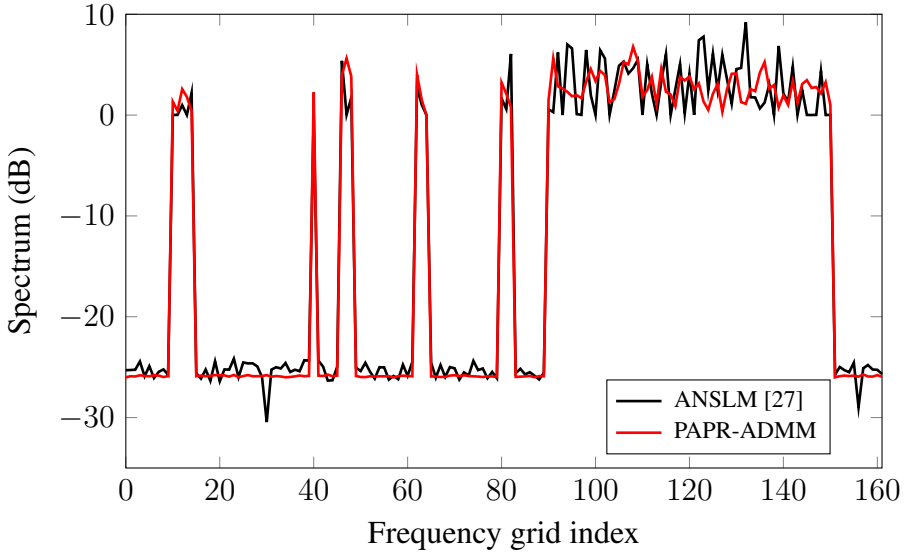


Figure 3.6: Spectrum comparison with ANSLM for radar waveform design.

3.6.4 Wideband Transmit Beamforming

We consider a sonar system having $N = 144$ elements uniform rectangular array (URA). The distance between neighboring sensors is $\lambda/2$ for 150 kHz with the propagation speed of 1500 m/s. The center frequency f_c is 140 kHz and the quality factor Q_f is 14. Using (3.23), we compute Δ_f as 10 kHz. Then, we take frequency samples $f_i \in \{135, 137.5, 140, 142.5, 145\}$ kHz, $i = 1, \ldots, 5$. Setting PAPR $= \sigma \in \{1,$

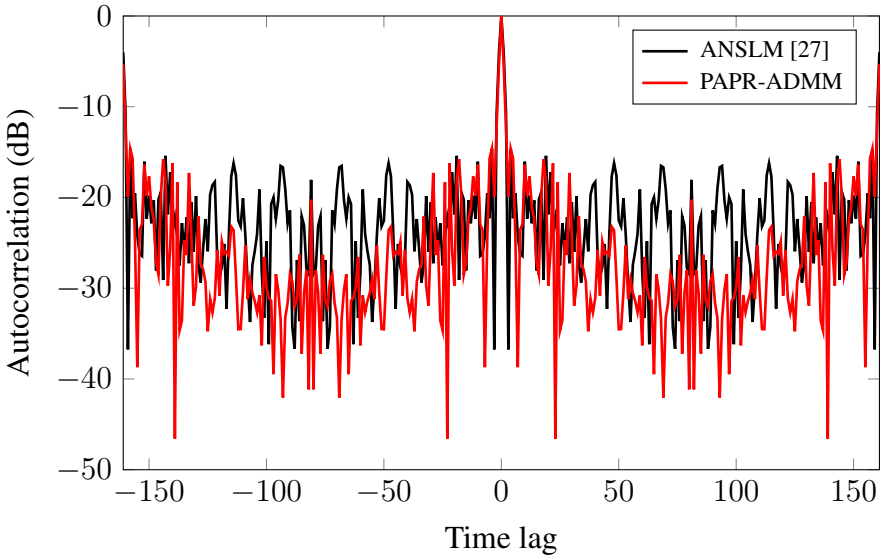
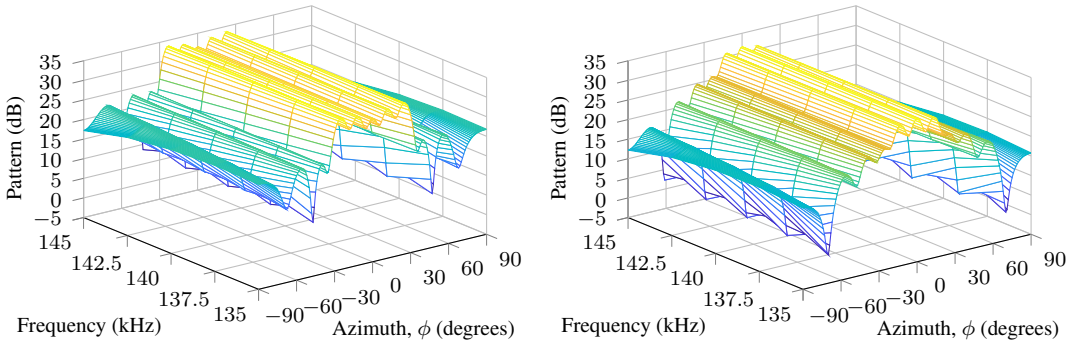


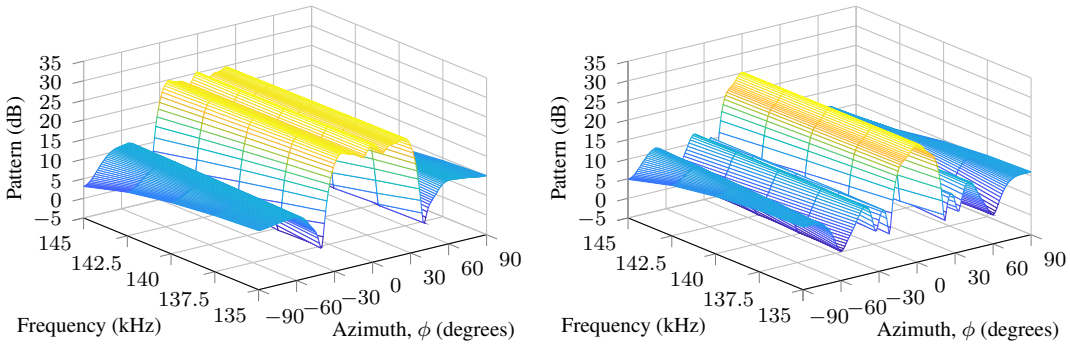
Figure 3.7: Autocorrelation function comparison with ANSLM for radar waveform design.

$2,\ 8,\ 36\}$, and using PAPR-ADMM initialized with the same vector $\mathbf{w}^{(1)}$, we find the transmitter weights $\mathbf{w}_\sigma$ that can be used to compute the two-dimensional flat-top beampatterns corresponding to these frequencies. The PSL of the desired beampattern is 15 dB. The main lobe is inside of an elliptical region centered at $(0°, 0°)$ and having semi-major and semi-minor axis of $30°$ and $15°$, respectively. The sidelobe is outside of an elliptical region centered at $(0°, 0°)$ and having semi-major and semi-minor axis of $40°$ and $20°$, respectively. The direction samples are taken by using a uniform grid spacing of $1°$. Since the array has uniform grid and the desired beampattern is symmetric in azimuth and elevation, we can assume that the transmitter weights are symmetric in horizontal and vertical directions. Therefore, we decrease the search space dimension from 144 to 36 and set σ as 36 at most.

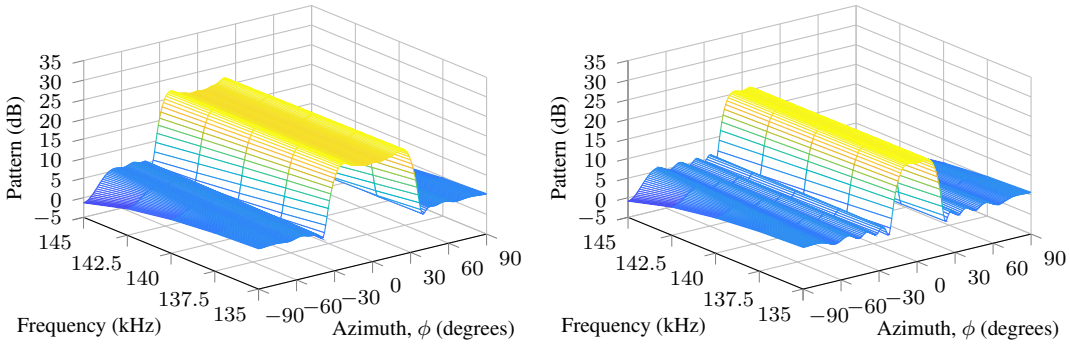
The beampatterns on cutting-planes, *i.e.*, $(\phi, 0°)$ and $(0°, \theta)$ planes, are shown in Figures 3.8 and 3.9. For $\sigma \in \{1, 2\}$, the main lobe has higher power swing than that of beampatterns corresponding to $\sigma \in \{8, 36\}$. Table 3.4 shows the performance metrics. The maximum of achievable PAPR is observed as $\mathrm{PAPR}\left(\mathbf{w}_\sigma\right) = 12.3001$ by relaxing the PAPR constraint as $\sigma = 36$.



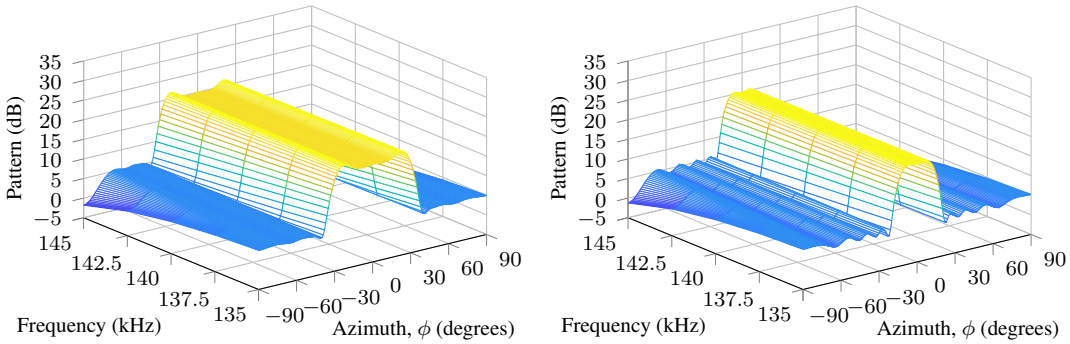
(a)



(b)



(c)



(d)

Figure 3.8: Beampatterns on cutting-planes for the wideband transmit beamforming with 144 elements URA and PAPR values, (a) $\sigma = 1$, (b) $\sigma = 2$, (c) $\sigma = 8$ and (d) $\sigma = 36$.

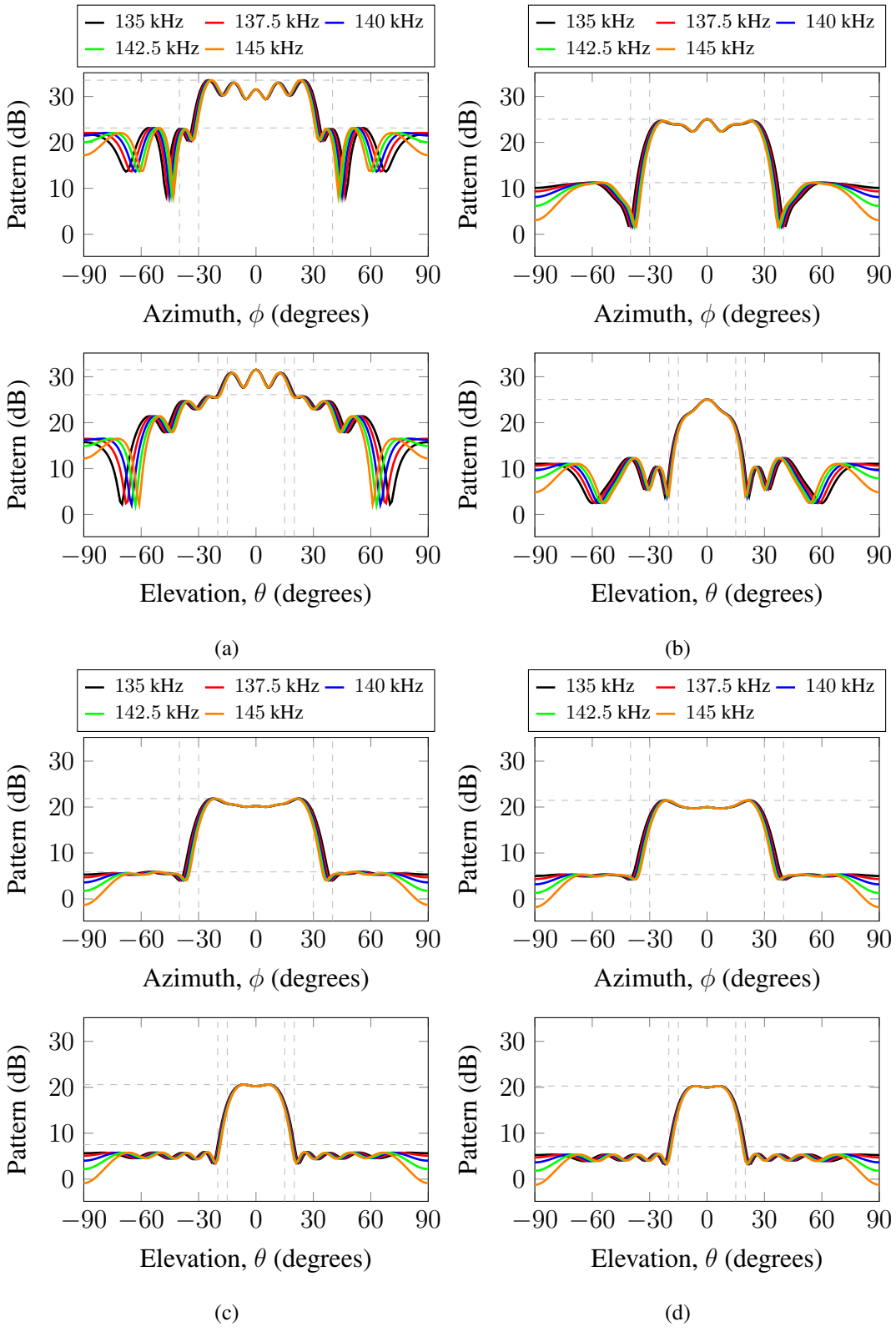


Figure 3.9: Two-dimensional viewed beampatterns on cutting-planes for the wide-band transmit beamforming with 144 elements URA and PAPR values, (a) $\sigma = 1$, (b) $\sigma = 2$, (c) $\sigma = 8$ and (d) $\sigma = 36$.

Table 3.4: Performance metrics for the wideband transmit beamforming with 144 elements URA and PAPR $\leq \sigma$.

σ	1	2	8	36
$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ (dB)	12.5855	8.6816	7.4218	7.1117
PSL (dB)	-5.1418	2.7020	5.6946	5.9029
$\mathrm{PAPR}\left(\mathbf{w}_\sigma\right)$	1	2	8	12.3001
$P_{\mathrm{eff}}(\mathbf{w}_\sigma)$ (%)	100.00	50.00	12.50	8.13

3.7 Conclusion

The main goal of this chapter is to examine the transmit beamformer design problem under the peak-to-average-power-ratio (PAPR) constraint. The PAPR constraint is included in the problem setting to provide a trade-off mechanism between the beampattern shape, average power in the main lobe and other classical metrics, such as the integrated sidelobe level and the peak sidelobe level. By using an alternating direction method of multipliers (ADMM) formulation, we combine the phase retrieval method PhareADMM [24] and the alternating projection method [17] to solve the problem. Our formulation includes the distinct penalty parameters in the augmented Lagrangian function for the beam shape and PAPR constraints. Thus, we can control the constraint violation of the beam shape and PAPR separately. Owing to the generality of formulation, the suggested method PAPR-ADMM can be applied to both narrowband and wideband beamforming.

CHAPTER 4

TRANSMIT BEAMFORMING PROBLEM WITH PAPR CONSTRAINT: CONSENSUS ADMM-BASED SOLUTION

In this chapter, we present a solution for the transmit beamformer design problem under the PAPR constraint after expressing the problem in a consensus form. The solution is based on the consensus ADMM. Before introducing the solution, we briefly describe the consensus ADMM in the next section.

4.1 Consensus ADMM

We let a convex optimization problem be

$$\begin{aligned} &\text{minimize} \quad \sum_{m=1}^{M} f_m(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{x} \in \mathcal{S} \end{aligned} \tag{4.1}$$

with variable $\mathbf{x} \in \mathbb{R}^N$, where the functions $f_m : \mathbb{R}^N \to \mathbb{R}$, $m = 1, \ldots, M$, are convex, and the set $\mathcal{S} \subseteq \mathbb{R}^N$ is nonempty, closed, and convex.

When the objective function and constraints of an optimization problem can be split into terms, as in $f(\mathbf{x}) = \sum_{m=1}^{M} f_m(\mathbf{x})$, one of the generic approaches is the consensus optimization, [28]. We can construct a consensus optimization problem equivalent to (4.1) as

$$\begin{aligned} &\text{minimize} \quad \mathcal{I}(\mathbf{z}) + \sum_{m=1}^{M} f_m(\mathbf{x}_m) \\ &\text{subject to} \quad \mathbf{x}_m = \mathbf{z}, \quad m = 1, \ldots, M, \end{aligned} \tag{4.2}$$

with variables $\mathbf{x}_1, \ldots, \mathbf{x}_M, \mathbf{z} \in \mathbb{R}^N$, where $\mathcal{I}_{\mathcal{S}} : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the indicator

function of $\mathcal{S}$, that is

$$\mathcal{I}(\mathbf{z}) = \begin{cases} 0 & \text{if } \mathbf{z} \in \mathcal{S} \\ +\infty & \text{otherwise.} \end{cases} \tag{4.3}$$

The ADMM steps for solving the consensus problem (4.2) can be determined from the augmented Lagrangian function given as

$$\mathcal{L}_\rho\big(\{\mathbf{x}_m, \mathbf{y}_m\}_{m=1}^M, \mathbf{z}\big) = \mathcal{I}(\mathbf{z}) + \sum_{m=1}^{M} \big(f_m(\mathbf{x}_m) + \mathbf{y}_m^{\mathsf{T}}(\mathbf{x}_m - \mathbf{z}) + (\rho/2)\|\mathbf{x}_m - \mathbf{z}\|_2^2\big), \tag{4.4}$$

where $\rho > 0$ is the step size and $\mathbf{y}_m \in \mathbb{R}^N$, $m = 1, \ldots, M$, are the dual variables. These steps are given as follows:

$$\big\{\mathbf{x}_m^{(k+1)}\big\}_{m=1}^M = \underset{\mathbf{x}_1, \ldots, \mathbf{x}_M \in \mathbb{R}^N}{\operatorname{argmin}} \quad \mathcal{L}_\rho\big(\big\{\mathbf{x}_m, \mathbf{y}_m^{(k)}\big\}_{m=1}^M, \mathbf{z}^{(k)}\big), \tag{4.5a}$$

$$\mathbf{z}^{(k+1)} = \underset{\mathbf{z} \in \mathbb{R}^N}{\operatorname{argmin}} \quad \mathcal{L}_\rho\big(\big\{\mathbf{x}_m^{(k+1)}, \mathbf{y}_m^{(k)}\big\}_{m=1}^M, \mathbf{z}\big), \tag{4.5b}$$

$$\mathbf{y}_m^{(k+1)} = \mathbf{y}_m^{(k)} + \rho\big(\mathbf{x}_m^{(k+1)} - \mathbf{z}^{(k+1)}\big), \quad m = 1, \ldots, M, \tag{4.5c}$$

After eliminating constants or completing linear and quadratic terms, and using $\mathbf{u}_m = \mathbf{y}_m/\rho$, $m = 1, \ldots, M$, the scaled form of ADMM steps for (4.2) are

$$\mathbf{x}_m^{(k+1)} = \underset{\mathbf{x}_m \in \mathbb{R}^N}{\operatorname{argmin}} \left(f_m(\mathbf{x}_m) + (\rho/2)\big\|\mathbf{x}_m - \mathbf{z}^{(k)} + \mathbf{u}_m^{(k)}\big\|_2^2\right), \; m = 1, \ldots, M, \tag{4.6a}$$

$$\mathbf{z}^{(k+1)} = \underset{\mathbf{z} \in \mathcal{S}}{\operatorname{argmin}} \quad \big\|\bar{\mathbf{x}}^{(k+1)} - \mathbf{z} + \bar{\mathbf{u}}^{(k)}\big\|_2^2, \tag{4.6b}$$

$$\mathbf{u}_m^{(k+1)} = \mathbf{u}_m^{(k)} + \big(\mathbf{x}_m^{(k+1)} - \mathbf{z}^{(k+1)}\big), \quad m = 1, \ldots, M, \tag{4.6c}$$

where $\bar{\mathbf{x}}^{(k+1)} = (1/M)\sum_{m=1}^M \mathbf{x}_m^{(k+1)}$ and $\bar{\mathbf{u}}^{(k)} = (1/M)\sum_{m=1}^M \mathbf{u}_m^{(k)}$. Note that the update (4.6b), which is a simplified expression equivalent to the actual problem

$$\mathbf{z}^{(k+1)} = \underset{\mathbf{z} \in \mathbb{R}^N}{\operatorname{argmin}} \left(\mathcal{I}(\mathbf{z}) + (M\rho/2)\big\|\bar{\mathbf{x}}^{(k+1)} - \mathbf{z} + \bar{\mathbf{u}}^{(k)}\big\|_2^2\right), \tag{4.7}$$

denotes the Euclidean projection of $\bar{\mathbf{x}}^{(k+1)} + \bar{\mathbf{u}}^{(k)}$ on $\mathcal{S}$. The updates (4.6a) and (4.6c) can be computed in parallel for $m = 1, \ldots, M$. The ADMM with update steps like (4.6) is referred to as consensus ADMM due to the applicability of the parallel computation, [28].

The transmit beampattern design problem (2.10) has a convenient form to apply the consensus ADMM steps similar to that given in (4.6). In the next section, we introduce a consensus ADMM-based solution to this problem.

4.2 Proposed Method

We suggest introducing local variables $\mathbf{w}_\ell \in \mathbb{C}^N$ with equality constraints $\mathbf{w}_\ell = \mathbf{v}$, $\ell = 1, \ldots, L$, where $\mathbf{v} \in \mathbb{C}^N$ is the common global vector, into (2.10) for constructing a consensus optimization problem as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{\ell=1}^{L} h_\ell \left(b_\ell - |\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell|\right)^2 \\ \text{subject to} \quad & \mathbf{w}_\ell = \mathbf{v}, \quad \ell = 1, \ldots, L \\ & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P \end{aligned} \tag{4.8}$$

with variables $\mathbf{w}_1, \ldots, \mathbf{w}_L, \mathbf{v} \in \mathbb{C}^N$. Here the common global vector $\mathbf{v}$ also prevents the appearance of the PAPR constraint argument in the objective function.

To circumvent the absolute value operator in the objective function, we introduce the auxiliary variables $\alpha_\ell \in \mathbb{R}_+$ and $\beta_\ell \in \mathbb{R}$ as

$$\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell = \alpha_\ell \, e^{j\beta_\ell}, \tag{4.9}$$

where

$$\alpha_\ell = |\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell| \geq 0 \text{ and } \beta_\ell = \angle(\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell),$$

for $\ell = 1, \ldots, L$. After using (4.9) and replacing $|\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell|$ with α_ℓ in (4.8), the problem becomes

$$\begin{aligned} \text{minimize} \quad & \sum_{\ell=1}^{L} h_\ell \left(b_\ell - \alpha_\ell\right)^2 \\ \text{subject to} \quad & \mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell = \alpha_\ell \, e^{j\beta_\ell}, \quad \ell = 1, \ldots, L \\ & \mathbf{w}_\ell = \mathbf{v}, \quad \ell = 1, \ldots, L \\ & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P \end{aligned} \tag{4.10}$$

with variables $\mathbf{w}_1, \ldots, \mathbf{w}_L, \mathbf{v} \in \mathbb{C}^N$, $\alpha_1, \ldots, \alpha_L \in \mathbb{R}_+$ and $\beta_1, \ldots, \beta_L \in \mathbb{R}$. Thus, the problem formulation in (2.10) is converted to (4.10) having a convex objective function and nonconvex constraints.

The augmented Lagrangian for the application of ADMM on (4.10) requires the def-

inition of penalty parameters. We present the augmented Lagrangian function as

$$
\begin{aligned}
&\mathcal{L}_{\rho_L,\rho_N}\big(\{\mathbf{w}_\ell,\boldsymbol{\lambda}_\ell\}_{\ell=1}^{L},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{v},\mathbf{y}\big)\\
&\quad=\sum_{\ell=1}^{L} h_\ell\,(b_\ell-\alpha_\ell)^2\\
&\qquad+\sum_{\ell=1}^{L}\mathsf{Re}\left\{y_\ell\right\}\mathsf{Re}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}_\ell-\alpha_\ell\,e^{j\beta_\ell}\right\}+\sum_{\ell=1}^{L}\mathsf{Im}\left\{y_\ell\right\}\mathsf{Im}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}_\ell-\alpha_\ell\,e^{j\beta_\ell}\right\}\\
&\qquad+\frac{\rho_L}{2}\sum_{\ell=1}^{L}\big(\mathsf{Re}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}_\ell-\alpha_\ell\,e^{j\beta_\ell}\right\}\big)^2+\frac{\rho_L}{2}\sum_{\ell=1}^{L}\big(\mathsf{Im}\left\{\mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}_\ell-\alpha_\ell\,e^{j\beta_\ell}\right\}\big)^2\\
&\qquad+\sum_{\ell=1}^{L}\sum_{n=1}^{N}\mathsf{Re}\left\{\lambda_{\ell,n}\right\}\mathsf{Re}\left\{w_{\ell,n}-v_n\right\}+\sum_{\ell=1}^{L}\sum_{n=1}^{N}\mathsf{Im}\left\{\lambda_{\ell,n}\right\}\mathsf{Im}\left\{w_{\ell,n}-v_n\right\}\\
&\qquad+\frac{\rho_N}{2}\sum_{\ell=1}^{L}\sum_{n=1}^{N}(\mathsf{Re}\left\{w_{\ell,n}-v_n\right\})^2+\frac{\rho_N}{2}\sum_{\ell=1}^{L}\sum_{n=1}^{N}(\mathsf{Im}\left\{w_{\ell,n}-v_n\right\})^2, \qquad (4.11)
\end{aligned}
$$

where $\rho_L>0$ and $\rho_N>0$ are two distinct penalty parameters corresponding to beampattern shape and PAPR constraints, respectively; $\mathbf{y}=[\,y_1\;\ldots\;y_L\,]^{\mathsf{T}}\in\mathbb{C}^L$ and $\boldsymbol{\lambda}_\ell=[\,\lambda_{\ell,1}\;\ldots\;\lambda_{\ell,N}\,]^{\mathsf{T}}\in\mathbb{C}^N$, $\ell=1,\ldots,L$, are the dual variables; $\boldsymbol{\alpha}=[\,\alpha_1\;\ldots\;\alpha_L\,]^{\mathsf{T}}\in\mathbb{R}_+^L$ and $\boldsymbol{\beta}=[\,\beta_1\;\ldots\;\beta_L\,]^{\mathsf{T}}\in\mathbb{R}^L$ are the auxiliary variables. In (4.11), the real and imaginary parts are separately considered due to the nature of ADMM, which is built on the real-valued and convex functions. Here, we utilize two different penalty parameters ρ_L and ρ_N in this application, as in PAPR-ADMM, rather than the usual definition of the augmented Lagrangian functions, where a single penalty parameter is assigned to all constraints.

We use ADMM steps to minimize (4.11) under the PAPR constraint with respect to variable sets $\{\mathbf{w}_\ell\}_{\ell=1}^{L}$ and $\{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{v}\}$ separately. The ADMM steps involved in each iteration are as follows:

Step-1: For given variables at the iteration number k, $\big\{\mathbf{w}_\ell^{(k)},\boldsymbol{\lambda}_\ell^{(k)}\big\}_{\ell=1}^{L}$ and $\mathbf{y}^{(k)}$, the auxiliary variables at $k+1$ can be determined by solving the problem

$$
\begin{aligned}
\{\boldsymbol{\alpha}^{(k+1)},\boldsymbol{\beta}^{(k+1)},\mathbf{v}^{(k+1)}\}=\;&\underset{}{\text{argmin}} \quad \mathcal{L}_{\rho_L,\rho_N}\big(\big\{\mathbf{w}_\ell^{(k)},\boldsymbol{\lambda}_\ell^{(k)}\big\}_{\ell=1}^{L},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{v},\mathbf{y}^{(k)}\big)\\
&\text{subject to}\quad \mathrm{PAPR}\,(\mathbf{v})\le\sigma\\
&\qquad\qquad\;\; \|\mathbf{v}\|_2^2=P
\end{aligned}
\qquad (4.12)
$$

with variables $\boldsymbol{\alpha} \in \mathbb{R}_+^L$, $\boldsymbol{\beta} \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{C}^N$.

We introduce vectors $\mathbf{z} = [\, z_1 \, \ldots \, z_L \,]^{\mathsf{T}} = \mathbf{y}/\rho_L \in \mathbb{C}^L$ and $\boldsymbol{\tau}_\ell = [\, \tau_{\ell,1} \, \ldots \, \tau_{\ell,N} \,]^{\mathsf{T}} = \boldsymbol{\lambda}_\ell/\rho_N \in \mathbb{C}^N$ for using the scaled form of ADMM. We also consider a function

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}) = \sum_{\ell=1}^{L} h_\ell \, (b_\ell - \alpha_\ell)^2 + \frac{\rho_L}{2} \sum_{\ell=1}^{L} \left| \mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell^{(k)} + z_\ell^{(k)} - \alpha_\ell \, e^{j\beta_\ell} \right|^2 + \frac{\rho_N L}{2} \|\bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} - \mathbf{v}\|_2^2, \quad (4.13)$$

where $\bar{\mathbf{w}}^{(k)} = (1/L) \sum_{\ell=1}^{L} \mathbf{w}_\ell^{(k)}$ and $\bar{\boldsymbol{\tau}}^{(k)} = (1/L) \sum_{\ell=1}^{L} \boldsymbol{\tau}_\ell^{(k)}$, such that

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}) \overset{c}{=} \mathcal{L}_{\rho_L, \rho_N} \left(\left\{ \mathbf{w}_\ell^{(k)}, \boldsymbol{\lambda}_\ell^{(k)} \right\}_{\ell=1}^{L}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}, \mathbf{y}^{(k)} \right).$$

Thus, a problem equivalent to (4.12) can be given as

$$\begin{aligned} \{\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{v}^{(k+1)}\} = \operatorname{argmin} \quad & g(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{v}) \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{v}) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P \end{aligned} \quad (4.14)$$

with variables $\boldsymbol{\alpha} \in \mathbb{R}_+^L$, $\boldsymbol{\beta} \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{C}^N$.

In (4.14), $\mathbf{v}$ appears only in the last term of objective function, and therefore $\mathbf{v}^{(k+1)}$ can be determined by solving the equivalent problem

$$\begin{aligned} \mathbf{v}^{(k+1)} = \operatorname*{argmin}_{\mathbf{v} \in \mathbb{C}^N} \quad & \left\| \bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} - \mathbf{v} \right\|_2^2 \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{v}) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P, \end{aligned} \quad (4.15)$$

whose solution can be determined by following the steps given in [17, Algorithm 2], see also Algorithm 2.1. As with PAPR-ADMM, we suggest replacing P in (4.15) by $\left\| \bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} \right\|_2^2$ to obtain the problem

$$\begin{aligned} \mathbf{v}^{(k+1)} = \operatorname*{argmin}_{\mathbf{v} \in \mathbb{C}^N} \quad & \left\| \bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} - \mathbf{v} \right\|_2^2 \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{v}) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = \left\| \bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} \right\|_2^2, \end{aligned} \quad (4.16)$$

whose solution improves the performance of suggested method.

Similarly, $\boldsymbol{\beta}$ appears only in the middle term of the objective function of (4.14), and an equivalent problem of determining $\boldsymbol{\beta}^{(k+1)}$ is

$$\boldsymbol{\beta}^{(k+1)} = \operatorname*{argmin}_{\boldsymbol{\alpha} \in \mathbb{R}_+^L, \boldsymbol{\beta} \in \mathbb{R}^L} \quad \sum_{\ell=1}^{L} \left| \mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell^{(k)} + z_\ell^{(k)} - \alpha_\ell e^{j\beta_\ell} \right|^2. \quad (4.17)$$

The solution of (4.17), $\boldsymbol{\beta}^{(k+1)} = \left[\, \beta_1^{(k+1)} \;\ldots\; \beta_L^{(k+1)} \,\right]^{\mathsf{T}}$, has the entries

$$\beta_\ell^{(k+1)} = \angle\big(\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell^{(k)} + z_\ell^{(k)}\big), \quad \ell = 1, \ldots, L. \tag{4.18}$$

We substitute $\beta_\ell^{(k+1)}$ for β_ℓ in the objective function of (4.14) and obtain an equivalent problem of finding $\boldsymbol{\alpha}^{(k+1)}$ as follows:

$$\boldsymbol{\alpha}^{(k+1)} = \underset{\boldsymbol{\alpha} \in \mathbb{R}_+^L}{\operatorname{argmin}} \quad \left(\sum_{\ell=1}^{L} h_\ell \, (b_\ell - \alpha_\ell)^2 + (\rho_L/2) \sum_{\ell=1}^{L} (r_\ell - \alpha_\ell)^2 \right), \tag{4.19}$$

where $r_\ell = \big|\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell^{(k)} + z_\ell^{(k)}\big|$, $\ell = 1, \ldots, L$. Except the definition of r_ℓ, the expressions of (4.19) and (3.13) are identical. As with (3.13), the solution of (4.19) is

$$\alpha_\ell^{(k+1)} = \frac{\rho_L \, r_\ell + 2 h_\ell \, b_\ell}{\rho_L + 2 h_\ell}, \quad \ell = 1, \ldots, L \tag{4.20}$$

due to the parameter settings, $\rho_L > 0$, $r_\ell \geq 0$, $h_\ell \geq 0$ and $b_\ell \geq 0$.

Step-2: We update $\{\mathbf{w}_\ell\}_{\ell=1}^L$ by solving

$$\big\{\mathbf{w}_\ell^{(k+1)}\big\}_{\ell=1}^L = \underset{\{\mathbf{w}_\ell\}_{\ell=1}^L \in \mathbb{C}^N}{\operatorname{argmin}} \quad \mathcal{L}_{\rho_L, \rho_N}\big(\big\{\mathbf{w}_\ell, \boldsymbol{\tau}_\ell^{(k)}\big\}_{\ell=1}^L, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{v}^{(k+1)}, \mathbf{z}^{(k)}\big). \tag{4.21}$$

According to (4.21), all variables except $\{\mathbf{w}_\ell\}_{\ell=1}^L$ are fixed in (4.11). Thus, a problem equivalent to (4.21) can be given as

$$\big\{\mathbf{w}_\ell^{(k+1)}\big\}_{\ell=1}^L = \underset{\mathbf{w}_1, \ldots, \mathbf{w}_L \in \mathbb{C}^N}{\operatorname{argmin}} \quad \left(\sum_{\ell=1}^{L} \big|\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell + z_\ell^{(k)} - \gamma_\ell^{(k+1)}\big|^2 + (\rho_N/\rho_L) \sum_{\ell=1}^{L} \big\|\mathbf{w}_\ell + \boldsymbol{\tau}_\ell^{(k)} - \mathbf{v}^{(k+1)}\big\|_2^2 \right), \tag{4.22}$$

where $\gamma_\ell^{(k+1)} = \alpha_\ell^{(k+1)} e^{j\beta_\ell^{(k+1)}}$, $\ell = 1, \ldots, L$.

The equivalent problems for $\mathbf{w}_\ell^{(k+1)}$, $\ell = 1, \ldots, L$, can be given as

$$\mathbf{w}_\ell^{(k+1)} = \underset{\mathbf{w}_\ell \in \mathbb{C}^N}{\operatorname{argmin}} \quad \left(\big|\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell + z_\ell^{(k)} - \gamma_\ell^{(k+1)}\big|^2 + \frac{\rho_N}{\rho_L} \big\|\mathbf{w}_\ell + \boldsymbol{\tau}_\ell^{(k)} - \mathbf{v}^{(k+1)}\big\|_2^2 \right). \tag{4.23}$$

For $\ell = 1, \ldots, L$, by setting the gradient of the objective function in (4.23) with respect to $\mathbf{w}_\ell^*$ (conjugate of $\mathbf{w}_\ell$) equal to zero at $\mathbf{w}_\ell^{(k+1)}$ [36], we get

$$0 = \mathbf{a}_\ell \big(\mathbf{a}_\ell^{\mathsf{H}} \mathbf{w}_\ell^{(k+1)} + z_\ell^{(k)} - \gamma_\ell^{(k+1)}\big) + \frac{\rho_N}{\rho_L} \big(\mathbf{w}_\ell^{(k+1)} + \boldsymbol{\tau}_\ell^{(k)} - \mathbf{v}^{(k+1)}\big),$$

and find

$$\begin{aligned}\mathbf{w}_\ell^{(k+1)} &= \left(\mathbf{a}_\ell\mathbf{a}_\ell^{\mathsf{H}} + \frac{\rho_N}{\rho_L}\mathbf{I}_N\right)^{-1}\left(\mathbf{a}_\ell\big(\gamma_\ell^{(k+1)} - z_\ell^{(k)}\big) + \frac{\rho_N}{\rho_L}\big(\mathbf{v}^{(k+1)} - \boldsymbol{\tau}_\ell^{(k)}\big)\right)\\ &= \frac{\rho_L}{\rho_N}\left(\mathbf{I}_N - \frac{\mathbf{a}_\ell\mathbf{a}_\ell^{\mathsf{H}}}{\frac{\rho_N}{\rho_L} + \mathbf{a}_\ell^{\mathsf{H}}\mathbf{a}_\ell}\right)\left(\mathbf{a}_\ell\big(\gamma_\ell^{(k+1)} - z_\ell^{(k)}\big) + \frac{\rho_N}{\rho_L}\big(\mathbf{v}^{(k+1)} - \boldsymbol{\tau}_\ell^{(k)}\big)\right).\end{aligned} \tag{4.24}$$

Step-3: The dual updates are as follows:

$$z_\ell^{(k+1)} = z_\ell^{(k)} + \mathbf{a}_\ell^{\mathsf{H}}\mathbf{w}_\ell^{(k+1)} - \gamma_\ell^{(k+1)}, \quad \ell = 1,\ldots,L, \tag{4.25}$$

$$\boldsymbol{\tau}_\ell^{(k+1)} = \boldsymbol{\tau}_\ell^{(k)} + \mathbf{w}_\ell^{(k+1)} - \mathbf{v}^{(k+1)}, \quad \ell = 1,\ldots,L. \tag{4.26}$$

We have described the consensus ADMM steps to compute the transmitter weights satisfying the PAPR constraint. These steps are iteratively applied until the condition

$$\frac{\left\|\bar{\mathbf{w}}^{(k)} - \bar{\mathbf{w}}^{(k-1)}\right\|_2}{\left\|\bar{\mathbf{w}}^{(k-1)}\right\|_2} \leq 10^{-6}$$

or the maximum iteration number $k = k_{\max}$ is satisfied. After termination of iterations, the solution is determined from

$$\begin{aligned}\mathbf{w}^\star = \underset{\mathbf{w}\in\mathbb{C}^N}{\operatorname{argmin}} \quad & \|\bar{\mathbf{w}}^{\text{final}} + \bar{\boldsymbol{\tau}}^{\text{final}} - \mathbf{w}\|_2^2 \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{w}) \leq \sigma \\ & \|\mathbf{w}\|_2^2 = P,\end{aligned} \tag{4.27}$$

where the vectors $\bar{\mathbf{w}}^{\text{final}}$ and $\bar{\boldsymbol{\tau}}^{\text{final}}$ are formed by the final entries of mean vectors $\bar{\mathbf{w}}$ and $\bar{\boldsymbol{\tau}}$, respectively, after terminating the iterations. It should be noted that (4.27) has the form of the common global variable $\mathbf{v}$ update given in (4.16), where the objective function depends on the mean vector $\bar{\boldsymbol{\tau}}$ of the dual variables $\{\boldsymbol{\tau}_\ell\}_{\ell=1}^L$, in contrast to the objective function given in (3.22) for the ADMM-based formulation, where $\mathbf{v}$ is an auxiliary variable.

The consensus ADMM-based method, PAPR-cADMM, for computing the transmitter weights under the PAPR constraint is outlined in Algorithm 4.1.

The convergence of PAPR-cADMM is similar to that of PAPR-ADMM discussed in Section 3.4. We suggest using the formulation (4.16) instead of (4.15) for improving the performance. This suggestion is examined via an experiment given in

Algorithm 4.1: Proposed method PAPR-cADMM for the narrowband transmit beamformer design under a PAPR constraint.

Input : $\bar{\mathbf{w}}^{(1)} \in \mathbb{C}^N$, $\mathbf{A} \in \mathbb{C}^{N \times L}$, $\mathbf{b} \in \mathbb{R}_+^L$, $\mathbf{h} \in \mathbb{R}_+^L$, $\rho_L > 0$, $\rho_N > 0$, $\sigma \geq 1$, $P \in \mathbb{R}_{++}$, $k_{\max} \in \{1, 2, \ldots\}$

Output : $\mathbf{w}^\star \in \mathbb{C}^N$

Initialization: $\bar{\mathbf{w}}^{(0)} = 2\bar{\mathbf{w}}^{(1)}$, $\left\{\mathbf{w}_\ell^{(1)} = \bar{\mathbf{w}}^{(1)}, \boldsymbol{\lambda}_\ell^{(1)} = \mathbf{0}_N\right\}_{\ell=1}^L$, $\mathbf{y}^{(1)} = \mathbf{0}_L$, $k = 1$

Assign $\mathbf{z}^{(1)} = \mathbf{y}^{(1)}/\rho_L$, and $\boldsymbol{\tau}_\ell^{(1)} = \boldsymbol{\lambda}_\ell^{(1)}/\rho_N$, $\ell = 1, \ldots, L$

while $k \leq k_{\max}$ and $\left(\left\|\bar{\mathbf{w}}^{(k)} - \bar{\mathbf{w}}^{(k-1)}\right\|_2 / \left\|\bar{\mathbf{w}}^{(k-1)}\right\|_2\right) > 10^{-6}$ **do**

// Step-1

Determine $\mathbf{v}^{(k+1)}$ from (4.16)

Compute $\boldsymbol{\beta}^{(k+1)}$ from (4.18)

Compute $\boldsymbol{\alpha}^{(k+1)}$ from (4.20)

// Step-2

Determine $\mathbf{w}_\ell^{(k+1)}$ from (4.24) by using $\gamma_\ell^{(k+1)} = \alpha_\ell^{(k+1)} e^{j\beta_\ell^{(k+1)}}$, $\ell = 1, \ldots, L$

// Step-3

Compute $\mathbf{z}^{(k+1)}$ from (4.25)

Compute $\boldsymbol{\tau}_\ell^{(k+1)}$, $\ell = 1, \ldots, L$, from (4.26)

$\bar{\mathbf{w}}^{(k+1)} = (1/L) \sum_{\ell=1}^L \mathbf{w}_\ell^{(k+1)}$

$\bar{\boldsymbol{\tau}}^{(k+1)} = (1/L) \sum_{\ell=1}^L \boldsymbol{\tau}_\ell^{(k+1)}$

$k = k + 1$

end

Determine $\mathbf{w}^\star$ from (4.27) by using $\bar{\mathbf{w}}^{\text{final}} = \bar{\mathbf{w}}^{(k)}$ and $\bar{\boldsymbol{\tau}}^{\text{final}} = \bar{\boldsymbol{\tau}}^{(k)}$

return $\mathbf{w}^\star$

Section 4.5.1. Next, we present an extension of PAPR-cADMM to the wideband transmit beamformer design.

4.3 An Extension of the Proposed Method to Wideband Beamforming

As described in Section 3.3, the main task of a wideband beamforming application is to set the sensor weights yielding the desired beampattern within the negligible difference for different operating frequencies/wavelengths. With a minor modification of (3.24) to invoke the consensus optimization, the wideband beamforming problem with PAPR constraint is given as

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{F} \sum_{\ell=1}^{L} h_{i,\ell} \left(b_{i,\ell} - |\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}| \right)^2 \\ \text{subject to} \quad & \mathrm{PAPR}\left(\mathbf{w}\right) \leq \sigma \\ & \|\mathbf{w}\|_2^2 = P \end{aligned} \tag{4.28}$$

with variable $\mathbf{w} \in \mathbb{C}^N$.

Similar to (4.8), we introduce the local variables $\{\mathbf{w}_i\}_{i=1}^F \in \mathbb{C}^N$ and let $\mathbf{v} \in \mathbb{C}^N$ be a common global vector in the consensus optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{F} \sum_{\ell=1}^{L} h_{i,\ell} \left(b_{i,\ell} - |\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i| \right)^2 \\ \text{subject to} \quad & \mathbf{w}_i = \mathbf{v}, \quad i = 1, \ldots, F \\ & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P \end{aligned} \tag{4.29}$$

with variables $\mathbf{w}_1, \ldots, \mathbf{w}_F, \mathbf{v} \in \mathbb{C}^N$.

For $i = 1, \ldots, F$ and $\ell = 1, \ldots, L$, we let the auxiliary variables $\alpha_{i,\ell} \in \mathbb{R}_+$ and $\beta_{i,\ell} \in \mathbb{R}$ be

$$\alpha_{i,\ell} = |\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i| \text{ and } \beta_{i,\ell} = \angle(\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i),$$

respectively, to have

$$\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i = \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}. \tag{4.30}$$

By using (4.30) and replacing $|\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i|$ with $\alpha_{i,\ell}$ in (4.29), we construct a problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{F} \sum_{\ell=1}^{L} h_{i,\ell} \left(b_{i,\ell} - \alpha_{i,\ell} \right)^2 \\ \text{subject to} \quad & \mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i = \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}, \quad i = 1, \ldots, F, \text{ and } \ell = 1, \ldots, L \\ & \mathbf{w}_i = \mathbf{v}, \quad i = 1, \ldots, F \\ & \mathrm{PAPR}\left(\mathbf{v}\right) \leq \sigma \\ & \|\mathbf{v}\|_2^2 = P \end{aligned} \tag{4.31}$$

with variables $\mathbf{w}_1, \ldots, \mathbf{w}_F, \mathbf{v} \in \mathbb{C}^N$, $\alpha_{1,1}, \ldots, \alpha_{F,L} \in \mathbb{R}_+$ and $\beta_{1,1}, \ldots, \beta_{F,L} \in \mathbb{R}$. Thus, the problem (4.28) has been converted to (4.31) with a convex objective function and nonconvex constraints.

The augmented Lagrangian function for (4.31) is given as

$$\begin{aligned}
&\mathcal{L}_{\rho_L,\rho_N}\big(\{\mathbf{w}_i, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \mathbf{y}_i, \boldsymbol{\lambda}_i\}_{i=1}^F, \mathbf{v}\big) \\
&\quad = \sum_{i=1}^{F}\sum_{\ell=1}^{L} h_{i,\ell}\left(b_{i,\ell} - \alpha_{i,\ell}\right)^2 + \sum_{i=1}^{F}\sum_{\ell=1}^{L} \mathsf{Re}\left\{y_{i,\ell}\right\} \mathsf{Re}\left\{\mathbf{a}_{i,\ell}^{\mathsf{H}}\mathbf{w}_i - \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}\right\} \\
&\qquad + \sum_{i=1}^{F}\sum_{\ell=1}^{L} \mathsf{Im}\left\{y_{i,\ell}\right\} \mathsf{Im}\left\{\mathbf{a}_{i,\ell}^{\mathsf{H}}\mathbf{w}_i - \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}\right\} \\
&\qquad + \frac{\rho_L}{2}\sum_{i=1}^{F}\sum_{\ell=1}^{L}\big(\mathsf{Re}\left\{\mathbf{a}_{i,\ell}^{\mathsf{H}}\mathbf{w}_i - \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}\right\}\big)^2 \\
&\qquad + \frac{\rho_L}{2}\sum_{i=1}^{F}\sum_{\ell=1}^{L}\big(\mathsf{Im}\left\{\mathbf{a}_{i,\ell}^{\mathsf{H}}\mathbf{w}_i - \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}\right\}\big)^2 \\
&\qquad + \sum_{i=1}^{F}\sum_{n=1}^{N} \mathsf{Re}\left\{\lambda_{i,n}\right\} \mathsf{Re}\left\{w_{i,n} - v_n\right\} + \sum_{i=1}^{F}\sum_{n=1}^{N} \mathsf{Im}\left\{\lambda_{i,n}\right\} \mathsf{Im}\left\{w_{i,n} - v_n\right\} \\
&\qquad + \frac{\rho_N}{2}\sum_{i=1}^{F}\sum_{n=1}^{N}(\mathsf{Re}\left\{w_{i,n} - v_n\right\})^2 + \frac{\rho_N}{2}\sum_{i=1}^{F}\sum_{n=1}^{N}(\mathsf{Im}\left\{w_{i,n} - v_n\right\})^2, \qquad (4.32)
\end{aligned}$$

where $\rho_L > 0$ and $\rho_N > 0$ are two distinct penalty parameters for beampattern shape and PAPR constraints, respectively, $\mathbf{y}_i = [\, y_{i,1} \ \ldots \ y_{i,L} \,]^{\mathsf{T}} \in \mathbb{C}^L$ and $\boldsymbol{\lambda}_i = [\, \lambda_{i,1} \ \ldots \ \lambda_{i,N} \,]^{\mathsf{T}} \in \mathbb{C}^N$ are the dual variables, and $\boldsymbol{\alpha}_i = [\, \alpha_{i,1} \ \ldots \ \alpha_{i,L} \,]^{\mathsf{T}} \in \mathbb{R}_+^L$ and $\boldsymbol{\beta}_i = [\, \beta_{i,1} \ \ldots \ \beta_{i,L} \,]^{\mathsf{T}} \in \mathbb{R}^L$ are the auxiliary variables for $i = 1, \ldots, F$.

The consensus ADMM steps for wideband transmit beamforming are similar to those given in the previous section.

Step-1: For given $\left\{\mathbf{w}_i^{(k)}, \mathbf{y}_i^{(k)}, \boldsymbol{\lambda}_i^{(k)}\right\}_{i=1}^F$, the remaining variables at $k+1$ can be determined by solving the problem

$$\begin{aligned}
\left\{\left\{\boldsymbol{\alpha}_i^{(k+1)}, \boldsymbol{\beta}_i^{(k+1)}\right\}_{i=1}^F, \mathbf{v}^{(k+1)}\right\} = \operatorname{argmin} \quad & g(\{\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i\}_{i=1}^F, \mathbf{v}) \qquad (4.33) \\
\text{subject to} \quad & \mathrm{PAPR}\left(\mathbf{v}\right) \le \sigma \\
& \|\mathbf{v}\|_2^2 = P
\end{aligned}$$

with variables $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_F \in \mathbb{R}_+^L$, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_F \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{C}^N$, where

$$\begin{aligned} g(\{\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i\}_{i=1}^F, \mathbf{v}) = & \sum_{i=1}^{F} \sum_{\ell=1}^{L} h_{i,\ell} \left(b_{i,\ell} - \alpha_{i,\ell}\right)^2 \\ & + \frac{\rho_L}{2} \sum_{i=1}^{F} \sum_{\ell=1}^{L} \left|\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i^{(k)} + z_{i,\ell}^{(k)} - \alpha_{i,\ell}\, e^{j\beta_{i,\ell}}\right|^2 \\ & + \frac{\rho_N F}{2} \|\bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} - \mathbf{v}\|_2^2 \\ \overset{c}{=} & \mathcal{L}_{\rho_L, \rho_N}\left(\left\{\mathbf{w}_i^{(k)}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \mathbf{y}_i^{(k)}, \boldsymbol{\lambda}_i^{(k)}\right\}_{i=1}^F, \mathbf{v}\right) \end{aligned} \tag{4.34}$$

with the scaled dual variables $\mathbf{z}_i = [\, z_{i,1} \ \ldots \ z_{i,L} \,]^{\mathsf{T}} = \mathbf{y}_i/\rho_L \in \mathbb{C}^L$ and $\boldsymbol{\tau}_i = [\, \tau_{i,1} \ \ldots \ \tau_{i,N} \,]^{\mathsf{T}} = \boldsymbol{\lambda}_i/\rho_N \in \mathbb{C}^N$, $i = 1, \ldots, F$, and the mean vectors $\bar{\mathbf{w}}^{(k)} = (1/F) \sum_{i=1}^F \mathbf{w}_i^{(k)}$ and $\bar{\boldsymbol{\tau}}^{(k)} = (1/F) \sum_{i=1}^F \boldsymbol{\tau}_i^{(k)}$.

Following the suggested formulation (4.16), instead of (4.15), to improve the performance, we use $\|\mathbf{v}\|_2^2 = \|\bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)}\|_2^2$ for the global variable update, that is

$$\begin{aligned} \mathbf{v}^{(k+1)} = \underset{\mathbf{v} \in \mathbb{C}^N}{\operatorname{argmin}} \quad & \|\bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)} - \mathbf{v}\|_2^2 \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{v}) \le \sigma \\ & \|\mathbf{v}\|_2^2 = \|\bar{\mathbf{w}}^{(k)} + \bar{\boldsymbol{\tau}}^{(k)}\|_2^2. \end{aligned} \tag{4.35}$$

which can be solved by applying [17, Algorithm 2], see also Algorithm 2.1.

The variable $\boldsymbol{\beta}_i$ appears only in the middle term of the objective function of (4.33), and an equivalent problem of determining $\left\{\boldsymbol{\beta}_i^{(k+1)}\right\}_{i=1}^F$ is

$$\boldsymbol{\beta}_i^{(k+1)} = \underset{\boldsymbol{\alpha}_i \in \mathbb{R}_+^L, \boldsymbol{\beta}_i \in \mathbb{R}^L}{\operatorname{argmin}} \ \sum_{\ell=1}^{L} \left|\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i^{(k)} + z_{i,\ell}^{(k)} - \alpha_{i,\ell} e^{j\beta_{i,\ell}}\right|^2, \quad i = 1, \ldots, F. \tag{4.36}$$

Letting $\mathbf{A}_i = [\, \mathbf{a}_{i,1} \ \ldots \ \mathbf{a}_{i,L} \,] \in \mathbb{C}^{N \times L}$ and $\boldsymbol{\gamma}_i = \boldsymbol{\alpha}_i \odot e^{j\boldsymbol{\beta}_i} \in \mathbb{C}^L$, the objective function of (4.36) can be expressed as

$$\sum_{\ell=1}^{L} \left|\mathbf{a}_{i,\ell}^{\mathsf{H}} \mathbf{w}_i^{(k)} + z_{i,\ell}^{(k)} - \alpha_{i,\ell} e^{j\beta_{i,\ell}}\right|^2 = \left\|\mathbf{A}_i^{\mathsf{H}} \mathbf{w}_i^{(k)} + \mathbf{z}_i^{(k)} - \boldsymbol{\gamma}_i\right\|_2^2. \tag{4.37}$$

As $\boldsymbol{\beta}_i = \angle \boldsymbol{\gamma}_i$ in (4.37), the solution of (4.36) is

$$\boldsymbol{\beta}_i^{(k+1)} = \angle\left(\mathbf{A}_i^{\mathsf{H}} \mathbf{w}_i^{(k)} + \mathbf{z}_i^{(k)}\right), \quad i = 1, \ldots, F. \tag{4.38}$$

After replacing $\beta_{i,\ell}$ in the objective function of (4.33) by $\beta_{i,\ell}^{(k+1)}$, we obtain an equivalent problem of finding $\left\{\boldsymbol{\alpha}_i^{(k+1)}\right\}_{i=1}^F$ as follows:

$$\boldsymbol{\alpha}_i^{(k+1)} = \underset{\boldsymbol{\alpha}_i \in \mathbb{R}_+^L}{\operatorname{argmin}} \ \left(\sum_{\ell=1}^{L} h_{i,\ell} \left(b_{i,\ell} - \alpha_{i,\ell}\right)^2 + (\rho_L/2) \sum_{\ell=1}^{L} (r_{i,\ell} - \alpha_{i,\ell})^2 \right), \tag{4.39}$$

where $r_{i,\ell} = \left|\mathbf{a}_{i,\ell}^{\mathsf{H}}\mathbf{w}_i^{(k)} + z_{i,\ell}^{(k)}\right|$ for $i = 1, \ldots, F$, and $\ell = 1, \ldots, L$. After taking the derivative of the objective function with respect to $\alpha_{i,\ell}$ and zeroing at $\alpha_{i,\ell}^{(k+1)}$, we find

$$\alpha_{i,\ell}^{(k+1)} = \frac{\rho_L\, r_{i,\ell} + 2h_{i,\ell}\, b_{i,\ell}}{\rho_L + 2h_{i,\ell}}, \quad i = 1, \ldots, F, \text{ and } \ell = 1, \ldots, L, \tag{4.40}$$

due to our parameter settings, $\rho_L > 0$, $r_{i,\ell} \geq 0$, $h_{i,\ell} \geq 0$ and $b_{i,\ell} \geq 0$.

Step-2: We update $\{\mathbf{w}_i\}_{i=1}^F$ by solving

$$\mathbf{w}_i^{(k+1)} = \underset{\mathbf{w}_i \in \mathbb{C}^N}{\operatorname{argmin}} \left(\left\| \mathbf{A}_i^{\mathsf{H}}\mathbf{w}_i + \mathbf{z}_i^{(k)} - \boldsymbol{\gamma}_i^{(k+1)} \right\|_2^2 + \frac{\rho_N}{\rho_L} \left\| \mathbf{w}_i + \boldsymbol{\tau}_i^{(k)} - \mathbf{v}^{(k+1)} \right\|_2^2 \right), \tag{4.41}$$

whose objective function is quadratic in $\mathbf{w}_i$. For $i = 1, \ldots, F$, we find

$$\mathbf{w}_i^{(k+1)} = \left(\mathbf{A}_i\mathbf{A}_i^{\mathsf{H}} + \frac{\rho_N}{\rho_L}\mathbf{I}_N \right)^{-1} \left(\mathbf{A}_i\big(\boldsymbol{\gamma}_i^{(k+1)} - \mathbf{z}_i^{(k)}\big) + \frac{\rho_N}{\rho_L}\big(\mathbf{v}^{(k+1)} - \boldsymbol{\tau}_i^{(k)}\big) \right). \tag{4.42}$$

Step-3: The dual updates are as follows:

$$\mathbf{z}_i^{(k+1)} = \mathbf{z}_i^{(k)} + \mathbf{A}_i^{\mathsf{H}}\mathbf{w}_i^{(k+1)} - \boldsymbol{\gamma}_i^{(k+1)}, \quad i = 1, \ldots, F, \tag{4.43}$$

$$\boldsymbol{\tau}_i^{(k+1)} = \boldsymbol{\tau}_i^{(k)} + \mathbf{w}_i^{(k+1)} - \mathbf{v}^{(k+1)}, \quad i = 1, \ldots, F. \tag{4.44}$$

Steps 1 to 3 are iteratively applied. The stopping criteria are identical to those given for PAPR-cADMM. After satisfying any of these criteria, the iteration is terminated and the solution is determined from

$$\begin{aligned} \mathbf{w}^\star = \underset{\mathbf{w} \in \mathbb{C}^N}{\operatorname{argmin}} \quad & \|\bar{\mathbf{w}}^{\text{final}} + \bar{\boldsymbol{\tau}}^{\text{final}} - \mathbf{w}\|_2^2 \\ \text{subject to} \quad & \operatorname{PAPR}(\mathbf{w}) \leq \sigma \\ & \|\mathbf{w}\|_2^2 = P, \end{aligned} \tag{4.45}$$

where the vectors $\bar{\mathbf{w}}^{\text{final}}$ and $\bar{\boldsymbol{\tau}}^{\text{final}}$ are formed by the final entries of mean vectors $\bar{\mathbf{w}}$ and $\bar{\boldsymbol{\tau}}$, respectively.

Similar to (3.21), we suggest an expression for adjusting ρ_N as follows:

$$\rho_N = \rho_L \xi \max_{i \in \{1, \ldots, F\}} \|\mathbf{A}_i\|_2, \tag{4.46}$$

where $\xi > 0$ is a coefficient. The selection of ξ is discussed in Section 4.5.3.

Algorithm 4.2: Proposed method PAPR-cADMM-wb for the wideband transmit beamformer design under a PAPR constraint.

Input : $\bar{\mathbf{w}}^{(1)} \in \mathbb{C}^N$, $\mathbf{A} \in \mathbb{C}^{N\times L}$, $\mathbf{b} \in \mathbb{R}_+^L$, $\mathbf{h} \in \mathbb{R}_+^L$, $\rho_L > 0$, $\xi > 0$, $\sigma \geq 1$, $P \in \mathbb{R}_{++}$, $k_{\max} \in \{1, 2, \ldots\}$

Output : $\mathbf{w}^\star \in \mathbb{C}^N$

Initialization: $\bar{\mathbf{w}}^{(0)} = 2\bar{\mathbf{w}}^{(1)}$, $\{\mathbf{w}_i^{(1)} = \bar{\mathbf{w}}^{(1)}, \boldsymbol{\lambda}_i^{(1)} = \mathbf{0}_N, \mathbf{y}_i^{(1)} = \mathbf{0}_L\}_{i=1}^F$, $k = 1$

Compute ρ_N from (4.46)

Assign $\mathbf{z}_i^{(1)} = \mathbf{y}_i^{(1)}/\rho_L$ and $\boldsymbol{\tau}_i^{(1)} = \boldsymbol{\lambda}_i^{(1)}/\rho_N$, $i = 1, \ldots, F$

while $k \leq k_{\max}$ and $\left(\left\|\bar{\mathbf{w}}^{(k)} - \bar{\mathbf{w}}^{(k-1)}\right\|_2 / \left\|\bar{\mathbf{w}}^{(k-1)}\right\|_2\right) > 10^{-6}$ **do**

  // Step-1

  Determine $\mathbf{v}^{(k+1)}$ from (4.35)

  Compute $\boldsymbol{\beta}_i^{(k+1)}$, $i = 1, \ldots, F$, from (4.38)

  Compute $\boldsymbol{\alpha}_i^{(k+1)}$, $i = 1, \ldots, F$, from (4.40)

  // Step-2

  Determine $\mathbf{w}_i^{(k+1)}$ from (4.42) by using $\boldsymbol{\gamma}_i^{(k+1)} = \boldsymbol{\alpha}_i^{(k+1)} \odot e^{j\boldsymbol{\beta}_i^{(k+1)}}$, $i = 1, \ldots, F$

  // Step-3

  Compute $\mathbf{z}_i^{(k+1)}$, $i = 1, \ldots, F$, from (4.43)

  Compute $\boldsymbol{\tau}_i^{(k+1)}$, $i = 1, \ldots, F$, from (4.44)

  $\bar{\mathbf{w}}^{(k+1)} = (1/F) \sum_{i=1}^F \mathbf{w}_i^{(k+1)}$

  $\bar{\boldsymbol{\tau}}^{(k+1)} = (1/F) \sum_{i=1}^F \boldsymbol{\tau}_i^{(k+1)}$

  $k = k + 1$

end

Determine $\mathbf{w}^\star$ from (4.45) by using $\bar{\mathbf{w}}^{\text{final}} = \bar{\mathbf{w}}^{(k)}$ and $\bar{\boldsymbol{\tau}}^{\text{final}} = \bar{\boldsymbol{\tau}}^{(k)}$

return $\mathbf{w}^\star$

For the wideband transmit beamformer design under a PAPR constraint, we have presented a consensus ADMM-based method. The method is referred to as PAPR-cADMM-wb and is outlined in Algorithm 4.2. When comparing (4.38), (4.42) and (4.43) with (3.12), (3.18) and (3.19), respectively, PAPR-cADMM-wb can be considered to have a similar formulation with PAPR-ADMM.

4.4 Computational Complexity Considerations

We present Algorithms 4.1 and 4.2 in this chapter. To follow these algorithms step-by-step, we either compute a closed-form expression or utilize Algorithm 2.1. Similar to PAPR-ADMM given in Algorithm 3.1, these algorithms use Algorithm 2.1 in their Step-1 of each iteration and one more time after their iterations terminate. Algorithm 2.1 requires a sorting operation with the complexity of $\mathcal{O}\left(N \log N\right)$ operations and computes a vector for the worst case complexity of $\mathcal{O}\left(N^2\right)$.

In Step-1 of PAPR-cADMM given in Algorithm 4.1, the required updates are implemented through the equations (4.18) and (4.20) requiring $\mathcal{O}\left(N\right)$ multiplications. In Step-2, (4.24) requires $\mathcal{O}\left(N^2\right)$. In Step-3, (4.25) and (4.26) require $\mathcal{O}\left(N\right)$. Thus, the overall complexity of each iteration is $\mathcal{O}\left(N^2\right)$ multiplications.

In Step-1 of PAPR-cADMM-wb given in Algorithm 4.2, (4.38) and (4.40) require $\mathcal{O}\left(LN\right)$ multiplications. In Step-2, (4.42) requires $\mathcal{O}\left(LN^2\right)$ that can be reduced to $\mathcal{O}\left(N^2\right)$ by computing the matrix inverse in (4.42) offline and storing the result. In Step-3, (4.43) and (4.44) require $\mathcal{O}\left(LN\right)$ and $\mathcal{O}\left(N\right)$, respectively. By the storage option, the overall complexity of each iteration is $\mathcal{O}\left(N^2 + LN\right)$ multiplications.

Note that $\mathcal{O}\left(N^2\right)$ and $\mathcal{O}\left(N^2 + LN\right)$ multiplications for Algorithms 4.1 and 4.2, respectively, are given by considering the parallel processing. If the consensus ADMM steps are processed within loops by a single processor instead of the parallel computation with multiple processors, Algorithms 4.1 and 4.2 have the overall complexities $\mathcal{O}\left(LN^2\right)$ and $\mathcal{O}\left(FN^2 + FLN\right)$ multiplications, respectively, per iteration. For $L \gg N$, the overall complexity of PAPR-cADMM without parallel computation is higher than that of PAPR-ADMM with offline computation of the matrix inverse.

4.5 Numerical Results

We compare PAPR-cADMM with PAPR-ADMM through the narrowband transmit beamformer design with the 32 elements ULA, whose specifications are given in Section 3.6.1, and its wideband extension. We use a uniform grid spacing of 1° for the direction samples. The parameter settings in Section 3.6, except for ρ_N, are used.

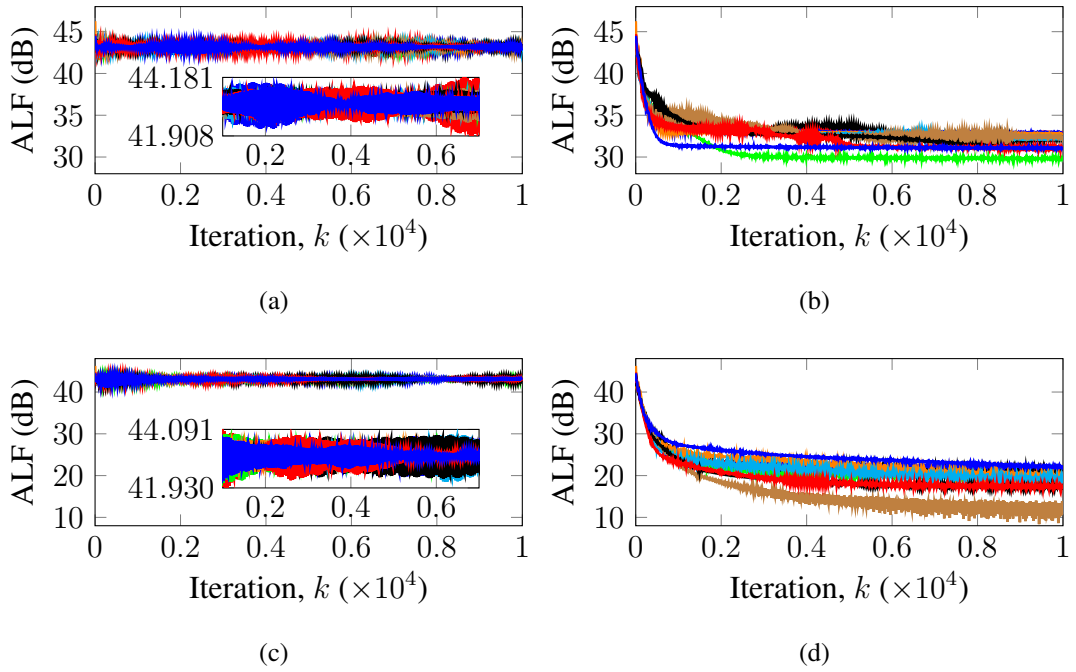


Figure 4.1: Augmented Lagrangian function (4.11) versus iteration number, $\mathcal{L}_{\rho_L,\rho_N}\big(\{\mathbf{w}_\ell^{(k)},\boldsymbol{\lambda}_\ell^{(k)}\}_{\ell=1}^L,\boldsymbol{\alpha}^{(k)},\boldsymbol{\beta}^{(k)},\mathbf{v}^{(k)},\mathbf{y}^{(k)}\big)$, for the narrowband transmit beamforming with 32 elements ULA: (a) PAPR $=$ 1 using (4.15), (b) PAPR $=$ 1 using (4.16), (c) PAPR $=$ 2 using (4.15), and (d) PAPR $=$ 2 using (4.16).

4.5.1 Effect of the Fixed Total Power

We aim to examine the performance of PAPR-cADMM when its Step-1 is coupled with (4.15) instead of (4.16). We run PAPR-cADMM to design a narrowband transmit beamformer with 32 elements ULA for $\rho_N = 50$ and 10 realizations of the initial weight vector. Each realization is used with PAPR $= \sigma \in \{1, 2\}$, and the formulations (4.15), where $P = N$, (fixed total power) and (4.16) (varying total power).

The plots in Figure 4.1 show the augmented Lagrangian function (ALF) value (4.11) versus iteration number k on a case by case basis. Comparing Figures 4.1a and 4.1b (Figures 4.1c and 4.1d), we see that the varying total power formulation (4.16) is more effective than the fixed one (4.15) for minimizing ALF. Furthermore, Table 4.1 shows that the varying total power formulation (4.16) usually leads to a solution yielding less power swing in the main lobe, higher PSL than that of (4.15). Hence, we suggest the formulation of Step-1 with (4.16) for the performance improvement.

Table 4.1: Performance metrics of 10 realizations for the narrowband transmit beamforming with 32 elements ULA and PAPR $\leq \sigma$ by using (4.15) with $P = 32$ and (4.16).

	$\sigma = 1$				$\sigma = 2$			
	$P^{\text{m}}_{\text{p-p}}$ (dB)		PSL (dB)		$P^{\text{m}}_{\text{p-p}}$ (dB)		PSL (dB)	
#	(4.15)	(4.16)	(4.15)	(4.16)	(4.15)	(4.16)	(4.15)	(4.16)
1	36.6299	15.2172	−16.8912	14.7920	18.8737	2.5639	1.5319	26.7935
2	23.5012	12.2477	−3.6153	16.9157	26.9968	2.1811	−6.6780	28.1264
3	25.0891	13.0257	−5.3465	17.8068	17.4444	3.5058	2.9216	26.2687
4	27.3730	6.6815	−7.7870	22.7151	39.4854	2.3997	−19.2836	27.9000
5	33.9836	12.5039	−12.9631	16.8245	33.0596	1.6521	−12.4666	28.1812
6	31.5445	8.8537	−10.9930	19.0053	35.1723	1.9843	−15.5809	26.3707
7	16.5522	29.1457	2.0605	0.4463	24.4567	0.8877	−4.0839	28.7702
8	31.8635	9.4145	−11.6401	19.0342	28.3707	1.6264	−8.3680	28.3336
9	23.2363	14.6915	−3.0456	13.9807	33.7948	1.7801	−13.6902	28.4878
10	42.5626	8.9313	−22.4913	19.0073	33.8891	2.9863	−14.0109	26.0119

4.5.2 Narrowband Transmit Beamforming

We compare two methods, PAPR-ADMM and PAPR-cADMM using ρ_N from (3.21) and $\rho_N = 50$, respectively, and starting at the same point, for the narrowband transmit beamforming with 32 elements ULA. Within 100 Monte Carlo (MC) runs, the mean and variance of PSL and $P^{\text{m}}_{\text{p-p}}$ (peak-to-peak power swing in the main lobe) values are given in Table 4.2. The corresponding PSL-$P^{\text{m}}_{\text{p-p}}$ pairs are visualized in Figure 4.2.

We let an optimal value

$$p^{\star}_{\text{nb}} = \sum_{\ell=1}^{L} h_{\ell} \left(b_{\ell} - |\mathbf{a}_{\ell}^{\mathsf{H}} \mathbf{w}^{\star}|\right)^2 \tag{4.47}$$

Table 4.2: Comparison of the mean and variance of PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs for the narrowband transmit beamforming experiment with 100 MC runs.

	Mean		Variance	
Method	PSL (dB)	$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ (dB)	PSL (dB)	$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ (dB)
PAPR-ADMM	17.8829	11.7817	28.7748	16.2247
PAPR-cADMM	18.2797	12.0222	31.7405	18.9551

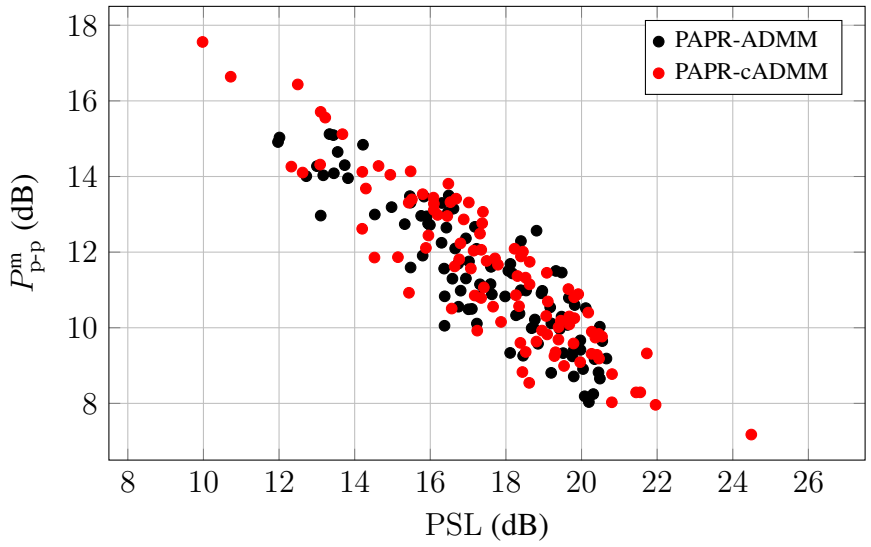


Figure 4.2: The PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs for the narrowband transmit beamforming experiment with 100 MC runs.

denote the value of the objective function evaluated at $\mathbf{w}^\star$ for the narrowband transmit beamformer design problem (2.10), and compare the histogram of the dB-scaled optimal values, $10\log_{10}(p^\star_{\mathrm{nb}})$, as shown in Figure 4.3. Figure 4.4 shows beampatterns for the one where PAPR-cADMM gives its largest PSL and smallest $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ values out of 100 MC runs. Note that a weight vector yielding better performance metrics (say, lower $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ or higher PSL) than that of another weight vector may not achieve a lower objective function value, as shown in Figure 4.4. One reason is that PAPR-ADMM and PAPR-cADMM minimize the convex functions, which are two distinct augmented Lagrangian functions given in (3.4) and (4.11), respectively, instead of the nonconvex objective function of (2.10), in different ways. Another reason is that

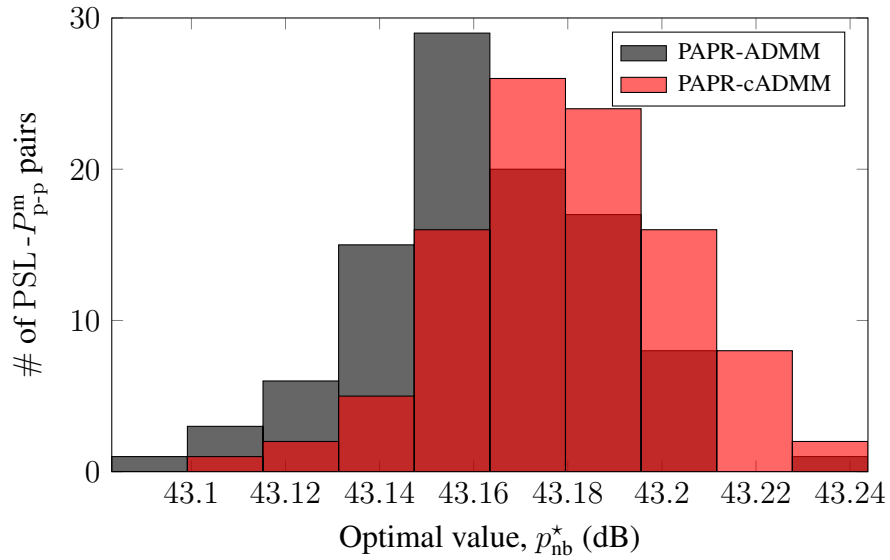


Figure 4.3: Histogram of optimal values of (2.10) for the narrowband transmit beamforming experiment with 100 MC runs.

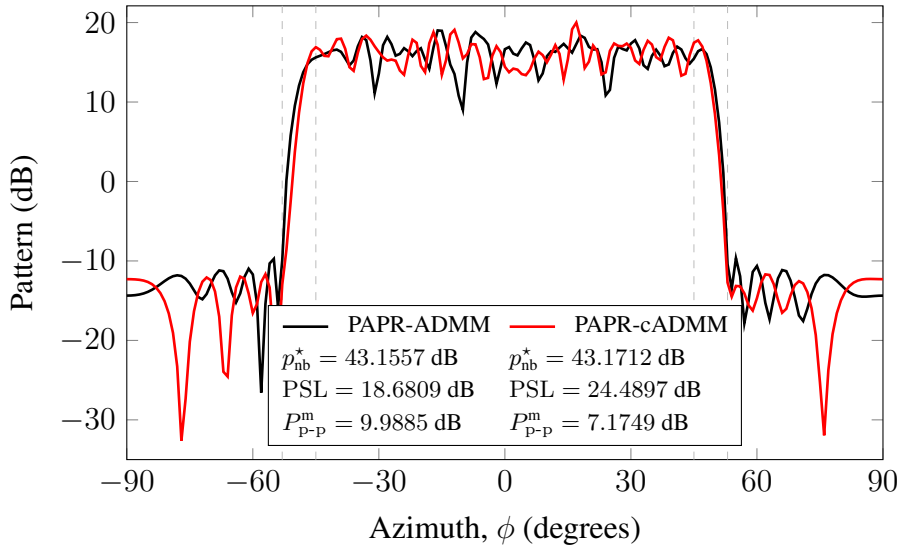


Figure 4.4: Beampattern comparison for the one where PAPR-cADMM gives its largest PSL and smallest $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ values out of 100 MC runs.

the PAPR constraint is highly nonlinear and difficult to characterize, as discussed in Section 2.3, and therefore, a minor change in penalty parameter ρ_N can result in a significant performance improvement or degradation rather than changing the rate of convergence. In brief, the nonconvex structure of problem makes the performances of PAPR-ADMM and PAPR-cADMM hard to characterize.

Table 4.3: Comparison of the mean and variance of PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs for the wideband transmit beamforming experiment with 100 MC runs.

	Mean		Variance	
Method	PSL (dB)	$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ (dB)	PSL (dB)	$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ (dB)
PAPR-ADMM	10.8346	17.6425	21.4170	44.6496
PAPR-cADMM-wb	11.4611	15.3020	22.8014	30.6496

4.5.3 Wideband Transmit Beamforming

In wideband transmit beamforming experiment, we consider the specifications given in Section 3.6.1 for the narrowband transmit beamformer design with the 32 elements ULA, and set PAPR $= 1$. The wideband beamforming parameters are $f_c = 140$ kHz and $Q_f = 14$, which are the center frequency and the quality factor, respectively. From (3.23), the bandwidth Δ_f is equal to 10 kHz. The frequency samples are $f_i \in \{135, 137.5, 140, 142.5, 145\}$ kHz, $i = 1, \ldots, 5$.

We compare PAPR-ADMM and PAPR-cADMM for the wideband transmit beamforming with 32 elements ULA by using the same initial point. For PAPR-ADMM and PAPR-cADMM, ρ_N is set by using (3.21) and (4.46) with $\xi = 0.27$, respectively. Here, ξ is empirically determined to obtain the objective function values yielding a multimodal distribution.

Table 4.3 shows the mean and variance of PSL and $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ values corresponding to 100 MC runs. In Figure 4.5, PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs are visualized.

We let an optimal value $p^{\star}_{\mathrm{wb}}$ be equal to the value of the objective function evaluated at $\mathbf{w}^{\star}$ for the wideband transmit beamformer design problem (3.24), that is

$$p^{\star}_{\mathrm{wb}} = \sum_{\ell=1}^{FL} h_{\ell} \left(b_{\ell} - \left| \mathbf{a}_{\ell}^{\mathsf{H}} \mathbf{w}^{\star} \right| \right)^2, \tag{4.48}$$

and compare the histogram of the dB-scaled optimal values, $10 \log_{10}(p^{\star}_{\mathrm{wb}})$, as shown in Figure 4.6. Both of the histograms are bimodal. However, whether the histogram of the optimal values achieved by PAPR-cADMM is multimodal or not depends on

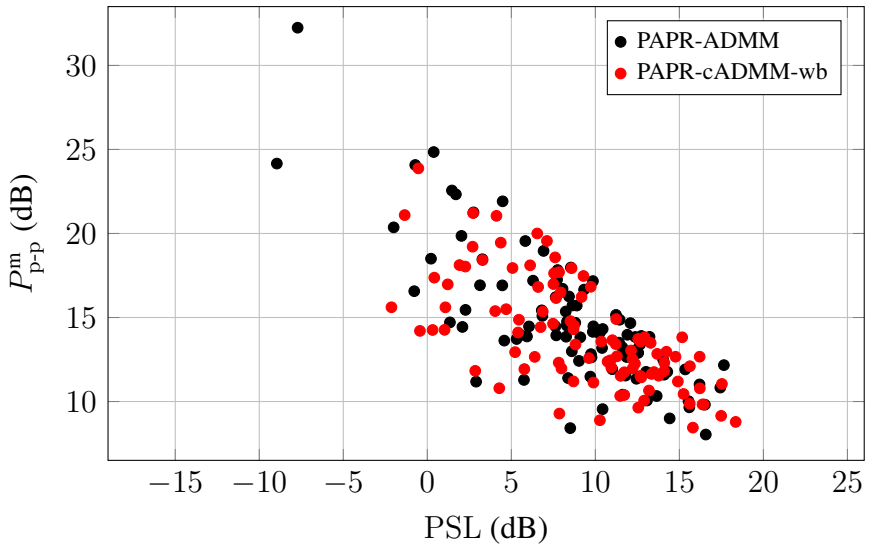


Figure 4.5: The PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs for the wideband transmit beamforming experiment with 100 MC runs.

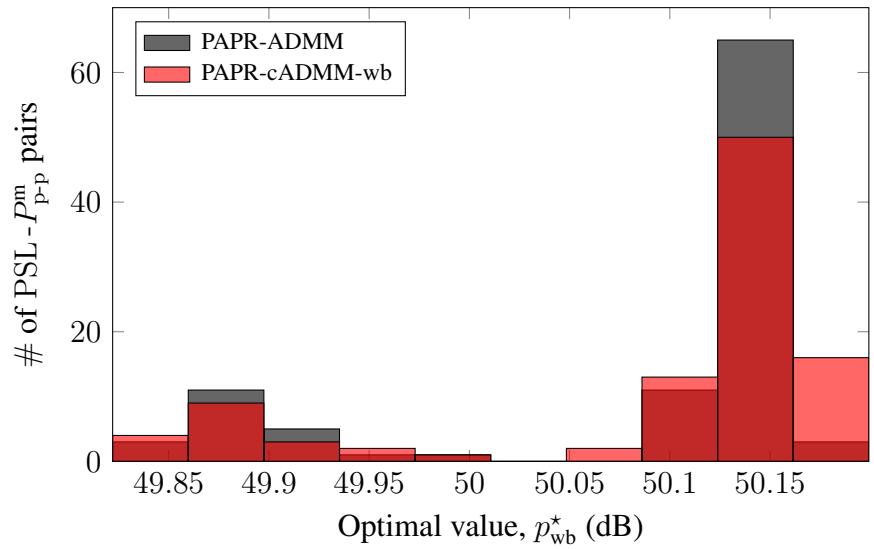
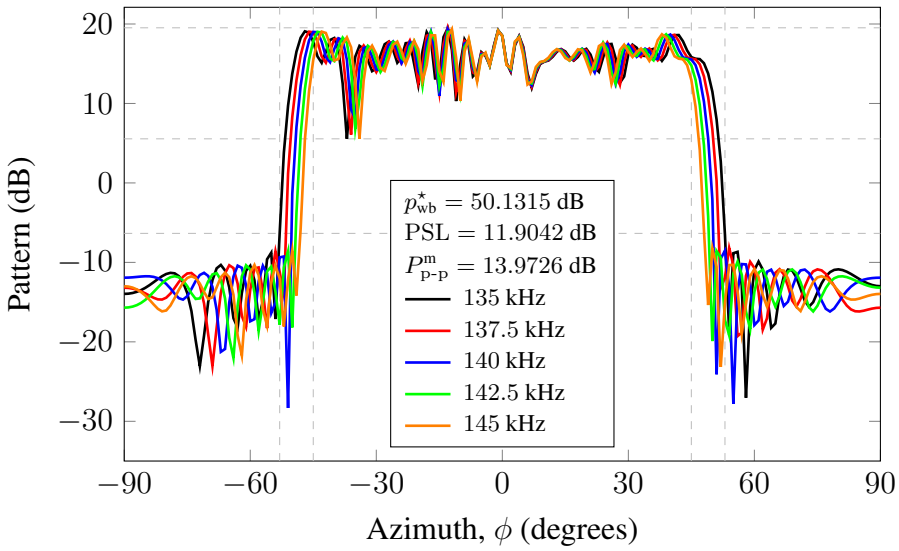


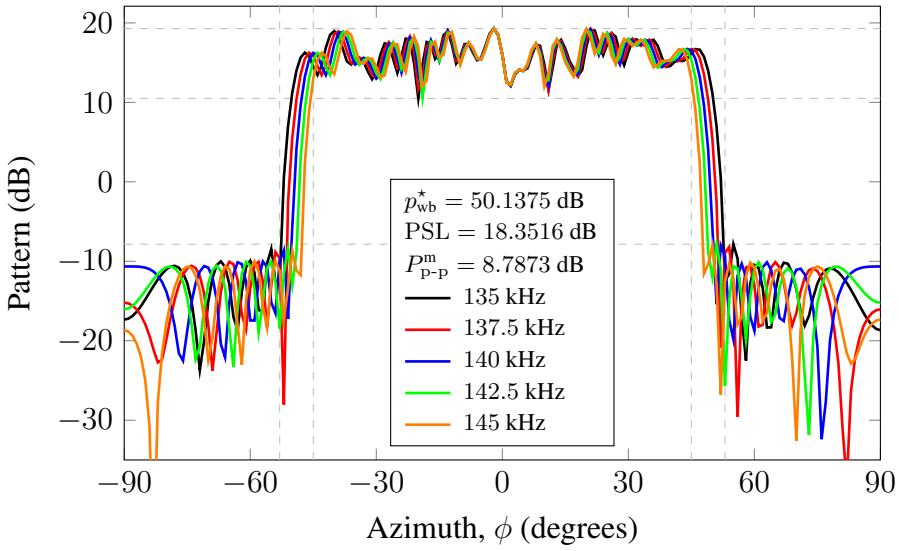
Figure 4.6: Histogram of optimal values of (3.24) for the wideband transmit beamforming experiment with 100 MC runs.

the selection of ξ.

Figures 4.7 and 4.8 shows the beampatterns corresponding to the PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs, where PAPR-cADMM-wb gives its largest PSL and smallest $P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ values, respectively. In addition to PSL-$P^{\mathrm{m}}_{\mathrm{p\text{-}p}}$ pairs, these figures show $p^{\star}_{\mathrm{wb}}$ values.

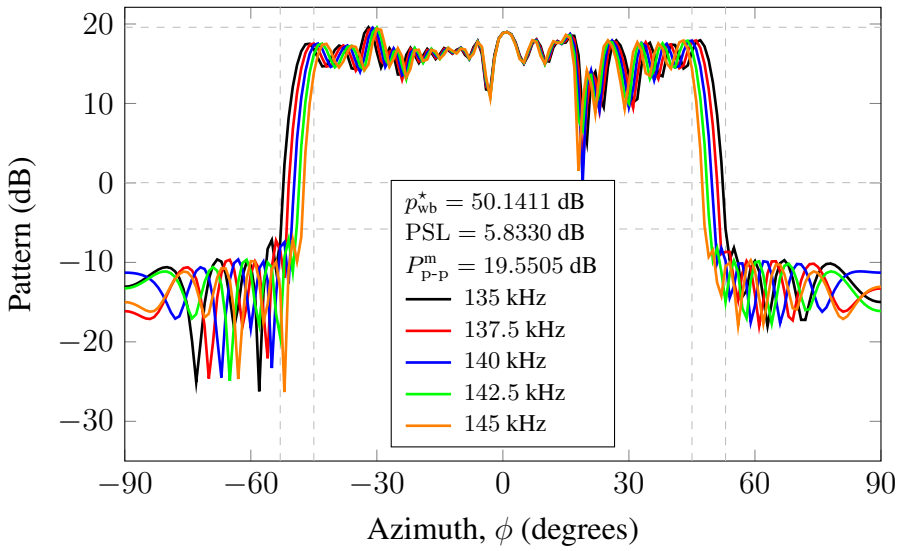


(a)

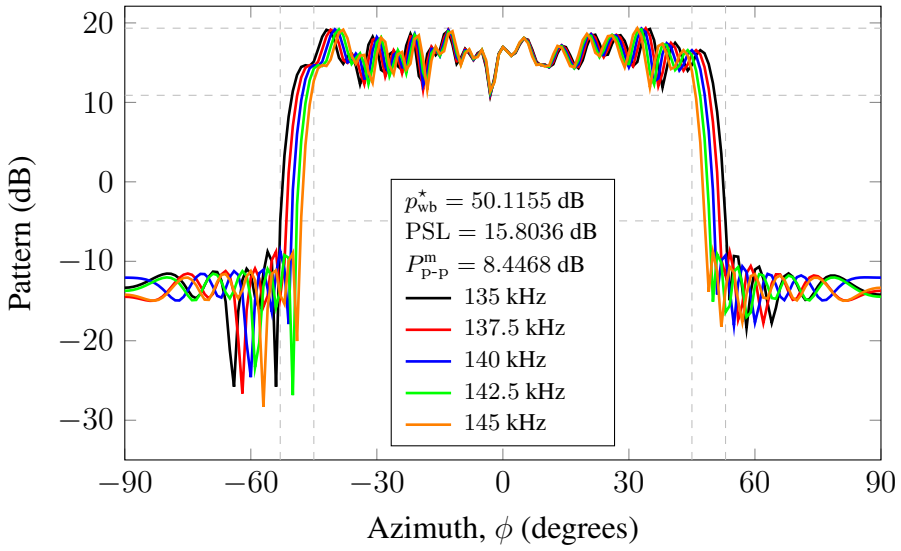


(b)

Figure 4.7: Beampattern comparison for the one where PAPR-cADMM-wb gives its largest PSL value out of 100 MC runs: (a) PAPR-ADMM and (b) PAPR-cADMM-wb.



(a)



(b)

Figure 4.8: Beampattern comparison for the one where PAPR-cADMM-wb gives its smallest $P^{\text{m}}_{\text{p-p}}$ value out of 100 MC runs: (a) PAPR-ADMM and (b) PAPR-cADMM-wb.

4.6 Conclusion

The main goal of this chapter is to present a solution for the transmit beamformer design problem under the PAPR constraint after expressing the problem in a consensus form. Our solution uses the consensus ADMM, which allows parallel computation. The suggested methods, PAPR-cADMM and PAPR-cADMM-wb, can be used to design narrowband and wideband beamformers, respectively. Both methods are applicable via parallel processing. The formulations of PAPR-cADMM and PAPR-cADMM-wb allow L (number of direction samples) and F (number of frequencies) individual updates in parallel, respectively. The main consideration in parallel computation is the number of processors (or cores). When the number of processors is less than the number of individual updates, all individual updates cannot be simultaneously processed, and some of these updates wait for idle processors. Since $L \gg F$ typically holds, the parallel computation in the narrowband case may not be as practical as in the wideband case.

Numerical results reveal that PAPR-ADMM and PAPR-cADMM have comparable performance, especially when the mean values given in Tables 4.2 and 4.3 are compared. Considering the formulations of PAPR-ADMM and PAPR-cADMM, both methods have computational advantages depending on applications, see Section 4.4. When the offline computation and the storage options are available for the inverse matrix required in (3.18), PAPR-ADMM becomes preferable to PAPR-cADMM in the narrowband case with L higher than the number of processors, which is a typical case. To design a wideband transmit beamformer, the consensus ADMM-based method PAPR-cADMM-wb behaves as PAPR-ADMM applied in a parallel manner. For an F value, which ensures the simultaneous processing of all individual updates, PAPR-cADMM-wb with $\mathcal{O}\left(N^2 + LN\right)$ is preferable to PAPR-ADMM with $\mathcal{O}\left(N^2 + FLN\right)$.

CHAPTER 5

MAXIMUM LIKELIHOOD ESTIMATION OF NOISY AUTOREGRESSIVE MODEL PARAMETER WITH INDEPENDENT SNAPSHOTS

In this chapter, we study the noisy autoregressive (AR) parameter estimation problem in the presence of multiple independent snapshots. We solely focus on the parameter estimation problem for the scalar AR processes. We apply the expectation-maximization (EM) method [41] by assigning the noise-free snapshots as the missing data to develop a solution. In the maximization step (M-step) of EM method, we transform the AR parameter estimation problem into a form that can be solved by using an approach similar to the recent work in [9] which is a two-stage method utilized for AR parameter estimation for a single noise-free snapshot. For the calculation of the expectation step (E-step), we describe an approximate yet highly efficient method to reduce the computational load.

5.1 Related Works and Contributions

Model parameter estimation of an AR process observed under noise, namely noisy AR parameter estimation problem, is prone to estimator bias and statistical efficiency problems when modeling assumptions are not carefully taken into account. For instance, AR parameter estimates obtained from Yule-Walker (YW) equations are typically biased due to the bias of the zero-lag term of the autocorrelation introduced by white noise [42]. Furthermore, when the variance of the noise corrupting the AR process is not known, an asymptotic Cramér-Rao bound (CRB) study by Weruaga *et al.* reveals that the joint estimation of the autoregressive signal variance and noise variance is not a well-conditioned problem [43, 44]. In spite of these setbacks,

several practical solutions have been developed for the noisy AR parameter estimation problem in the literature. For example, since the autoregressive moving average (ARMA) model also characterizes the noisy AR processes, it is possible to apply ARMA modeling approaches, such as the maximum likelihood [45], the modified YW [46], and the recursive prediction error [47], for the solution of noisy AR parameter estimation problem. In addition, a number of improved least-squares (LS) solutions are introduced to compensate for the bias in the parameter estimates due to the measurement noise [48–50]. The main challenge for the bias compensating solutions is the estimation of the measurement noise variance. The solutions based on the eigendecomposition [49] and the inverse filtering coupled with YW equations [50] have been suggested for this purpose. Among other solutions, we can list a subspace-based solution [51], an errors-in-variables approach utilizing both low and high order YW equations [52], a nonlinear optimization (for estimating the measurement noise variance) solution [53], a solution with two interacting Kalman filters [54] and some adaptive filtering type solutions [55–57]. A particularly interesting solution is the method presented in [58], based on the approach developed by Mehra [59], that avoids the estimation of the process and measurement noise variances. Recently, four novel methods have been proposed in [60]. The first one utilizes the null space of AR parameter vector, the second one solves a constrained LS problem, the third one reduces the parameter estimation problem for an AR(P) process to a problem of estimating two parameters, and the fourth one is based on the eigendecomposition of the enlarged autocorrelation matrix.

The EM method has been previously applied to the noisy AR parameter estimation problem in [61], and some computational simplifications in the M-step have been suggested. Different from [61], we formulate the M-step such that it is possible to extend the maximum likelihood like estimator given in [9] to the multiple-snapshot setting and also describe some novel computational load reduction methods for the E-step. In [62], a related EM-based method using Kalman filters is presented for the colored Gaussian noise. It is well known that the performance of EM algorithm is sensitive to initialization, that is, EM iterations can converge to a local maximum, instead of the global maximum, due to poor initialization [63]. We present an initialization method for the suggested method (see Section 5.3), consider the cases of known/unknown

measurement noise variance individually, and present detailed comparisons with the alternative estimators and CRBs derived in [43, 44].

In the literature, there are several works, including [64–66], which are based on the Whittle likelihood [67], a frequency-domain approximation to the exact likelihood function. While Whittle likelihood maximization is computationally easier, the resulting parameter estimates with finite sample sizes are biased, especially for short data records [66]. As the sample size increases, time- and frequency-domain solutions yield similar results [68]. Further discussions on time- and frequency-domain approaches can be found in [44]. Here, we focus on the time-domain approach in relation to our main goal of extending the maximum likelihood like estimator in [9] to the noisy, multiple-snapshot setting.

The AR modeling has several important applications in the speech processing area. In this area, the typical data size can be hundreds of samples, and frequency domain methods can be utilized in this large sample-size regime without any performance worries. In some other applications, such as radar signal processing applications, the data size can be much smaller, and the data collection mechanism can operate intermittently, in contrast to the continuous data-collection modality in speech processing, leading to multiple, short data-length (snapshot) observations [69–71]. For example, in the clutter (the unwanted echoes received by radar systems [69]) cancellation application of radar signal processing, the clutter power spectral density is estimated from a collection of snapshots [69]. Each entry of the snapshot vector is formed by a radar pulse return from a particular range cell. The number of transmitted pulses, which is the dimension of the snapshot vector, affects all subsequent radar operations, and it can be as few as 10-20 pulses due to other constraints [72]. For this application, the structured estimation of the clutter power spectrum from small-dimensional multiple snapshots becomes a necessity. Some solutions to this problem, in addition to the examined maximum likelihood solution, are the multiple-snapshot version of Burg's method [70] or the multiple-snapshot version of any other AR parameter estimation method given in [73]. In this chapter, we consider the AR parameter estimation problem specifically for small-dimensional multiple snapshots and pursue an exact time-domain maximum likelihood parameter estimation solution.

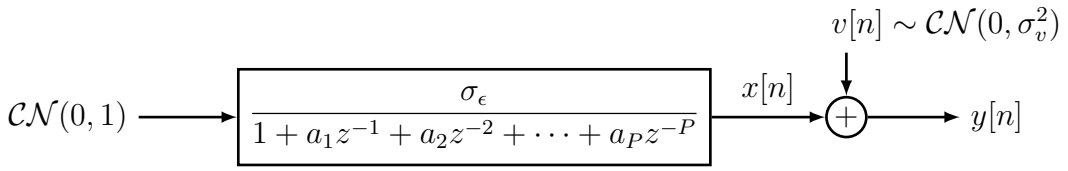


Figure 5.1: Signal model block diagram for the noisy AR parameter estimation problem: AR process samples $x[n]$ corrupted by the observation noise $v[n]$ to form the observed sequence $y[n]$.

This chapter presents the following main contributions.

- Expressing the conventional EM formulation as a multiple-snapshot, noise-free AR parameter estimation problem (Section 5.2.1).
- Extension of the efficient single snapshot noise-free AR parameter estimation method given in [9] to the multiple-snapshot case and its application in the solution of EM problem (Section 5.2.2).
- An approximate but efficient version of the proposed solution by using a matrix inversion-free Kalman smoother [74, Sec. 5.2.4] and the Gohberg-Semencul formula [73, Sec. 3.9.4] (Section 5.2.3).

5.2 Noisy AR Parameter Estimation Problem

Consider the transfer function given below for the generation of an AR(P) process.

$$H(z) = \frac{\sigma_\epsilon}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}} = \frac{\sigma_\epsilon}{A(z)}, \tag{5.1}$$

where $a_1, a_2, \ldots, a_P$ can be either the real- or complex-valued constants, and σ_ϵ is a real-valued constant scaling the input.

The filter $H(z)$ is assumed to be excited with zero-mean, unit variance complex-valued (circular symmetric) white Gaussian noise, as shown in Figure 5.1. The filter $H(z)$ is also assumed to be stable so that the filter output, *i.e.*, the AR(P) process, is wide-sense stationary. The output of the filter at steady-state is denoted as $x[n]$. The lth snapshot $\mathbf{x}_l$ is an N-dimensional vector formed by concatenating the N-consecutive $x[n]$ samples. It is assumed that a total number of L snapshot vectors

are collected where each snapshot is independent and identically distributed (i.i.d.) Gaussian vector.

The autocorrelation matrix of the snapshot $\mathbf{x}_l$ is denoted as $\mathbf{R} = \mathbf{R}_{f,N}\sigma_\epsilon^2$ where $\mathbf{R}_{f,N}$ is an $N \times N$ Hermitian Toeplitz matrix given as

$$\mathbf{R}_{f,N} = \begin{bmatrix} r_f[0] & r_f[-1] & \dots & r_f[-N+1] \\ r_f[1] & r_f[0] & \dots & r_f[-N+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_f[N-1] & r_f[N-2] & \dots & r_f[0] \end{bmatrix} \tag{5.2}$$

with the first column entries

$$r_f[k] = \mathsf{E}\{x[n]\, x^*[n-k]\}/\sigma_\epsilon^2, \quad k = 0, \dots, N-1.$$

Here, it is assumed that snapshot vectors are observed under independent additive white Gaussian noise, *i.e.*, the lth observation vector is $\mathbf{y}_l = \mathbf{x}_l + \mathbf{v}_l$ as $y[n] = x[n] + v[n]$ in Figure 5.1, where $\mathbf{v}_l$ vectors are an i.i.d. circular symmetric Gaussian distributed vector with zero-mean and covariance matrix $\sigma_v^2\mathbf{I}_N$. Hence, the sample signal-to-noise ratio is

$$\text{SNR} = \frac{\sigma_\epsilon^2 r_f[0]}{\sigma_v^2}.$$

The general AR parameter estimation problem is the estimation of the unknown non-random parameter vector $\boldsymbol{\theta} = \left[\, \sigma_v^2 \;\; \sigma_\epsilon^2 \;\; \mathbf{a}^\mathsf{T} \,\right]^\mathsf{T}$, where $\mathbf{a} = [\, a_1 \; \dots \; a_P \,]^\mathsf{T}$, given the observation vectors $\mathbf{y}_l$, $l = 1, \dots, L$. If the noise variance σ_v^2 is known, the unknown parameter vector becomes $\boldsymbol{\theta} = \left[\, \sigma_\epsilon^2 \;\; \mathbf{a}^\mathsf{T} \,\right]^\mathsf{T}$. For this case, the true value for the noise variance can be substituted for its estimate in the formulation given below. If the noise variance is known to be zero, the problem becomes the AR parameter estimation problem with multiple noise-free snapshots. The solution of this problem can be retrieved as the limiting case of the discussion as $\sigma_v^2 \to 0$.

5.2.1 Expectation-Maximization Formulation for the Noisy AR Parameter Estimation With Multiple Snapshots

For the solution of the problem, we apply the expectation-maximization method. To do that, we define the $N \times L$ matrices $\mathbf{X} = [\, \mathbf{x}_1 \; \dots \; \mathbf{x}_L \,]$ and $\mathbf{Y} = [\, \mathbf{y}_1 \; \dots \; \mathbf{y}_L \,]$

to denote the noise-free AR process snapshots and observations, respectively. The snapshot matrix $\mathbf{X}$ is the latent variable of the problem. The $N \times 2L$ dimensional

$$\mathbf{Z} = [\,\mathbf{X}\;\;\mathbf{Y}\,]$$

is the complete data matrix. The log-likelihood of the complete data matrix

$$\begin{aligned}\Lambda(\mathbf{Z};\boldsymbol{\theta}) &= \log \mathsf{f}(\mathbf{X},\mathbf{Y};\boldsymbol{\theta}) \\ &= \log \mathsf{f}(\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}) + \log \mathsf{f}(\mathbf{X};\boldsymbol{\theta})\end{aligned}$$

can be written as

$$\Lambda(\mathbf{Z};\boldsymbol{\theta}) \stackrel{c}{=} -\sum_{l=1}^{L}\left\{ N\log(\sigma_v^2) + \frac{\|\mathbf{y}_l - \mathbf{x}_l\|_2^2}{\sigma_v^2} + \log\det\left(\sigma_\epsilon^2 \mathbf{R}_{f,N}\right) + \frac{\mathbf{x}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\mathbf{x}_l}{\sigma_\epsilon^2}\right\}, \tag{5.3}$$

where the first two and last two terms of the summation correspond to $\log \mathsf{f}(\mathbf{Y}|\mathbf{X};\boldsymbol{\theta})$ and $\log \mathsf{f}(\mathbf{X};\boldsymbol{\theta})$, respectively. The symbol $\stackrel{c}{=}$ denotes the equality of both sides apart from constant terms.

The EM method has two steps: In the expectation step, the expected value of the complete log-likelihood is calculated with respect to the posterior density of the latent variables. The expectation operation can be written as

$$\mathcal{J}(\boldsymbol{\theta}) = \mathsf{E}\{\Lambda(\mathbf{Z};\boldsymbol{\theta})|\mathbf{Y};\boldsymbol{\theta}^{\text{old}}\}. \tag{5.4}$$

Here the vector $\boldsymbol{\theta}^{\text{old}} = \left[\,(\sigma_v^2)^{\text{old}}\;\;(\sigma_\epsilon^2)^{\text{old}}\;\;(\mathbf{a}^{\text{old}})^{\mathsf{T}}\,\right]^{\mathsf{T}}$ contains the current estimates for the unknown parameters to be updated. In the second step, the expectation result is maximized with respect to the unknown parameters to update the unknown parameters, that is $\boldsymbol{\theta}^{\text{new}} = \text{argmax}_{\boldsymbol{\theta}}\,\mathcal{J}(\boldsymbol{\theta})$. The algorithm is initiated with a proper $\boldsymbol{\theta}^{\text{old}}$ value for the implementation of the first step (posterior calculation) and iteratively run by using $\boldsymbol{\theta}^{\text{new}}$ of an earlier iteration as the $\boldsymbol{\theta}^{\text{old}}$ of the next iteration until the convergence of the estimates.

To execute the expectation step, we need the posterior density of latent variables, $\mathsf{f}(\mathbf{x}_l|\mathbf{y}_l;\boldsymbol{\theta}^{\text{old}})$, $l = 1,\ldots,L$. It is well known that the posterior density $\mathsf{f}(\mathbf{x}_l|\mathbf{y}_l;\boldsymbol{\theta}^{\text{old}})$ is the Gaussian vector with mean vector

$$\widehat{\mathbf{x}}_l = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})\left((\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) + (\sigma_v^2)^{\text{old}}\mathbf{I}_N\right)^{-1}\mathbf{y}_l \tag{5.5}$$

and covariance matrix

$$\mathbf{K} = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) - (\sigma_\epsilon^4)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})\big((\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) + (\sigma_v^2)^{\text{old}}\mathbf{I}_N\big)^{-1}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}). \quad (5.6)$$

Taking the expectation of the complete log-likelihood function in (5.3) with respect to the posterior density results in (5.4) that is

$$\mathcal{J}(\sigma_v^2, \sigma_\epsilon^2, \mathbf{a}) \stackrel{c}{=} -\sum_{l=1}^{L}\left\{N\log(\sigma_v^2) + \frac{\|\mathbf{y}_l - \widehat{\mathbf{x}}_l\|_2^2 + \operatorname{tr}(\mathbf{K})}{\sigma_v^2}\right\} - \sum_{l=1}^{L}\left\{\log\det\left(\sigma_\epsilon^2\mathbf{R}_{f,N}\right) + \frac{\widehat{\mathbf{x}}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l + \operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)}{\sigma_\epsilon^2}\right\}, \quad (5.7)$$

which completes the first step (expectation step) of EM method.

The maximization step of EM begins with the partial derivative calculation of $\mathcal{J}(\sigma_v^2, \sigma_\epsilon^2, \mathbf{a})$ given in (5.7) with respect to σ_v^2. Setting the result equal to zero, we get the update expression for the measurement noise variance σ_v^2 estimate

$$(\sigma_v^2)^{\text{new}} = \frac{1}{LN}\left(L\operatorname{tr}(\mathbf{K}) + \sum_{l=1}^{L}\|\mathbf{y}_l - \widehat{\mathbf{x}}_l\|_2^2\right). \quad (5.8)$$

Similarly, by taking the partial derivative of $\mathcal{J}(\sigma_v^2, \sigma_\epsilon^2, \mathbf{a})$ with respect to σ_ϵ^2 and equating the result to zero, we get the update expression for the process noise variance σ_ϵ^2 estimate

$$(\sigma_\epsilon^2)^{\text{new}} = \frac{1}{LN}\left(L\operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) + \sum_{l=1}^{L}\widehat{\mathbf{x}}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l\right). \quad (5.9)$$

Substituting $(\sigma_v^2)^{\text{new}}$ and $(\sigma_\epsilon^2)^{\text{new}}$ given in (5.8) and (5.9), respectively, into (5.7), we get the compressed expected log-likelihood function as

$$\mathcal{J}\big((\sigma_v^2)^{\text{new}}, (\sigma_\epsilon^2)^{\text{new}}, \mathbf{a}\big) \stackrel{c}{=} -\log\det(\mathbf{R}_{f,N}) - N\log\left(L\operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) + \sum_{l=1}^{L}\widehat{\mathbf{x}}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l\right), \quad (5.10)$$

where $\mathbf{R}_{f,N}$ is a function of the unknown parameter vector $\mathbf{a} = [\,a_1 \ \ldots \ a_P\,]^{\mathsf{T}}$. The maximization of (5.10) with respect to $\mathbf{a}$ is more challenging than earlier steps and is the main challenge of the problem. Fortunately, by expressing $L\operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ as a quadratic term, the function (5.10) can be converted into a form similar to that of

the single noise-free snapshot case given in [9] and the approach given therein can be utilized to maximize (5.10) with respect to $\mathbf{a}$.

To express $L \operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ as a quadratic term, we introduce the eigendecomposition of the covariance matrix $\mathbf{K} = \sum_{n=1}^{N} \lambda_n \mathbf{e}_n \mathbf{e}_n^{\mathsf{H}}$ into the problem. Here λ_n and $\mathbf{e}_n$ are an eigenvalue and an associated unit norm eigenvector of matrix $\mathbf{K}$, respectively. It should be noted that we can find an orthonormal set of the eigenvectors $\mathbf{e}_n$, since the matrix $\mathbf{K}$ is Hermitian. With the eigendecomposition, it is possible to express $\operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ as

$$\operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) = \sum_{n=1}^{N} \lambda_n \mathbf{e}_n^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \mathbf{e}_n. \tag{5.11}$$

By introducing the scaled versions of eigenvectors $\mathbf{e}_n$ as $\bar{\mathbf{e}}_n = \sqrt{L\lambda_n}\mathbf{e}_n$, we can write

$$L \operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) = \sum_{n=1}^{N} \bar{\mathbf{e}}_n^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \bar{\mathbf{e}}_n. \tag{5.12}$$

Hence, the maximization of compressed likelihood relation in (5.10) is equivalent to the following minimization problem

$$\begin{aligned} \mathbf{a}^{\text{new}} &= \underset{\mathbf{a}}{\operatorname{argmin}} \quad (1/N)\log\det\left(\mathbf{R}_{f,N}\right) + \log\left(\sum_{n=1}^{N} \bar{\mathbf{e}}_n^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \bar{\mathbf{e}}_n + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \widehat{\mathbf{x}}_l\right) \\ &= \underset{\mathbf{a}}{\operatorname{argmin}} \quad (1/N)\log\det\left(\mathbf{R}_{f,N}\right) + \log\left(\sum_{\ell=1}^{L_a=L+N} \mathbf{f}_\ell^{\mathsf{H}} \mathbf{R}_{f,N}^{-1}(\mathbf{a}) \mathbf{f}_\ell\right). \end{aligned} \tag{5.13}$$

In the second line of (5.13), we combine the sums in the argument of logarithm by augmenting the set of $\{\widehat{\mathbf{x}}_l\}_{l=1}^{L}$ vectors with $\{\bar{\mathbf{e}}_n\}_{n=1}^{N}$ vectors to form a set of vectors with $L_a = L + N$ elements where $\mathbf{f}_{L+n} = \bar{\mathbf{e}}_n$.

Assuming that the solution of (5.13), that is $\mathbf{a}^{\text{new}}$, is available; the remaining unknown parameter can be estimated from (5.9) as

$$\begin{aligned} (\sigma_\epsilon^2)^{\text{new}} &= \frac{1}{LN}\left(L \operatorname{tr}\left(\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{K}\right) + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^{\mathsf{H}} \mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}}) \widehat{\mathbf{x}}_l\right) \\ &= \frac{1}{LN} \sum_{\ell=1}^{L_a} \mathbf{f}_\ell^{\mathsf{H}} \mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}}) \mathbf{f}_\ell. \end{aligned} \tag{5.14}$$

Hence, the crux of the parameter problem is the solution of the optimization problem given in (5.13).

When compared with the minimization problem [9, Eq. 11], whose cost function is

$$(1/N)\log\det\left(\mathbf{R}_{f,N}\right) + \log\left(\mathbf{x}^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \mathbf{x}\right),$$

we see that cost function of the minimization problem in (5.13) differs from the earlier one with the inclusion of

$$L_a = L + N$$

snapshots instead of a single one. We present the details for this extension in the following section. We reiterate that some of the snapshots in this formulation are generated from the eigendecomposition of $\mathbf{K}$ matrix and augmented to actual snapshots, called observation vectors, to facilitate a solution similar to the one in [9]. In Section 5.2.3, we present a reduced complexity implementation alternative for this solution. Different methods of initialization for the EM method are provided in Section 5.3.

5.2.2 AR Parameter Estimation Problem With Multiple Noise-free Snapshots

The method in [9] is an efficient method for AR parameter estimation which is developed for a single snapshot under the noiseless observation scenario. This method can be considered as an alternative for numerical search based maximum likelihood estimator having much higher complexity, and it is shown that the method performs very similar to the maximum likelihood estimator in many scenarios. In this section, we present the multiple-snapshot extension of this method.

5.2.2.1 The First Stage: Weighted Forward-Backward Prediction With the Increased Number of Snapshots

Following [9], we ignore $\log\det\left(\mathbf{R}_{f,N}\right)$ term in (5.13) and use the weighted forward-backward prediction method to generate the first stage estimate of $\mathbf{a}$, $\mathbf{a}_{\text{FS}}$, as follows:

$$\mathbf{a}_{\text{FS}} = \underset{\mathbf{a}}{\operatorname{argmin}} \sum_{\ell=1}^{L_a} \left(\sum_{n=1}^{N-P} w_b[n] \left| e_{\ell,b}[n] \right|^2 + \sum_{n=P+1}^{N} w_f[n] \left| e_{\ell,f}[n] \right|^2 \right), \tag{5.15}$$

where

$$e_{\ell,f}[n] = f_{\ell,n} + \mathbf{a}^{\mathsf{T}}\mathbf{f}_{\ell,n-1:-1:n-P} \text{ and } e_{\ell,b}[n] = f_{\ell,n} + \mathbf{a}^{\mathsf{H}}\mathbf{f}_{\ell,n+1:n+P}$$

are the forward and backward prediction errors with weights

$$w_f[n] = n - P \text{ and } w_b[n] = N - P + 1 - n,$$

respectively. The problem in (5.15) can be solved by introducing the linear equation systems

$$\mathbf{A}_{\ell,f}\mathbf{a} = -\mathbf{b}_{\ell,f} \text{ and } \mathbf{A}_{\ell,b}\mathbf{a} = -\mathbf{b}_{\ell,b}$$

given below for $\ell = 1, \ldots, L_a$, which generate the forward and backward prediction errors:

$$\underbrace{\begin{bmatrix} f_{\ell,P} & f_{\ell,P-1} & \cdots & f_{\ell,1} \\ f_{\ell,P+1} & f_{\ell,P} & \cdots & f_{\ell,2} \\ \vdots & \vdots & \cdots & \vdots \\ f_{\ell,N-1} & f_{\ell,N-2} & \cdots & f_{\ell,N-P} \end{bmatrix}}_{\mathbf{A}_{\ell,f}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}} = -\underbrace{\begin{bmatrix} f_{\ell,P+1} \\ f_{\ell,P+2} \\ \vdots \\ f_{\ell,N} \end{bmatrix}}_{\mathbf{b}_{\ell,f}},$$

$$\underbrace{\begin{bmatrix} f^*_{\ell,N-P+1} & f^*_{\ell,N-P+2} & \cdots & f^*_{\ell,N} \\ f^*_{\ell,N-P} & f^*_{\ell,N-P+1} & \cdots & f^*_{\ell,N-1} \\ \vdots & \vdots & \cdots & \vdots \\ f^*_{\ell,2} & f^*_{\ell,3} & \cdots & f^*_{\ell,P+1} \end{bmatrix}}_{\mathbf{A}_{\ell,b}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}} = -\underbrace{\begin{bmatrix} f^*_{\ell,N-P} \\ f^*_{\ell,N-P-1} \\ \vdots \\ f^*_{\ell,1} \end{bmatrix}}_{\mathbf{b}_{\ell,b}}.$$

Using the introduced matrices, the final result of the first stage becomes

$$\mathbf{a}_{\text{FS}} = -\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}^{\mathsf{H}}_{\ell,f}\mathbf{W}\mathbf{A}_{\ell,f} + \mathbf{A}^{\mathsf{H}}_{\ell,b}\mathbf{W}\mathbf{A}_{\ell,b}\right)\right)^{-1}\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}^{\mathsf{H}}_{\ell,f}\mathbf{W}\mathbf{b}_{\ell,f} + \mathbf{A}^{\mathsf{H}}_{\ell,b}\mathbf{W}\mathbf{b}_{\ell,b}\right)\right), \tag{5.16}$$

where $\mathbf{W}$ is the diagonal matrix with the diagonal entries of $w_f[n]$, $n = P+1, P+2, \ldots, N$, that is

$$\mathbf{W} = \text{diag}\,(1, 2, \ldots, N-P).$$

5.2.2.2 The Second Stage: Maximizing Likelihood Around the First Stage Estimate

The second stage takes into account $\log\det(\mathbf{R}_{f,N})$ in (5.13), [9]. The nonlinear function is expanded into Taylor series at the operating point of $\mathbf{a} = \mathbf{a}_{\text{FS}}$ and a quadratic approximation for both terms of the sum forming the cost function of (5.13) is formed. We note that both $\det(\mathbf{R}_{f,N})$ and $\mathbf{R}^{-1}_{f,N}$ are highly nonlinear functions of $\mathbf{a}$.

For the *quadratic approximation* of $\log\det(\mathbf{R}_{f,N})$, we can follow the procedure given

in [9, Eq. 11]. The procedure uses the expression

$$\begin{aligned}\det\left(\mathbf{R}_{f,N}\right) &= \det\left(\mathbf{R}_{f,P}\right) \\ &= \prod_{i=1}^{P}(1-|k_i|^2)^{-i}, \quad N \geq P,\end{aligned}$$

[4, 75], that connects the autocorrelation matrix determinant of an AR(P) process to the reflection coefficients k_i, $i = 1, \ldots, P$, of its synthesis filter. The log-determinant

$$\log\det\left(\mathbf{R}_{f,N}\right) = -\sum_{i=1}^{P} i \log\left(1-|k_i|^2\right)$$

is approximated via Taylor series at the expansion point of the reflection coefficient vector $\mathbf{k} = [\, k_1 \; \ldots \; k_P \,]^{\mathsf{T}}$ that corresponds to the first stage estimate $\mathbf{a}_{\text{FS}}$ and introducing a perturbation vector $\boldsymbol{\delta}_{\mathbf{k}}$ as in [9, Eq. 16].

For the *quadratic approximation* of $\sum_{\ell=1}^{L_a} \mathbf{f}_\ell^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \mathbf{f}_\ell$, we can use [9, Eq. 21] directly

$$\begin{aligned}\sum_{\ell=1}^{L_a} \mathbf{f}_\ell^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \mathbf{f}_\ell = \sum_{\ell=1}^{L_a} \big(&\|\mathbf{b}_{\ell,1} + \mathbf{M}_{\ell,1}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{\text{c}}\boldsymbol{\delta}_{\mathbf{k}}^*)\|_2^2 \\ &- \|\mathbf{b}_{\ell,2} + \mathbf{M}_{\ell,2}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{\text{c}}\boldsymbol{\delta}_{\mathbf{k}}^*)\|_2^2\big), \quad (5.17)\end{aligned}$$

where the $N \times P$ dimensional $\mathbf{M}_{\ell,1}$ and the $P \times P$ dimensional $\mathbf{M}_{\ell,2}$ are Hankel and Toeplitz matrices, respectively, with the definitions of

$$\mathbf{M}_{\ell,1} = \begin{bmatrix} f_{\ell,2}^* & f_{\ell,3}^* & f_{\ell,4}^* & \cdots & f_{\ell,P+1}^* \\ f_{\ell,3}^* & f_{\ell,4}^* & f_{\ell,5}^* & \cdots & f_{\ell,P+2}^* \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ f_{\ell,N-1}^* & f_{\ell,N}^* & 0 & \ldots & 0 \\ f_{\ell,N}^* & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}, \quad (5.18)$$

$$\mathbf{M}_{\ell,2} = \begin{bmatrix} f_{\ell,N} & f_{\ell,N-1} & f_{\ell,N-2} & \cdots & f_{\ell,N-P+1} \\ 0 & f_{\ell,N} & f_{\ell,N-1} & \cdots & f_{\ell,N-P+2} \\ 0 & 0 & f_{\ell,N} & \ldots & f_{\ell,N-P+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & f_{\ell,N} \end{bmatrix}, \quad (5.19)$$

$\mathbf{b}_{\ell,1} = \mathbf{f}_{\ell,1:N}^* + \mathbf{M}_{\ell,1}\mathbf{a}_{\text{FS}}$ and $\mathbf{b}_{\ell,2} = \mathbf{M}_{\ell,2}\mathbf{a}_{\text{FS}}$ are constant vectors. The matrices $\mathbf{G}$ and $\mathbf{G}_{\text{c}}$ are the $P \times P$ Jacobian matrices having entries $[\mathbf{G}]_{ij} = \frac{\partial a_i}{\partial k_j}$ and $[\mathbf{G}_{\text{c}}]_{ij} = \frac{\partial a_i}{\partial k_j^*}$

evaluated at the expansion point of $\mathbf{a}_{\text{FS}}$ and $(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{\text{c}}\boldsymbol{\delta}_{\mathbf{k}}^{*})$ is the perturbation vector for $\mathbf{a}_{\text{FS}}$ such that $\mathbf{k} + \boldsymbol{\delta}_{\mathbf{k}}$ is the reflection coefficients corresponding to $\mathbf{a}_{\text{FS}} + (\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{\text{c}}\boldsymbol{\delta}_{\mathbf{k}}^{*})$. These matrices can be efficiently calculated via the inverse Levinson-Durbin recursion, as given in [9, Algorithm 2].

By the Taylor series expansion of

$$\log(A + Bx) \approx \log(A) + (B/A)\, x$$

for $|x| \ll 1$, we can approximate the second term of (5.13) as

$$\log \sum_{\ell=1}^{L_a} \mathbf{f}_{\ell}^{\mathsf{H}} \mathbf{R}_{f,N}^{-1} \mathbf{f}_{\ell} \approx \log(A) + \sum_{\ell=1}^{L_a} \frac{\|\mathbf{b}_{\ell,1} + \mathbf{M}_{\ell,1}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{\text{c}}\boldsymbol{\delta}_{\mathbf{k}}^{*})\|_2^2 - \|\mathbf{b}_{\ell,1}\|_2^2}{A} - \sum_{\ell=1}^{L_a} \frac{\|\mathbf{b}_{\ell,2} + \mathbf{M}_{\ell,2}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{\text{c}}\boldsymbol{\delta}_{\mathbf{k}}^{*})\|_2^2 - \|\mathbf{b}_{\ell,2}\|_2^2}{A}, \quad (5.20)$$

where

$$A = \sum_{\ell=1}^{L_a} \left(\|\mathbf{b}_{\ell,1}\|_2^2 - \|\mathbf{b}_{\ell,2}\|_2^2 \right)$$

is the value when $\boldsymbol{\delta}_{\mathbf{k}}$ is replaced by all zeros vector in (5.17). Thus, A can be expressed as

$$A = \sum_{\ell=1}^{L_a} \mathbf{f}_{\ell}^{\mathsf{H}} \mathbf{R}_{f,N}^{-1}(\mathbf{a}_{\text{FS}}) \mathbf{f}_{\ell},$$

where $\mathbf{f}_{\ell}^{\mathsf{H}} \mathbf{R}_{f,N}^{-1}(\mathbf{a}_{\text{FS}}) \mathbf{f}_{\ell}$ can be efficiently calculated by using [9, Algorithm 1], even without constructing $\mathbf{R}_{f,N}$.

Following the step of optimization, we reach the following equation system for the solution of $\boldsymbol{\delta}_{\mathbf{k}}$ and $\boldsymbol{\delta}_{\mathbf{k}}^{*}$, which are perturbation vectors for the complex-valued reflection coefficients corresponding to the initial reflection coefficient vector $\mathbf{k}$ generated from $\mathbf{a}_{\text{FS}}$:

$$\begin{bmatrix} \mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 & \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2 \\ \mathbf{Q}_2^{*} + \widetilde{\mathbf{Q}}_2^{*} & \mathbf{Q}_1^{*} + \widetilde{\mathbf{Q}}_1^{*} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_{\mathbf{k}} \\ \boldsymbol{\delta}_{\mathbf{k}}^{*} \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_1 + \widetilde{\mathbf{r}}_1 \\ \mathbf{r}_1^{*} + \widetilde{\mathbf{r}}_1^{*} \end{bmatrix}. \quad (5.21)$$

In the last equation, we have

$$\widetilde{\mathbf{Q}}_1 = \frac{(\mathbf{G}^{\mathsf{H}}\mathbf{P}\mathbf{G}_{\text{c}})^{*} + \mathbf{G}_{\text{c}}^{\mathsf{H}}\mathbf{P}\mathbf{G}}{A}, \quad (5.22)$$

$$\widetilde{\mathbf{Q}}_2 = \frac{(\mathbf{G}^{\mathsf{H}}\mathbf{P}\mathbf{G})^{*} + \mathbf{G}_{\text{c}}^{\mathsf{H}}\mathbf{P}\mathbf{G}_{\text{c}}}{A}, \quad (5.23)$$

and

$$\widetilde{\mathbf{r}}_1 = \frac{(\mathbf{G}^{\mathsf{H}}\mathbf{v})^* + \mathbf{G}_{\mathsf{c}}^{\mathsf{H}}\mathbf{v}}{A}, \tag{5.24}$$

where

$$\mathbf{P} = \sum_{\ell=1}^{L_a}(\mathbf{M}_{\ell,1}^{\mathsf{H}}\mathbf{M}_{\ell,1} - \mathbf{M}_{\ell,2}^{\mathsf{H}}\mathbf{M}_{\ell,2}) \text{ and } \mathbf{v} = \sum_{\ell=1}^{L_a}(\mathbf{M}_{\ell,1}^{\mathsf{H}}\mathbf{b}_{\ell,1} - \mathbf{M}_{\ell,2}^{\mathsf{H}}\mathbf{b}_{\ell,2}).$$

The diagonal matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have the diagonal entries

$$[\mathbf{Q}_1]_{ii} = \frac{i(k_i^*)^2}{N(1-|k_i|^2)^2} \text{ and } [\mathbf{Q}_2]_{ii} = \frac{i}{N(1-|k_i|^2)^2},$$

respectively, and $\mathbf{r}_1 = [\, r_{1,1} \ \ldots \ r_{1,P} \,]^{\mathsf{T}}$ has the entries

$$r_{1,i} = \frac{ik_i^*}{N(1-|k_i|^2)}$$

for $i = 1, \ldots, P$.

To express (5.21), the gradients of quadratic approximations of $(1/N)\log\det(\mathbf{R}_{f,N})$ and $\log\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$ with respect to $\boldsymbol{\delta}_{\mathbf{k}}^*$,

$$\nabla_{\boldsymbol{\delta}_{\mathbf{k}}^*}\big((1/N)\log\det(\mathbf{R}_{f,N})\big) = \mathbf{Q}_2^*\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{Q}_1^*\boldsymbol{\delta}_{\mathbf{k}}^* + \mathbf{r}_1^*,$$
$$\nabla_{\boldsymbol{\delta}_{\mathbf{k}}^*}\left(\log\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell\right) = \widetilde{\mathbf{Q}}_2^*\boldsymbol{\delta}_{\mathbf{k}} + \widetilde{\mathbf{Q}}_1^*\boldsymbol{\delta}_{\mathbf{k}}^* + \widetilde{\mathbf{r}}_1^*,$$

are used, as derived in [9].

For the real-valued processes, the reflection coefficients are also real-valued, *i.e.*, $\boldsymbol{\delta}_{\mathbf{k}} = \boldsymbol{\delta}_{\mathbf{k}}^*$, and hence, (5.21) can be simplified as

$$\left(\mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 + \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2\right)\boldsymbol{\delta}_{\mathbf{k}} = -(\mathbf{r}_1 + \widetilde{\mathbf{r}}_1). \tag{5.25}$$

The proposed method for the AR parameter estimation from multiple noise-free snapshots is outlined in Algorithm 5.1. It should be observed from Algorithm 5.1 that the second stage is iteratively applied by using the previous iteration result as the initial point of the following iteration.

5.2.3 An Efficient Implementation for the Suggested Solution

The expectation step of the suggested method requires the inversion of an $N \times N$ matrix for the calculation of posterior density parameters, *i.e.*, mean vector (5.5) and

Algorithm 5.1: Suggested method for the AR parameter estimation from multiple noise-free snapshots.

Input : $\mathbf{X} = [\,\mathbf{x}_1 \;\ldots\; \mathbf{x}_L\,]$, P

Output : $\widehat{\mathbf{a}}$, $\widehat{\sigma}_\epsilon^2$

// The first stage estimation by (5.16) for $L_a = L$ and $\mathbf{f}_l = \mathbf{x}_l$, $l = 1, \ldots, L$

$\mathbf{a}_{\text{FS}} = -\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{\ell,f} + \mathbf{A}_{\ell,b}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{\ell,b}\right)\right)^{-1}\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{\ell,f} + \mathbf{A}_{\ell,b}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{\ell,b}\right)\right)$

// The second stage estimation by using $L_a = L$ and $\mathbf{f}_l = \mathbf{x}_l$, $l = 1, \ldots, L$

$\mathbf{a}^{\text{old}} = \mathbf{a}_{\text{FS}}$

for iteration $\leftarrow 1$ **to** 10 **do**

  // The loop with 10 iterations yields good performance, see [9]

  $\mathbf{k} = \text{atog}\left(\left[\,1 \;\; (\mathbf{a}^{\text{old}})^{\mathsf{T}}\,\right]^{\mathsf{T}}\right)$ // $\text{atog}\,(\cdot)$: Step-down recursion, [4, p. 236]

  if $\mathbf{X}$ is real-valued **then**

    Solve $\left(\mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 + \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2\right)\boldsymbol{\delta}_{\mathbf{k}} = -(\mathbf{r}_1 + \widetilde{\mathbf{r}}_1)$ // See (5.25)

  else

    Solve $\begin{bmatrix} \mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 & \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2 \\ \mathbf{Q}_2^* + \widetilde{\mathbf{Q}}_2^* & \mathbf{Q}_1^* + \widetilde{\mathbf{Q}}_1^* \end{bmatrix}\begin{bmatrix} \boldsymbol{\delta}_{\mathbf{k}} \\ \boldsymbol{\delta}_{\mathbf{k}}^* \end{bmatrix} = -\begin{bmatrix} \mathbf{r}_1 + \widetilde{\mathbf{r}}_1 \\ \mathbf{r}_1^* + \widetilde{\mathbf{r}}_1^* \end{bmatrix}$ // See (5.21)

  end

  $\mathbf{k} = \mathbf{k} + \boldsymbol{\delta}_{\mathbf{k}}$ // Reflection coefficients update

  $\left[\,1 \;\; (\mathbf{a}^{\text{new}})^{\mathsf{T}}\,\right]^{\mathsf{T}} = \text{gtoa}\,(\mathbf{k})$ // $\text{gtoa}\,(\cdot)$: Step-up recursion, [4, p. 233]

  $\mathbf{a}^{\text{old}} = \mathbf{a}^{\text{new}}$

end

$(\sigma_\epsilon^2)^{\text{new}} = \left(1/(LN)\right)\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{f}_\ell$ // $\mathbf{f}_\ell^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{f}_\ell$ is calculated by applying [9, Algorithm 1]

return $\widehat{\mathbf{a}} = \mathbf{a}^{\text{new}}$, $\widehat{\sigma}_\epsilon^2 = (\sigma_\epsilon^2)^{\text{new}}$

error covariance matrix (5.6). In general, the observation vector length (N) is much greater than the order of AR process (P), and the implementation cost of the expectation step becomes a computational bottleneck. In this section, we present four approaches to reduce the computational load. We start with the disturbance smoother, a variation of the Kalman smoothers which does not require any matrix inversion for the estimation of the mean vector in the present problem setup [74, Sec. 5.2.4].

5.2.3.1 Efficient Calculation of the Mean Vector

It is well known that the Wiener filtering operation, which is the operation implemented with the set of equations (5.5) and (5.6), coincides with the Kalman filtering for the processing of a wide sense stationary input, with rational power spectral density, corrupted by independent measurement noise, [4]. To facilitate the recursive calculation via the Kalman smoothing, we introduce the following state space model:

$$\underbrace{\begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2] \\ \vdots \\ x[n-P+1] \end{bmatrix}}_{\mathbf{s}_n} = \underbrace{\begin{bmatrix} -a_1^{\text{old}} & -a_2^{\text{old}} & -a_3^{\text{old}} & \dots & -a_P^{\text{old}} \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x[n-1] \\ x[n-2] \\ x[n-3] \\ \vdots \\ x[n-P] \end{bmatrix}}_{\mathbf{s}_{n-1}} + \underbrace{\begin{bmatrix} \sigma_\epsilon^{\text{old}} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{b}} w_n,$$

$$y_n = \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}}_{\mathbf{c}} \mathbf{s}_n + v_n, \tag{5.26}$$

where w_n, y_n and v_n denote the process noise $w[n]$, the measurement $y[n]$ and the measurement noise $v[n]$, respectively.

Algorithm 5.2 outlines the disturbance smoother for the state space model of (5.26). The disturbance smoother is a three-pass operation, where in the first pass the Kalman filtering is applied in the forward direction, the disturbance is estimated in the second pass and the smoothed state vector is formed in the last pass, [74, Sec. 5.2.4]. The passes do not involve any matrix inversion operation. Hence, the mean vector calculation cost is effectively reduced from the $N \times N$ matrix inversion cost to the order of N complex multiplications. The computational savings becomes very significant when $N \geq 100$.

5.2.3.2 Approximating the Error Covariance Matrix

The matrix $\mathbf{K}$ in (5.6) corresponds to the error covariance matrix of the Wiener filter estimate. Unfortunately, it is not possible to retrieve an error covariance matrix of size $N \times N$ from a Kalman filtering implementation. To avoid the calculation of $\mathbf{K}$ matrix, we examine the limiting case of $N \to \infty$ that corresponds to the infinite impulse

Algorithm 5.2: The disturbance smoother, also see [76].

Input : $y_n, \quad n = 1, \ldots, N$

Output : $\mathbf{c}\breve{\mathbf{s}}_{n|N}, \quad n = 1, \ldots, N$, *i.e.*, the first element of $\breve{\mathbf{s}}_{n|N}$

// State Equation: $\mathbf{s}_n = \mathbf{A}\mathbf{s}_{n-1} + \mathbf{b}w_n, \quad (w_n \sim \mathcal{CN}(0,1))$

// $y_n = \mathbf{c}\mathbf{s}_n + v_n, \quad (v_n \sim \mathcal{CN}(0, (\sigma_v^2)^{\text{old}}))$

// Kalman Filtering

$\widehat{\mathbf{s}}_{0|0} = \mathbf{0}_N, \mathbf{P}_{0|0} = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})$ // see (5.2)

for $n \leftarrow 0$ **to** $N-1$ **do**

  $\widehat{\mathbf{s}}_{n+1|n} = \mathbf{A}\widehat{\mathbf{s}}_{n|n}$

  $\mathbf{P}_{n+1|n} = \mathbf{A}\mathbf{P}_{n|n}\mathbf{A}^{\mathsf{H}} + \mathbf{b}\mathbf{b}^{\mathsf{H}}$

  $\epsilon_{n+1} = y_{n+1} - \mathbf{c}\widehat{\mathbf{s}}_{n+1|n}$

  $\gamma_{n+1} = \mathbf{c}\mathbf{P}_{n+1|n}\mathbf{c}^{\mathsf{H}} + (\sigma_v^2)^{\text{old}}$

  $\mathbf{H}_{n+1} = \mathbf{P}_{n+1|n}\mathbf{c}^{\mathsf{H}}\gamma_{n+1}^{-1}$

  $\widehat{\mathbf{s}}_{n+1|n+1} = \widehat{\mathbf{s}}_{n+1|n} + \mathbf{H}_{n+1}\epsilon_{n+1}$

  $\mathbf{P}_{n+1|n+1} = \mathbf{P}_{n+1|n} - \mathbf{H}_{n+1}\mathbf{c}\mathbf{P}_{n+1|n}$

  $\mathbf{\Lambda}_{n+1} = \mathbf{A} - \mathbf{A}\mathbf{H}_{n+1}\mathbf{c}$

end

// Backward Smoothing

$\mathbf{p}_{N-1} = \mathbf{c}^{\mathsf{H}}\epsilon_N\gamma_N^{-1}$

$\widehat{e}_{N-1|N} = \mathbf{b}^{\mathsf{H}}\mathbf{p}_{N-1}$

for $n \leftarrow N-2$ **to** 1 **do**

  $\mathbf{p}_n = \mathbf{c}^{\mathsf{H}}\epsilon_{n+1}\gamma_{n+1}^{-1} + \mathbf{\Lambda}_{n+1}^{\mathsf{H}}\mathbf{p}_{n+1}$

  $\widehat{e}_{n|N} = \mathbf{b}^{\mathsf{H}}\mathbf{p}_n$

end

// Smoothed State Vector

$\breve{\mathbf{s}}_{1|N} = \mathbf{P}_{0|0}(\mathbf{c}^{\mathsf{H}}\epsilon_1\gamma_1^{-1} + \mathbf{\Lambda}_1^{\mathsf{H}}\mathbf{p}_1)$

for $n \leftarrow 2$ **to** $N-1$ **do**

  $\breve{\mathbf{s}}_{n|N} = \mathbf{A}\breve{\mathbf{s}}_{n-1|N} + \mathbf{b}\widehat{e}_{n-1|N}$

end

$\breve{\mathbf{s}}_{N|N} = \widehat{\mathbf{s}}_{N|N}$

return $\mathbf{c}\breve{\mathbf{s}}_{n|N}$

response (IIR) noncausal Wiener filtering operation, [4, Sec. 7.3]. As $N \to \infty$, the error covariance matrix approaches to a Hermitian Toeplitz matrix whose first column is sufficient to characterize the complete matrix, [77]. For the filtering application with the noncausal IIR Wiener filter $H^{\text{IIR-NC}}(z)$, the error autocorrelation sequence can be expressed as

$$r_e[k] = (\sigma_v^2)^{\text{old}} h^{\text{IIR-NC}}[k],$$

where $h^{\text{IIR-NC}}[k]$ is the impulse response of the noncausal IIR Wiener filter. We suggest using the residue theorem to evaluate the inverse z-transform of $H^{\text{IIR-NC}}(z)$ for the calculation of error correlation sequence:

$$r_e[k] = \frac{(\sigma_v^2)^{\text{old}}}{2\pi j} \oint_{C:|z|=1} H^{\text{IIR-NC}}(z)\, z^{k-1}\, dz, \quad k = 0, 1, \ldots, P, \tag{5.27}$$

where

$$H^{\text{IIR-NC}}(z) = \frac{(\sigma_\epsilon^2)^{\text{old}}}{(\sigma_\epsilon^2)^{\text{old}} + (\sigma_v^2)^{\text{old}}\, A^{\text{old}}(z) \left(A^{\text{old}}(1/z^*)\right)^*},$$

[4, Sec. 7.3] for

$$A^{\text{old}}(z) = 1 + a_1^{\text{old}} z^{-1} + a_2^{\text{old}} z^{-2} + \cdots + a_P^{\text{old}} z^{-P}.$$

We suggest calculating only $P+1$ error correlation lags given in (5.27) and constructing $\mathbf{K}$ as a banded matrix with P nonzero super/sub-diagonals.

5.2.3.3 Efficient Calculation of Trace Term

The compressed likelihood function in (5.10) requires the calculation of trace term, $\text{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$, where $\mathbf{R}_{f,N}^{-1}$ depends on the optimization variable $\mathbf{a}$ and $\mathbf{K}$ is a constant matrix. It has been suggested that this problem is solved by an eigenvalue decomposition of $\mathbf{K}$ and expressing $\text{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ as a sum of quadratic terms. We present an efficient method that avoids the computationally costly eigendecomposition step.

We utilize the Gohberg-Semencul (GS) formula from [73, Sec. 3.9.4] for the inverse of $\mathbf{R}_{f,N}$. The GS formula states that

$$\mathbf{R}_{f,N}^{-1} = \mathbf{A}_1\mathbf{A}_1^{\mathsf{H}} - \mathbf{A}_2\mathbf{A}_2^{\mathsf{H}},$$

where the $N \times N$ dimensional matrices $\mathbf{A}_1$ and $\mathbf{A}_1$ are given as follows:

$$\mathbf{A}_1 = \begin{bmatrix} 1 & & & & & & 0 \\ a_1 & 1 & & & & & \\ \vdots & a_1 & 1 & & & & \\ a_P & \ddots & a_1 & 1 & & & \\ 0 & a_P & \ddots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & 1 & \\ 0 & \dots & 0 & a_P & \dots & a_1 & 1 \end{bmatrix}, \tag{5.28}$$

$$\mathbf{A}_2 = \begin{bmatrix} 0 & & & & & & 0 \\ \vdots & 0 & & & & & \\ 0 & \ddots & 0 & & & & \\ a_P^* & \ddots & \ddots & \ddots & & & \\ a_{P-1}^* & a_P^* & \ddots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ a_1^* & \dots & a_{P-1}^* & a_P^* & 0 & \dots & 0 \end{bmatrix}. \tag{5.29}$$

By defining an $N \times N$ dimensional shift matrix $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} 0 & & & & 0 \\ 1 & 0 & & & \\ 0 & 1 & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix},$$

$\mathbf{A}_1$ matrix can be expressed as

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{a}_+ & \mathbf{S}\mathbf{a}_+ & \mathbf{S}^2\mathbf{a}_+ & \dots & \mathbf{S}^{N-1}\mathbf{a}_+ \end{bmatrix},$$

where $\mathbf{a}_+$ denotes the first column of $\mathbf{A}_1$ matrix from (5.28) and $\mathbf{S}^k\mathbf{a}_+$ is the $(k+1)$th column formed by shifting $\mathbf{a}_+$ vector k times. We also define

$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{a}_R^* & \mathbf{S}\mathbf{a}_R^* & \mathbf{S}^2\mathbf{a}_R^* & \dots & \mathbf{S}^{N-1}\mathbf{a}_R^* \end{bmatrix},$$

where $\mathbf{a}_R^*$ is the first column of $\mathbf{A}_2$ matrix. Using these definitions, $\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ can

be expressed as

$$\begin{aligned}\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) &= \mathrm{tr}\left((\mathbf{A}_1\mathbf{A}_1^{\mathsf{H}} - \mathbf{A}_2\mathbf{A}_2^{\mathsf{H}})\mathbf{K}\right) \\ &= \mathrm{tr}\left(\mathbf{A}_1^{\mathsf{H}}\mathbf{K}\mathbf{A}_1\right) - \mathrm{tr}\left(\mathbf{A}_2^{\mathsf{H}}\mathbf{K}\mathbf{A}_2\right) \\ &= \mathbf{a}_+^{\mathsf{H}}\left(\sum_{k=0}^{N-1}(\mathbf{S}^k)^{\mathsf{H}}\mathbf{K}\mathbf{S}^k\right)\mathbf{a}_+ - \mathbf{a}_R^{\mathsf{T}}\left(\sum_{k=0}^{N-1}(\mathbf{S}^k)^{\mathsf{H}}\mathbf{K}\mathbf{S}^k\right)\mathbf{a}_R^* \\ &= \mathbf{a}_+^{\mathsf{H}}\mathbf{V}\mathbf{a}_+ - \mathbf{a}_R^{\mathsf{T}}\mathbf{V}\mathbf{a}_R^*, \end{aligned} \tag{5.30}$$

where $\mathbf{V} = \sum_{k=0}^{N-1}(\mathbf{S}^k)^{\mathsf{H}}\mathbf{K}\mathbf{S}^k$. Hence, $\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ is a quadratic product of the unknown parameters.

5.2.3.4 Efficient Calculation of Optimal Perturbation Around the First Stage Estimate

The objective function (5.10) to be maximized includes $\det\left(\mathbf{R}_{f,N}\right)$ and $L\,\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ $+ \sum_{l=1}^{L}\widehat{\mathbf{x}}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l$, where the elements of matrix $\mathbf{R}_{f,N}$ are the functions of vector $\mathbf{a}$. While the determinant of $\mathbf{R}_{f,N}$ can be written as a quadratic product of the reflection coefficients corresponding to the vector $\mathbf{a}$, the other reduces to quadratic product in terms of $\mathbf{a}$. Following the approach given in [9], we convert the problem domain to the reflection coefficient domain and treat $\mathbf{a}$ as a vector-valued function of the reflection coefficients, and then, we expand it into a Taylor series around the reflection coefficients derived from the first stage result ($\mathbf{a}_{\mathrm{FS}}$) and keep only the first order term, *i.e.*, the Jacobian term, in the expansion. In brief, for

$$\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) = \mathbf{a}_+^{\mathsf{H}}\mathbf{V}\mathbf{a}_+ - \mathbf{a}_R^{\mathsf{T}}\mathbf{V}\mathbf{a}_R^*,$$

we express the unknown vectors $\mathbf{a}_+$ and $\mathbf{a}_R$ as follows:

$$\mathbf{a}_+ = \mathbf{a}_{\mathrm{FS},+} + \mathbf{G}_+\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{+\mathrm{c}}\boldsymbol{\delta}_{\mathbf{k}}^*,$$

$$\mathbf{a}_R = \mathbf{a}_{\mathrm{FS},R} + \mathbf{G}_R\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_{R\mathrm{c}}\boldsymbol{\delta}_{\mathbf{k}}^*,$$

where $\mathbf{a}_{\mathrm{FS},+} = \begin{bmatrix}1 & \mathbf{a}_{\mathrm{FS}}^{\mathsf{T}} & \mathbf{0}_{N-P-1}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$ and $\mathbf{a}_{\mathrm{FS},R} = \begin{bmatrix}\mathbf{0}_{N-P}^{\mathsf{T}} & \mathbf{a}_{\mathrm{FS},P:-1:1}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$. The vectors $\boldsymbol{\delta}_{\mathbf{k}}$ and its conjugate $\boldsymbol{\delta}_{\mathbf{k}}^*$ denote the unknown reflection coefficient perturbation vectors. The matrices $\mathbf{G}_+$, $\mathbf{G}_{+\mathrm{c}}$, $\mathbf{G}_R$ and $\mathbf{G}_{R\mathrm{c}}$ are the Jacobian matrices, for the vector-valued

functions $\mathbf{a}_{\mathrm{FS},+}$ and $\mathbf{a}_{\mathrm{FS},R}$ in terms of their reflection coefficients, with the definitions given below.

$$\begin{aligned}
\mathbf{G}_{+} &= \left[\, \mathbf{0}_{P} \;\; \mathbf{G}^{\mathsf{T}} \;\; \mathbf{0}_{P\times(N-P-1)} \,\right]^{\mathsf{T}}, \\
\mathbf{G}_{+\mathrm{c}} &= \left[\, \mathbf{0}_{P} \;\; \mathbf{G}_{\mathrm{c}}^{\mathsf{T}} \;\; \mathbf{0}_{P\times(N-P-1)} \,\right]^{\mathsf{T}}, \\
\mathbf{G}_{R} &= \left[\, \mathbf{0}_{P\times(N-P)} \;\; \widetilde{\mathbf{G}}^{\mathsf{T}} \,\right]^{\mathsf{T}}, \\
\mathbf{G}_{R\mathrm{c}} &= \left[\, \mathbf{0}_{P\times(N-P)} \;\; \widetilde{\mathbf{G}}_{\mathrm{c}}^{\mathsf{T}} \,\right]^{\mathsf{T}},
\end{aligned}$$

where $\mathbf{G}$ and $\mathbf{G}_{\mathrm{c}}$ are the Jacobian matrices corresponding to $\mathbf{a}_{\mathrm{FS}}$, as defined in the description of (5.17); $[\widetilde{\mathbf{G}}]_{ij} = [\mathbf{G}]_{pj}$ and $[\widetilde{\mathbf{G}}_{\mathrm{c}}]_{ij} = [\mathbf{G}_{\mathrm{c}}]_{pj}$ for $p = P - i + 1$ and $i, j \in \{1, \ldots, P\}$. With these definitions, the gradient of $\mathbf{a}_{+}^{\mathsf{H}}\mathbf{V}\mathbf{a}_{+}$ with respect to $\boldsymbol{\delta}_{\mathbf{k}}^{*}$ can be given as

$$\nabla_{\boldsymbol{\delta}_{\mathbf{k}}^{*}}(\mathbf{a}_{+}^{\mathsf{H}}\mathbf{V}\mathbf{a}_{+}) = \mathbf{Q}_{3}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{Q}_{3\mathrm{c}}\boldsymbol{\delta}_{\mathbf{k}}^{*} + \mathbf{r}_{3},$$

where

$$\begin{aligned}
\mathbf{Q}_{3} &= \mathbf{G}_{+}^{\mathsf{H}}\mathbf{V}\mathbf{G}_{+} + \left(\mathbf{G}_{+\mathrm{c}}^{\mathsf{H}}\mathbf{V}\mathbf{G}_{+\mathrm{c}}\right)^{*}, \\
\mathbf{Q}_{3\mathrm{c}} &= \mathbf{G}_{+}^{\mathsf{H}}\mathbf{V}\mathbf{G}_{+\mathrm{c}} + \left(\mathbf{G}_{+\mathrm{c}}^{\mathsf{H}}\mathbf{V}\mathbf{G}_{+}\right)^{*},
\end{aligned}$$

and

$$\mathbf{r}_{3} = \mathbf{G}_{+}^{\mathsf{H}}\mathbf{V}\mathbf{a}_{\mathrm{FS},+} + \left(\mathbf{G}_{+\mathrm{c}}^{\mathsf{H}}\mathbf{V}\mathbf{a}_{\mathrm{FS},+}\right)^{*}.$$

Similarly, the gradient of $\mathbf{a}_{R}^{\mathsf{T}}\mathbf{V}\mathbf{a}_{R}^{*}$ with respect to $\boldsymbol{\delta}_{\mathbf{k}}^{*}$ is

$$\nabla_{\boldsymbol{\delta}_{\mathbf{k}}^{*}}\left(\mathbf{a}_{R}^{\mathsf{T}}\mathbf{V}\mathbf{a}_{R}^{*}\right) = \mathbf{Q}_{4}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{Q}_{4\mathrm{c}}\boldsymbol{\delta}_{\mathbf{k}}^{*} + \mathbf{r}_{4},$$

where

$$\begin{aligned}
\mathbf{Q}_{4} &= \mathbf{G}_{R}^{\mathsf{T}}\mathbf{V}^{*}\mathbf{G}_{R\mathrm{c}}^{*} + \left(\mathbf{G}_{R\mathrm{c}}^{\mathsf{T}}\mathbf{V}^{*}\mathbf{G}_{R}^{*}\right)^{*}, \\
\mathbf{Q}_{4\mathrm{c}} &= \mathbf{G}_{R}^{\mathsf{T}}\mathbf{V}^{*}\mathbf{G}_{R}^{*} + \left(\mathbf{G}_{R\mathrm{c}}^{\mathsf{T}}\mathbf{V}^{*}\mathbf{G}_{R\mathrm{c}}^{*}\right)^{*},
\end{aligned}$$

and

$$\mathbf{r}_{4} = \mathbf{G}_{R}^{\mathsf{T}}\mathbf{V}^{*}\mathbf{a}_{\mathrm{FS},R}^{*} + \left(\mathbf{G}_{R\mathrm{c}}^{\mathsf{T}}\mathbf{V}^{*}\mathbf{a}_{\mathrm{FS},R}^{*}\right)^{*}.$$

Combining the gradient results of $\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ with the gradients of terms used for (5.21), we get the following equation system:

$$\begin{bmatrix}
\mathbf{Q}_{1} + \widetilde{\mathbf{Q}}_{1} + \mathbf{Q}_{\mathrm{tr},1} & \mathbf{Q}_{2} + \widetilde{\mathbf{Q}}_{2} + \mathbf{Q}_{\mathrm{tr},2} \\
\mathbf{Q}_{2}^{*} + \widetilde{\mathbf{Q}}_{2}^{*} + \mathbf{Q}_{\mathrm{tr},2}^{*} & \mathbf{Q}_{1}^{*} + \widetilde{\mathbf{Q}}_{1}^{*} + \mathbf{Q}_{\mathrm{tr},1}^{*}
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{\delta}_{\mathbf{k}} \\ \boldsymbol{\delta}_{\mathbf{k}}^{*}
\end{bmatrix}
= -
\begin{bmatrix}
\mathbf{r}_{1} + \widetilde{\mathbf{r}}_{1} + \mathbf{r}_{\mathrm{tr}} \\
\mathbf{r}_{1}^{*} + \widetilde{\mathbf{r}}_{1}^{*} + \mathbf{r}_{\mathrm{tr}}^{*}
\end{bmatrix}, \qquad (5.31)$$

where

$$\mathbf{Q}_{\mathrm{tr},1} = \frac{L}{A}(\mathbf{Q}_{3\mathrm{c}} - \mathbf{Q}_{4\mathrm{c}})^*,$$
$$\mathbf{Q}_{\mathrm{tr},2} = \frac{L}{A}(\mathbf{Q}_{3} - \mathbf{Q}_{4})^*,$$

and

$$\mathbf{r}_{\mathrm{tr}} = \frac{L}{A}(\mathbf{r}_3 - \mathbf{r}_4)^*$$

with

$$A = L\,\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right) + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l.$$

In (5.31), the terms $\mathbf{Q}_1$, $\mathbf{Q}_2$ and $\mathbf{r}_1$ corresponding to $(1/N)\log\det\left(\mathbf{R}_{f,N}\right)$ are the same as used in (5.21); the terms $\widetilde{\mathbf{Q}}_1$, $\widetilde{\mathbf{Q}}_2$ and $\widetilde{\mathbf{r}}_1$ corresponding to $\sum_{l=1}^{L}\widehat{\mathbf{x}}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l$ are calculated as given in (5.22), (5.23) and (5.24), respectively, by using

$$\mathbf{P} = \sum_{l=1}^{L}\left(\mathbf{M}_{l,1}^{\mathsf{H}}\mathbf{M}_{l,1} - \mathbf{M}_{l,2}^{\mathsf{H}}\mathbf{M}_{l,2}\right) \text{ and } \mathbf{v} = \sum_{l=1}^{L}\left(\mathbf{M}_{l,1}^{\mathsf{H}}\mathbf{b}_{l,1} - \mathbf{M}_{l,2}^{\mathsf{H}}\mathbf{b}_{l,2}\right),$$

where $\mathbf{b}_{l,1} = \widehat{\mathbf{x}}_{l,1:N}^* + \mathbf{M}_{l,1}\,\mathbf{a}_{\mathrm{FS}}$, $\mathbf{b}_{l,2} = \mathbf{M}_{l,2}\,\mathbf{a}_{\mathrm{FS}}$, and the matrices $\mathbf{M}_{l,1}$ and $\mathbf{M}_{l,2}$ are formed by using (5.18) and (5.19), respectively, with $\mathbf{f}_l = \widehat{\mathbf{x}}_l$, $l = 1, \ldots, L$.

To determine the optimal perturbation vector for the real-valued processes, by using the condition $\boldsymbol{\delta}_{\mathbf{k}} = \boldsymbol{\delta}_{\mathbf{k}}^*$, (5.31) can be simplified as

$$\left(\mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 + \mathbf{Q}_{\mathrm{tr},1} + \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2 + \mathbf{Q}_{\mathrm{tr},2}\right)\boldsymbol{\delta}_{\mathbf{k}} = -(\mathbf{r}_1 + \widetilde{\mathbf{r}}_1 + \mathbf{r}_{\mathrm{tr}}). \tag{5.32}$$

To solve the problem of AR parameter estimation with multiple noisy snapshots, the suggested second stage for the calculation of $\mathrm{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ without an eigendecomposition is outlined in Algorithm 5.3. The complete procedure including four cases is given in Algorithm 5.4. In Algorithm 5.4, the first case calculates $\mathbf{K}$ matrix exactly and assumes σ_v^2 is known by following Line 4 in Algorithm 5.5, and Lines 5 and 10 in Algorithm 5.6; the second case calculates $\mathbf{K}$ matrix approximately and assumes σ_v^2 is known by following Line 7 in Algorithm 5.5, and Lines 7 and 10 in Algorithm 5.6; the third and fourth cases are similar to the first and second cases, respectively, except σ_v^2 is estimated via Line 12 in Algorithm 5.6.

Algorithm 5.3: Efficient calculation of the optimal perturbation around $\mathbf{a}_{\text{FS}}$.

Input : $\widehat{\mathbf{X}} = \left[\, \widehat{\mathbf{x}}_1 \ \ldots \ \widehat{\mathbf{x}}_L \,\right]$, $\mathbf{K}$, $\mathbf{a}_{\text{FS}}$

Output : $\mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{new}}$

$\mathbf{a}^{\text{old}} = \mathbf{a}_{\text{FS}}$

for iteration $\leftarrow 1$ **to** 10 **do**

  // The loop with 10 iterations yields good performance

  $\mathbf{k} = \text{atog}\left(\left[\, 1 \ \ (\mathbf{a}^{\text{old}})^{\mathsf{T}} \,\right]^{\mathsf{T}}\right)$ // $\text{atog}\,(\cdot)$: Step-down recursion, [4, p. 236]

  if $\widehat{\mathbf{X}}$ is real-valued **then**

    Solve $\left(\mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 + \mathbf{Q}_{\text{tr},1} + \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2 + \mathbf{Q}_{\text{tr},2}\right)\boldsymbol{\delta}_{\mathbf{k}} = -(\mathbf{r}_1 + \widetilde{\mathbf{r}}_1 + \mathbf{r}_{\text{tr}})$

    // See (5.32), where $\text{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ is computed by the GS formula for $\mathbf{R}_{f,N}^{-1}$

  else

    Solve

$$\begin{bmatrix} \mathbf{Q}_1 + \widetilde{\mathbf{Q}}_1 + \mathbf{Q}_{\text{tr},1} & \mathbf{Q}_2 + \widetilde{\mathbf{Q}}_2 + \mathbf{Q}_{\text{tr},2} \\ \mathbf{Q}_2^* + \widetilde{\mathbf{Q}}_2^* + \mathbf{Q}_{\text{tr},2}^* & \mathbf{Q}_1^* + \widetilde{\mathbf{Q}}_1^* + \mathbf{Q}_{\text{tr},1}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_{\mathbf{k}} \\ \boldsymbol{\delta}_{\mathbf{k}}^* \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_1 + \widetilde{\mathbf{r}}_1 + \mathbf{r}_{\text{tr}} \\ \mathbf{r}_1^* + \widetilde{\mathbf{r}}_1^* + \mathbf{r}_{\text{tr}}^* \end{bmatrix}$$

    // See (5.31), where $\text{tr}\left(\mathbf{R}_{f,N}^{-1}\mathbf{K}\right)$ is computed by the GS formula for $\mathbf{R}_{f,N}^{-1}$

  end

  $\mathbf{k} = \mathbf{k} + \boldsymbol{\delta}_{\mathbf{k}}$ // Reflection coefficients update

  $\left[\, 1 \ \ (\mathbf{a}^{\text{new}})^{\mathsf{T}} \,\right]^{\mathsf{T}} = \text{gtoa}\,(\mathbf{k})$ // $\text{gtoa}\,(\cdot)$: Step-up recursion, [4, p. 233]

  $\mathbf{a}^{\text{old}} = \mathbf{a}^{\text{new}}$

end

$(\sigma_\epsilon^2)^{\text{new}} = \left(1/(LN)\right)\left(L\,\text{tr}\left(\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{K}\right) + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^{\mathsf{H}} \mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\widehat{\mathbf{x}}_l\right)$

// $\text{tr}\left(\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{K}\right)$ is computed by the GS formula for $\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})$

return $\mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{new}}$

Algorithm 5.4: Proposed AR parameter estimation method, also see [76].

Input : $\mathbf{Y} = [\,\mathbf{y}_1 \ \ldots \ \mathbf{y}_L\,]$, P, $t_{\max}$: the maximum number of iterations, σ_v^2 (optional)

Output : $\widehat{\mathbf{a}}$, $\widehat{\sigma}_\epsilon^2$, $\widehat{\sigma}_v^2$ (or σ_v^2 if exists)

Initialization: $\mathbf{a}^{\text{old}} = \mathbf{a}^{\text{init}}$, $(\sigma_\epsilon^2)^{\text{old}} = (\sigma_\epsilon^2)^{\text{init}}$ and $(\sigma_v^2)^{\text{old}} = (\sigma_v^2)^{\text{init}}$ by applying Algorithm 5.7

for $t \leftarrow 1$ **to** $t_{\max}$ **do**

  // Expectation step of EM, see (5.5) and (5.6)

  Compute $\widehat{\mathbf{X}}$ and $\mathbf{K}$ by applying Algorithm 5.5

  // Maximization step of EM

  Compute $\mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{new}}$ and $(\sigma_v^2)^{\text{new}}$ by applying Algorithm 5.6

  $\mathbf{a}^{\text{old}} = \mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{old}} = (\sigma_\epsilon^2)^{\text{new}}$ and $(\sigma_v^2)^{\text{old}} = (\sigma_v^2)^{\text{new}}$

end

return $\widehat{\mathbf{a}} = \mathbf{a}^{\text{new}}$, $\widehat{\sigma}_\epsilon^2 = (\sigma_\epsilon^2)^{\text{new}}$, $\widehat{\sigma}_v^2 = (\sigma_v^2)^{\text{new}}$

Algorithm 5.5: Expectation step of EM for the proposed AR parameter estimation method given in Algorithm 5.4.

Input : $\mathbf{a}^{\text{old}}$, $(\sigma_\epsilon^2)^{\text{old}}$, $(\sigma_v^2)^{\text{old}}$

Output : $\widehat{\mathbf{X}} = [\,\widehat{\mathbf{x}}_1 \ \ldots \ \widehat{\mathbf{x}}_L\,]$, $\mathbf{K}$

if the exact calculation of the covariance matrix $\mathbf{K}$ is desired **then**

  $\mathbf{Q} = \big((\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) + (\sigma_v^2)^{\text{old}}\mathbf{I}_N\big)^{-1}$ // $N \times N$ matrix inversion

  Compute $\widehat{\mathbf{X}} = [\,\widehat{\mathbf{x}}_1 \ \ldots \ \widehat{\mathbf{x}}_L\,]$, where $\widehat{\mathbf{x}}_l = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})\mathbf{Q}\mathbf{y}_l$ is the mean vector of the posterior density $\mathsf{f}(\mathbf{x}_l|\mathbf{y}_l;\boldsymbol{\theta}^{\text{old}})$ for $l = 1, \ldots, L$

  $\mathbf{K} = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) - (\sigma_\epsilon^4)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})\mathbf{Q}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})$

else

  Compute $\widehat{\mathbf{X}} = [\,\widehat{\mathbf{x}}_1 \ \ldots \ \widehat{\mathbf{x}}_L\,]$ by applying Algorithm 5.2

  Construct $\mathbf{K}$ matrix as a banded matrix with P nonzero super/sub-diagonals by using $P+1$ error correlation lags,
  $r_e[k] = ((\sigma_v^2)^{\text{old}}/(2\pi j)) \oint_{C:|z|=1} H^{\text{IIR-NC}}(z)\, z^{k-1}\, dz$, $k = 0, 1, \ldots, P$, where
  $H^{\text{IIR-NC}}(z) = (\sigma_\epsilon^2)^{\text{old}}/\big((\sigma_\epsilon^2)^{\text{old}} + (\sigma_v^2)^{\text{old}} A^{\text{old}}(z)\big(A^{\text{old}}(1/z^*)\big)^*\big)$ and
  $A^{\text{old}}(z) = 1 + a_1^{\text{old}} z^{-1} + a_2^{\text{old}} z^{-2} + \cdots + a_P^{\text{old}} z^{-P}$.

end

return $\widehat{\mathbf{X}}$, $\mathbf{K}$

Algorithm 5.6: Maximization step of EM for the proposed AR parameter estimation method given in Algorithm 5.4.

Input : $\widehat{\mathbf{X}} = [\,\widehat{\mathbf{x}}_1 \;\ldots\; \widehat{\mathbf{x}}_L\,]$, $\mathbf{K}$

Output : $\mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{new}}$, $(\sigma_v^2)^{\text{new}}$

// The first stage estimation by (5.16) for $L_a = L$ and $\mathbf{f}_l = \widehat{\mathbf{x}}_l$, $l = 1, \ldots, L$

$$\mathbf{a}_{\text{FS}} = -\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{\ell,f} + \mathbf{A}_{\ell,b}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{\ell,b}\right)\right)^{-1}\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{\ell,f} + \mathbf{A}_{\ell,b}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{\ell,b}\right)\right)$$

// The second stage estimation

if the exact calculation of the covariance matrix $\mathbf{K}$ is desired **then**

  $\mathbf{K} = \sum_{n=1}^{N} \lambda_n \mathbf{e}_n \mathbf{e}_n^{\mathsf{H}}$ // The eigendecomposition of $N \times N$ matrix

  Form the set of $L_a = L + N$ vectors $\mathbf{f}_\ell$, $\ell = 1, \ldots, L_a$, where $\mathbf{f}_l = \widehat{\mathbf{x}}_l$, $l = 1, \ldots, L$, and $\mathbf{f}_{L+n} = \sqrt{L\lambda_n}\mathbf{e}_n$, $n = 1, \ldots, N$

  Compute $\mathbf{a}^{\text{new}}$ and $(\sigma_\epsilon^2)^{\text{new}}$ by following Lines 2 to 14 in Algorithm 5.1

else

  Compute $\mathbf{a}^{\text{new}}$ and $(\sigma_\epsilon^2)^{\text{new}}$ by applying Algorithm 5.3

end

if σ_v^2 exists **then**

  $(\sigma_v^2)^{\text{new}} = \sigma_v^2$

else

  $$(\sigma_v^2)^{\text{new}} = \left(1/(LN)\right)\left(L\,\mathrm{tr}\,(\mathbf{K}) + \sum_{l=1}^{L}\|\mathbf{y}_l - \widehat{\mathbf{x}}_l\|_2^2\right)$$

end

return $\mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{new}}$, $(\sigma_v^2)^{\text{new}}$

5.2.4 Computational Complexity Considerations

In noise-free multiple-snapshot case given in Algorithm 5.1, $\mathbf{a}_{\text{FS}}$ is computed with $\mathcal{O}\left(P^3\right)$ multiplications in the first stage, and the optimal perturbation vector $\boldsymbol{\delta}_{\mathbf{k}}$ is computed with $\mathcal{O}\left(P^3\right)$ and $\mathcal{O}\left(8P^3\right)$ multiplications for the real- and complex-valued processes, respectively, in the second stage. In (5.21), $A = \sum_{\ell=1}^{L_a} \mathbf{f}_\ell^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$, which is required to get $\widetilde{\mathbf{Q}}_1$, $\widetilde{\mathbf{Q}}_2$ and $\widetilde{\mathbf{r}}_1$, is calculated efficiently by using [9, Algorithm 1], which decreases $\mathcal{O}\left(N^3\right)$ operations to $\mathcal{O}\left(NP\right)$ multiplications to calculate $\mathbf{f}_\ell^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$. The noise-free case requires $L_a = L$ snapshots and corresponds to $\mathcal{O}\left(LNP + P^3\right)$

and $\mathcal{O}\left(LNP + 8P^3\right)$ multiplications per EM iteration for the real- and complex-valued processes, respectively.

In noisy multiple-snapshot case with the exact calculation of $\mathbf{K}$ matrix given in Algorithm 5.4, both of the inversion and eigendecomposition of $N \times N$ matrix required in expectation step of EM and the second part of the maximization step of EM, respectively, are computed with $\mathcal{O}\left(N^3\right)$ operations. Owing to Line 5, Algorithm 5.1 is also utilized for this case with $L_a = L + N$ instead of $L_a = L$ corresponding to the noise-free case. Hence, the overall complexity is $\mathcal{O}\left(N^3 + (L+N)NP + P^3\right)$ and $\mathcal{O}\left(N^3 + (L+N)NP + 8P^3\right)$ operations for the real- and complex-valued processes, respectively, in each EM iteration.

When $\mathbf{K}$ matrix is calculated approximately in Algorithm 5.4, the inversion of $N \times N$ matrix is not required due to the disturbance smoother, and the eigendecomposition of $\mathbf{K}$ matrix is eliminated by using the GS formula for the second stage estimation given in Algorithm 5.3. Similar to Algorithm 5.1, Algorithm 5.3 has the complexity of $\mathcal{O}\left(LNP + P^3\right)$ and $\mathcal{O}\left(LNP + 8P^3\right)$ multiplications per EM iteration for the real- and complex-valued processes, respectively.

5.3 Initialization of the Suggested Algorithm

The proposed method requires the initial estimates for $\mathbf{a}$, σ_ϵ^2 and σ_v^2, see (5.5) and (5.6). By applying the weighted forward-backward prediction approach to the set of $\{\mathbf{y}_l\}_{l=1}^{L}$ vectors, as applied to the set of $\{\mathbf{f}_\ell\}_{\ell=1}^{L_a}$ vectors in (5.15), the initial estimate of $\mathbf{a}$, $\mathbf{a}^{\text{init}}$, is calculated as

$$\mathbf{a}^{\text{init}} = -\left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{l,f} + \mathbf{A}_{l,b}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{l,b}\right)\right)^{-1}\left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{l,f} + \mathbf{A}_{l,b}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{l,b}\right)\right) \tag{5.33}$$

where

$$\mathbf{A}_{l,f} = \begin{bmatrix} y_{l,P} & y_{l,P-1} & \cdots & y_{l,1} \\ y_{l,P+1} & y_{l,P} & \cdots & y_{l,2} \\ \vdots & \vdots & \cdots & \vdots \\ y_{l,N-1} & y_{l,N-2} & \cdots & y_{l,N-P} \end{bmatrix}, \quad \mathbf{b}_{l,f} = \begin{bmatrix} y_{l,P+1} \\ y_{l,P+2} \\ \vdots \\ y_{l,N} \end{bmatrix},$$

$$\mathbf{A}_{l,b} = \begin{bmatrix} y^*_{l,N-P+1} & y^*_{l,N-P+2} & \cdots & y^*_{l,N} \\ y^*_{l,N-P} & y^*_{l,N-P+1} & \cdots & y^*_{l,N-1} \\ \vdots & \vdots & \ldots & \vdots \\ y^*_{l,2} & y^*_{l,3} & \cdots & y^*_{l,P+1} \end{bmatrix}, \quad \mathbf{b}_{l,b} = \begin{bmatrix} y^*_{l,N-P} \\ y^*_{l,N-P-1} \\ \vdots \\ y^*_{l,1} \end{bmatrix},$$

and $\mathbf{W} = \mathrm{diag}\,(1, 2, \ldots, N-P)$. Then, the initial values of σ_ϵ^2 and σ_v^2 are determined by using YW equations and $\mathbf{a}^{\text{init}}$.

According to the block diagram shown in Figure 5.1, the YW equations for the autocorrelation sequence of the AR(P) process $x[n]$ can be written as

$$r_x[k] = -\sum_{p=1}^{P} a_p r_x[k-p] + \sigma_\epsilon^2 \delta[k], \quad k \geq 0, \tag{5.34}$$

where $r_x[k] = \mathsf{E}\{x[n]\, x^*[n-k]\}$, and the YW equations for the autocorrelation sequence of $y[n]$ can be expressed as

$$r_y[k] = r_x[k] + \sigma_v^2 \delta[k], \quad k \geq 0, \tag{5.35}$$

which implies that $r_y[k] = r_x[k]$ for $k > 0$, [4]. Using (5.34) and (5.35), we get

$$(\mathbf{R}_y - \sigma_v^2 \mathbf{I}_P)\, \mathbf{a} = -\mathbf{r}_y, \tag{5.36}$$

where

$$\mathbf{R}_y = \begin{bmatrix} r_y[0] & r_y[-1] & \ldots & r_y[-P+1] \\ r_y[1] & r_y[0] & \ldots & r_y[-P+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_y[P-1] & r_y[P-2] & \ldots & r_y[0] \end{bmatrix} \text{ and } \mathbf{r}_y = \begin{bmatrix} r_y[1] \\ r_y[2] \\ \vdots \\ r_y[P] \end{bmatrix}.$$

In (5.36), $\mathbf{R}_y$ and $\mathbf{r}_y$ are replaced with $\widehat{\mathbf{R}}_y$ and $\widehat{\mathbf{r}}_y$ estimates formed by using

$$\widehat{r}_y[k] = \frac{1}{L} \sum_{l=1}^{L} \widehat{r}_{y_l}[k],$$

where

$$\widehat{r}_{y_l}[k] = \frac{1}{N} \sum_{n=k}^{N-1} y_l[n]\, y_l^*[n-k], \quad k = 0, 1, \ldots, P \text{ and } l = 1, \ldots, L,$$

and $\widehat{r}_y[-k] = \widehat{r}_y^*[k]$. Replacing $\mathbf{a}$ with $\mathbf{a}^{\text{init}}$, the LS solution of

$$(\widehat{\mathbf{R}}_y - \sigma_v^2 \mathbf{I}_P)\, \mathbf{a}^{\text{init}} = -\widehat{\mathbf{r}}_y$$

for σ_v^2 is the initial value of σ_v^2, that is

$$(\sigma_v^2)^{\text{init}} = \frac{\mathsf{Re}\left\{(\mathbf{a}^{\text{init}})^{\mathsf{H}}\big(\widehat{\mathbf{r}}_y + \widehat{\mathbf{R}}_y\mathbf{a}^{\text{init}}\big)\right\}}{\|\mathbf{a}^{\text{init}}\|_2^2}, \tag{5.37}$$

see [52, Eq. 46] and [60, Eq. 23] for the real-valued AR process. Using YW equations (5.34) and (5.35) for $k = 0$, the initial value of σ_ϵ^2 is calculated as

$$\begin{aligned}(\sigma_\epsilon^2)^{\text{init}} &= \widehat{r}_y[0] + (\mathbf{a}^{\text{init}})^{\mathsf{T}}\widehat{\mathbf{r}}_y^{*} - (\sigma_v^2)^{\text{init}} \\ &= \widehat{r}_y[0] + \widehat{\mathbf{r}}_y^{\mathsf{H}}\mathbf{a}^{\text{init}} - (\sigma_v^2)^{\text{init}}.\end{aligned} \tag{5.38}$$

For the initialization of σ_ϵ^2 and σ_v^2, another suggestion can be made by assuming that

$$(\sigma_v^2)^{\text{init}} = (\sigma_\epsilon^2)^{\text{init}}$$

and following the approach for the estimate of

$$\widehat{\sigma}_\epsilon^2 = \big(\mathbf{x}^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}\mathbf{x}\big)/N$$

given in [9] such that

$$(\sigma_v^2)^{\text{init}} + (\sigma_\epsilon^2)^{\text{init}} = \frac{1}{LN}\sum_{l=1}^{L}\mathbf{y}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{init}})\mathbf{y}_l. \tag{5.39}$$

Thus, we have

$$(\sigma_v^2)^{\text{init}} = (\sigma_\epsilon^2)^{\text{init}} = \frac{1}{2LN}\sum_{l=1}^{L}\mathbf{y}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{init}})\mathbf{y}_l. \tag{5.40}$$

If the measurement noise variance σ_v^2 is known, say the unknown parameter vector is changed from $\boldsymbol{\theta} = \begin{bmatrix}\sigma_v^2 & \sigma_\epsilon^2 & \mathbf{a}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$ to $\boldsymbol{\theta} = \begin{bmatrix}\sigma_\epsilon^2 & \mathbf{a}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$, then the true value σ_v^2 can be substituted for $(\sigma_v^2)^{\text{init}}$ in (5.39). Therefore, we get

$$(\sigma_\epsilon^2)^{\text{init}} = \frac{1}{LN}\sum_{l=1}^{L}\mathbf{y}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{init}})\mathbf{y}_l - \sigma_v^2. \tag{5.41}$$

The initialization methods are summarized in Algorithm 5.7.

Algorithm 5.7: Initialization of the proposed AR parameter estimation method, also see [76].

Input : $\mathbf{Y} = [\,\mathbf{y}_1 \,\ldots\, \mathbf{y}_L\,]$, P, σ_v^2 (optional)

Output : $\mathbf{a}^{\text{init}}$, $(\sigma_\epsilon^2)^{\text{init}}$, $(\sigma_v^2)^{\text{init}}$

$\mathbf{a}^{\text{init}} = -\left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{l,f} + \mathbf{A}_{l,b}^{\mathsf{H}}\mathbf{W}\mathbf{A}_{l,b}\right)\right)^{-1}\left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{l,f} + \mathbf{A}_{l,b}^{\mathsf{H}}\mathbf{W}\mathbf{b}_{l,b}\right)\right)$

if σ_v^2 exists **then**

  $(\sigma_v^2)^{\text{init}} = \sigma_v^2$ and $(\sigma_\epsilon^2)^{\text{init}} = \left(1/(LN)\right)\sum_{l=1}^{L}\mathbf{y}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{init}})\mathbf{y}_l - \sigma_v^2$

else

  $(\sigma_v^2)^{\text{init}} = \mathsf{Re}\left\{(\mathbf{a}^{\text{init}})^{\mathsf{H}}(\widehat{\mathbf{r}}_y + \widehat{\mathbf{R}}_y\mathbf{a}^{\text{init}})\right\}/\|\mathbf{a}^{\text{init}}\|_2^2$ and
  $(\sigma_\epsilon^2)^{\text{init}} = \widehat{r}_y[0] + \widehat{\mathbf{r}}_y^{\mathsf{H}}\mathbf{a}^{\text{init}} - (\sigma_v^2)^{\text{init}}$,

  or $(\sigma_v^2)^{\text{init}} = (\sigma_\epsilon^2)^{\text{init}} = \left(1/(2LN)\right)\sum_{l=1}^{L}\mathbf{y}_l^{\mathsf{H}}\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{init}})\mathbf{y}_l$

end

return $\mathbf{a}^{\text{init}}$, $(\sigma_\epsilon^2)^{\text{init}}$, $(\sigma_v^2)^{\text{init}}$

5.4 Numerical Results

We present a performance comparison of the suggested AR parameter estimation method under different noise conditions including the noise-free case. The performance comparison of AR parameter estimation methods requires an application specific fidelity criterion such as spectral peak location, spurious peak avoidance, filter coefficient/pole accuracy etc., as discussed at length in [75].

In this study, we use objective metrics such as total mean square error (MSE)

$$\mathsf{E}\{\|\mathbf{a} - \widehat{\mathbf{a}}\|_2^2\},$$

attained likelihood value and the Hellinger distance between the true and estimated Gaussian random vector densities. The Hellinger distance between zero-mean Gaussian vectors $\mathbf{p}$ and $\mathbf{q}$ with covariance matrices Σ_1 and Σ_2, [78], is given as

$$\mathcal{H}(\mathbf{p},\mathbf{q}) = \sqrt{1 - \left(|\Sigma_1|^{1/4}\,|\Sigma_2|^{1/4}/|(\Sigma_1+\Sigma_2)/2|^{1/2}\right)}. \tag{5.42}$$

The Hellinger distance is a true metric satisfying positivity, symmetry, triangle inequality axioms. The Hellinger distance is utilized to upper bound the detection error

of equally likely hypotheses $\mathbf{p}$ and $\mathbf{q}$ with $p_e \leq \big(1-\mathcal{H}^2(\mathbf{p},\mathbf{q})\big)/2$, and used as a robust measure for the distance between distributions, [79]. All experiments are conducted by using 100 Monte Carlo runs.

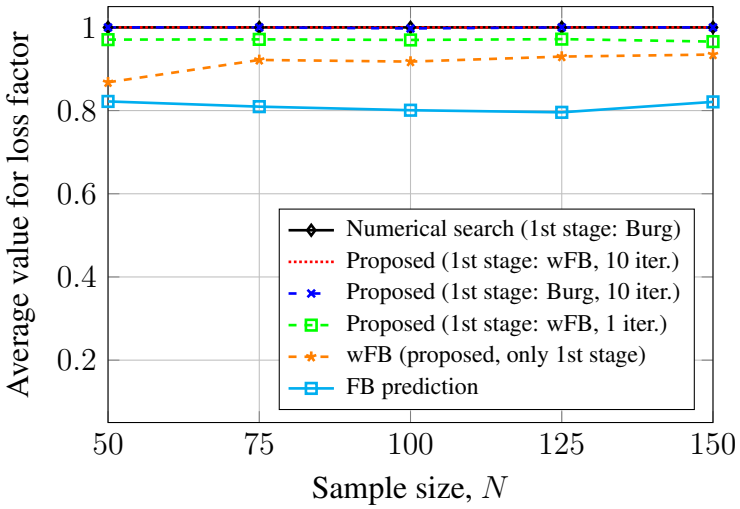
Experiment 1 - Multiple Noise-free Snapshots: This experiment compares the likelihood values attained by the estimates of different methods including the forward-backward (FB) prediction, the weighted forward-backward (wFB) prediction, which performs only the first stage estimation presented in Section 5.2.2.1, and the numerical search method initialized with Burg's method. For comparison purposes, one of the cases presented in [9] that is the estimation of the real- and complex-valued AR(6) parameters randomly generated from the uniformly distributed reflection coefficients and $\sigma_\epsilon^2 = 0.36$ for a single snapshot scenarios is repeated for $L = 10$ snapshots. The results in Figure 5.2 closely follow the earlier conclusions and reveal that the proposed approach yields likelihood values which are almost identical to the ones attained by the numerical search having much higher complexity.

Experiment 2 - Multiple Noisy Snapshots: This experiment examines the estimation accuracy of the method on an AR(4) system with

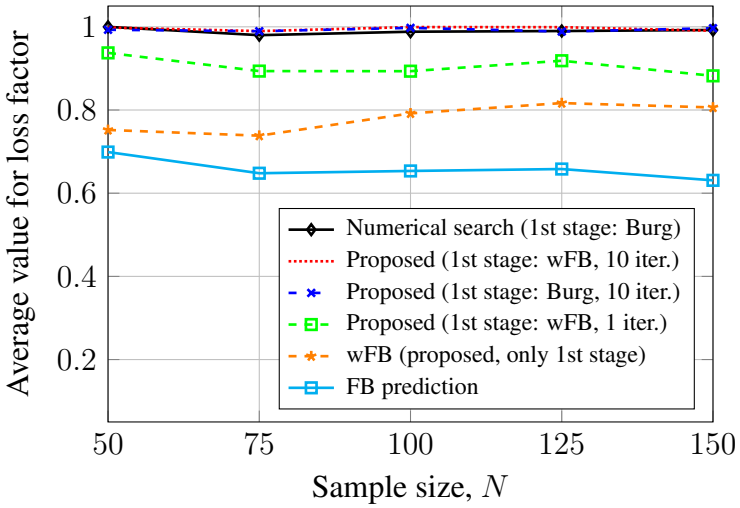
$$A(z) = 1 + 0.1z^{-1} + 0.2z^{-2} + 0.3z^{-3} + 0.4z^{-4}$$

under unity measurement noise variance, $\sigma_v^2 = 1$. The parameter σ_ϵ^2 in (5.1) is varied according to the experiment SNR. Figure 5.3 shows the Hellinger distance and total MSE comparisons for different L (number of snapshots) values, where each of the L snapshots is a vector of length $N = 50$. Four estimators, for the cases of $\mathbf{K}$ is calculated exactly or approximately and σ_v^2 is known exactly or estimated, are compared. Figure 5.3 also includes the estimator performance in the noise-free scenario. The estimator for the noise-free case has $\widehat{\mathbf{x}}_l = \mathbf{y}_l$ and $\mathbf{K} = \mathbf{0}_{N\times N}$. These relations can be also retrieved from (5.5) and (5.6) as $\sigma_v^2 \to 0$. The asymptotic CRB (ACRB) [43, 44] is provided as a performance benchmark for total MSE comparisons. The results of a similar setup constructed to compare the estimation accuracy for the complex-valued AR(4) system with

$$A(z) = 1 + (0.1 + j0.2)z^{-1} + (0.2 - j0.3)z^{-2} + (0.3 + j0.1)z^{-3} + (0.4 - j0.2)z^{-4}$$



(a)

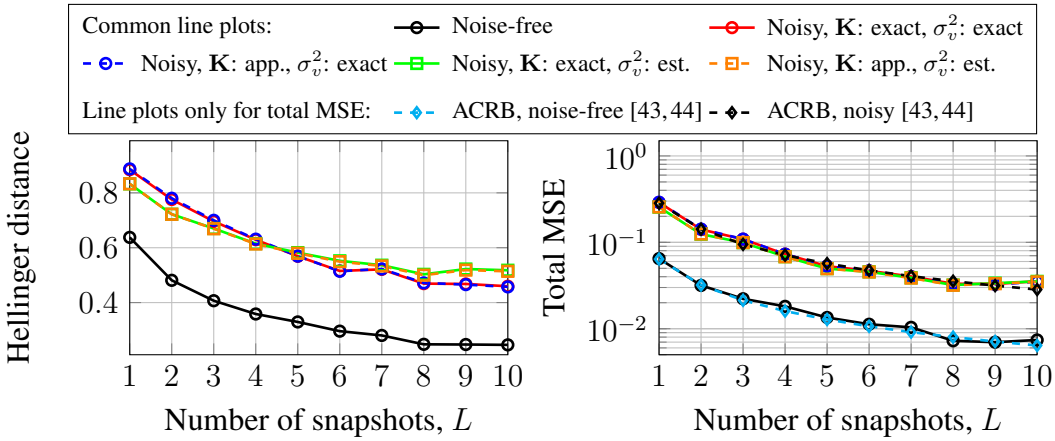


(b)

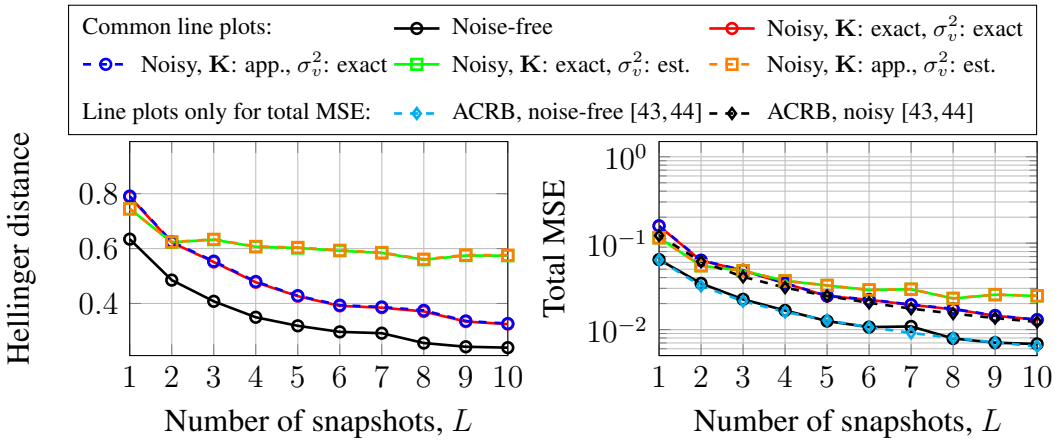
Figure 5.2: Experiment 1 - Multiple noise-free snapshots: Likelihood value comparison for AR(6) parameters randomly generated from the uniformly distributed reflection coefficients and $\sigma_\epsilon^2 = 0.36$; (a) real- and (b) complex-valued AR processes.

and $\sigma_v^2 = 1$ are shown in Figure 5.4.

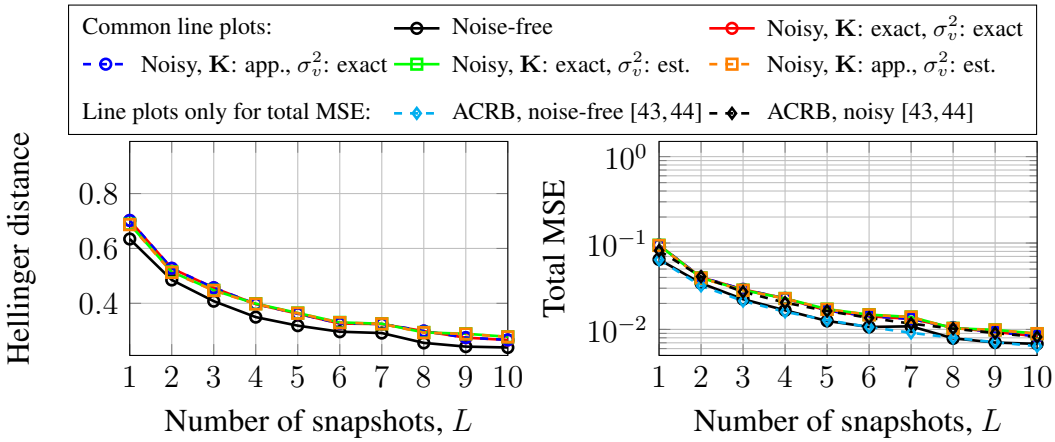
The results shown in Figure 5.3 with SNR $= 0$ dB and Figure 5.4 with SNR $= 0$ dB and SNR $= 5$ dB are obtained by initializing the suggested method with (5.40), instead of (5.37) and (5.38) used in Algorithm 5.4. Further details on the algorithm initialization are given in Section 5.3.



(a) SNR $= 0$ dB

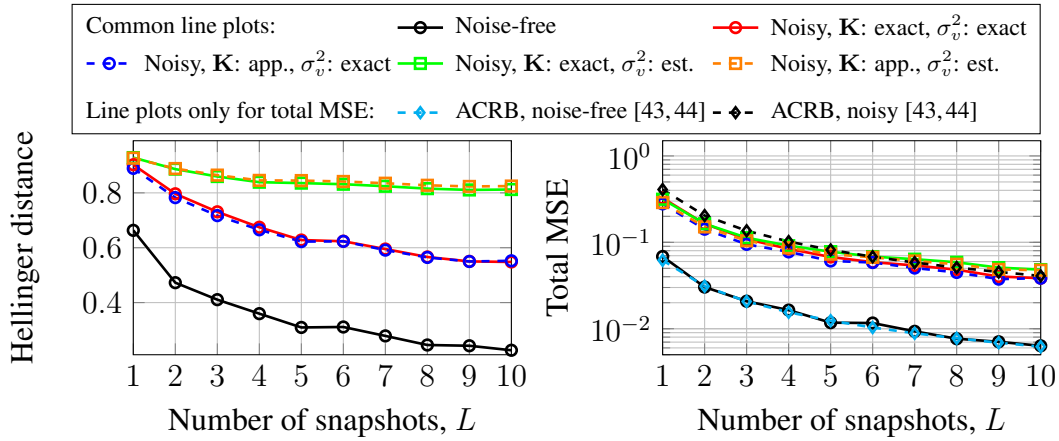


(b) SNR $= 5$ dB

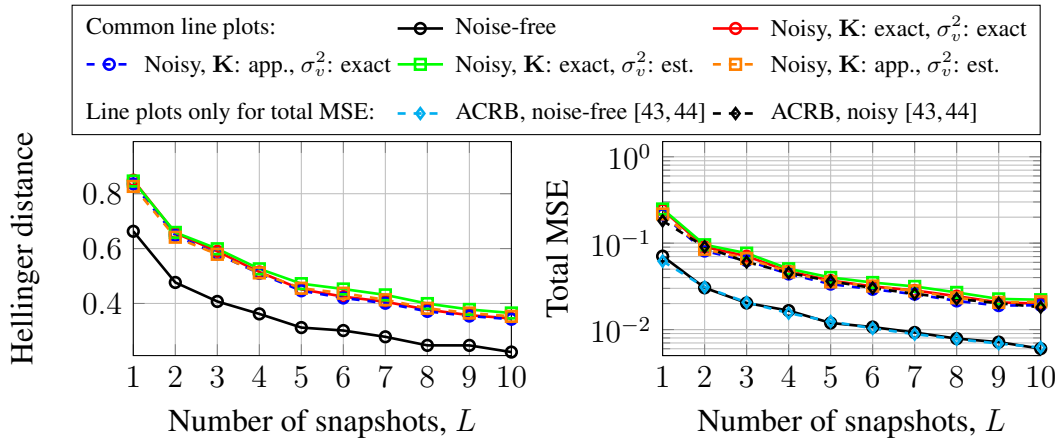


(c) SNR $= 10$ dB

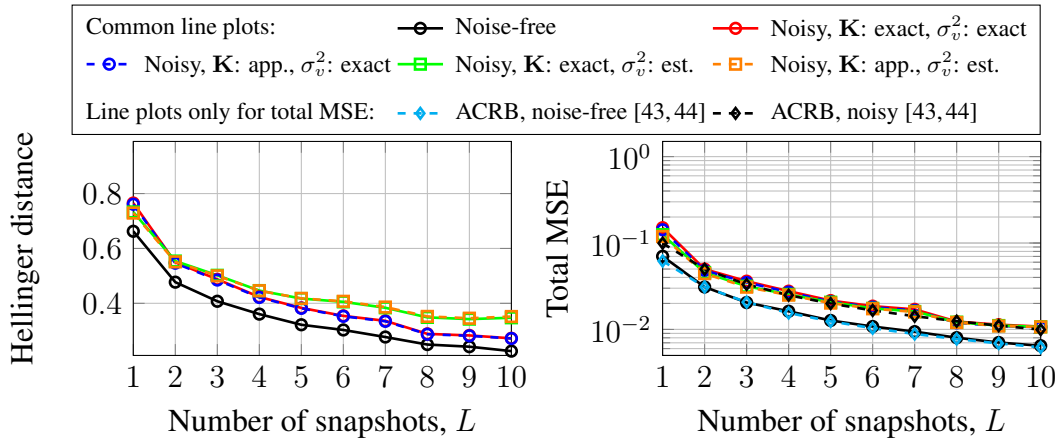
Figure 5.3: Experiment 2 - Hellinger distance and total MSE comparisons of the proposed method at different numbers of snapshots for the real-valued AR(4).



(a) SNR $= 0$ dB



(b) SNR $= 5$ dB



(c) SNR $= 10$ dB

Figure 5.4: Experiment 2 - Hellinger distance and total MSE comparisons of the proposed method at different numbers of snapshots for the complex-valued AR(4).

Table 5.1: Hellinger distance comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the real-valued AR(4) process given in Experiment 2.

		σ_v^2: exact		σ_v^2: est.	
N	L	$\mathbf{K}$: exact	$\mathbf{K}$: app.	$\mathbf{K}$: exact	$\mathbf{K}$: app.
50	1	0.7017	0.7022	0.6884	0.6869
50	5	0.3624	0.3627	0.3559	0.3551
50	10	0.2602	0.2604	0.2642	0.2643
100	1	0.7041	0.7043	0.6870	0.6866
100	5	0.3799	0.3801	0.4270	0.4272
100	10	0.2680	0.2681	0.3580	0.3585
150	1	0.6905	0.6906	0.6870	0.6869
150	5	0.3794	0.3796	0.4705	0.4706
150	10	0.2590	0.2591	0.4288	0.4291
500	1	0.7080	0.7081	0.8247	0.8247
500	5	0.3652	0.3653	0.7804	0.7804
500	10	0.2645	0.2645	0.7771	0.7771
1000	1	0.6882	0.6882	0.9210	0.9210
1000	5	0.3733	0.3733	0.9295	0.9295
1000	10	0.2631	0.2631	0.9222	0.9222

We note from Figures 5.3 and 5.4 that there is a significant performance gap between the cases of known and unknown measurement noise variance. Yet, in both cases the total MSE value for vector $\mathbf{a}$ coincides the ACRB associated with the problem. More detailed numerical results for SNR of 10 dB are given in Tables 5.1 to 5.4.

In Tables 5.2 and 5.4, the column labeled as "CRB" is the nonasymptotic (exact)

Table 5.2: Total MSE comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the real-valued AR(4) process given in Experiment 2.

		σ_v^2: exact		σ_v^2: est.			σ_v^2: exact
N	L	$\mathbf{K}$: exact	$\mathbf{K}$: app.	$\mathbf{K}$: exact	$\mathbf{K}$: app.	ACRB[43, 44]	CRB
50	1	0.0946	0.0948	0.0944	0.0938	0.0817	0.0858
50	5	0.0163	0.0163	0.0164	0.0164	0.0163	0.0165
50	10	0.0085	0.0086	0.0090	0.0090	0.0082	0.0082
100	1	0.0420	0.0420	0.0424	0.0424	0.0409	0.0418
100	5	0.0089	0.0089	0.0099	0.0099	0.0082	0.0082
100	10	0.0043	0.0043	0.0053	0.0053	0.0041	0.0041
150	1	0.0280	0.0280	0.0283	0.0283	0.0272	0.0277
150	5	0.0060	0.0060	0.0078	0.0078	0.0054	0.0055
150	10	0.0028	0.0028	0.0046	0.0046	0.0027	0.0027
500	1	0.0087	0.0087	0.0114	0.0114	0.0082	0.0082
500	5	0.0017	0.0017	0.0048	0.0048	0.0016	0.0016
500	10	0.0008	0.0008	0.0041	0.0041	0.0008	0.0008
1000	1	0.0043	0.0043	0.0074	0.0074	0.0041	0.0041
1000	5	0.0009	0.0009	0.0046	0.0046	0.0008	0.0008
1000	10	0.0004	0.0004	0.0038	0.0038	0.0004	0.0004

bound for the problem by numerical differentiation expressions [80, Eq. (3.31), p. 47] and [80, Eq. (15.52), p. 525] corresponding to the real- and complex-valued AR processes, respectively.

Table 5.3: Hellinger distance comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the complex-valued AR(4) process given in Experiment 2.

		σ_v^2: exact		σ_v^2: est.	
N	L	$\mathbf{K}$: exact	$\mathbf{K}$: app.	$\mathbf{K}$: exact	$\mathbf{K}$: app.
50	1	0.7654	0.7610	0.7329	0.7290
50	5	0.3872	0.3860	0.4247	0.4251
50	10	0.2782	0.2778	0.3379	0.3419
100	1	0.7138	0.7122	0.7221	0.7218
100	5	0.3791	0.3779	0.5876	0.5888
100	10	0.2749	0.2755	0.5516	0.5531
150	1	0.7328	0.7319	0.7883	0.7883
150	5	0.3737	0.3730	0.7022	0.7027
150	10	0.2742	0.2743	0.7002	0.7007
500	1	0.7336	0.7332	0.9700	0.9700
500	5	0.3716	0.3715	0.9705	0.9705
500	10	0.2547	0.2548	0.9676	0.9676
1000	1	0.6911	0.6909	0.9979	0.9979
1000	5	0.3613	0.3613	0.9985	0.9985
1000	10	0.2646	0.2644	0.9985	0.9985

Table 5.4: Total MSE comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots on the complex-valued AR(4) process given in Experiment 2.

		σ_v^2: exact		σ_v^2: est.			σ_v^2: exact
N	L	$\mathbf{K}$: exact	$\mathbf{K}$: app.	$\mathbf{K}$: exact	$\mathbf{K}$: app.	ACRB[43, 44]	CRB
50	1	0.1504	0.1424	0.1250	0.1191	0.0996	0.1034
50	5	0.0199	0.0194	0.0189	0.0186	0.0199	0.0199
50	10	0.0107	0.0105	0.0107	0.0106	0.0100	0.0099
100	1	0.0477	0.0468	0.0414	0.0412	0.0498	0.0504
100	5	0.0112	0.0110	0.0121	0.0121	0.0100	0.0099
100	10	0.0049	0.0049	0.0075	0.0075	0.0050	0.0049
150	1	0.0363	0.0358	0.0307	0.0306	0.0332	0.0333
150	5	0.0064	0.0063	0.0094	0.0094	0.0066	0.0066
150	10	0.0034	0.0034	0.0077	0.0077	0.0033	0.0033
500	1	0.0108	0.0108	0.0131	0.0131	0.0100	0.0099
500	5	0.0020	0.0020	0.0074	0.0074	0.0020	0.0020
500	10	0.0009	0.0010	0.0062	0.0062	0.0010	0.0010
1000	1	0.0049	0.0048	0.0098	0.0098	0.0050	0.0049
1000	5	0.0010	0.0010	0.0066	0.0066	0.0010	0.0010
1000	10	0.0004	0.0004	0.0063	0.0063	0.0005	0.0005

Experiment 3 - Single Noisy Snapshot: This experiment compares the estimation accuracy of the proposed method with those of the state-of-the-art methods, SS [51], EIV [52], IFILS [50], XZ [53], and four methods (EVK-1, EVK-2, EVK-3, EVK-4) given in [60] for the unknown noise variance. The experiment is conducted on the single snapshot real-valued AR(4) process defined in Experiment 2. Similar to Example 1 given in [60], the parameter settings are as follows: $q = 2$ and $\delta = 0.001$ for IFILS; $\delta_1 = 0.01$ when SNR $= 0$ dB and $\delta_1 = 0.001$ when SNR $= 5$ dB or SNR $= 10$ dB, $\delta_2 = 0.01$ and $\eta = 0.96$ for XZ; $q = 3$ and $\delta = 0.1$ for EVK-1 and EVK-2; $q = 4$ and $\delta = 0.1$ for EVK-3; $q = 3$ and $m = 8$ for EVK-4. (The reader can consult [60] for the parameter descriptions.)

Tables 5.5 and 5.6 show that the suggested method using either the exact or approximated error covariance matrix $\mathbf{K}$ provides more accurate AR parameter estimates in terms of Hellinger distance and total MSE in comparison to other methods. Moreover, Table 5.7 shows information on the number of unstable AR systems out of 100 trials. (The results for SNR $= 0$ dB are obtained by initializing with (5.40).)

Experiment 4 - Average Computational Time: This experiment compares the average computational time (in seconds) of the proposed method for different N (sample size) and L (number of snapshots) values in Experiment 2 setup at SNR $= 10$ dB. Tables 5.8 and 5.9 show that the suggested method with approximate error covariance matrix $\mathbf{K}$ (eliminating the inversion of an $N \times N$ matrix via the disturbance smoother in expectation step and the eigendecomposition via the GS formula in maximization step of EM) requires significantly less CPU time than the suggested method with exact $\mathbf{K}$ matrix. In Tables 5.8 and 5.9, the column labeled "Ratio" contains the ratios of average computational times.

Table 5.5: Hellinger distance comparison for the single snapshot real-valued AR(4) process given in Experiment 2.

N	SNR	$\mathbf{K}$: exact	$\mathbf{K}$: app.	SS [52]	EIV [51]	IFILS [50]	XZ [53]	EVK-1 [60]	EVK-2 [60]	EVK-3 [60]	EVK-4 [60]
50	0 dB	**0.8312**	0.8315	0.9326	0.9387	0.9402	0.9515	0.9355	0.9441	0.9630	0.8967
50	5 dB	0.7451	**0.7446**	0.9299	0.8501	0.8589	0.9472	0.8465	0.9359	0.9034	0.8925
50	10 dB	0.6884	**0.6869**	0.9226	0.8177	0.8109	0.9149	0.7843	0.9057	0.8659	0.9076
100	0 dB	0.8558	**0.8555**	0.9596	0.9535	0.9602	0.9858	0.9606	0.9781	0.9782	0.9259
100	5 dB	**0.8297**	0.8299	0.9314	0.8687	0.9177	0.9417	0.8595	0.9487	0.9436	0.9623
100	10 dB	0.6870	**0.6866**	0.9190	0.8027	0.8405	0.9137	0.7917	0.9048	0.8869	0.9755
150	0 dB	0.8703	**0.8698**	0.9594	0.9623	0.9704	0.9703	0.9602	0.9877	0.9752	0.9576
150	5 dB	0.8893	0.8895	0.9351	**0.8788**	0.9105	0.9286	0.8885	0.9347	0.9432	0.9767
150	10 dB	0.6870	**0.6869**	0.9291	0.8233	0.8753	0.8919	0.8122	0.8950	0.8914	0.9855

Table 5.6: Total MSE comparison for the single snapshot real-valued AR(4) process given in Experiment 2.

N	SNR	$\mathbf{K}$: exact	$\mathbf{K}$: app.	SS [52]	EIV [51]	IFILS [50]	XZ [53]	EVK-1 [60]	EVK-2 [60]	EVK-3 [60]	EVK-4 [60]
50	0 dB	**0.2520**	0.2534	0.6938	0.3245	0.2554	0.3685	0.5761	0.6164	0.6421	0.3814
50	5 dB	0.1156	**0.1152**	0.3225	0.1777	0.1691	0.2993	0.2921	0.3943	0.4148	0.2310
50	10 dB	0.0944	**0.0938**	0.2672	0.1486	0.1456	0.2075	0.2017	0.2301	0.3716	0.1888
100	0 dB	**0.1265**	0.1266	0.9673	0.2867	0.2675	0.4038	0.5159	0.6870	0.5805	0.3014
100	5 dB	0.0620	**0.0620**	0.2090	0.1213	0.1450	0.1772	0.1587	0.2430	0.3508	0.1452
100	10 dB	0.0424	**0.0424**	0.1943	0.0867	0.0884	0.1860	0.1572	0.2482	0.3240	0.1193
150	0 dB	0.0849	**0.0848**	0.2577	0.1860	0.1711	0.2945	0.4021	0.4964	0.4038	0.2034
150	5 dB	**0.0502**	0.0502	0.1090	0.0764	0.1030	0.1918	0.1825	0.2439	0.2787	0.1171
150	10 dB	0.0283	**0.0283**	0.1235	0.0460	0.0668	0.1647	0.1052	0.1628	0.2125	0.0881

Table 5.7: Comparison of the number of unstable system estimates out of 100 runs for the single snapshot real-valued AR(4) process given in Experiment 2.

N	SNR	$\mathbf{K}$: exact	$\mathbf{K}$: app.	SS [52]	EIV [51]	IFILS [50]	XZ [53]	EVK-1 [60]	EVK-2 [60]	EVK-3 [60]	EVK-4 [60]
50	0 dB	–	–	70	10	13	36	31	35	19	–
50	5 dB	–	–	62	10	8	20	19	22	12	–
50	10 dB	–	–	60	6	11	26	14	14	8	–
100	0 dB	–	–	60	10	14	26	15	21	10	–
100	5 dB	–	–	33	4	7	30	10	16	2	–
100	10 dB	–	–	34	1	5	17	8	11	5	–
150	0 dB	–	–	46	8	8	29	15	19	7	–
150	5 dB	–	–	19	3	6	16	9	14	7	–
150	10 dB	–	–	23	–	1	15	2	5	3	–

Table 5.8: The average computational time (in seconds) comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots for the real-valued AR(4) process in Experiment 2.

		σ_v^2: exact			σ_v^2: est.		
N	L	$\mathbf{K}$: exact	$\mathbf{K}$: app.	Ratio	$\mathbf{K}$: exact	$\mathbf{K}$: app.	Ratio
50	1	0.2810	0.0372	7.5514	0.2802	0.0370	7.5788
50	5	0.3008	0.0654	4.5955	0.3009	0.0654	4.5989
50	10	0.3270	0.0941	3.4736	0.3270	0.0942	3.4703
100	1	0.5801	0.0427	13.5718	0.5806	0.0428	13.5522
100	5	0.6071	0.0810	7.4917	0.6072	0.0812	7.4793
100	10	0.6327	0.1152	5.4907	0.6348	0.1151	5.5170
150	1	0.9734	0.0519	18.7452	0.9723	0.0518	18.7583
150	5	0.9845	0.0976	10.0890	0.9863	0.0976	10.1026
150	10	1.0143	0.1377	7.3640	1.0160	0.1380	7.3642
500	1	5.6314	0.1071	52.5841	5.8269	0.1070	54.4819
500	5	5.7710	0.2515	22.9437	5.9788	0.2525	23.6791
500	10	5.9106	0.3766	15.6936	6.1155	0.3774	16.2040
1000	1	18.9031	0.2096	90.1922	19.0119	0.2093	90.8282
1000	5	20.2822	0.5896	34.4003	20.5374	0.5886	34.8913
1000	10	21.2807	0.9352	22.7548	21.6083	0.9370	23.0603

Table 5.9: The average computational time (in seconds) comparison of the proposed method at 10 dB SNR for different sample sizes and numbers of snapshots for the complex-valued AR(4) process in Experiment 2.

		σ_v^2: exact			σ_v^2: est.		
N	L	$\mathbf{K}$: exact	$\mathbf{K}$: app.	Ratio	$\mathbf{K}$: exact	$\mathbf{K}$: app.	Ratio
50	1	0.5379	0.0562	9.5631	0.5351	0.0552	9.6881
50	5	0.5829	0.1083	5.3835	0.5812	0.1082	5.3716
50	10	0.6369	0.1628	3.9128	0.6380	0.1638	3.8959
100	1	1.2242	0.0729	16.7909	1.2054	0.0725	16.6221
100	5	1.2338	0.1328	9.2888	1.2373	0.1361	9.0880
100	10	1.2879	0.1951	6.6009	1.2800	0.1940	6.5986
150	1	1.9246	0.0811	23.7213	1.9184	0.0814	23.5769
150	5	1.9972	0.1590	12.5615	1.9967	0.1607	12.4283
150	10	2.0345	0.2273	8.9494	2.0287	0.2285	8.8783
500	1	13.7240	0.1844	74.4286	13.7681	0.1850	74.4232
500	5	14.2745	0.4096	34.8524	14.2137	0.4110	34.5829
500	10	14.8254	0.6022	24.6197	14.7588	0.6093	24.2244
1000	1	48.2300	0.3524	136.8763	48.0578	0.3531	136.0939
1000	5	48.8622	0.8417	58.0521	48.3051	0.8422	57.3534
1000	10	51.6543	1.2907	40.0217	50.9763	1.2935	39.4101

5.5 Conclusion

An expectation-maximization based solution for the noisy AR parameter estimation problem and its efficient implementation are given in this chapter. The formulation given in [9] for a single snapshot likelihood maximization of AR parameter estimation problem is extended to the multiple snapshots. Our formulation also covers the problem of AR parameter estimation under the effect of white noise with the unknown variance given multiple independent snapshots. Furthermore, an approximate implementation of the suggested method is given. With a significant less computational load, the approximate version shows a performance almost identical to that of the exact version. We present the MATLAB codes of the proposed method in [76] for further exploration.

CHAPTER 6

CONCLUSIONS

This thesis presents algorithms for two problems of interest in signal processing area. These problems are as follows:

1. Transmit beamformer design subject to a peak-to-average power ratio (PAPR) constraint.
2. Maximum likelihood autoregressive (AR) model parameter estimation from the independent snapshots corrupted by white Gaussian measurement noise.

Both problems have multiple application specific variations, which can be solved by the proposed general algorithms.

In Chapters 2 and 3, we cover the transmit beamformer design problem under a PAPR constraint. The PAPR constraint is included in the problem setting to provide a trade-off mechanism between the beam shape, average power in the main lobe and other classical metrics, such as the integrated sidelobe level, the peak-sidelobe level, etc. Typically, a flat-top beampattern and maximum average power transmission over the sector of interest are jointly desired. Unfortunately, these two objectives are contradictory, and the introduced PAPR constraint provides a mean of trading one objective with the other, see Section 3.6.2. For the problem solution, we use the alternating direction method of multipliers (ADMM), which is capable of converging on a solution for the problems with convex objective function and nonconvex constraints. We combine the phase retrieval method PhareADMM [24] and the alternating projection method [17] through an ADMM formulation. For the beam shape and PAPR constraints, we make use of distinct penalty parameters in the augmented Lagrangian function. Thus, we can control the constraint violation of the beam shape and PAPR

separately. Owing to the generality of formulation, the suggested method PAPR-ADMM can be applied to both narrowband and wideband beamforming. Furthermore, depending on the application, a fast implementation with FFT (fast Fourier transform) can also be used. We invite readers to explore the MATLAB [37], Python [38], R [39], and C++ [40] implementations of the PAPR-ADMM method.

In Chapter 4, we present a solution for the transmit beamformer design problem of interest after expressing the problem in a consensus form. Our solution is based on the consensus ADMM, which is suitable for parallel computation. Numerical results indicate that PAPR-ADMM and its consensus-based counterpart PAPR-cADMM yield comparable performance. Considering their formulations, both methods have computational advantages depending on applications, see Section 4.4. The PAPR-cADMM and its wideband version PAPR-cADMM-wb allow L (number of direction samples) and F (number of frequencies) individual updates in parallel, respectively. However, L is generally much higher than the number of available processors (or cores) for parallel computation. In such a case, all of the L individual updates cannot be simultaneously processed, which makes some of these updates wait for idle processors. Hence, PAPR-ADMM becomes preferable to PAPR-cADMM in the narrowband case for higher values of L. For wideband beamforming, the value of F typically fulfills the simultaneous processing of updates, and PAPR-cADMM-wb is preferred.

In Chapter 5, we describe an expectation-maximization based solution for noisy AR parameter estimation problem and its efficient implementation. The heart of the method contains an extension of the formulation given in [9] for a single snapshot likelihood maximization of AR parameter estimation problem to the multiple snapshots. In addition to this, the current formulation examines the problem of AR parameter estimation under the effect of white noise with the unknown variance given multiple independent snapshots. A highly efficient, yet approximate, implementation of the suggested method is also given. The performance of the approximate version is almost identical to the exact version; but the approximate version eliminates $N^3 + N^2P$ multiplications per EM iteration (N is the snapshot vector length) that results in significant cost savings in both computation and memory. We present the ready-to-use MATLAB codes reproducing the presented numerical results for further exploration in [76].

REFERENCES

[1] M. I. Skolnik, *Radar Handbook*, 3rd ed. McGraw-Hill, 2008.

[2] A. Mutapcic, S.-J. Kim, and S. Boyd, "Beamforming with uncertain weights," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 348–351, May 2007, doi: 10.1109/LSP.2006.888102.

[3] H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 526–532, Mar. 1997, doi: 10.1109/78.558465.

[4] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.

[5] P. Wang, H. Li, and B. Himed, "A Bayesian parametric test for multichannel adaptive signal detection in nonhomogeneous environments," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 351–354, Apr. 2010, doi: 10.1109/LSP.2009.2039380.

[6] J. Petitjean, R. Diversi, E. Grivel, R. Guidorzi, and P. Roussilhe, "Recursive errors-in-variables approach for ar parameter estimation from noisy observations. application to radar sea clutter rejection," in *2009 IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 5, Apr. 2009, pp. 3401–3404, doi: 10.1109/ICASSP.2009.4960355.

[7] Y. Shekofteh and F. Almasganj, "Autoregressive modeling of speech trajectory transformed to the reconstructed phase space for ASR purposes," *Digit. Signal Process.*, vol. 23, no. 6, pp. 1923–1932, Dec. 2013, doi: 10.1016/j.dsp.2013.06.011.

[8] M. Aktaruzzaman and R. Sassi, "Parametric estimation of sample entropy in heart rate variability analysis," *Biomed. Signal Process. Control*, vol. 14, no. 1, pp. 141–147, Nov. 2014, doi: 10.1016/j.bspc.2014.07.011.

[9] Ç. Candan, "Making linear prediction perform like maximum likelihood in Gaussian autoregressive model parameter estimation," *Signal Process.*, vol. 166, p. 107256, Jan. 2020, doi: 10.1016/j.sigpro.2019.107256.

[10] H. He, J. Li, and P. Stoica, *Waveform Design for Active Sensing Systems: A Computational Approach*. Cambridge University Press, 2012, doi: 10.1017/CBO9781139095174.

[11] H. He, P. Stoica, and J. Li, "Wideband MIMO systems: Signal design for transmit beampattern synthesis," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 618–628, Feb. 2011, doi: 10.1109/TSP.2010.2091410.

[12] M. Soltanalian, B. Tang, J. Li, and P. Stoica, "Joint design of the receive filter and transmit sequence for active sensing," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 423–426, May 2013, doi: 10.1109/LSP.2013.2250279.

[13] Y. Tang, Y. D. Zhang, M. G. Amin, and W. Sheng, "Wideband multiple-input multiple-output radar waveform design with low peak-to-average ratio constraint," *IET Radar, Sonar Navig.*, vol. 10, no. 2, pp. 325–332, Feb. 2016, doi: 10.1049/iet-rsn.2015.0189.

[14] H. Esmaeili-Najafabadi, M. Ataei, and M. F. Sabahi, "Designing sequence with minimum psl using Chebyshev distance and its application for chaotic MIMO radar waveform design," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 690–704, Feb. 2017, doi: 10.1109/TSP.2016.2621728.

[15] Z. Cheng, Z. He, B. Liao, and M. Fang, "MIMO radar waveform design with PAPR and similarity constraints," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 968–981, Feb. 2018, doi: 10.1109/TSP.2017.2780052.

[16] M. Soltanalian, M. M. Naghsh, and P. Stoica, "A fast algorithm for designing complementary sets of sequences," *Signal Process.*, vol. 93, no. 7, pp. 2096–2102, Jul. 2013, doi: 10.1016/j.sigpro.2013.02.008.

[17] J. Tropp, I. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005, doi: 10.1109/TIT.2004.839492.

[18] L. Tang, Y. Zhu, and Q. Fu, "Designing PAR-constrained periodic/aperiodic sequence via the gradient-based method," *Signal Process.*, vol. 147, pp. 11–22, Jun. 2018, doi: 10.1016/j.sigpro.2018.01.006.

[19] H. Griffiths, L. Cohen, S. Watts, E. Mokole, C. Baker, M. Wicks, and S. Blunt, "Radar spectrum engineering and management: Technical and regulatory issues," *Proc. IEEE*, vol. 103, no. 1, pp. 85–102, Jan. 2015, doi: 10.1109/JPROC.2014.2365517.

[20] J. Song, P. Babu, and D. P. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, Aug. 2015, doi: 10.1109/TSP.2015.2425808.

[21] J. Liang, H. C. So, C. S. Leung, J. Li, and A. Farina, "Waveform design with unit modulus and spectral shape constraints via Lagrange programming neural network," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 8, pp. 1377–1386, Dec. 2015, doi: 10.1109/JSTSP.2015.2464178.

[22] J. Song, P. Babu, and D. P. Palomar, "Sequence design to minimize the weighted integrated and peak sidelobe levels," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2051–2064, Apr. 2016, doi: 10.1109/TSP.2015.2510982.

[23] J. Liang, H. C. So, J. Li, and A. Farina, "Unimodular sequence design based on alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5367–5381, Oct. 2016, doi: 10.1109/TSP.2016.2597123.

[24] J. Liang, P. Stoica, Y. Jing, and J. Li, "Phase retrieval via the alternating direction method of multipliers," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 5–9, Jan. 2018, doi: 10.1109/LSP.2017.2767826.

[25] J. Liang, H. So, J. Li, A. Farina, and D. Zhou, "On optimizations with magnitude constraints on frequency or angular responses," *Signal Process.*, vol. 145, pp. 214–224, Apr. 2018, doi: 10.1016/j.sigpro.2017.12.009.

[26] W. Rowe, P. Stoica, and J. Li, "Spectrally constrained waveform design [sp tips & tricks]," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 157–162, May 2014, doi: 10.1109/MSP.2014.2301792.

[27] Y. Jing, J. Liang, D. Zhou, and H. C. So, "Spectrally constrained unimodular sequence design without spectral level mask," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 1004–1008, Jul. 2018, doi: 10.1109/LSP.2018.2836219.

[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jun. 2011, doi: 10.1561/2200000016.

[29] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, "A general analysis of the convergence of ADMM," in *Proc. 32nd Int. Conf. Mach. Learn.*, F. Bach and D. Blei, Eds., vol. 37. PMLR, 2015, pp. 343–352.

[30] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed. Academic Press, 1969, pp. 283–298.

[31] D. P. Bertsekas, "Multiplier methods: A survey *," *IFAC Proc. Vol.*, vol. 8, no. 1, pp. 351–363, Aug. 1975, doi: 10.1016/S1474-6670(17)67759-0.

[32] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. John Wiley & Sons, 1987, doi: 10.1002/9781118723203.

[33] M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, no. 5, pp. 303–320, Nov. 1969, doi: 10.1007/BF00927673.

[34] D. P. Bertsekas, "On penalty and multiplier methods for constrained minimization," in *Nonlinear Program. 2*, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, Eds. Elsevier, 1975, pp. 165–191, doi: 10.1016/B978-0-12-468650-2.50010-9.

[35] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming*. John Wiley & Sons, 2006, doi: 10.1002/0471787779.

[36] C. Candan, "Properly handling complex differentiation in optimization and approximation problems," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 117–124, Mar. 2019, doi: 10.1109/MSP.2018.2876761.

[37] Ö. Çayır and Ç. Candan, "PAPR-ADMM: A method for transmit beamformer

design under PAPR constraint (MATLAB code)," 2020, doi: 10.24433/CO.7886035.v1.

[38] ——, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (Python code)," 2022, doi: 10.24433/CO.2603813.v1.

[39] ——, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (R code)," 2022, doi: 10.24433/CO.1947438.v1.

[40] ——, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (C++ code)," 2022, doi: 10.24433/CO.4042158.v1.

[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–22, Sep. 1977, doi: 10.1111/j.2517-6161.1977.tb01600.x.

[42] S. M. Kay, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. Acoust.*, vol. 28, no. 3, pp. 292–303, Jun. 1980, doi: 10.1109/TASSP.1980.1163406.

[43] L. Weruaga and O. M. Melko, "Asymptotic Cramér-Rao bound for noise-compensated autoregressive analysis," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 59, no. 9, pp. 2017–2024, Sep. 2012, doi: 10.1109/TCSI.2012.2185277.

[44] L. Weruaga and L. Dimitrov, "The spectral nature of maximum likelihood noise compensated linear prediction," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 21, no. 8, pp. 1760–1765, Aug. 2013, doi: 10.1109/TASL.2013.2255277.

[45] H. Tong, "Autoregressive model fitting with noisy data by akaike's information criterion (corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 4, pp. 476–480, Jul. 1975, doi: 10.1109/TIT.1975.1055402.

[46] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Prentice Hall, 1988.

[47] A. Nehorai and P. Stoica, "Adaptive algorithms for constrained ARMA signals in the presence of noise," *IEEE Trans. Acoust.*, vol. 36, no. 8, pp. 1282–1291, Aug. 1988, doi: 10.1109/29.1656.

[48] W. X. Zheng, "Autoregressive parameter estimation from noisy data," *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 47, no. 1, pp. 71–75, Jan. 2000, doi: 10.1109/82.818897.

[49] ——, "On estimation of autoregressive signals in the presence of noise," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 53, no. 12, pp. 1471–1475, Dec. 2006, doi: 10.1109/TCSII.2006.883094.

[50] A. Mahmoudi and M. Karimi, "Inverse filtering based method for estimation of noisy autoregressive signals," *Signal Process.*, vol. 91, no. 7, pp. 1659–1664, Jul. 2011, doi: 10.1016/j.sigpro.2011.01.008.

[51] C. Davila, "A subspace approach to estimation of autoregressive parameters from noisy measurements," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 531–534, Feb. 1998, doi: 10.1109/78.655442.

[52] R. Diversi, R. Guidorzi, and U. Soverini, "Identification of autoregressive models in the presence of additive noise," *Int. J. Adapt. Control Signal Process.*, vol. 22, no. 5, pp. 465–481, Jun. 2008, doi: 10.1002/acs.989.

[53] Y. Xia and W. X. Zheng, "Novel parameter estimation of autoregressive signals in the presence of noise," *Automatica*, vol. 62, pp. 98–105, Dec. 2015, doi: 10.1016/j.automatica.2015.09.008.

[54] D. Labarre, E. Grivel, Y. Berthoumieu, E. Todini, and M. Najim, "Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters," *Signal Process.*, vol. 86, no. 10, pp. 2863–2876, Oct. 2006, doi: 10.1016/j.sigpro.2005.12.001.

[55] J. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust.*, vol. 27, no. 1, pp. 53–62, Feb. 1979, doi: 10.1109/TASSP.1979.1163200.

[56] W.-R. Wu and P.-C. Chen, "Adaptive AR modeling in white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1184–1192, May 1997, doi: 10.1109/78.575693.

[57] Y. Zhang, C. Wen, and Y. C. Soh, "Unbiased LMS filtering in the presence of white measurement noise with unknown power," *IEEE Trans. Circuits Syst. II*

Analog Digit. Signal Process., vol. 47, no. 9, pp. 968–972, Sep. 2000, doi: 10.1109/82.868469.

[58] M. Gabrea, E. Grivel, and M. Najun, "A single microphone Kalman filter-based noise canceller," *IEEE Signal Process. Lett.*, vol. 6, no. 3, pp. 55–57, Mar. 1999, doi: 10.1109/97.744623.

[59] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. 15, no. 2, pp. 175–184, Apr. 1970, doi: 10.1109/TAC.1970.1099422.

[60] M. Esfandiari, S. A. Vorobyov, and M. Karimi, "New estimation methods for autoregressive process in the presence of white observation noise," *Signal Process.*, vol. 171, p. 107480, Jun. 2020, doi: 10.1016/j.sigpro.2020.107480.

[61] M. Deriche, "AR parameter estimation from noisy data using the EM algorithm," in *Proc. ICASSP '94. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. iv, 1994, pp. IV/69–IV/72, doi: 10.1109/ICASSP.1994.389874.

[62] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 373–385, Jul. 1998, doi: 10.1109/89.701367.

[63] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, Mar. 1983, doi: 10.1214/aos/1176346060.

[64] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 411–423, Feb. 1991, doi: 10.1109/78.80824.

[65] L. Weruaga, "Frequency-selective noise-compensated autoregressive estimation," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 58, no. 10, pp. 2469–2476, Oct. 2011, doi: 10.1109/TCSI.2011.2142830.

[66] A. M. Sykulski, S. C. Olhede, A. P. Guillaumin, J. M. Lilly, and J. J. Early, "The debiased Whittle likelihood," *Biometrika*, vol. 106, no. 2, pp. 251–266, Jun. 2019, doi: 10.1093/biomet/asy071.

[67] P. Whittle, "Estimation and information in stationary time series," *Ark. för Mat.*, vol. 2, no. 5, pp. 423–434, Aug. 1953, doi: 10.1007/BF02590998.

[68] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*. Springer-Verlag, 2006, doi: 10.1007/1-84628-329-9.

[69] S. Haykin, B. Currie, and S. Kesler, "Maximum-entropy spectral analysis of radar clutter," *Proc. IEEE*, vol. 70, no. 9, pp. 953–962, Sep. 1982, doi: 10.1109/PROC.1982.12426.

[70] S. de Waele and P. Broersen, "The Burg algorithm for segments," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2876–2880, Oct. 2000, doi: 10.1109/78.869039.

[71] S. de Waele and P. M. T. Broersen, "Spectral analysis of segmented data," in *Proc. 39th IEEE Conf. Decis. Control*, vol. 1, 2000, pp. 189–190, doi: 10.1109/CDC.2000.912756.

[72] M. A. Richards, *Fundamentals of Radar Signal Processing*, 1st ed. McGraw-Hill Education, 2005.

[73] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[74] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. Springer New York, 2005, doi: 10.1007/0-387-28982-8.

[75] S. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. Acoust.*, vol. 31, no. 1, pp. 56–65, Feb. 1983, doi: 10.1109/TASSP.1983.1164050.

[76] Ö. Çayır and Ç. Candan, "Maximum likelihood estimator for noisy autoregressive model parameter estimation problem with multiple snapshots (MATLAB code)," 2021, doi: 10.24433/CO.7257794.v1.

[77] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends® Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, Jan. 2006, doi: 10.1561/0100000006.

[78] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer New York, 2009, doi: 10.1007/b13794.

[79] D. Donoho and X. Huo, “Large-sample modulation classification using Hellinger representation,” in *First IEEE Signal Process. Work. Signal Process. Adv. Wirel. Commun.*, 1997, pp. 133–136, doi: 10.1109/SPAWC.1997.630175.

[80] S. M. Kay, *Fundamentals of Statistical Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

APPENDIX A

PERFORMANCE METRICS FOR TRANSMIT BEAMFORMING

Let the function $\mathcal{P}_\sigma(\phi,\theta) : \mathbb{R}^2 \to \mathbb{R}$ to compute the transmitted power in the direction pair $(\phi,\theta) \in \mathbb{R}^2$ by the transmitter weights $\mathbf{w}_\sigma \in \mathbb{C}^N$ having $1 \leq \mathrm{PAPR}\,(\mathbf{w}_\sigma) \leq \sigma$ be

$$\mathcal{P}_\sigma(\phi,\theta) = \left|\mathbf{a}(\phi,\theta)^{\mathsf{H}}\mathbf{w}_\sigma\right|^2, \tag{A.1}$$

where the steering function $\mathbf{a}(\phi,\theta) : \mathbb{R}^2 \to \mathbb{C}^N$ is given as

$$\mathbf{a}(\phi,\theta) = [\, a_1(\phi,\theta) \;\ldots\; a_N(\phi,\theta)\,]^{\mathsf{T}} \tag{A.2}$$

with

$$a_n(\phi,\theta) = e^{j\mathbf{k}(\phi,\theta)^{\mathsf{T}}\mathbf{p}_n}, \quad n = 1,\ldots,N, \tag{A.3}$$

for the wavenumber $\mathbf{k}$, as a function of (ϕ,θ)

$$\mathbf{k}(\phi,\theta) = (2\pi/\lambda)\,[\,\cos\phi\cos\theta\mathbf{u}_x \;\; \sin\phi\cos\theta\mathbf{u}_y \;\; \sin\theta\mathbf{u}_z\,]^{\mathsf{T}}. \tag{A.4}$$

The integrated sidelobe level (ISL) can be expressed as

$$\mathrm{ISL} = \iint\limits_{(\phi,\theta)\in\mathcal{S}} \mathcal{P}_\sigma(\phi,\theta)\,d\phi\,d\theta, \tag{A.5}$$

where

$$\mathcal{S} = \Phi_{\mathrm{s}} \times \Theta_{\mathrm{s}} = \left\{(\phi,\theta)\middle|\phi \in \Phi_{\mathrm{s}}, \theta \in \Theta_{\mathrm{s}}\right\}$$

denotes the set of direction pairs for the sidelobe.

The average of ISL over the set $\mathcal{S}$ is

$$\overline{\mathrm{ISL}} = \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} \mathcal{P}_\sigma(\phi,\theta)\,d\phi\,d\theta. \tag{A.6}$$

Substituting (A.1) into (A.6), we get

$$\begin{aligned}
\overline{\mathrm{ISL}} &= \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} \mathbf{w}_\sigma^{\mathsf{H}} \mathbf{a}(\phi,\theta)\mathbf{a}(\phi,\theta)^{\mathsf{H}} \mathbf{w}_\sigma \, d\phi \, d\theta \\
&= \mathbf{w}_\sigma^{\mathsf{H}} \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} \mathbf{a}(\phi,\theta)\mathbf{a}(\phi,\theta)^{\mathsf{H}} \, d\phi \, d\theta \, \mathbf{w}_\sigma \\
&= \mathbf{w}_\sigma^{\mathsf{H}} \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} \begin{bmatrix} a_1(\phi,\theta) \\ \vdots \\ a_N(\phi,\theta) \end{bmatrix} [\, a_1^*(\phi,\theta) \, \ldots \, a_N^*(\phi,\theta) \,] \, d\phi \, d\theta \, \mathbf{w}_\sigma \\
&= \mathbf{w}_\sigma^{\mathsf{H}} \mathbf{R}_{\mathcal{S}} \mathbf{w}_\sigma
\end{aligned} \tag{A.7}$$

where the entries of the $N \times N$ matrix $\mathbf{R}_{\mathcal{S}}$ are

$$\begin{aligned}
[\mathbf{R}_{\mathcal{S}}]_{mn} &= \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} a_m(\phi,\theta) a_n^*(\phi,\theta) \, d\phi \, d\theta \\
&= \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} e^{j\mathbf{k}(\phi,\theta)^{\mathsf{T}}\mathbf{p}_m} e^{-j\mathbf{k}(\phi,\theta)^{\mathsf{T}}\mathbf{p}_n} \, d\phi \, d\theta \\
&= \frac{1}{\mathrm{card}\,(\mathcal{S})} \iint\limits_{(\phi,\theta)\in\mathcal{S}} e^{j\mathbf{k}(\phi,\theta)^{\mathsf{T}}(\mathbf{p}_m-\mathbf{p}_n)} \, d\phi \, d\theta
\end{aligned} \tag{A.8}$$

for $m, n = 1, \ldots, N$.

Similarly, to integrate the power in the main lobe, the $N \times N$ matrix $\mathbf{R}_{\mathcal{M}}$ is given as

$$\mathbf{R}_{\mathcal{M}} = \frac{1}{\mathrm{card}\,(\mathcal{M})} \iint\limits_{(\phi,\theta)\in\mathcal{M}} \mathbf{a}(\phi,\theta)\mathbf{a}(\phi,\theta)^{\mathsf{H}} \, d\phi \, d\theta \tag{A.9}$$

with the entries

$$[\mathbf{R}_{\mathcal{M}}]_{mn} = \frac{1}{\mathrm{card}\,(\mathcal{M})} \iint\limits_{(\phi,\theta)\in\mathcal{M}} e^{j\mathbf{k}(\phi,\theta)^{\mathsf{T}}(\mathbf{p}_m-\mathbf{p}_n)} \, d\phi \, d\theta, \quad m, n = 1, \ldots, N, \tag{A.10}$$

where

$$\mathcal{M} = \Phi_{\mathrm{m}} \times \Theta_{\mathrm{m}} = \big\{ (\phi,\theta) \big| \phi \in \Phi_{\mathrm{m}}, \theta \in \Theta_{\mathrm{m}} \big\}$$

denotes the set of direction pairs for the main lobe.

By using these definitions, the performance metrics related to the power computations are given as follows:

- The average power in the main lobe is

$$P_{\text{avg}}^{\text{m}}(\mathbf{w}_\sigma) = \mathbf{w}_\sigma^{\mathsf{H}} \mathbf{R}_{\mathcal{M}} \mathbf{w}_\sigma. \tag{A.11}$$

- The minimum power in the main lobe is

$$P_{\text{min}}^{\text{m}} = \min_{(\phi,\theta)\in\mathcal{M}} \mathcal{P}_\sigma(\phi,\theta). \tag{A.12}$$

- The maximum power in the main lobe is

$$P_{\text{max}}^{\text{m}} = \max_{(\phi,\theta)\in\mathcal{M}} \mathcal{P}_\sigma(\phi,\theta). \tag{A.13}$$

- The peak-to-peak power swing in the main lobe is

$$P_{\text{p-p}}^{\text{m}} = \frac{P_{\text{max}}^{\text{m}}}{P_{\text{min}}^{\text{m}}}. \tag{A.14}$$

- The power efficiency is

$$\begin{aligned} P_{\text{eff}}(\mathbf{w}_\sigma) &= \frac{\text{PAPR}\left(\mathbf{w}_1\right)}{\text{PAPR}\left(\mathbf{w}_\sigma\right)} \\ &= \left(\text{PAPR}\left(\mathbf{w}_\sigma\right)\right)^{-1}, \end{aligned} \tag{A.15}$$

  where $\mathbf{w}_1$ denotes the weight satisfying $\text{PAPR} = 1$ constraint, *i.e.*, the unimodular condition.

- The average power in the sidelobe is

$$P_{\text{avg}}^{\text{s}}(\mathbf{w}_\sigma) = \mathbf{w}_\sigma^{\mathsf{H}} \mathbf{R}_{\mathcal{S}} \mathbf{w}_\sigma. \tag{A.16}$$

- The maximum power in the sidelobe is

$$P_{\text{max}}^{\text{s}} = \max_{(\phi,\theta)\in\mathcal{S}} \mathcal{P}_\sigma(\phi,\theta). \tag{A.17}$$

- The peak sidelobe level is

$$\text{PSL} = \frac{P_{\text{min}}^{\text{m}}}{P_{\text{max}}^{\text{s}}}. \tag{A.18}$$

- The ratio of average powers in the main lobe and sidelobe is

$$P_{\text{avg}}^{\text{m/s}}(\mathbf{w}_\sigma) = \frac{\mathbf{w}_\sigma^{\mathsf{H}} \mathbf{R}_{\mathcal{M}} \mathbf{w}_\sigma}{\mathbf{w}_\sigma^{\mathsf{H}} \mathbf{R}_{\mathcal{S}} \mathbf{w}_\sigma}. \tag{A.19}$$

- The maximum range ratio is

$$R_{\max}(\mathbf{w}_\sigma) = \left(\frac{P^{\mathrm{m}}_{\mathrm{avg}}(\mathbf{w}_\sigma)}{P^{\mathrm{m}}_{\mathrm{avg}}(\mathbf{w}_1)}\right)^{1/4} = \left(\frac{\mathbf{w}_\sigma^{\mathsf{H}}\mathbf{R}_{\mathcal{M}}\mathbf{w}_\sigma}{\mathbf{w}_1^{\mathsf{H}}\mathbf{R}_{\mathcal{M}}\mathbf{w}_1}\right)^{1/4} \tag{A.20}$$

  from the relation [1] that the maximum range is proportional to

$$(\text{average power} \times \text{aperture})^{1/4}.$$

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Çayır, Ömer

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.S.	Electrical and Electronics Eng., METU	2014
B.S.	Electrical and Electronics Eng., Hacettepe Uni.	2011

PROFESSIONAL EXPERIENCE

Year	**Place**	**Enrollment**
2013-2021	Electrical and Electronics Eng. Dept., METU	Research Assistant

PUBLICATIONS

1. Ö. Çayır and Ç. Candan, "Maximum likelihood autoregressive model parameter estimation with noise corrupted independent snapshots," *Signal Process.*, vol. 186, p. 108118, Sep. 2021, doi: 10.1016/j.sigpro.2021.108118.

2. Ö. Çayır and Ç. Candan, "Transmit beamformer design with a PAPR constraint to trade-off between beampattern shape and power efficiency," *Digit. Signal Process.*, vol. 99, p. 102674, Apr. 2020, doi: 10.1016/j.dsp.2020.102674.

3. Ö. Çayır and Ç. Candan, "Performance improvement of time-balance radar schedulers through decision policies," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1679–1691, Aug. 2018, doi: 10.1109/TAES.2018.2798418.

SOFTWARE

1. Ö. Çayır and Ç. Candan, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (C++ code)," 2022, doi: 10.24433/CO.4042158.v1.

2. Ö. Çayır and Ç. Candan, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (R code)," 2022, doi: 10.24433/CO.1947438.v1.

3. Ö. Çayır and Ç. Candan, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (Python code)," 2022, doi: 10.24433/CO.2603813.v1.

4. Ö. Çayır, Ş. Bilgi Akdemir, and Ç. Candan, "Frequency estimation of complex exponentials under signal intermittency and noise (MATLAB code)," 2022, doi: 10.24433/CO.8142587.v1.

5. Ö. Çayır and Ç. Candan, "Maximum likelihood estimator for noisy autoregressive model parameter estimation problem with multiple snapshots (MATLAB code)," 2021, doi: 10.24433/CO.7257794.v1.

6. Ö. Çayır and Ç. Candan, "PAPR-ADMM: A method for transmit beamformer design under PAPR constraint (MATLAB code)," 2020, doi: 10.24433/CO.7886035.v1.