Min-Max Framework for Majorization-Minimization Algorithms in Signal Processing Applications: An Overview

Other titles in Foundations and Trends® in Signal Processing

Causal Deep Learning: Encouraging Impact on Real-World Problems Through Causality
Jeroen Berrevoets, Krzysztof Kacprzyk, Zhaozhi Qian and Mihaela van der Schaar
ISBN: 978-1-63828-400-0

Model-Based Deep Learning
Nir Shlezinger and Yonina C. Eldar
ISBN: 978-1-63828-264-8

Generalizing Graph Signal Processing: High Dimensional Spaces, Models and Structures
Xingchao Jian, Feng Ji and Wee Peng Tay
ISBN: 978-1-63828-150-4

Learning with Limited Samples: Meta-Learning and Applications to Communication Systems
Lisha Chen, Sharu Theresa Jose, Ivana Nikoloska, Sangwoo Park, Tianyi Chen and Osvaldo Simeone
ISBN: 978-1-63828-136-8

Min-Max Framework for Majorization-Minimization Algorithms in Signal Processing Applications: An Overview

Astha Saini
Indian Institute of Technology Delhi
astha.saini@care.iitd.ac.in

Petre Stoica
Uppsala University
ps@it.uu.se

Prabhu Babu
Indian Institute of Technology Delhi
Prabhu.Babu@care.iitd.ac.in

Aakash Arora
Indian Institute of Technology Delhi
aarora@care.iitd.ac.in

the essence of knowledge
Boston — Delft

Foundations and Trends® in Signal Processing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

A. Saini *et al.*. *Min-Max Framework for Majorization-Minimization Algorithms in Signal Processing Applications: An Overview*. Foundations and Trends® in Signal Processing, vol. 18, no. 4, pp. 310–389, 2024.

ISBN: 978-1-63828-467-3

Foundations and Trends® in Signal Processing
Volume 18, Issue 4, 2024

Editorial Scope

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multirate signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing
  - Classification and detection
  - Estimation and regression
  - Tree-structured methods

Information for Librarians

Foundations and Trends® in Signal Processing, 2024, Volume 18, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Contents

1 Introduction **3**
1.1 MM Summary 4
1.2 Need for MM4MM 5
1.3 Organization 6
1.4 Notation 7

2 Max Formulation **9**
2.1 Conjugate Function 10
2.2 Max Formulation for Certain Non-Convex Functions 11
2.3 Illustrative Examples of Max Formulations 12

3 Min-Max Framework for Majorization Minimization **23**
3.1 Minimax Theorem 23
3.2 MM4MM 24

4 Special Cases **29**
4.1 Constrained Minimization Problem 29
4.2 Worst-Case Minimization Problem 30

5 Signal Processing Applications of MM4MM **31**
5.1 Total Variation Filtering 31
5.2 Phase Retrieval for Poisson Noise 33
5.3 Outlier Robust Phase Retrieval 37

5.4 RSS Based Source Localization 39
5.5 Optimal Sensor Placement 42
5.6 Independent Vector Analysis 45
5.7 Dual-Function Beamforming Design 47
5.8 Optimal Experiment Design 52
5.9 Quantum State Discrimination 55
5.10 Fair PCA 57

6 Conclusions **62**

Appendix **64**

References **67**

Min-Max Framework for Majorization-Minimization Algorithms in Signal Processing Applications: An Overview

Astha Saini[1], Petre Stoica[2], Prabhu Babu[3] and Aakash Arora[4]

[1]*Indian Institute of Technology Delhi, India; astha.saini@care.iitd.ac.in*
[2]*Uppsala University, Sweden; ps@it.uu.se*
[3]*Indian Institute of Technology Delhi, India; Prabhu.Babu@care.iitd.ac.in*
[4]*Indian Institute of Technology Delhi, India; aarora@care.iitd.ac.in*

ABSTRACT

This monograph presents a theoretical background and a broad introduction to the **M**in-**M**ax Framework **for M**ajorization-**M**inimization (MM4MM), an algorithmic methodology for solving minimization problems by formulating them as min-max problems and then employing majorization-minimization. The monograph lays out the mathematical basis of the approach used to reformulate a minimization problem as a min-max problem. With the prerequisites covered, including multiple illustrations of the formulations for convex and non-convex functions, this work serves as a guide for developing MM4MM-based algorithms for solving non-convex optimization problems in various areas of signal processing. As special cases, we discuss using the

Astha Saini, Petre Stoica, Prabhu Babu and Aakash Arora (2024), "Min-Max Framework for Majorization-Minimization Algorithms in Signal Processing Applications: An Overview", Foundations and Trends® in Signal Processing: Vol. 18, No. 4, pp 310–389. DOI: 10.1561/2000000129.

majorization-minimization technique to solve min-max problems encountered in signal processing applications and min-max problems formulated using the Lagrangian. Lastly, we present detailed examples of using MM4MM in ten signal processing applications such as phase retrieval, source localization, independent vector analysis, beamforming and optimal sensor placement in wireless sensor networks. The devised MM4MM algorithms are free of hyper-parameters and enjoy the advantages inherited from the use of the majorization-minimization technique such as monotonicity.

Keywords: Conjugate function; min-max problem; majorization-minimization; non-convex optimization.

1

Introduction

The Majorization Minimization (MM) method, with its roots in the 1970s, is a generalization of the classical Expectation-Maximization (EM) method [33]. MM was earlier known by different names like space-alternating generalized EM (SAGE) [39], optimization transfer [6], and iterative majorization [56], and it was mainly used in the field of statistics [10], [29], [55], [56], [68] and image processing [30]–[32], [46], [47], [66], [101]. MM is a general algorithmic framework for convex and non-convex problems [59], [68]. [59], [67], [69], [132], [133] triggered the attention of researchers from the statistics and signal processing fields to the MM methodology. The work [59], in particular, sparked research interest in using the MM algorithmic framework for solving non-convex, non-smooth optimization problems encountered in applications such as compressive sensing [13], [17], [73], [107], covariance estimation [109]–[111], [123], non-negative matrix factorization [40]–[42], [113], sequence design [104]–[106], [115], [127], localization in sensor networks [26], [61], [82], and image processing [11], [43], [44], [70], [83], [125].

We will always abbreviate majorization-minimization as MM. However, the abbreviation MM4MM has different (but related) meanings depending on the context:

- In Section 3.2.1, we use a min-max formulation to derive an MM algorithm and there MM4MM = min-max for MM.
- In Section 3.2.2, we swap the min and max operators and therefore MM4MM = max-min for MM.
- Finally in Section 4, we get the min-max formulation for free (without any need for a max formulation of the function to be minimized), consequently there MM4MM = MM for min-max.

1.1 MM Summary

Employing MM involves two basic steps: majorizing the objective function to create a surrogate function and then minimizing the surrogate function iteratively to solve the min problem. Consider the following optimization problem:

$$\min_{\mathbf{x}\in\mathbb{X}} f(\mathbf{x}), \tag{1.1}$$

where $\mathbb{X} \subseteq \mathbb{R}^n$ is a non-empty convex set and $f(\mathbf{x})$: $\mathbb{X} \to \mathbb{R}$ is a continuous function that is bounded below. To solve (1.1) using MM, a suitable surrogate function $f_s(\mathbf{x} \mid \mathbf{x}^t)$ needs to be constructed such that given a feasible point $\mathbf{x}^t \in \mathbb{X}$, the following inequality holds:

$$f_s(\mathbf{x} \mid \mathbf{x}^t) \geq f(\mathbf{x}) + c_t \quad \forall \;\; \mathbf{x} \in \mathbb{X}. \tag{1.2}$$

The constant $c_t \in \mathbb{R}$ ensures that equality is satisfied in (1.2) at $\mathbf{x} = \mathbf{x}^t$. In other words, the surrogate function $f_s(\mathbf{x} \mid \mathbf{x}^t)$ majorizes $f(\mathbf{x})$ and is equal to $f(\mathbf{x}^t) + c_t$ at $\mathbf{x} = \mathbf{x}^t$.

At the second step, the surrogate function is minimized yielding the next iterate point of the algorithm, that is,

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}\in\mathbb{X}} \; f_s(\mathbf{x} \mid \mathbf{x}^t). \tag{1.3}$$

This implies,

$$f_s(\mathbf{x}^{t+1} \mid \mathbf{x}^t) \leq f_s(\mathbf{x} \mid \mathbf{x}^t) \quad \forall\, \mathbf{x} \in \mathbb{X}. \tag{1.4}$$

Using (1.4) and (1.2), we get:

$$\begin{aligned} f(\mathbf{x}^{t+1}) &\leq f_s(\mathbf{x}^{t+1} \mid \mathbf{x}^t) - c_t \\ &\leq f_s(\mathbf{x}^t \mid \mathbf{x}^t) - c_t = f(\mathbf{x}^t). \end{aligned} \tag{1.5}$$

From (1.5) it follows that the sequence of function values $\{f(\mathbf{x}^t)\}$ is monotonically non-increasing. What plays a crucial role in developing a computationally efficient MM algorithm for an optimization problem is the construction of a suitable surrogate function. The surrogate function needs to be such that it follows the shape of the objective function as closely as possible and the computational cost of minimizing it is low. To construct a surrogate function, different techniques can be used such as Jensen's inequality, arithmetic-geometric mean inequality, and linearizing a concave function using a first-order Taylor expansion. Many of these techniques are discussed in [112], which also provides a general overview of the MM algorithmic framework along with applications of MM in signal processing, communication and machine learning. The MM technique has been used to solve a wide variety of optimization problems from areas such as signal processing [1], [25], [62], [91], communication [3], [28], [48], [65], [85], radar and sonar [7], [36], [98], [124], machine learning and computer vision [16], [22], [45], [57], [72], [121], [129], image recovery [18], [75], [90], intelligent transportation systems [58], [80], [130], graph learning [37], [63], [64], [71], [77], [92], [128], biomedical signal processing [23], [34], [78], [79], and neuroimaging applications [49], [52], [131].

1.2 Need for MM4MM

There exist minimization problems for which surrogate functions are difficult to derive or, even if such functions can be found, they are not convenient to deal with from a computational standpoint. Indeed for functions in $\mathbf{x} \in \mathbb{R}^n$ like $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}\log(\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x})$ where $\mathbf{A} \succ \mathbf{0}$, and $(\log\|\mathbf{x}-\mathbf{a}\|)^2$ where $\mathbf{a} \in \mathbb{R}^n$ finding a suitable surrogate is difficult. On the other hand, for $\log|\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{X}|$ (where $\mathbf{X} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{\Sigma} \in \mathbb{S}^n_{++}$) a surrogate can be found using a first-order Taylor expansion (below $\mathbf{Y} = \mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{X}$):

$$\log|\mathbf{Y}| \leq \log|\mathbf{Y}_t| + \mathrm{Tr}(\mathbf{Y}_t^{-1}(\mathbf{Y} - \mathbf{Y}_t)). \tag{1.6}$$

However, the cost of computing $\mathbf{Y}_t^{-1} = (\mathbf{X}_t^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{X}_t)^{-1}$ at each iteration is usually high.

To deal with cases such as those above, the proposed MM4MM framework expresses the function $f(\mathbf{x})$ as maximum of an augmented function $g(\mathbf{x}, \mathbf{z})$, where $\mathbf{z}$ is an auxiliary variable:

$$f(\mathbf{x}) = \max_{\mathbf{z}} \; g(\mathbf{x}, \mathbf{z}). \tag{1.7}$$

We call this representation of $f(\mathbf{x})$ the *max formulation*. Finding the function $g(\mathbf{x}, \mathbf{z})$ requires ingenuity. This function should be such that suitable surrogate function(s) for the non-convex term(s) of $g(\mathbf{x}, \mathbf{z})$ are easy to obtain, making it possible to come up with an efficient MM algorithm. The problem of minimizing $f(\mathbf{x})$ is reformulated as a min-max problem with $g(\mathbf{x}, \mathbf{z})$ as the objective function:

$$\min_{\mathbf{x}} \; \max_{\mathbf{z}} \; g(\mathbf{x}, \mathbf{z}). \tag{1.8}$$

The problem in (1.8) can be solved using MM as we explain in Section 3. We show how the proposed MM4MM algorithmic framework can be used to find suitable surrogate functions for the examples mentioned in the previous paragraph in Section 5.3 (for $\mathbf{x}^\top \mathbf{A}\mathbf{x} \log(\mathbf{x}^\top \mathbf{A}\mathbf{x})$), Section 5.4 (for $(\log \|\mathbf{x} - \mathbf{a}\|)^2$), and Section 5.5 (for $\log|\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X}|$).

One of the first algorithms using the basic ideas of the MM4MM framework is the PDMM (Primal-Dual Majorization Minimization) [38] algorithm that solves the inverse problem of phase retrieval for Poisson noise. [95] proposes a special case of MM4MM algorithm using a Lagrangian min-max formulation of the problem of E-optimal experiment design (see Section 4.1). Also, [99] uses the max formulation of the objective function and then employs MM for the problem of total variation filtering. However, this problem is convex, and thus [99] uses the max formulation only to deal with non-differentiability of ℓ_1 norm penalty. More details on the applications in [4], [38], [86], [95]–[97], [99], [114] and on the way in which these works use the MM4MM framework can be found in Section 5.

1.3 Organization

Before describing the Min-Max framework for MM, we present some preliminary results in Sections 2 and 3.1. In Section 2, we describe the max formulation for convex and non-convex functions, along with ten illustrative examples for functions which are found in various signal processing problems. The minimax theorem is discussed in Section 3.1. These sections cover the prerequisites for the MM4MM framework, which is described in Section 3.2. Section 4 describes two special cases of the MM4MM framework. The bulk of the monograph is Section 5 in which we present detailed derivations of the MM4MM algorithms for ten signal processing applications.

1.4 Notation

Italic letters (x), lower case bold letters ($\mathbf{x}$) and upper case bold letters ($\mathbf{X}$) denote scalars, vectors and matrices respectively. $|x|$ is the absolute value of x and $|\mathbf{x}|$ is the element-wise absolute value of vector $\mathbf{x}$. The ith element of a vector $\mathbf{x}$ is denoted x_i. $\mathbf{x}_i$ denotes the ith column and X_{ij} the (i,j)th element of the matrix $\mathbf{X}$. $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ the zero matrix. $\mathrm{Tr}(\mathbf{X})$ and $|\mathbf{X}|$ denote the trace and determinant of the matrix $\mathbf{X}$. $\mathbf{X} \succ \mathbf{0}$ ($\mathbf{X} \succeq \mathbf{0}$) denotes the positive definiteness (positive semi-definiteness) of the matrix $\mathbf{X}$. $\mathrm{vec}(\mathbf{X})$ is the vectorization operator reshaping a matrix $\mathbf{X}$ of size $m \times n$ in a vector of size of $mn \times 1$.

The sets of real numbers and complex numbers are denoted $\mathbb{R}$ and $\mathbb{C}$. $\mathbb{R}_+$ is the set of non-negative real numbers. $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ represent the sets of positive semi-definite matrices and positive definite matrices of size $n \times n$. A subspace is represented by $\mathcal{X}$. The indicator function $I_{\mathbb{X}}(\mathbf{x})$ is zero 0 for all $\mathbf{x}$ lying in the set $\mathbb{X}$ and ∞ otherwise. $\mathcal{N}(\mu, \sigma^2)$ denotes Gaussian distribution with mean μ and variance σ^2. The notation $\mathbf{x} \sim \mathrm{Poisson}(\lambda)$ means that $\mathbf{x}$ follows a Poisson distribution with mean and variance λ. The natural logarithm is denoted $\log(\cdot)$ and the logarithm to the base 10 is denoted $\log_{10}(\cdot)$. $f'(x)$ is the derivative of the function $f(x)$. The symbols $\odot$ and $\oslash$ denote the element-wise multiplication and division of two vectors. $\otimes$ denotes the Kronecker product of two matrices. Infimal convolution between two functions $f_1(\mathbf{x})$ and $f_2(\mathbf{y})$ is defined as:

$$(f_1 \Box f_2)(\mathbf{x}) = \min_{\mathbf{y}} \; f_1(\mathbf{x} - \mathbf{y}) + f_2(\mathbf{y}). \tag{1.9}$$

The proximal operator of a scaled function $\mathbf{prox}_{\alpha f}(\mathbf{z})$ is defined as follows:

$$\mathbf{prox}_{\alpha f}(\mathbf{z}) = \arg \min_{\mathbf{x}} \left(f(\mathbf{x}) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{z}\|_2^2 \right). \tag{1.10}$$

Superscripts $(\cdot)^\mathsf{T}$, $(\cdot)^\mathsf{H}$ and $(\cdot)^{-1}$ denote the transpose, conjugate transpose and inverse operations respectively. $(\mathbf{X})^{\frac{1}{2}}$ or $\sqrt{\mathbf{X}}$ refers to the matrix square root of the positive semi-definite matrix $\mathbf{X} \in \mathbb{S}_+^n$ such that $(\mathbf{X})^{\frac{1}{2}}(\mathbf{X})^{\frac{1}{2}} = \mathbf{X}$. The symbol $\|\mathbf{x}\|_p$ denotes the ℓ_p norm of vector $\mathbf{x} \in \mathbb{R}^n$ defined as:

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}, \tag{1.11}$$

which has the following special cases:

$$\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|, \tag{1.12}$$

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, \ldots, |x_n|\}. \tag{1.13}$$

The dual norm of any norm $\|\mathbf{x}\|$ is represented by $\|\mathbf{z}\|_*$ and is defined as:

$$\|\mathbf{z}\|_* = \max\{\mathbf{z}^\mathsf{T}\mathbf{x} \mid \|\mathbf{x}\| \leq 1\}. \tag{1.14}$$

Huber norm $\|\mathbf{x}\|_\mathrm{H}$ of vector $\mathbf{x}$ is given as follows:

$$\|\mathbf{x}\|_\mathrm{H} = \sum_{i=1}^{n} f_\alpha(x_i); \quad f_\alpha(x_i) = \begin{cases} \dfrac{|x_i|^2}{2\alpha} & \text{if } |x_i| \leq \alpha \\ |x_i| - \dfrac{\alpha}{2} & \text{if } |x_i| > \alpha. \end{cases} \tag{1.15}$$

Unless otherwise stated, $\|\mathbf{x}\|$ will be used to denote the ℓ_2 norm of the vector $\mathbf{x}$, i.e., $\|\mathbf{x}\|_2$. For matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$, the $\|\mathbf{X}\|_1$ norm is given by:

$$\|\mathbf{X}\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |X_{ij}|. \tag{1.16}$$

The nuclear norm, defined as sum of the singular values and denoted $\|\mathbf{X}\|_{2*}$, is given by:

$$\|\mathbf{X}\|_{2*} = \sigma_1(\mathbf{X}) + \cdots + \sigma_r(\mathbf{X}) = \mathrm{Tr}(\mathbf{X}^\mathsf{T}\mathbf{X})^{\frac{1}{2}}, \tag{1.17}$$

where $\{\sigma_i\}$ are the singular values and r is the rank of the matrix $\mathbf{X}$. The spectral norm, defined as the maximum singular value and denoted $\|\mathbf{X}\|_2$, is given by:

$$\|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X}). \tag{1.18}$$

References

[1] R. B. Abdallah, A. Breloy, M. N. El Korso, and D. Lautru, "Bayesian signal subspace estimation with compound Gaussian sources," *Signal Processing*, vol. 167, 2020.

[2] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[3] A. Arora, C. G. Tsinos, B. S. M. R. Rao, S. Chatzinotas, and B. Ottersten, "Hybrid transceivers design for large-scale antenna arrays using majorization-minimization algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 701–714, 2020.

[4] P. Babu and P. Stoica, "Fair principal component analysis (PCA): Minorization-maximization algorithms for Fair PCA, Fair Robust PCA and Fair Sparse PCA," *arXiv preprint arXiv:2305.05963*, 2023.

[5] J. Bae and L.-C. Kwek, "Quantum state discrimination and its applications," *Journal of Physics A: Mathematical and Theoretical*, vol. 48, no. 8, 2015.

[6] M. P. Becker, I. Yang, and K. Lange, "EM algorithms without missing data," *Statistical Methods in Medical Research*, vol. 6, no. 1, pp. 38–54, 1997.

[7] R. Ben Abdallah, A. Mian, A. Breloy, A. Taylor, M. N. El Korso, and D. Lautru, "Detection methods based on structured covariance matrices for multivariate SAR images processing," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 7, pp. 1160–1164, 2019.

[8] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*, vol. 1. Athena Scientific, 2003.

[9] L. Bian, J. Suo, J. Chung, X. Ou, C. Yang, F. Chen, and Q. Dai, "Fourier ptychographic reconstruction using Poisson maximum likelihood and truncated Wirtinger gradient," *Scientific Reports*, vol. 6, no. 1, 2016.

[10] C. C. J. H. Bijleveld and J. De Leeuw, "Fitting longitudinal reduced-rank regression models by alternating least squares," *Psychometrika*, vol. 56, pp. 433–447, 1991.

[11] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: A majorization-minimization approach," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, pp. II–II, 2006.

[12] A. N. Bishop, B. Fidan, B. D. Anderson, K. Doğançay, and P. N. Pathirana, "Optimality analysis of sensor-target localization geometries," *Automatica*, vol. 46, no. 3, pp. 479–492, 2010.

[13] T. Blumensath, M. Yaghoobi, and M. E. Davies, "Iterative hard thresholding and ℓ_0 regularisation," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, IEEE, vol. 3, pp. III–877, 2007.

[14] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[16] A. Breloy, S. Kumar, Y. Sun, and D. P. Palomar, "Majorization-minimization on the Stiefel manifold with application to robust sparse PCA," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1507–1520, 2021.

[17] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[18] M. Chalvidal, E. Chouzenoux, J.-B. Fest, and C. Lefort, "Block delayed majorize-minimize subspace algorithm for large scale image restoration," *Inverse Problems*, vol. 39, no. 4, 2023.

[19] A. Chambolle, "Total variation minimization and a class of binary MRF models," in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, Springer, pp. 136–152, 2005.

[20] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011.

[21] A. Chan, H. Silva, S. Lim, T. Kozuno, A. R. Mahmood, and M. White, "Greedification operators for policy optimization: Investigating forward and reverse KL divergences," *Journal of Machine Learning Research*, vol. 23, no. 253, pp. 1–79, 2022.

[22] L. Chapel, R. Flamary, H. Wu, C. Févotte, and G. Gasso, "Unbalanced optimal transport through non-negative penalized linear regression," *Advances in Neural Information Processing Systems*, vol. 34, pp. 23 270–23 282, 2021.

[23] S. Chatterjee, R. S. Thakur, R. N. Yadav, and L. Gupta, "Sparsity-based modified wavelet de-noising autoencoder for ECG signals," *Signal Processing*, vol. 198, 2022.

[24] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Communications on Pure and Applied Mathematics*, vol. 70, no. 5, pp. 822–883, 2017.

[25] E. Chouzenoux and J.-C. Pesquet, "A stochastic majorize-minimize subspace algorithm for online penalized least squares estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 18, pp. 4770–4783, 2017.

[26] P. J. Chung and J. F. Bohme, "Comparative convergence analysis of EM and SAGE algorithms in DOA estimation," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 2940–2949, 2001.

[27] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 2010.

[28] J. Dai, A. Liu, and V. K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2584–2599, 2018.

[29] J. De Leeuw, "Block-relaxation algorithms in statistics," in *Information Systems and Data Analysis: Prospects—Foundations—Applications*, Springer, 1994, pp. 308–324.

[30] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, pp. 328–333, 1993.

[31] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 14, no. 1, pp. 132–137, 1995.

[32] A. R. De Pierro, "On the convergence of an EM-type algorithm for penalized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 762–765, 1995.

[33] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[34] S.-W. Deng and J.-Q. Han, "Adaptive overlapping-group sparse denoising for heart sound signals," *Biomedical Signal Processing and Control*, vol. 40, pp. 49–57, 2018.

[35] Y. C. Eldar, A. Megretski, and G. C. Verghese, "Designing optimal quantum detectors via semidefinite programming," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 1007–1012, 2003.

[36] W. Fan, J. Liang, G. Yu, H. C. So, and G. Lu, "MIMO radar waveform design for quasi-equiripple transmit beampattern synthesis via weighted l_p-minimization," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3397–3411, 2019.

[37] G. Fatima, A. Arora, P. Babu, and P. Stoica, "Learning sparse graphs via majorization-minimization for smooth node signals," *IEEE Signal Processing Letters*, vol. 29, pp. 1022–1026, 2022.

[38] G. Fatima, Z. Li, A. Arora, and P. Babu, "PDMM: A novel primal-dual majorization-minimization algorithm for poisson phase-retrieval problem," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1241–1255, 2022.

[39] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, 1994.

[40] C. Févotte, "Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1980–1983, 2011.

[41] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4810–4819, 2015.

[42] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[43] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980–2991, 2007.

[44] M. A. T. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," in *2006 International Conference on Image Processing*, pp. 2633–2636, 2006.

[45] J. Geiping and M. Moeller, "Parametric majorization for data-driven energy minimization methods," in *Proceedings of the IEEE/CVF International Conference on Computer Vision,* pp. 10262–10273, 2019.

[46] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

[47] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[48] S. Gong, C. Xing, V. K. N. Lau, S. Chen, and L. Hanzo, "Majorization-minimization aided hybrid transceivers for MIMO interference channels," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4903–4918, 2020.

[49] J. M. Górriz, J. Ramirez, J. Suckling, MRC AIMS Consortium, *et al.*, "On the computation of distribution-free performance bounds: Application to small sample sizes in neuroimaging," *Pattern Recognition*, vol. 93, pp. 1–13, 2019.

[50] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming,* version 2.1, URL: http://cvxr.com/cvx, 2014.

[51] C. Ha, "A noncompact minimax theorem," *Pacific Journal of Mathematics*, vol. 97, no. 1, pp. 115–117, 1981.

[52] A. Hashemi, Y. Gao, C. Cai, S. Ghosh, K.-R. Müller, S. Nagarajan, and S. Haufe, "Efficient hierarchical Bayesian inference for spatio-temporal regression models in neuroimaging," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24 855–24 870, 2021.

[53] A. Hassanien, M. G. Amin, E. Aboutanios, and B. Himed, "Dual-function radar communication systems: A solution to the spectrum congestion problem," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 115–126, 2019.

[54] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Signaling strategies for dual-function radar communications: An overview," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 10, pp. 36–45, 2016.

[55] W. J. Heiser, "Correspondence analysis with least absolute residuals," *Computational Statistics & Data Analysis*, vol. 5, no. 4, pp. 337–356, 1987.

[56] W. J. Heiser, "Convergent computation by iterative majorization," *Recent Advances in Descriptive Multivariate Analysis*, pp. 157–189, 1995.

[57] Z. Hu, F. Nie, R. Wang, and X. Li, "Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding," *Information Fusion*, vol. 55, pp. 251–259, 2020.
[58] Y. Huang, S. Hu, S. Ma, Z. Liu, and M. Xiao, "Designing low-PAPR waveform for OFDM-based RadCom systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 6979–6993, 2022.
[59] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
[60] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2411–2422, 2007.
[61] S. Jeong, O. Simeone, A. Haimovich, and J. Kang, "Beamforming design for joint localization and data transmission in distributed antenna system," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 62–76, 2015.
[62] S. Koyama, T. Amakasu, N. Ueno, and H. Saruwatari, "Amplitude matching: Majorization–minimization algorithm for sound field control only with amplitude constraint," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 411–415, 2021.
[63] S. Kumar, J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "A unified framework for structured graph learning via spectral constraints," *Journal of Machine Learning Research*, vol. 21, no. 22, pp. 1–60, 2020.
[64] S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. Palomar, "Structured graph learning via Laplacian spectral constraints," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
[65] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided MISO communications: From LMMSE to deep learning solutions," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 471–487, 2021.

[66] K. Lange and J. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1430–1438, 1995.
[67] K. Lange, *MM Optimization Algorithms*. SIAM, 2016.
[68] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, 2000.
[69] K. Lange and H. Zhou, "MM algorithms for geometric and signomial programming," *Mathematical Programming*, vol. 143, no. 1–2, pp. 339–356, 2014.
[70] S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 983–995, 2012.
[71] X. Lin, C. Ding, J. Zhang, Y. Zhan, and D. Tao, "Ru-net: Regularized unrolling network for scene graph generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19 457–19 466, 2022.
[72] Z. Lin, C. Xu, and H. Zha, "Robust matrix factorization by majorization minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 208–220, 2018.
[73] M. Malek-Mohammadi, A. Koochakzadeh, M. Babaie-Zadeh, M. Jansson, and C. R. Rojas, "Successive concave sparsity approximation for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5657–5671, 2016.
[74] A. Mandal, W. K. Wong, and Y. Yu, "Algorithmic searches for optimal designs," *Handbook of Design and Analysis of Experiments*, pp. 755–783, 2015.
[75] Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, and J.-C. Pesquet, "Majorize–minimize adapted metropolis–hastings algorithm," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2356–2369, 2020.
[76] A. W. Marshall, I. Olkin, and B. C. Arnold, "Inequalities: Theory of majorization and its applications," 1979.

[77] R. Mirzaeifard, V. C. Gogineni, N. K. D. Venkategowda, and S. Werner, "Dynamic graph topology learning with non-convex penalties," in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 682–686, 2022.

[78] N. Mourad, "ECG denoising algorithm based on group sparsity and singular spectrum analysis," *Biomedical Signal Processing and Control*, vol. 50, pp. 62–71, 2019.

[79] P. R. Muduli, A. K. Mandal, and A. Mukherjee, "An antinoise-folding algorithm for the recovery of biomedical signals from noisy measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 11, pp. 2909–2916, 2017.

[80] A. Ndikumana, N. H. Tran, D. H. Kim, K. T. Kim, and C. S. Hong, "Deep learning based caching for self-driving cars in multi-access edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2862–2877, 2021.

[81] G. Neumann, "Variational inference for policy search in changing situations," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 817–824, 2011.

[82] P. Oguz-Ekim, J. P. Gomes, J. Xavier, and P. Oliveira, "Robust localization of nodes and time-recursive tracking in sensor networks using noisy range measurements," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3930–3942, 2011.

[83] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization–minimization approach," *Signal Processing*, vol. 89, no. 9, pp. 1683–1693, 2009.

[84] N. Ono, "Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions," in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, IEEE, pp. 1–4, 2012.

[85] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, 2020.

[86] K. Panwar, G. Fatima, and P. Babu, "Optimal sensor placement for hybrid source localization using fused TOA–RSS–AOA measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 2, pp. 1643–1657, 2023.

[87] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

[88] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[89] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[90] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 52–67, 2019.

[91] J. Ren, T. Zhang, J. Li, and P. Stoica, "Sinusoidal parameter estimation from signed measurements via majorization–minimization based RELAX," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2173–2186, 2019.

[92] S. Rey, T. M. Roddenberry, S. Segarra, and A. G. Marques, "Enhanced graph-learning schemes driven by similar distributions of Motifs," *IEEE Transactions on Signal Processing*, vol. 71, pp. 3014–3027, 2023.

[93] R. T. Rockafellar, *Convex Analysis*, vol. 11. Princeton University Press, 1997.

[94] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.

[95] N. Sahu and P. Babu, "A new monotonic algorithm for the E-optimal experiment design problem," *Statistics & Probability Letters*, vol. 174, 2021.

[96] A. Saini, A. Arora, and P. Babu, “A novel iterative algorithm for outlier robust phase retrieval via majorization minimization technique,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[97] A. Saini, P. Stoica, and P. Babu, “Maximum likelihood method for received signal strength-based source localization,” *IEEE Transactions on Aerospace and Electronic Systems,* Under Review, 2024.

[98] S. P. Sankuru and P. Babu, “Designing unimodular sequence with good auto-correlation properties via block majorization-minimization method,” *Signal Processing*, vol. 176, 2020.

[99] I. W. Selesnick and I. Bayram, “Total variation filtering,” *White Paper*, 2010.

[100] K. Shen and W. Yu, “Fractional programming for communication systems—part i: Power control and beamforming,” *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

[101] L. A. Shepp and Y. Vardi, “Maximum likelihood reconstruction for emission tomography,” *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

[102] M. Sion, “On general minimax theorems,” *Pacific Journal of Mathematics*, vol. 8, pp. 171–176, 1958.

[103] H. C. So and L. Lin, “Linear least squares approach for accurate received signal strength based source localization,” *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 4035–4040, 2011.

[104] J. Song, P. Babu, and D. P. Palomar, “Optimization methods for designing sequences with low autocorrelation sidelobes,” *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3998–4009, 2015.

[105] J. Song, P. Babu, and D. P. Palomar, “Sequence design to minimize the weighted integrated and peak sidelobe levels,” *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2051–2064, 2016.

[106] J. Song, P. Babu, and D. P. Palomar, "Sequence set design with good correlation properties via majorization-minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2866–2879, 2016.

[107] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, 2012.

[108] P. Stoica, R. L. Moses, *et al.*, *Spectral Analysis of Signals*, vol. 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[109] Y. Sun, P. Babu, and D. P. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, 2014.

[110] Y. Sun, P. Babu, and D. P. Palomar, "Regularized robust estimation of mean and covariance matrix under heavy-tailed distributions," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3096–3109, 2015.

[111] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3576–3590, 2016.

[112] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.

[113] V. Y. F. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the (β)-divergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1592–1605, 2013.

[114] B. Tang, D. Li, W. Wu, A. Saini, P. Babu, and P. Stoica, "Dual-function beamforming design for multi-target localization and reliable communications," *IEEE Transactions on Signal Processing,* Under Review, 2024.

[115] B. Tang, M. M. Naghsh, and J. Tang, "Relative entropy-based waveform design for MIMO radar detection in the presence of clutter and interference," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3783–3796, 2015.

[116] P. Tarrio, A. M. Bernardos, J. A. Besada, and J. R. Casar, "A new positioning technique for RSS-based localization based on a weighted least squares estimator," in *2008 IEEE International Symposium on Wireless Communication Systems*, IEEE, pp. 633–637, 2008.

[117] D. Tian, D. Xue, and D. Wang, "A fractional-order adaptive regularization primal–dual algorithm for image denoising," *Information Sciences*, vol. 296, pp. 147–159, 2015.

[118] N. Vaswani and P. Narayanamurthy, "Static and dynamic robust PCA and matrix completion: A review," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1359–1379, 2018.

[119] R. Vidal, "Subspace Clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2011.

[120] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 773–794, 2017.

[121] S. Wang, Y.-C. Wu, M. Xia, R. Wang, and H. V. Poor, "Machine intelligence at the edge with learning centric power allocation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7293–7308, 2020.

[122] W. Wengang, D. Guohua, and L. Mingshan, "Minimum-error quantum state discrimination based on semidefinite programming," in *2008 27th Chinese Control Conference*, pp. 521–524, 2008.

[123] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29–38, 2012.

[124] L. Wu, P. Babu, and D. P. Palomar, "Transmit waveform/receive filter design for MIMO radar with multiple waveform constraints," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1526–1540, 2018.

[125] Q. Yuan, L. Zhang, and H. Shen, "Multiframe super-resolution employing a spatially weighted total variation model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp. 379–392, 2012.

[126] H. Zhang, Y. Chi, and Y. Liang, "Median-truncated nonconvex approach for phase retrieval with outliers," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7287–7310, 2018.

[127] L. Zhao, J. Song, P. Babu, and D. P. Palomar, "A unified framework for low autocorrelation sequence design via majorization–minimization," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 438–453, 2017.

[128] L. Zhao, Y. Wang, S. Kumar, and D. P. Palomar, "Optimization algorithms for graph Laplacian estimation via ADMM and MM," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4231–4244, 2019.

[129] S. Zheng, "A majorization-minimization scheme for ℓ_2 support vector regression," *Journal of Statistical Computation and Simulation*, vol. 91, no. 15, pp. 3087–3107, 2021.

[130] B. Zhou, A. Liu, and H. C. So, "Visible light communication-enabled simultaneous position and orientation detection for harnessing multipath interference and random fading," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1938–1952, 2024.

[131] G. Zhou, D. Tward, and K. Lange, "A majorization-minimization algorithm for neuroimage registration," *SIAM Journal on Imaging Sciences*, vol. 17, no. 1, pp. 273–300, 2024.

[132] H. Zhou, L. Hu, J. Zhou, and K. Lange, "MM algorithms for variance components models," *Journal of Computational and Graphical Statistics*, vol. 28, no. 2, pp. 350–361, 2019.

[133] H. Zhou and K. Lange, "MM algorithms for some discrete multivariate distributions," *Journal of Computational and Graphical Statistics*, vol. 19, no. 3, pp. 645–665, 2010.