

Compact Formulations for Sparse Reconstruction in Fully and Partly Calibrated Sensor Arrays

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Dipl.-Ing. Christian Steffens
geboren am 24. Mai 1981 in Norden, Deutschland

Referent:	Prof. Dr.-Ing. Marius Pesavento
Korreferent:	Prof. Dr. Marc Pfetsch
Tag der Einreichung:	20. Juni 2017
Tag der mündlichen Prüfung:	25. September 2017

Darmstadt 2018
D 17

Acknowledgments

First and foremost, I wish to express my gratitude to Prof. Dr.-Ing. Marius Pesavento, for his guidance and friendship during the recent years. He has given me the chance to pursue my doctoral degree, and the time and freedom to follow my research interests. His enthusiasm, persistence and support have played an important role in my professional and personal development. Similarly, I would like to thank Prof. Dr. Marc Pfetsch for the careful revision of this thesis and his kind and encouraging comments. I would also like to extend my thanks to my doctoral examiners, Prof. Dr. Florian Steinke, Prof. Dr.-Ing. Rolf Jakoby and Prof. Dr.-Ing. Jürgen Adamy.

During my doctoral studies I have been in the fortunate position to work with many great colleagues. I would like to thank all my former colleagues and friends from the Communication Systems Group and Darmstadt University of Technology for the active, helpful and open-minded work atmosphere, and for the lots of good and fun times during work and outside of work. I also very much enjoyed working with my colleagues from other institutes and different fields of research in the Cocoon and EXPRESS projects. Besides my former colleagues, I would like to thank the many excellent and inspiring students that I was able to work with. A great thanks also goes to Marlis Gorecki for her support in any organizational and administrative matters. For their careful review of this thesis I am grateful to Dima Taleb, Ganapati Hegde, Alexander Sorg and Minh Trinh Hoang. Furthermore, I want to give special thanks to Dr. Pouyan Parvazi and Jens Steinwandt for the fruitful cooperation and the many inspiring conversations on research, writing and the less profound things in life.

I would like to acknowledge the financial support of the Cocoon project, a LOEWE Research Priority Program funded by the state of Hesse, and the EXPRESS project, funded by the DFG priority program (DFG-SPP 1798) on “Compressed Sensing in Information Processing” (CoSIP).

Finally, I want to express my appreciation for the support and encouragement from my friends and family. Above all, I am thankful to my parents Hermann and Dörte, and my brother Hilko, for their constant love and support. And last but not least, I am grateful to have you in my life, Sabiha.

Kurzfassung

Die Sensorgruppensignalverarbeitung ist ein klassisches Feld der digitalen Signalverarbeitung und bietet sowohl vielfältige Anwendungen in der Praxis, wie etwa die Richtungsschätzung oder die Rekonstruktion überlagerter Signale, als auch umfangreiche theoretische Grundlagen mit einer Vielzahl von Schätzverfahren und statistischen Grenzen der erreichbaren Schätzgüte. Ein vergleichsweise neues Feld der Signalverarbeitung ist die Rekonstruktion dünnbesetzter Signale (RDS), welches sich seit einiger Zeit großer Aufmerksamkeit in der Forschungsgemeinde erfreut und zahlreiche attraktive Anwendungsfelder in der Signalverarbeitung bietet. In der vorliegenden Dissertation wird die Anwendung der RDS in vollständig kalibrierten Sensorgruppen sowie in teilweise kalibrierten Sensorgruppen untersucht. Die Hauptbeiträge dieser Arbeit bestehen in einem neuen RDS-Verfahren für die Anwendung in teilweise kalibrierten Sensorgruppen sowie in kompakten Formulierungen für das RDS-Problem, wobei spezielles Augenmerk auf die Ausnutzung von spezifischer Struktur in den Signalen und Sensorgruppen gelegt wird. Die Ausnutzung von Struktur wurde besonders im Rahmen des Projekts “EXploiting structure in comPREssed Sensing using Side constraints” (EXPRESS) untersucht, welches innerhalb des Schwerpunktprojekts “Compressed Sensing in Information Processing” (SSP CoSIP) von der “Deutschen Forschungsgemeinschaft” (DFG) gefördert wurde. Der Aspekt der kooperativen Signalverarbeitung in verteilten, teilweise kalibrierten Sensorgruppen wurde im Rahmen des Projekts “Cooperative Sensor Communication” (Cocoon) erforscht, einem LOEWE-Forschungsschwerpunkt des Landes Hessen.

Eine zentrale Gruppe von Verfahren der Sensorgruppensignalverarbeitung basiert auf der Signalunterraumschätzung. Solche Verfahren erfordern vergleichsweise geringen Rechenaufwand und erzielen eine optimale Schätzgüte bei einer hohen Anzahl an zeitlichen Messungen oder hohem Signal-zu-Rausch-Verhältnis. Im Falle einer geringen Anzahl zeitlicher Messungen oder bei hoher Korrelation der auf die Sensorgruppe einfallenden Signale kann bei diesen Verfahren jedoch eine verminderte Schätzgüte auftreten. RDS-Verfahren sind verhältnismäßig robust gegenüber solchen Einschränkungen, was eine Anwendung in der Sensorgruppensignalverarbeitung motiviert.

Wie bereits erwähnt, ist die Richtungsschätzung eine klassische Anwendung in der Sensorgruppensignalverarbeitung. Unter der Annahme statischer Einfallrichtungen zeigt sich dabei eine Gruppenstruktur in der zeitlichen und räumlichen Darstellung der einfallenden Signale. Basierend auf dieser Gruppenstruktur wird in der vorliegenden Arbeit eine kompakte Formulierung für die RDS in vollständig kalibrierten Sensorgruppen hergeleitet, welche den Rechenaufwand gegenüber bestehenden Verfahren deutlich verringert und gleichzeitig neue mathematische Beziehungen zwischen bestehenden Verfahren aufzeigt. Wird weiterhin eine spezielle Struktur in der Sensorgruppe angenommen, wie etwa eine gleichförmige oder translationsinvariante Struktur, so bringt dies weitere Vorteile für die effiziente Implementierung der RDS auf Basis der vorgestellten kompakten Formulierung.

Bei der Sensorgruppensignalverarbeitung ist es gemeinhin von großem Interesse eine hohe Auflösung in der Richtungsschätzung zu erzielen sowie eine hohe Anzahl von Signalen zu identifizieren. Solche Eigenschaften können durch Sensorgruppen mit großer Apertur und einer hohen Anzahl an Sensoren erreicht werden. Diese Art von Sensorgruppen benötigt jedoch eine aufwendige Kalibrierung, so dass in vergangenen Jahren vermehrt die Anwendung von teilweise kalibrierten Sensorgruppen untersucht wurde, bei welchen die gesamte Sensorgruppe in einzelne Untergruppen aufgeteilt wird. Die Untergruppen erlauben dabei eine einfache Kalibrierung, während eine Kalibrierung der gesamten Sensorgruppe vermieden wird. In der vorliegenden Arbeit wird ein neues Verfahren für RDS in dieser Art von teilweise kalibrierten Sensorgruppen vorgestellt. Während aktuelle RDS-Verfahren für teilweise kalibrierte Sensorgruppen auf inkohärenter Verarbeitung der Untergruppensignale basieren, beruht das in dieser Dissertation vorgestellte RDS-Verfahren auf der kohärenten Verarbeitung der Untergruppensignale. Im Vergleich zur inkohärenten Verarbeitung führt das vorgestellte RDS-Verfahren zu einer deutlichen Verbesserung der Schätzgüte, wie anhand numerischer Experimente demonstriert wird. Für die effiziente Implementierung des neuen RDS-Verfahrens wird weiterhin eine kompakte Formulierung hergeleitet, welche für teilweise kalibrierte Sensorgruppen von beliebiger Struktur anwendbar ist. Der spezielle Fall von Sensorgruppen mit gleichförmigen oder translationsinvarianten Untergruppen ermöglicht zudem eine Implementierung der kompakten Formulierung mit zusätzlich verringertem Rechenaufwand.

Abstract

Sensor array processing is a classical field of signal processing which offers various applications in practice, such as direction of arrival estimation or signal reconstruction, as well as a rich theory, including numerous estimation methods and statistical bounds on the achievable estimation performance. A comparably new field in signal processing is given by sparse signal reconstruction (SSR), which has attracted remarkable interest in the research community during the last years and similarly offers plentiful fields of application. This thesis considers the application of SSR in fully calibrated sensor arrays as well as in partly calibrated sensor arrays. The main contributions are a novel SSR method for application in partly calibrated arrays as well as compact formulations for the SSR problem, where special emphasis is given on exploiting specific structure in the signals as well as in the array topologies. The aspect of specific structure has been investigated in the context of the project “EXploiting structure in comPREssed Sensing using Side constraints” (EXPRESS), which was funded within the priority program on “Compressed Sensing in Information Processing” (SSP CoSIP) by the German Research Foundation (“Deutsche Forschungsgemeinschaft”, DFG). The part of cooperative signal processing in partly calibrated sensor arrays has been investigated in the project on “Cooperative Sensor Communication” (Cocoon), a LOEWE Research Priority Program funded by the state of Hesse.

A central class of methods in sensor array processing is based on signal-subspace estimation. Such subspace-based methods have comparably low computational cost and have been shown to achieve optimal estimation performance for a large number of signal snapshots or high signal-to-noise ratio. In the case of low number of signal snapshots or in the case of correlated source signals, however, these methods might suffer from degraded estimation performance. In contrast to that, SSR methods are known to be robust to this kind of conditions, motivating the application in sensor array processing.

As mentioned above, direction of arrival estimation is a classical application in sensor array processing. Assuming static source directions, the source signals impinging on a sensor array admit a joint sparse representation in the spatio-temporal domain. In this thesis, the joint sparse signal structure is exploited to derive a compact formulation for joint sparse reconstruction from multiple signal snapshots in fully calibrated arrays. In comparison with existing works, the compact formulation has significantly reduced computational cost and reveals interesting links between different methods for SSR. Besides the structure in the source signal representation, additional structure might be contained in the array topology, such as a uniform linear or a shift-invariant structure, which brings additional benefits in the efficient implementation of the proposed compact formulation.

In sensor array processing it is commonly desired to achieve high angular resolution and to identify a large number of signals. Such characteristics can be achieved by sensor arrays with a large aperture and a large number of sensors. However, this type

of sensor arrays requires complex calibration schemes for practical application, such that partly calibrated arrays have been investigated as an alternative model in recent years. In partly calibrated arrays, the overall array is partitioned into smaller subarrays, which are themselves easy to calibrate, while calibration of the entire array is avoided. This thesis provides a novel approach for joint sparse reconstruction in partly calibrated arrays. While existing methods for SSR in partly calibrated arrays rely on incoherent processing of the subarray signals, the proposed method is based on coherent processing of the subarray signals. As shown by numerical experiments, the proposed method shows significant improvement in estimation performance as compared to state of the art methods for SSR under incoherent processing. Similar to the results for fully calibrated arrays, a compact formulation for the proposed SSR method in partly calibrated arrays is derived. While the presented compact formulation is applicable to partly calibrated arrays of arbitrary topologies and relies on spectrum search, the case of subarrays with uniform linear or shift-invariant structure admits a search-free implementation of the compact formulation, which admits additional reduction in computational cost.

Contents

1	Introduction	1
2	Signal Model	7
2.1	Fully Calibrated Arrays	7
2.1.1	Uniform Linear Arrays	9
2.1.2	Thinned Linear Arrays	10
2.2	Partly Calibrated Arrays	10
2.2.1	Coherent Processing	11
2.2.2	Incoherent Processing	13
2.2.3	Identical Uniform Linear Subarrays	14
2.2.4	Thinned Linear Subarrays	14
2.2.5	Shift-Invariant Arrays	15
2.2.6	Thinned Shift-Invariant Arrays	17
3	State of the Art	19
3.1	Subspace Estimation	19
3.2	MUSIC Algorithms	21
3.3	RARE Algorithms	23
3.4	ESPRIT Algorithms	26
4	Sparse Reconstruction for Fully Calibrated Arrays	29
4.1	Sparse Reconstruction from Single Snapshots	30
4.1.1	ℓ_1 Norm Minimization	32
4.1.2	Atomic Norm Minimization	35
4.2	Joint Sparse Reconstruction from Multiple Snapshots	37
4.2.1	$\ell_{2,1}$ Mixed-Norm Minimization	39
4.2.2	Atomic Norm Minimization	41
4.3	A Compact Formulation for $\ell_{2,1}$ Minimization	43
4.3.1	Gridless SPARROW	45
4.3.2	Off-Grid SPARROW	50
4.3.3	Related Work: SPICE	52
4.4	Numerical Experiments	57
4.4.1	Regularization Parameter Selection	57
4.4.2	Resolution Performance and Estimation Bias	58
4.4.3	Varying Number of Snapshots	61
4.4.4	Correlated Signals	63
4.4.5	Off-Grid SPARROW	63
4.4.6	Computation Time of SDP Formulations	66
4.5	Chapter Summary	68
5	Sparse Reconstruction for Partly Calibrated Arrays	71
5.1	Incoherent Sparse Reconstruction	72
5.1.1	$\ell_{F,1}$ Mixed-Norm Minimization	74
5.2	Coherent Sparse Reconstruction	75

5.2.1	$\ell_{*,1}$ Mixed-Norm Minimization	77
5.3	A Compact Formulation for $\ell_{*,1}$ Minimization	79
5.3.1	Gridless COBRAS	81
5.4	Coherent Sparse Reconstruction for Shift-Invariant Arrays	87
5.5	Numerical Results	90
5.5.1	Regularization Parameter Selection	91
5.5.2	Resolution Performance and Estimation Bias	92
5.5.3	Arbitrary Array Topologies and Grid-Based Estimation	95
5.5.4	Correlated Signals	97
5.5.5	Array Calibration Performance	97
5.5.6	Computation Time of SDP Formulations	99
5.6	Chapter Summary	101
6	Low-Complexity Algorithms for Sparse Reconstruction	103
6.1	Algorithms for Sparse Reconstruction in FCAs	104
6.1.1	Coordinate Descent Method for $\ell_{2,1}$ Minimization	104
6.1.2	STELA Method for $\ell_{2,1}$ Minimization	106
6.1.3	Coordinate Descent Method for SPARROW	109
6.2	Algorithms for Sparse Reconstruction in PCAs	112
6.2.1	Coordinate Descent Method for $\ell_{*,1}$ Minimization	112
6.2.2	STELA Method for $\ell_{*,1}$ Minimization	113
6.2.3	Coordinate Descent Method for COBRAS	115
6.3	Numerical Experiments	117
6.3.1	Sparse Reconstruction in FCAs	117
6.3.2	Sparse Reconstruction in PCAs	121
6.4	Chapter Summary	121
7	Conclusion and Outlook	123
	Appendix	127
A	Regularization Parameter for $\ell_{2,1}$ Minimization	127
B	Equivalence of SPARROW and $\ell_{2,1}$ Minimization	128
C	Convexity of the SPARROW Problem	130
D	Equivalence of SPARROW and AST	131
E	Dual Problem of the SPARROW Formulation	132
F	Regularization Parameter for $\ell_{*,1}$ Minimization	134
G	Equivalence of COBRAS and $\ell_{*,1}$ Minimization	137
H	SDP Form of the Matrix Polynomial Constraint	139
I	Dual Problem of the COBRAS Formulation	140
	List of Acronyms	143
	List of Symbols	145
	Bibliography	149
	Lebenslauf	163

Chapter 1

Introduction

Sensor array signal processing is a classical, yet active, field of signal processing with a rich theory on estimation methods and estimation bounds, and various fields of application, such as wireless communications, radar and astronomy, to name a few [KV96, vT02, TF09]. Prominent applications in array processing include beamforming and direction of arrival (DOA) estimation, where beamforming considers the problem of signal reconstruction in the presence of noise and interference, while DOA estimation falls within the more general concept of signal parameter estimation.

Compared to array processing, sparse signal reconstruction (SSR) constitutes a relatively new field in signal processing, and is sometimes also referred to as sparse recovery, compressed sensing or compressive sensing. Since SSR provides many fields of application, such as parameter estimation, spectral analysis, image processing, or machine learning, it has obtained substantial research interest in recent years [Tib96, CDS98, DE03, CT05, Don06, CRT06a, CRT06b, CR06]. The classical SSR problem considers the reconstruction of a high-dimensional sparse signal vector from a low-dimensional measurement vector, which is characterized by an underdetermined system of linear equations. It has been shown that exploiting prior knowledge on the sparse structure of the signal admits a unique solution to the underdetermined system [DE03, CT05, Don06, CRT06a, CRT06b, CR06]. In the signal processing context, this implies that far fewer samples than postulated by the Shannon-Nyquist sampling theorem for bandlimited signals are required for perfect signal reconstruction [TLD⁺10], whereas, in the parameter estimation context, this indicates that SSR methods exhibit the superresolution property [Don92].

Ideally, recovery of the sparsest signal vector would be performed by solving an ℓ_0 minimization problem, where the ℓ_0 quasi-norm of a vector is the number of its non-zero elements. However, ℓ_0 minimization requires combinatorial search and becomes intractable for large problem dimensions, such that a plethora of methods has been proposed to approximately solve the SSR problem with reduced computational cost [TW10, FR13]. Some of the most prominent of these methods are based on convex relaxation in terms of ℓ_1 norm minimization [Tib96, CDS98], which makes the SSR problem computationally tractable while providing good reconstruction performance, or greedy methods [MZ93], which have low computational cost but provide reduced reconstruction performance.

Regarding signal reconstruction and parameter estimation, the objectives of array processing and SSR are quite similar. In fact, there are many links between the two fields such that recent results in SSR have sparked new research in sensor array processing and vice versa. As stated above, a standard problem in array processing is estimation of the DOAs of multiple sources signals impinging on a sensor array. Some of the

most prominent estimation methods for this application are based on subspace estimation [GRP10], e.g., the MUSIC method [Bar83,Sch86] which has been shown to perform asymptotically optimal and offers the super-resolution property at tractable computational cost [SA89]. However, these subspace-based methods often suffer from problems in practical scenarios. First, correlated source signals, e.g., in multipath environments, can significantly reduce the estimation performance. Second, subspace-based methods yield poor performance in the case of a low number of signal snapshots, e.g., in fast changing environments. SSR techniques provide an attractive alternative for these scenarios. From an SSR perspective, the DOA estimation problem can be modeled by an overcomplete representation of the source signals in the DOA domain. Given multiple signal snapshots and static DOAs, such an overcomplete representation exhibits a joint sparse structure of the source signals vectors. Approximate methods for the joint sparse reconstruction from multiple signal snapshots include convex relaxation by means of $\ell_{2,1}$ mixed-norm minimization [MÇW05, YL06]. Generally, the computational cost of these methods grows with the number of snapshots and the required DOA resolution, such that alternative low-complexity approaches are of great interest. In Chapter 4 of this thesis, a novel compact formulation for joint sparse reconstruction from multiple signal snapshots is introduced, which is equivalent to the prominent $\ell_{2,1}$ mixed-norm minimization approach and referred to as SPARse ROW-norm reconstruction (SPARROW). As compared to $\ell_{2,1}$ mixed-norm minimization, the compact SPARROW formulation has significantly reduced number of optimization parameters and employs the sample covariance matrix, such that its computational cost is independent of the number of signal snapshots. Besides the joint sparse structure in the source signals, additional structure in the sensor array might be available [HPREK14]. As will be shown for the case of uniform linear arrays, this additional structure in the array topology can be exploited for efficient implementation of the proposed compact reformulation. The results in Chapter 4 are based on the following publications:

- [1] C. Steffens, M. Pesavento, and M. E. Pfetsch, “A Compact Formulation for the $\ell_{2,1}$ Mixed-Norm Minimization Problem,” *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1483-1497, Mar. 2018.
- [2] W. Suleiman, C. Steffens, A. Sorg, and M. Pesavento, “Gridless Compressed Sensing for Fully Augmentable Arrays,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, pp. 1986-1990, Kos Island, Greece, Aug. 2017.
- [3] C. Steffens, M. Pesavento, and M. E. Pfetsch, “A Compact Formulation for the $\ell_{2,1}$ Mixed-Norm Minimization Problem,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4730 - 4734, New Orleans, USA, Mar. 2017.
- [4] A. Colonna Walewski, C. Steffens, and M. Pesavento, “Off-Grid Parameter Estimation Based on Joint Sparse Regularization,” in *Proceedings of the International ITG Conference on Systems, Communications and Coding (SCC)*, pp. 1-6, Hamburg, Germany, Feb. 2017.

In many applications in sensor array processing it is desired to resolve a large number of source signals with a high angular resolution. These characteristics can be achieved by sensor arrays with a large number of sensors and a large aperture [KV96, vT02]. However, larger aperture size makes it more difficult to maintain the array calibration. Therefore, research on partly calibrated arrays (PCAs) has gained increasing interest in recent years. In PCAs, the overall sensor array is partitioned into smaller subarrays which are assumed to be perfectly calibrated while the calibration among different subarrays is not available. The measurements obtained at the different subarrays can be processed either in a coherent or incoherent fashion, depending on the amount of synchronization available in the PCA. In coherent processing, the measurements of all the subarrays are jointly processed on a signal sample basis to obtain a global estimate, which requires precise synchronization in time and frequency among the different subarrays [SORK92, SSJ01, PGW02, SG04, PP11, SPZ13, SPPZ14]. Contrary, in incoherent processing, only the measurements within each subarray are coherently processed while the measurements of different subarrays are processed in an incoherent fashion, e.g., by first computing local estimates of the DOAs or the covariance matrices which are further combined in the entire PCA to obtain an improved global estimate [WK85, SNS95, SP14]. Hence, incoherent processing has significantly relaxed requirements in synchronization among the different subarrays, while coherent processing usually achieves better estimation performance since more information among the different subarrays can be exploited.

The most prominent DOA estimation techniques for PCAs in the literature are based on coherent processing and subspace separation. Similar to the case of a single, fully calibrated array (FCA), as discussed above, subspace-based methods for PCAs are shown to perform asymptotically optimal at low computational cost, but may suffer from reduced estimation performance in the case of correlated signals or low number of snapshots, such that SSR methods provide an attractive alternative in these difficult scenarios. So far, state of the art techniques for SSR in PCAs have been considered only under incoherent processing. In Chapter 5 of this thesis, a novel convex optimization problem formulation for sparse reconstruction in PCAs with coherent processing is devised. The proposed formulation is based on minimization of a mixed nuclear and ℓ_1 norm to fulfill a low-rank criterion, resulting from the specific structure of the subarray signal model, and to impose block-sparsity on the signal matrix, respectively. Similar to the case of $\ell_{2,1}$ mixed-norm minimization for FCAs discussed in Chapter 4, the SSR method presented in Chapter 5 is applicable to arbitrary PCA topologies and admits a compact and equivalent reformulation, termed as COmpact Block- and RAnk-Sparse reconstruction (COBRAS). Efficient gridless implementations based on the compact COBRAS formulation are presented for the special case of PCAs composed of subarrays with special structure, such as subarrays with a common baseline and subarrays with shift-invariant topologies. The following papers have been published in this context:

- [5] C. Steffens and M. Pesavento, “Block- and Rank-Sparse Recovery for Direction Finding in Partly Calibrated Arrays,” *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 384–399, Jan. 2018.

- [6] C. Steffens, W. Suleiman, A. Sorg, and M. Pesavento, “Gridless Compressed Sensing Under Shift-Invariant Sampling,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4735 - 4739, New Orleans, USA, Mar. 2017.
- [7] C. Steffens, P. Parvazi, and M. Pesavento, “Direction Finding and Array Calibration Based on Sparse Reconstruction in Partly Calibrated Arrays,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 21-24, A Coruna, Spain, Jun. 2014.

The approach of mixed nuclear and ℓ_1 norm minimization exploits specific structure in the PCA signal model, which might similarly be available in other applications besides PCAs. Such applications include, for example, DOA estimation for non-circular signals or multidimensional channel parameter estimation, and mixed nuclear and ℓ_1 norm minimization can likewise be employed in these applications as shown in the following publications:

- [8] J. Steinwandt, F. Römer, C. Steffens, M. Haardt, and M. Pesavento, “Gridless Super-Resolution Direction Finding for Strictly Non-Circular Sources Based on Atomic Norm Minimization,” in *Proceedings of the Annual Asilomar Conference on Signals, Systems and Computers*, pp. 1518 - 1522, Pacific Grove, California, USA, Nov. 2016.
- [9] C. Steffens, Y. Yang, and M. Pesavento, “Multidimensional Sparse Recovery for MIMO Channel Parameter Estimation,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, pp. 66-70, Budapest, Hungary, Sep. 2016.
- [10] J. Steinwandt, C. Steffens, M. Pesavento, and M. Haardt, “Sparsity-Aware Direction Finding for Strictly Non-Circular Sources Based on Rank Minimization,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1-5, Rio de Janeiro, Brasil, Jul. 2016.

Note that the compact formulation derived in Chapter 5 can as well be employed in the applications given in publications [8]-[10], admitting gridless and computationally efficient implementations of the corresponding SSR problems. For sake of brevity, these applications will, however, not be further discussed in this thesis.

The organization of this thesis is as follows: Chapter 2 provides the signal model for fully and partly calibrated arrays, including the characterization of specific array structures. A short review on subspace-based state of the art methods is provided in Chapter 3. Joint sparse reconstruction from multiple signal snapshots in fully calibrated sensor arrays is presented in Chapter 4, where the novel compact SPARROW formulation is presented in Section 4.3 as one of the major contributions of this thesis. The application of SSR in partly calibrated arrays is investigated in Chapter 5 of this thesis.

The major contributions of Chapter 5 are the mixed nuclear and ℓ_1 norm minimization problem formulation for coherent SSR in PCAs, given in Section 5.2, and the equivalent COBRAS formulation of the SSR problem, provided in Section 5.3. For efficient implementation of the presented SSR approaches, low-complexity algorithms are provided in Chapter 6. The major contributions in Chapter 6 are a low-complexity implementation of the SPARROW method in Section 6.1.3, and novel algorithms for mixed nuclear and ℓ_1 norm minimization as well as the COBRAS formulation in Sections 6.2.1-6.2.3. Concluding remarks and an outlook on future work are provided in Chapter 7.

Chapter 2

Signal Model

This chapter provides the general signal model for the direction of arrival (DOA) estimation problem in fully calibrated arrays (FCAs) and partly calibrated arrays (PCAs). The first part of this chapter considers the FCA case with two special topologies in the form of uniform linear arrays and thinned linear arrays. The second part of this chapter discusses PCAs, which can be employed with coherent and incoherent processing. Although the focus of this thesis is on coherent processing in PCAs, the model for incoherent processing is presented briefly to highlight the main differences between the coherent and incoherent signal model and to put the results into the general context of DOA estimation in PCAs. Furthermore, two special subarray topologies for PCAs are introduced, namely PCAs composed of uniform linear subarrays with a common baseline as well as thinned linear arrays with a common baseline. Another special PCA topology is the class of shift-invariant arrays. It should be noted that shift-invariant arrays can similarly be treated in the context of FCAs, e.g., a fully calibrated uniform linear array also exhibits shift-invariances [ROSK88, RK89]. However, the benefits of shift-invariant arrays become most apparent in the context of PCAs [SPZ16].

For ease of presentation this thesis only considers the case of linear array topologies. Assuming that the sensors and signal sources are located in the same plane, the concepts in this thesis can easily be extended to planar arrays, as discussed in [SPP14]. In the case when sensors and sources are not located in a common plane, e.g., for some planar or volume arrays, two-dimensional DOA estimation is required, i.e., estimation of the azimuth and elevation angle of arrival. Although the methods discussed in this thesis can be extended to two-dimensional estimation, analogous to [Pes05, SYP16], for sake of clarity the discussion is limited to linear arrays which only require one-dimensional estimation.

2.1 Fully Calibrated Arrays

Consider a set of L source signals impinging on a linear array of M omnidirectional and identical sensors with positions $r_1, \dots, r_M \in \mathbb{R}$, expressed in half signal wavelength and relative to sensor 1, i.e., $r_1 = 0$. The L sources are located in angular directions $\theta_1, \dots, \theta_L$, summarized as the vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_L]^\top$, as depicted in Figure 2.1. Define the spatial frequency μ_l corresponding to the angular direction θ_l as

$$\mu_l = \cos \theta_l \in [-1, 1), \quad (2.1)$$

for $l = 1, \dots, L$, comprising the vector $\boldsymbol{\mu} = [\mu_1, \dots, \mu_L]^\top$. All sources are transmitting electromagnetic signals in the same frequency band and the following assumptions are expected to apply [KV96, vT02]:

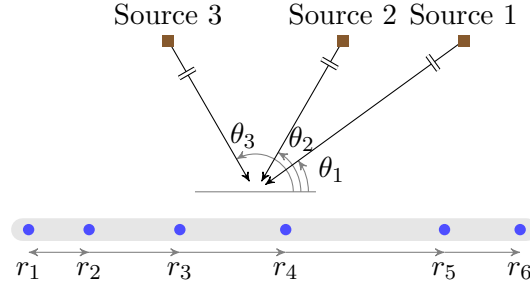


Figure 2.1: Sensor array of $M = 6$ sensors and $L = 3$ impinging source signals

- A1)** Farfield sources, i.e., the sources are located in the farfield region of the sensor array such that the electromagnetic waves arriving at the array approximately have a planar wavefront.
- A2)** Point sources assumption, i.e., the physical dimensions of the emitting sources are negligibly small such that each source is associated with a distinct angular direction.
- A3)** Narrowband signals, i.e., the radio frequency bandwidth of the source signals is negligible, such that the signal wavefront does not decorrelate significantly while traveling across the array aperture.
- A4)** The source signals have identical and known radio center frequency, such that sensor measurements can be properly converted to the baseband.
- A5)** The source and sensor positions are stationary during the time of observation, such that the DOAs remain constant.

Under assumptions A1-A5, the vector $\mathbf{y}(t) \in \mathbb{C}^M$ containing the sensor measurements in the baseband at time instant t is modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\mu}) \boldsymbol{\psi}(t) + \mathbf{n}(t), \quad (2.2)$$

with $\boldsymbol{\psi}(t) \in \mathbb{C}^L$ denoting the baseband source signal vector and $\mathbf{n}(t) \in \mathbb{C}^M$ representing additive circular and spatio-temporal white Gaussian sensor noise with covariance matrix $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma_N^2 \mathbf{I}_M$, where \mathbf{I}_M and σ_N^2 denote the $M \times M$ identity matrix and the noise power, respectively. For further considerations it is assumed that

- A6)** the source signals in $\boldsymbol{\psi}(t)$ and the sensor noise in $\mathbf{n}(t)$ are uncorrelated, i.e., $\mathbb{E}\{\boldsymbol{\psi}(t)\mathbf{n}^H(t)\} = \mathbf{0}$.

The array steering matrix $\mathbf{A}(\boldsymbol{\mu}) \in \mathbb{C}^{M \times L}$ in (2.2) is given by

$$\mathbf{A}(\boldsymbol{\mu}) = [\mathbf{a}(\mu_1), \dots, \mathbf{a}(\mu_L)], \quad (2.3)$$

where $\mathbf{a}(\mu_l)$ denotes the perfectly known array steering vector for spatial frequency μ_l . Under the given assumptions, the array response vector is modeled as

$$\mathbf{a}(\mu) = [1, e^{-j\pi\mu r_2}, \dots, e^{-j\pi\mu r_M}]^\top. \quad (2.4)$$

Throughout this thesis it is assumed that $L \leq M$ and that

A7) the steering matrix $\mathbf{A}(\boldsymbol{\mu}) \in \mathbb{C}^{M \times L}$ has full column rank.

Neglecting the noise, i.e., $\mathbf{n}(t) = \mathbf{0}$, Assumption A7 ensures that the signals $\boldsymbol{\psi}(t)$ in the model (2.2) can be uniquely recovered from the measurements $\mathbf{y}(t)$.

Let $\mathbf{Y} = [\mathbf{y}(t_1), \dots, \mathbf{y}(t_N)] \in \mathbb{C}^{M \times N}$ denote the matrix containing N snapshots of the sensor measurements in the baseband, where $[\mathbf{Y}]_{m,n}$ denotes the output at sensor m in time instant t_n , for $m = 1, \dots, M$ and $n = 1, \dots, N$. The multiple sensor measurement vectors are modeled in compact notation as

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\mu}) \boldsymbol{\Psi} + \mathbf{N}, \quad (2.5)$$

where $\boldsymbol{\Psi} = [\boldsymbol{\psi}(t_1), \dots, \boldsymbol{\psi}(t_N)] \in \mathbb{C}^{L \times N}$ is the baseband source signal matrix, with $[\boldsymbol{\Psi}]_{l,n}$ denoting the signal transmitted by source l in time instant t_n , and $\mathbf{N} = [\mathbf{n}(t_1), \dots, \mathbf{n}(t_N)] \in \mathbb{C}^{M \times N}$ is the sensor noise matrix. Equation (2.5) is referred to as the multi snapshot signal model [KV96, vT02]. Given full rank signal and noise matrices $\boldsymbol{\Psi}$ and \mathbf{N} , and under Assumptions A6 and A7, it can be concluded that the sensor measurement matrix \mathbf{Y} in (2.5) has full column rank with probability 1.

2.1.1 Uniform Linear Arrays

The case of uniform linear arrays (ULAs), as illustrated in Figure 2.2 a), is a special type of fully calibrated arrays which exhibits attractive structure that can be exploited in the DOA estimation process. In ULAs, the sensors are positioned equidistantly on a line, according to $r_m = (m - 1) \Delta \in \mathbb{R}$, for $m = 1, \dots, M$, with Δ denoting the array baseline. The array response vector for the ULA is expressed as

$$\mathbf{a}(z) = [1, z, \dots, z^{M-1}]^\top, \quad (2.6)$$

where $z = e^{-j\pi\mu\Delta}$, such that the steering matrix $\mathbf{A}(\mathbf{z}) = [\mathbf{a}(z_1), \dots, \mathbf{a}(z_L)]$ has a Vandermonde structure, where $\mathbf{z} = [z_1, \dots, z_L]^\top$, with $z_l = e^{-j\pi\mu_l\Delta}$ for $l = 1, \dots, L$.



Figure 2.2: a) Uniform linear array of $M = 7$ sensors and b) thinned linear array of $M = 4$ sensors, obtained by removing sensors from a virtual uniform linear array of $M_0 = 7$ sensors

2.1.2 Thinned Linear Arrays

Large array apertures are attractive since these admit high resolution in the spatial frequency. However, a straightforward realization of a large aperture by a uniform linear array requires a large number of sensors, which results in high hardware costs. Thinned linear arrays (TLAs) form a compromise between large array aperture and the number of sensors and are derived from a virtual ULA by removing sensors, as depicted in Figure 2.2 b). In this case the positions of the M sensors are given as $r_m = (\ddot{m}_m - 1) \Delta$, for $\{\ddot{m}_1, \dots, \ddot{m}_M\} \subseteq \{1, \dots, M_0\}$, where $M_0 \Delta$ represents the largest sensor lag in the array. Let \mathbf{J} denote a selection matrix of size $M_0 \times M$, selecting the corresponding sensors from a virtual ULA, then the response vector of the TLA can be described by the response vector of the corresponding virtual ULA according to

$$\mathbf{a}(z) = [1, z^{\ddot{m}_2}, \dots, z^{\ddot{m}_{M-1}}]^\top = \mathbf{J}^\top [1, z, \dots, z^{M_0-1}]^\top. \quad (2.7)$$

Well known TLA types with a large degree of freedom include, e.g., minimum redundancy arrays [Mof68], nested arrays [PV10] and co-prime arrays [VP11].

2.2 Partly Calibrated Arrays

Consider a linear array of arbitrary topology, composed of M omnidirectional sensors and partitioned into P subarrays with M_p sensors in subarray p , for $p = 1, \dots, P$, such that $M = \sum_{p=1}^P M_p$, as depicted in Figure 2.3. The vector $\boldsymbol{\eta} = [\eta^{(2)}, \dots, \eta^{(P)}]^\top$ contains the $P-1$ inter-subarray displacements $\eta^{(2)}, \dots, \eta^{(K)}$, expressed in half signal wavelength and relative to the first subarray, i.e., $\eta^{(1)} = 0$. Furthermore, let $\rho_m^{(p)}$ denote the intra-subarray position of the m th sensor of subarray p , for $m = 1, \dots, M_p$ and $p = 1, \dots, P$, expressed in half signal wavelength and relative to the first sensor in the respective subarray, hence $\rho_1^{(p)} = 0$. Consequently, the position of sensor m in subarray p , relative to the first sensor in the first subarray, can be expressed as [PGW02, SG04, Par12]

$$r_m^{(p)} = \rho_m^{(p)} + \eta^{(p)}, \quad (2.8)$$

for $m = 1, \dots, M_p$ and $p = 1, \dots, P$. For ease of notation, the following ordering of sensors is assumed throughout this thesis:

A8) Let $r_m^{(p)}$ denote the position of sensor m in subarray p , for $m = 1, \dots, M_p$ and $p = 1, \dots, P$. Whenever special order of the sensors is required, i.e., for the sensor measurements or the array response, the corresponding vector/matrix elements are sorted in increasing order by subarray index p and sensor index m according to

$$r_1^{(1)}, r_2^{(1)}, \dots, r_{M_1-1}^{(1)}, r_{M_1}^{(1)}, r_1^{(2)}, r_2^{(2)}, \dots, r_{M_P-1}^{(P)}, r_{M_P}^{(P)}. \quad (2.9)$$

In the context of partly calibrated arrays it is commonly assumed that the intra-subarray positions $\rho_m^{(p)}$, for $m = 1, \dots, M_p$ and $p = 1, \dots, P$, are perfectly known while the inter-subarray displacements $\eta^{(2)}, \dots, \eta^{(K)}$ are unknown, hence only the subarray responses are available for the estimation process while the overall array response is not.

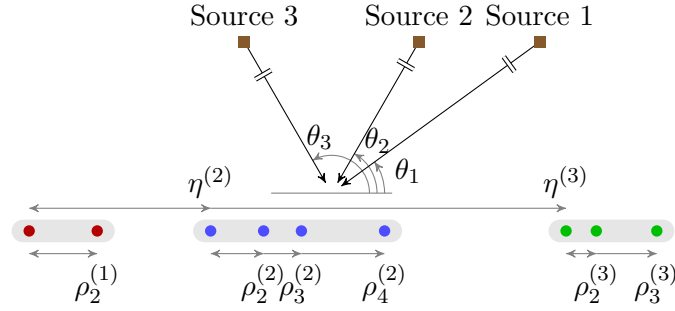


Figure 2.3: Partly calibrated array of $M = 9$ sensors partitioned in $P = 3$ subarrays, and $L = 3$ impinging source signals

2.2.1 Coherent Processing

Assume that all the subarrays are synchronized in time and frequency and that a total of N signal snapshots are obtained at the output of each subarray p , which are collected in the subarray measurement matrix $\mathbf{Y}^{(p)} = [\mathbf{y}^{(p)}(t_1), \dots, \mathbf{y}^{(p)}(t_N)] \in \mathbb{C}^{M_p \times N}$, for $p = 1, \dots, P$, where $[\mathbf{Y}^{(p)}]_{m,n}$ denotes the output of sensor m in subarray p and time instant t_n . The subarray measurement matrices are collected in the $M \times N$ array measurement matrix $\mathbf{Y} = [\mathbf{Y}^{(1)\top}, \dots, \mathbf{Y}^{(P)\top}]^\top \in \mathbb{C}^{M \times N}$, which is modeled as

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \boldsymbol{\Psi} + \mathbf{N}, \quad (2.10)$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{L \times N}$ is the source signal matrix and $\mathbf{N} \in \mathbb{C}^{M \times N}$ denotes the sensor noise matrix, defined in correspondence with the sensor measurements in \mathbf{Y} . The $M \times L$ array steering matrix $\mathbf{A}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ in (2.10) is given by

$$\mathbf{A}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = [\mathbf{a}(\mu_1, \boldsymbol{\alpha}, \boldsymbol{\eta}), \dots, \mathbf{a}(\mu_L, \boldsymbol{\alpha}, \boldsymbol{\eta})], \quad (2.11)$$

and represents the steering matrix of the entire array, where $\mathbf{a}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ denotes the array response vector for unknown subarray perturbations $\boldsymbol{\alpha} = [\alpha^{(2)}, \dots, \alpha^{(P)}]^\top \in \mathbb{C}^{P-1}$ and subarray displacements $\boldsymbol{\eta} = [\eta^{(2)}, \dots, \eta^{(P)}]^\top \in \mathbb{R}^{P-1}$ relative to subarray 1, respectively, i.e., $\alpha^{(1)} = 1$ and $\eta^{(1)} = 0$. Based on the sensor position notation in (2.8), the array response vector can be factorized according to

$$\mathbf{a}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \mathbf{B}(\boldsymbol{\mu})\boldsymbol{\varphi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \quad (2.12)$$

where the $M \times P$ block-diagonal matrix

$$\mathbf{B}(\boldsymbol{\mu}) = \text{blkdiag}(\mathbf{a}^{(1)}(\boldsymbol{\mu}), \dots, \mathbf{a}^{(P)}(\boldsymbol{\mu})) \quad (2.13)$$

contains the perfectly known subarray response vectors

$$\mathbf{a}^{(p)}(\boldsymbol{\mu}) = [1, e^{-j\pi\mu\rho_2^{(p)}}, \dots, e^{-j\pi\mu\rho_{M_p}^{(p)}}]^\top, \quad (2.14)$$

for $p = 1, \dots, P$, on its diagonal. The P element vector

$$\boldsymbol{\varphi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = [1, \alpha^{(2)}e^{-j\pi\mu\eta^{(2)}}, \dots, \alpha^{(P)}e^{-j\pi\mu\eta^{(P)}}]^\top \quad (2.15)$$

takes account of the subarray displacement shifts $e^{-j\pi\mu\eta^{(p)}}$, for $p = 2, \dots, P$, depending on the spatial frequencies in $\boldsymbol{\mu}$ and the subarray displacements in $\boldsymbol{\eta}$, and possibly additional unknown shifts $\boldsymbol{\alpha}$, e.g., gain/phase or timing offsets among the subarrays [PGW02,SG04,Par12]. In relation to (2.11), define the $M \times PL$ subarray steering block matrix

$$\mathbf{B}(\boldsymbol{\mu}) = [\mathbf{B}(\mu_1), \dots, \mathbf{B}(\mu_L)] \quad (2.16)$$

containing all the subarray response block matrices for the spatial frequencies in $\boldsymbol{\mu}$, and the $PL \times L$ block-diagonal matrix

$$\boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \text{blkdiag}(\boldsymbol{\varphi}(\mu_1, \boldsymbol{\alpha}, \boldsymbol{\eta}), \dots, \boldsymbol{\varphi}(\mu_L, \boldsymbol{\alpha}, \boldsymbol{\eta})), \quad (2.17)$$

composed of the subarray shift vectors in (2.15). Using (2.16) and (2.17), the overall array steering matrix (2.11) can be factorized as

$$\mathbf{A}(\boldsymbol{\mu}, \boldsymbol{\eta}) = \mathbf{B}(\boldsymbol{\mu}) \boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}), \quad (2.18)$$

such that the overall array measurement matrix in (2.10) is equivalently modeled as

$$\mathbf{Y} = \mathbf{B}(\boldsymbol{\mu}) \boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \boldsymbol{\Psi} + \mathbf{N}, \quad (2.19)$$

providing an extended factorization of the coherent signal model for PCAs [SPP14], which will be employed in Section 5.2.

2.2.2 Incoherent Processing

Proper synchronization in time and frequency among the subarrays might be difficult to achieve in practical applications, where different subarrays might sample the signals at different time instants. Another effect that might occur in PCAs is that the increased size of the array aperture might also lead to a violation of the narrowband assumption A3, leading to decorrelated signals at the different subarrays.

For further discussion consider the case of imperfect timing synchronization and let $t_1^{(p)}, \dots, t_N^{(p)}$ denote the N sampling instants of subarray p , for $p = 1, \dots, P$. The measurement matrix $\tilde{\mathbf{Y}}^{(p)} = [\mathbf{y}^{(p)}(t_1^{(p)}), \dots, \mathbf{y}^{(p)}(t_N^{(p)})] \in \mathbb{C}^{M_p \times N}$ obtained at the output of subarray p can be modeled as [WK85, DSB⁺05, LYJ⁺15]

$$\tilde{\mathbf{Y}}^{(p)} = \mathbf{A}^{(p)}(\boldsymbol{\mu}) \tilde{\boldsymbol{\Psi}}^{(p)} + \tilde{\mathbf{N}}^{(p)}, \quad (2.20)$$

where

$$\tilde{\boldsymbol{\Psi}}^{(p)} = \alpha^{(p)} \text{diag}\left(1, e^{-j\pi\mu_1\eta^{(p)}}, \dots, e^{-j\pi\mu_L\eta^{(p)}}\right) \left[\boldsymbol{\psi}(t_1^{(p)}), \dots, \boldsymbol{\psi}(t_N^{(p)})\right] \quad (2.21)$$

denotes the source signals as observed by sensor 1 in subarray p , for $p = 1, \dots, P$. The matrix $\tilde{\mathbf{N}}^{(p)}$ represents the sensor noise matrix under incoherent processing and is defined in correspondence with the subarray measurements in $\tilde{\mathbf{Y}}^{(p)}$. Furthermore,

$$\mathbf{A}^{(p)}(\boldsymbol{\mu}) = [\mathbf{a}^{(p)}(\mu_1), \dots, \mathbf{a}^{(p)}(\mu_L)] \quad (2.22)$$

is the steering matrix of subarray p , composed of the subarray steering vectors as defined in (2.14). The incoherently sampled subarray measurements and sensor noise of the overall array are summarized as $\tilde{\mathbf{Y}} = [\tilde{\mathbf{Y}}^{(1)\top}, \dots, \tilde{\mathbf{Y}}^{(P)\top}]^\top$ and $\tilde{\mathbf{N}} = [\tilde{\mathbf{N}}^{(1)\top}, \dots, \tilde{\mathbf{N}}^{(P)\top}]^\top$, respectively. Furthermore, let the $L \times N$ source signal matrix observed by subarray p be constructed as $\tilde{\boldsymbol{\Psi}}^{(p)} = [\tilde{\boldsymbol{\psi}}_1^{(p)}, \dots, \tilde{\boldsymbol{\psi}}_L^{(p)}]^\top$, for $p = 1, \dots, P$, and introduce the array source signal matrix under incoherent processing as

$$\tilde{\boldsymbol{\Psi}} = [\tilde{\boldsymbol{\psi}}_1^{(1)}, \dots, \tilde{\boldsymbol{\psi}}_1^{(P)}, \tilde{\boldsymbol{\psi}}_2^{(1)}, \dots, \tilde{\boldsymbol{\psi}}_L^{(P)}]^\top. \quad (2.23)$$

With above definitions, the PCA signal model under incoherent processing can be formulated as

$$\tilde{\mathbf{Y}} = \mathbf{B}(\boldsymbol{\mu}) \tilde{\boldsymbol{\Psi}} + \tilde{\mathbf{N}}, \quad (2.24)$$

where $\mathbf{B}(\boldsymbol{\mu})$ denotes the subarray steering block matrix as defined in (2.16) and used in the coherent signal model (2.19). The difference of the models in (2.19) and (2.24) lies in the signal representation, given by the product $\boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \boldsymbol{\Psi}$ and the matrix $\tilde{\boldsymbol{\Psi}}$, respectively. In contrast to the incoherent signal matrix $\tilde{\boldsymbol{\Psi}}$, the matrix product $\boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \boldsymbol{\Psi}$ exhibits an attractive low-rank block structure that can be exploited for signal reconstruction, as will be discussed in more detail in Section 5. The model in (2.24) will be further considered for incoherent signal reconstruction in Section 5.1

2.2.3 Identical Uniform Linear Subarrays

Similar to fully calibrated uniform linear arrays as introduced in Section 2.1.1, partly calibrated arrays composed of identical uniform linear subarrays of M_0 sensors per subarray provide attractive structure that can be exploited for efficient signal reconstruction. In this scenario, the M_0 sensors within each subarray are located at integer multiples of a baseline Δ , i.e., $\rho_m^{(p)} = (m-1)\Delta$ with $m = 1, \dots, M_0$ and $p = 1, \dots, P$. Upon defining $z = e^{-j\pi\mu\Delta}$, the subarray response block matrix in (2.13) takes the block-Vandermonde form

$$\mathbf{B}(z) = \mathbf{I}_P \otimes [1 \quad z \quad \dots \quad z^{M_0-1}]^\top \quad (2.25)$$

for a PCA of identical uniform linear subarrays.

2.2.4 Thinned Linear Subarrays

The concept of fully calibrated thinned linear arrays introduced in Section 2.1.2 can similarly be transferred to PCAs. Consider a PCA of $M = \sum_{p=1}^P M_p$ sensors partitioned into P subarrays of M_p sensors, for $p = 1, \dots, P$. It is assumed that the PCAs have a thinned linear topology with a common baseline Δ , i.e., the intra-subarray sensor positions are given by $\rho_m^{(p)} = (\ddot{m}_m^{(p)} - 1)\Delta$, for $m = 1, \dots, M_p$ and $p = 1, \dots, P$, and with $\{\ddot{m}_1^{(p)}, \dots, \ddot{m}_{M_p}^{(p)}\} \subseteq \{1, \dots, M_0\}$, where $M_0\Delta$ represents the largest sensor lag in all subarrays. From this definition, a PCA of thinned linear subarrays can similarly be formed by removing sensors from a virtual PCA composed of P identical uniform linear subarrays of M_0 sensors, as illustrated in Figure 2.4. Thus, the subarray response block matrix according to (2.14) can be derived from a block-Vandermonde matrix according to

$$\mathbf{B}(z) = \mathbf{J}^\top (\mathbf{I}_P \otimes [1 \quad z \quad \dots \quad z^{M_0-1}]^\top), \quad (2.26)$$

where \mathbf{J} is a proper selection matrix of size $PM_0 \times M$, selecting the corresponding sensors from the underlying virtual PCA of identical uniform linear subarrays, and \otimes denotes the Kronecker product.

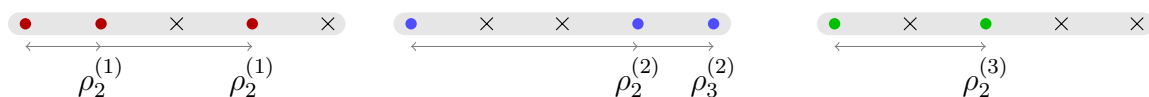


Figure 2.4: PCA composed of thinned linear subarrays, obtained by removing sensors from virtual PCA of $P = 3$ identical uniform linear subarrays of $M_0 = 5$ sensors per subarray

2.2.5 Shift-Invariant Arrays

Consider a linear PCA of $M = PM_0$ sensors partitioned into P identical subarrays of M_0 sensors and with arbitrary subarray topology, as displayed in Figure 2.5. The given array topology exhibits a shift-invariance property, expressed as

$$r_m^{(p)} = r_m^{(1)} + \eta^{(p)}, \quad p = 2, \dots, P, \quad m = 1, \dots, M_0, \quad (2.27a)$$

$$r_m^{(p)} = r_1^{(p)} + \rho_m, \quad p = 1, \dots, P, \quad m = 2, \dots, M_0, \quad (2.27b)$$

where the shifts $\eta^{(p)}$, for $p = 2, \dots, P$, denote the unknown inter-subarray displacements while the shifts ρ_m , for $m = 2, \dots, M_0$, denote the perfectly known relative sensor positions within the subarrays, as introduced in (2.8).

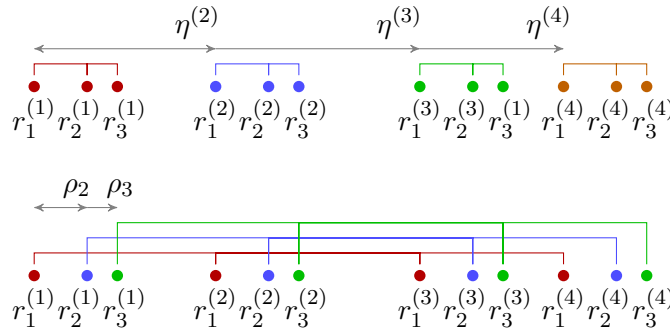


Figure 2.5: Illustration of multiple shift-invariances in array topology

In correspondence with (2.12), the array response vector is given by

$$\mathbf{a}(\mu) = \mathbf{B}(\mu)\boldsymbol{\varphi}(\mu, \boldsymbol{\eta}). \quad (2.28)$$

Similar to the case of identical uniform linear subarrays introduced in Section 2.2.3, the subarray response block matrix introduced in (2.13) can be represented as

$$\mathbf{B}(\mu) = \mathbf{I}_P \otimes \mathbf{a}^{(0)}(\mu) \quad (2.29)$$

with $\mathbf{a}^{(0)}(\mu) = [1, e^{-j\pi\mu\rho_2}, \dots, e^{-j\pi\mu\rho_{M_0}}]^\top$ denoting the subarray response vector, which is identical for all subarrays. Furthermore, consider the subarray displacement shift vector

$$\boldsymbol{\varphi}(\mu, \boldsymbol{\eta}) = [1, e^{-j\pi\mu\eta^{(2)}}, \dots, e^{-j\pi\mu\eta^{(P)}}]^\top \quad (2.30)$$

defined in correspondence to (2.15), where no additional subarray perturbations are assumed in this scenario, i.e., $\alpha^{(p)} = 1$ for $p = 2, \dots, P$ in (2.15). The information of no additional perturbation can be exploited in the signal model, as discussed in the following.

Let $\mathbf{J}^{(p)}$ denote the $PM_0 \times M_0$ matrix selecting the sensors in the p th subarray, for $p = 1, \dots, P$, and \mathbf{J}_m denote the $PM_0 \times P$ matrix selecting the m th sensor in each subarray, for $m = 1, \dots, M_0$, i.e.,

$$\mathbf{J}^{(p)} = \mathbf{e}_{P,p} \otimes \mathbf{I}_{M_0} \quad (2.31a)$$

$$\mathbf{J}_m = \mathbf{I}_P \otimes \mathbf{e}_{M_0,m}, \quad (2.31b)$$

where $\mathbf{e}_{P,p} = [0, \dots, 0, 1, 0, \dots, 0]^\top$ denotes the P -dimensional basis vector with a single 1 in the p th element and zero entries elsewhere. By the shift-invariance property (2.27) the steering vector in (2.28) fulfills the identities

$$\mathbf{J}^{(p)\top} \mathbf{a}(\boldsymbol{\mu}) = e^{-j\pi\boldsymbol{\mu}\boldsymbol{\eta}^{(p)}} \mathbf{J}^{(1)\top} \mathbf{a}(\boldsymbol{\mu}) \quad (2.32a)$$

$$\mathbf{J}_m^\top \mathbf{a}(\boldsymbol{\mu}) = e^{-j\pi\boldsymbol{\mu}\boldsymbol{\rho}_m} \mathbf{J}_1^\top \mathbf{a}(\boldsymbol{\mu}) \quad (2.32b)$$

and correspondingly the steering matrix fulfills

$$\mathbf{J}^{(p)\top} \mathbf{A}(\boldsymbol{\mu}) = \mathbf{J}^{(1)\top} \mathbf{A}(\boldsymbol{\mu}) \bar{\boldsymbol{\Phi}}^{\boldsymbol{\eta}^{(p)}}(\boldsymbol{\mu}) \quad (2.33a)$$

$$\mathbf{J}_m^\top \mathbf{A}(\boldsymbol{\mu}) = \mathbf{J}_1^\top \mathbf{A}(\boldsymbol{\mu}) \bar{\boldsymbol{\Phi}}^{\boldsymbol{\rho}_m}(\boldsymbol{\mu}), \quad (2.33b)$$

for $m = 1, \dots, M_0$ and $p = 1, \dots, P$, where the $L \times L$ unitary diagonal matrix

$$\bar{\boldsymbol{\Phi}}(\boldsymbol{\mu}) = \text{diag}(e^{-j\pi\boldsymbol{\mu}_1}, \dots, e^{-j\pi\boldsymbol{\mu}_L}) \quad (2.34)$$

contains the complex phase shifts for the spatial frequencies in $\boldsymbol{\mu}$ on its main diagonal. The shift-invariances displayed in (2.33) represent constraints on the signal model which can be exploited in the signal reconstruction process, as discussed in Sections 3.4 and 5.4.

Note that besides the shift-invariances illustrated in Figure 2.5, there exist other forms of shift-invariances such as overlapping shift-invariant groups or centro-symmetric groups, that can be exploited [ROSK88, SORK92]. The class of ULAs discussed in Section 2.1 represents a special case of shift-invariant arrays. As illustrated in Figure 2.6 a), the ULA can similarly be described by two overlapping groups of sensors which are invariant under linear translation.

Similarly, centro-symmetric arrays, as displayed in Figure 2.6 b), exhibit a structure that can be exploited by an additional constraint. Let the reference point of the array be given by the array center, e.g., $r_4 = 0$ for the example in Figure 2.6 b). Due to the centro-symmetry of the array, the steering matrix fulfills

$$\mathbf{A}(\boldsymbol{\mu}) = \bar{\mathbf{J}}^\top \mathbf{A}^*(\boldsymbol{\mu}), \quad (2.35)$$

where $\mathbf{A}^*(\boldsymbol{\mu})$ denotes the complex conjugate of the steering matrix $\mathbf{A}(\boldsymbol{\mu})$ and $\bar{\mathbf{J}}$ denotes the permutation matrix with ones on the anti-diagonal and zeros elsewhere.

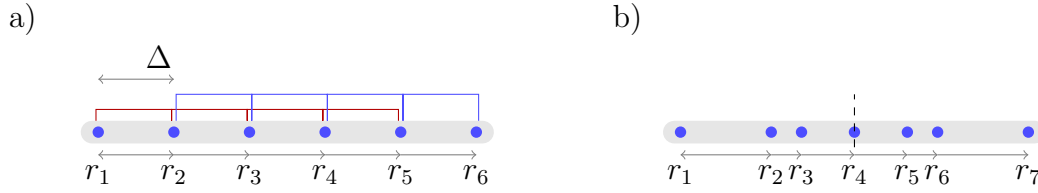


Figure 2.6: a) Uniform linear array with two overlapping shift-invariant groups of sensors and b) centro-symmetric group of sensors, i.e., the sensors are symmetrically positioned around the array center

2.2.6 Thinned Shift-Invariant Arrays

The concept of thinned arrays, introduced in the context of fully calibrated uniform linear arrays in Section 2.1.2 and partly calibrated arrays composed of uniform linear subarrays in Section 2.2.4, can likewise be applied to shift-invariant arrays. Similar to the discussion in Section 2.2.4, a thinned shift-invariant array of M sensors in P subarrays is modeled by removing sensors from an underlying virtual shift-invariant array of PM_0 sensors in P identical subarrays, as illustrated in Figure 2.7. The resulting subarray response block matrix is described by

$$\mathbf{B}(\mu) = \mathbf{J}^T (\mathbf{I}_P \otimes \mathbf{a}^{(0)}(\mu)), \quad (2.36)$$

where $\mathbf{a}^{(0)}(\mu) \in \mathbb{C}^{M_0}$ denotes the response vector of the identical virtual subarrays and \mathbf{J} denotes an appropriate selection matrix of dimensions $PM_0 \times M$.



Figure 2.7: Thinned shift-invariant array of $M = 9$, obtained by removing sensors from the virtual shift-invariant array of $P = 4$ sensors with $M_0 = 3$ sensors per subarray, given in Figure 2.5

Chapter 3

State of the Art

In the past decades, research on sensor array processing has provided a rich theory on estimation methods and statistical bounds on the achievable performance [KV96, vT02, TF09]. One of the objectives in array processing is reconstruction of impinging signals in the presence of noise and interference, which is achieved, e.g., by beamforming methods, also referred to as spatial filtering. Prominent beamforming methods are the Bartlett beamformer [Bar48] or the Capon beamformer [Cap69]. Proper application of beamforming methods usually requires knowledge of the directions of arrival (DOAs) of the impinging signals, which can be obtained by appropriate parameter estimation methods.

Given a proper model to describe the sensor measurements, maximum likelihood (ML) estimation [Böh84, Wax85, Böh86, Wax91, OVS93] provides several attractive features for DOA estimation. On the one hand, ML estimation has been shown to be asymptotically efficient, i.e., for a large number of samples the estimator converges to the Cramér-Rao lower bound [SLG01]. Moreover, ML estimation can be used both in uncorrelated and coherent source scenarios, has excellent threshold and asymptotic performances, and is unbiased and consistent. On the other hand, ML estimation involves highly nonlinear and nonconvex multidimensional optimization and is sensitive to model-errors. For this reason alternative methods with reduced computational cost are desired.

This chapter provides a short review of subspace-based estimation methods [GRP10, HPREK14], which have been shown to perform asymptotically optimal [SA89] while having comparably low computational cost. The chapter starts with a brief introduction of the concept of subspace separation, followed by the presentation of the well-known Multiple Signal Classification (MUSIC) method [Sch86] and its search-free implementation, the root-MUSIC method [Bar83]. The MUSIC method and its variants are only applicable to fully-calibrated arrays, when the array response is known. For application in partly-calibrated arrays (PCAs), alternative DOA estimation techniques in form of the Rank Reduction (RARE) method [PGW02, SG04], the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [RK89] and Multiple-Invariance (MI-) ESPRIT [SORK92] are considered.

3.1 Subspace Estimation

Subspace-based methods rely on estimating the signal and noise-subspace in the measurements $\mathbf{Y} = [\mathbf{y}(t_1), \dots, \mathbf{y}(t_N)] \in \mathbb{C}^{M \times N}$, obtained for the FCA (2.5) or PCA

case (2.19). Under Assumption A6, the signal covariance matrix can be represented as [KV96, vT02, TF09]

$$\begin{aligned} \mathbf{R} &= \mathbb{E}\{\mathbf{y}(t)\mathbf{y}^H(t)\} \\ &= \mathbf{A}(\boldsymbol{\mu})\mathbb{E}\{\boldsymbol{\psi}(t)\boldsymbol{\psi}^H(t)\}\mathbf{A}^H(\boldsymbol{\mu}) + \mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} \\ &= \mathbf{A}(\boldsymbol{\mu})\boldsymbol{\Pi}\mathbf{A}^H(\boldsymbol{\mu}) + \sigma_N^2\mathbf{I}_M \in \mathbb{C}^{M \times M} \end{aligned} \quad (3.1)$$

with $\boldsymbol{\Pi} = \mathbb{E}\{\boldsymbol{\psi}(t)\boldsymbol{\psi}^H(t)\} \in \mathbb{C}^{L \times L}$ denoting the source covariance matrix and $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma_N^2\mathbf{I}_M \in \mathbb{C}^{M \times M}$ representing the covariance matrix of the sensor noise, where it is assumed that $\mathbf{n}(t)$ is spatially and temporarily white complex circular Gaussian noise. In the case of uncorrelated source signals, the source covariance matrix $\boldsymbol{\Pi}$ is a full-rank diagonal matrix of size $L \times L$. Assuming further that the steering matrix $\mathbf{A}(\boldsymbol{\mu}) \in \mathbb{C}^{M \times L}$ has full column rank, according to Assumption A7, the matrix product $\mathbf{A}(\boldsymbol{\mu})\boldsymbol{\Pi}\mathbf{A}^H(\boldsymbol{\mu})$ is of full-rank and the eigenvalues of the covariance matrix \mathbf{R} in (3.1) fulfill [KV96, vT02, TF09]

$$\xi_1 \geq \dots \geq \xi_L > \xi_{L+1} = \dots = \xi_M, \quad (3.2)$$

where the $M - L$ smallest eigenvalues $\xi_{L+1} = \dots = \xi_M = \sigma_N^2$ correspond to the noise power σ_N^2 . Let $\boldsymbol{\Xi}_S = \text{diag}(\xi_1, \dots, \xi_L) \in \mathbb{C}^{L \times L}$ denote the diagonal matrix containing the L signal eigenvalues and let $\mathbf{U}_S \in \mathbb{C}^{M \times L}$ denote the matrix containing the corresponding eigenvectors, representing the signal subspace. Furthermore, let $\mathbf{U}_N \in \mathbb{C}^{M \times M-L}$ denote the matrix containing the eigenvectors corresponding to the $M - L$ noise eigenvalues, representing the noise subspace, such that $\mathbf{U}_N^H \mathbf{U}_S = \mathbf{0}$. The eigendecomposition of the signal covariance matrix can be expressed as

$$\begin{aligned} \mathbf{R} &= \mathbf{U}_S \boldsymbol{\Xi}_S \mathbf{U}_S^H + \mathbf{U}_N \boldsymbol{\Xi}_N \mathbf{U}_N^H \\ &= \mathbf{U}_S \boldsymbol{\Xi}_S \mathbf{U}_S^H + \sigma_N^2 \mathbf{U}_N \mathbf{U}_N^H, \end{aligned} \quad (3.3)$$

where $\boldsymbol{\Xi}_N = \sigma_N^2 \mathbf{I}_{M-L}$. Comparing (3.1) and (3.3) it becomes apparent that the steering matrix $\mathbf{A}(\boldsymbol{\mu})$ is orthogonal to the noise subspace spanned by \mathbf{U}_N [Bar83, Sch86], i.e.,

$$\mathbf{U}_N^H \mathbf{A}(\boldsymbol{\mu}) = \mathbf{0}. \quad (3.4)$$

Analogously, it can be observed that the signal subspace, represented by \mathbf{U}_S , is spanned by the steering vectors in $\mathbf{A}(\boldsymbol{\mu})$, such that

$$\mathbf{A}(\boldsymbol{\mu})\mathbf{T} = \mathbf{U}_S, \quad (3.5)$$

for some non-singular matrix $\mathbf{T} \in \mathbb{C}^{L \times L}$ [ROSK88, RK89]. The properties (3.4) and (3.5) of the noise and signal subspaces are exploited in subspace-based parameter estimation methods such as MUSIC [Bar83, Sch86], RARE [PGW02, SG04] and ESPRIT [ROSK88, RK89], as discussed in the following sections.

In practice the signal covariance matrix \mathbf{R} is not available and is approximated by the sample covariance matrix [KV96, vT02, TF09]

$$\hat{\mathbf{R}} = \mathbf{Y}\mathbf{Y}^H/N. \quad (3.6)$$

Similar as in (3.3), the signal and noise subspaces are estimated from the eigendecomposition of the sample covariance matrix according to

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_S \hat{\mathbf{\Sigma}}_S \hat{\mathbf{U}}_S^H + \hat{\mathbf{U}}_N \hat{\mathbf{\Sigma}}_N \hat{\mathbf{U}}_N^H, \quad (3.7)$$

where $\hat{\mathbf{\Sigma}}_S$ is an $L \times L$ diagonal matrix containing the L largest eigenvalues $\hat{\xi}_1, \dots, \hat{\xi}_L$ of the sample covariance matrix $\hat{\mathbf{R}}$, and $\hat{\mathbf{U}}_S$ contains the L principal eigenvectors of the sample covariance matrix, representing the estimated signal subspace. Correspondingly, the $(M - L) \times (M - L)$ diagonal matrix $\hat{\mathbf{\Sigma}}_N$ contains the estimated noise eigenvalues $\hat{\xi}_{L+1}, \dots, \hat{\xi}_M$, whereas $\hat{\mathbf{U}}_N$ contains the estimated noise eigenvectors, providing an estimate of the noise subspace.

In the asymptotic case of large number of snapshots $N \rightarrow \infty$, the sample covariance matrix approaches the true covariance matrix $\hat{\mathbf{R}} \rightarrow \mathbf{R}$. However, in the non-asymptotic case this is not true and the quality of the signal and noise subspace estimates $\hat{\mathbf{U}}_S$ and $\hat{\mathbf{U}}_N$ depends on the number of signal snapshots N . Note that for proper separation of the L -dimensional signal subspace \mathbf{U}_S and the $(M - L)$ -dimensional noise subspace \mathbf{U}_N , it is required to obtain $N \geq L$ linearly independent measurement vectors, such that conditions (3.4) and (3.5) can be fulfilled for the signal and noise subspace estimates. Another difficulty in the subspace separation occurs in the case of correlated or coherent source signals, leading to a rank-deficient source covariance matrix $\mathbf{\Pi}$. In this case the matrix product $\mathbf{A}(\boldsymbol{\mu})\mathbf{\Pi}\mathbf{A}^H(\boldsymbol{\mu})$ is rank deficient and the subspace separation by means of the eigenvalues, according to (3.2) and (3.3), can no longer be applied. In the case of ULAs, special methods like forward-backward averaging and spatial smoothing [SWK85, PK89, GE95] can be applied to circumvent the problem of a rank-deficient source matrix $\mathbf{\Pi}$. However, for arbitrary array topologies these techniques might not be applicable, leading to a potential degradation in estimation performance.

3.2 MUSIC Algorithms

The MULTiple Signal Classification (MUSIC) method [Sch86] relies on the observation in (3.4), that the noise subspace \mathbf{U}_N is orthogonal to the steering vectors, i.e.,

$$\mathbf{U}_N^H \mathbf{a}(\mu_l) = \mathbf{0}, \quad (3.8)$$

for spatial frequency μ_l corresponding to source l , for $l = 1, \dots, L$, or equivalently

$$\mathbf{a}^H(\mu_l) \mathbf{U}_N \mathbf{U}_N^H \mathbf{a}(\mu_l) = 0. \quad (3.9)$$

In the MUSIC method, the spatial frequency is sampled in K points ν_1, \dots, ν_K and the criterion (3.9) is evaluated for each of the sampled frequencies. In practice the noise subspace estimate $\hat{\mathbf{U}}_N$ is used, for which the criterion in (3.9) is usually not fulfilled with equality. Therefore, the function

$$f_{\text{MUSIC}}(\nu_k) = \mathbf{a}^H(\nu_k) \hat{\mathbf{U}}_N \hat{\mathbf{U}}_N^H \mathbf{a}(\nu_k) \quad (3.10)$$

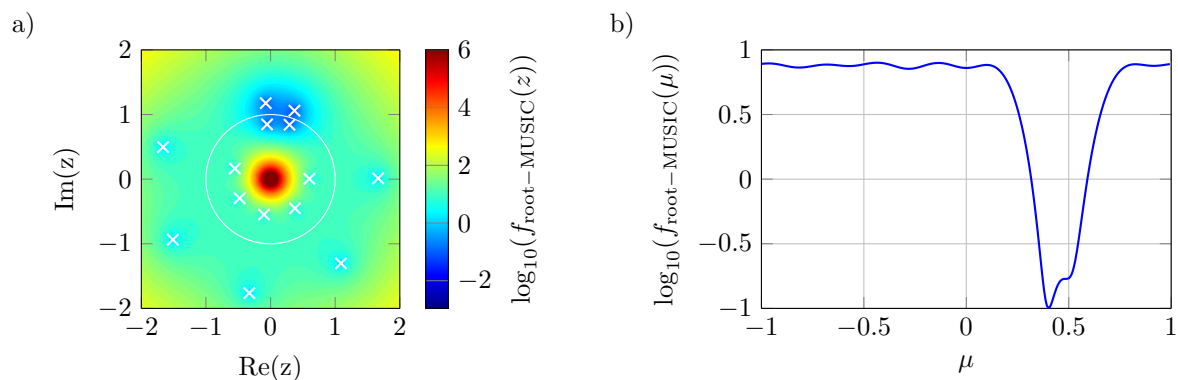


Figure 3.1: Exemplary MUSIC spectrum a) in the complex z -plane (white marks indicate root locations) and b) on the unit circle, for a ULA of $M = 8$ sensors and $L = 2$ source signals with spatial frequencies $\boldsymbol{\mu} = [-0.4, -0.5]^T$, $N = 30$ snapshots and SNR = 0 dB

is evaluated instead, providing the null-spectrum for the MUSIC method. The L smallest local minima of (3.10) w.r.t. ν_k provide estimates of the spatial frequencies corresponding to the L source signals.

In the case of ULAs, special structure in the steering vectors can be exploited, leading to the well-known root-MUSIC method [Bar83]. To see this, consider the array response vector $\mathbf{a}(z) = [1, z, \dots, z^{M-1}]^T$ as defined in (2.6), with $z = e^{-j\pi\mu\Delta}$ mapping the continuous spatial frequency $\mu \in [-1, 1)$ to the unit circle in the complex plane. On the unit circle $|z| = 1$ the array steering vector $\mathbf{a}(z)$ fulfills the identity

$$\mathbf{a}^H(z) = \mathbf{a}^T(1/z). \quad (3.11)$$

Using (3.11) in (3.9) results in the condition

$$\mathbf{a}^T(1/z)\mathbf{U}_N\mathbf{U}_N^H\mathbf{a}(z) = 0, \quad (3.12)$$

where the left-hand side constitutes a polynomial of degree $2(M-1)$ in z , according to

$$f_{\text{root-MUSIC}}(z) = \mathbf{a}^T(1/z)\mathbf{U}_N\mathbf{U}_N^H\mathbf{a}(z) = \sum_{m=-M+1}^{M-1} c_m z^{-m}. \quad (3.13)$$

The polynomial coefficients are given by

$$c_m = \text{Tr}(\boldsymbol{\Theta}_m\mathbf{U}_N\mathbf{U}_N^H), \quad (3.14)$$

for $m = -M+1, \dots, M-1$, where $\boldsymbol{\Theta}_m$ is the $M \times M$ elementary Toeplitz matrix with ones on the m th diagonal and zeros elsewhere, e.g.,

$$\boldsymbol{\Theta}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Theta}_{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (3.15)$$

for $M = 4$. Given the polynomial coefficients c_m , the polynomial roots can be found by, e.g., computing the eigenvalues of the corresponding companion matrix [HJ90]. The $2(M - 1)$ complex roots \hat{z}_m , for $m = 1, \dots, 2(M - 1)$, of (3.13) occur in conjugate pairs and the roots of interest lie on the unit circle, with the spatial frequencies corresponding to the source signals determined by $\hat{\mu}_l = -\frac{1}{\pi\Delta} \arg(\hat{z}_l)$. Similar to the spectral MUSIC function in (3.10), in practice the noise subspace \mathbf{U}_N in (3.13) is replaced by its estimate $\hat{\mathbf{U}}_N$ and the L roots of the polynomial

$$f_{\text{root-MUSIC}}(z) = \mathbf{a}^\top(1/z) \hat{\mathbf{U}}_N \hat{\mathbf{U}}_N^H \mathbf{a}(z) \quad (3.16)$$

which are closest to the unit circle are selected for frequency reconstruction. In other words, the root-MUSIC method computes the roots of (3.16) in the entire complex plane, while the MUSIC method according to (3.10) limits the search to minima on the unit circle, as illustrated in Figure 3.1. This difference in the two approaches results in better resolution performance of the root-MUSIC method as compared to the MUSIC method [RH89, PGH00], as demonstrated by numerical experiments in Section 4.4.2. Another advantage of the root-MUSIC method is the reduction in computational cost, since the roots for (3.16) can be computed efficiently instead of evaluating the MUSIC criterion (3.10) for multiple frequencies ν_1, \dots, ν_K . On the other hand, an advantage of MUSIC over root-MUSIC is that the array response for different spatial frequencies can be found by calibration, i.e., no analytical description of the array response is required, making MUSIC applicable to arbitrary sensor systems [RK89].

In the case of a thinned linear array of M sensors with maximum lag ΔM_0 , the steering vector can be expressed as $\mathbf{a}(z) = \mathbf{J}^\top \mathbf{a}_0(z)$, where $\mathbf{a}_0(z)$ represents the steering vector of a virtual ULA of M_0 sensors, and \mathbf{J} denotes a proper selection matrix of size $M_0 \times M$, as discussed in Section 2.1.2. In this case, the root-MUSIC polynomial is given as

$$\begin{aligned} f_{\text{root-MUSIC}}(z) &= \mathbf{a}^\top(1/z) \hat{\mathbf{U}}_N \hat{\mathbf{U}}_N^H \mathbf{a}(z) \\ &= \mathbf{a}_0^\top(1/z) \mathbf{J} \hat{\mathbf{U}}_N \hat{\mathbf{U}}_N^H \mathbf{J}^\top \mathbf{a}_0(z) \end{aligned} \quad (3.17)$$

and the $2M_0 - 1$ polynomial coefficients are computed as

$$c_m = \text{Tr}(\mathbf{\Theta}_m \mathbf{J} \hat{\mathbf{U}}_N \hat{\mathbf{U}}_N^H \mathbf{J}^\top), \quad (3.18)$$

in relation to (3.14). The approach of computing the polynomial coefficients as given in (3.14) and (3.18) will similarly be considered in the context of sparse reconstruction in Section (4.3.1).

3.3 RARE Algorithms

The MUSIC method discussed in the previous section is applicable to fully calibrated arrays, when the array response $\mathbf{a}(\mu)$ is perfectly known. In the case of partly calibrated arrays the MUSIC approach has to be modified, leading to the RANk

REduction (RARE) method presented in [PGW02, SG04]. Using the factorization $\mathbf{a}(\mu, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \mathbf{B}(\mu)\boldsymbol{\varphi}(\mu, \boldsymbol{\alpha}, \boldsymbol{\eta})$ of the steering vector, as given in (2.12), the MUSIC criterion in (3.9) can be rewritten as

$$\begin{aligned} \mathbf{a}^H(\mu_l, \boldsymbol{\alpha}, \boldsymbol{\eta})\mathbf{U}_N\mathbf{U}_N^H\mathbf{a}(\mu_l, \boldsymbol{\alpha}, \boldsymbol{\eta}) &= \boldsymbol{\varphi}^H(\mu_l, \boldsymbol{\alpha}, \boldsymbol{\eta})\mathbf{B}^H(\mu_l)\mathbf{U}_N\mathbf{U}_N^H\mathbf{B}(\mu_l)\boldsymbol{\varphi}(\mu_l, \boldsymbol{\alpha}, \boldsymbol{\eta}) \\ &= 0. \end{aligned} \quad (3.19)$$

For a displacement shift vector $\boldsymbol{\varphi}(\mu, \boldsymbol{\alpha}, \boldsymbol{\eta}) \neq \mathbf{0}$, the criterion in (3.19) can only be fulfilled if the matrix product $\mathbf{B}^H(\mu_l)\mathbf{U}_N\mathbf{U}_N^H\mathbf{B}(\mu_l)$ is rank-deficient, or equivalently, if

$$\det(\mathbf{B}^H(\mu_l)\mathbf{U}_N\mathbf{U}_N^H\mathbf{B}(\mu_l)) = 0. \quad (3.20)$$

In contrast to the criterion in (3.19), the criterion in (3.20) does not require any knowledge of the displacement shifts in $\boldsymbol{\varphi}(\mu, \boldsymbol{\alpha}, \boldsymbol{\eta})$. Only knowledge of the known subarray responses in $\mathbf{B}(\mu)$ and of the noise subspace \mathbf{U}_N are required, where in practice the latter is approximated by its estimate $\hat{\mathbf{U}}_N$. Similar as for the MUSIC method, the RARE method relies on scanning the frequency space in K points ν_1, \dots, ν_K and evaluating the function

$$f_{\text{RARE}}(\nu_k) = \det(\mathbf{B}^H(\nu_k)\hat{\mathbf{U}}_N\hat{\mathbf{U}}_N^H\mathbf{B}(\nu_k)). \quad (3.21)$$

The corresponding frequency estimates are obtained by finding the L smallest local minima of the null-spectrum $f_{\text{RARE}}(\nu_k)$.

Consider the special case of a PCA of $M = PM_0$ sensors, composed of P identical ULAs with baseline Δ and M_0 sensors per subarray, as introduced in Section 2.2.3. Upon defining $z = e^{-j\pi\mu\Delta}$ and using the subarray response block matrix $\mathbf{B}(z)$ as given in (2.25), it can be seen that the matrix product $\mathbf{B}^H(z)\mathbf{U}_N\mathbf{U}_N^H\mathbf{B}(z)$ corresponds to a matrix polynomial of degree $2(M_0 - 1)$ according to

$$\mathbf{F}_{\text{root-RARE}}(z) = \mathbf{B}^H(z)\mathbf{U}_N\mathbf{U}_N^H\mathbf{B}(z) = \sum_{m=-M_0+1}^{M_0-1} \mathbf{C}_m z^{-m}. \quad (3.22)$$

The $2M_0 - 1$ matrix coefficients \mathbf{C}_m , with $m = -M_0 + 1, \dots, M_0 - 1$, can be computed from the matrix product $\mathbf{U}_N\mathbf{U}_N^H$ by extending the concepts in Section 3.2, which will similarly be applied for sparse reconstruction in Section 5.3.1. To this end, define the $M \times P$ matrix

$$\boldsymbol{\Omega}(z) = [1 \quad z \quad z^2 \quad \dots \quad z^{M_0-1}]^T \otimes \mathbf{I}_P \quad (3.23)$$

and introduce the $M \times M$ permutation matrix \mathbf{J} such that the subarray steering block matrix can be expressed as

$$\begin{aligned} \mathbf{B}(z) &= \mathbf{I}_P \otimes [1 \quad z \quad z^2 \quad \dots \quad z^{M_0-1}]^T \\ &= \mathbf{J}^T \boldsymbol{\Omega}(z). \end{aligned} \quad (3.24)$$

Inserting (3.24) in (3.22) yields

$$\begin{aligned}\mathbf{F}_{\text{root-RARE}}(z) &= \mathbf{B}^H(z)\mathbf{U}_N\mathbf{U}_N^H\mathbf{B}(z) \\ &= \boldsymbol{\Omega}^H(z)\mathbf{J}\mathbf{U}_N\mathbf{U}_N^H\mathbf{J}^T\boldsymbol{\Omega}(z) \\ &= \boldsymbol{\Omega}^H(z)\mathbf{G}\boldsymbol{\Omega}(z),\end{aligned}\tag{3.25}$$

where $\mathbf{G} = \mathbf{J}\mathbf{U}_N\mathbf{U}_N^H\mathbf{J}^T$ is of size $M \times M$ and is composed of the $P \times P$ blocks $\mathbf{G}_{i,j}$, for $i, j = 1, \dots, M_0$, as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \cdots & \mathbf{G}_{1,M_0} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{M_0,1} & \cdots & \mathbf{G}_{M_0,M_0} \end{bmatrix}.\tag{3.26}$$

Equation (3.25) is also referred to as the Gram matrix representation of the matrix polynomial $\mathbf{F}_{\text{root-RARE}}(z)$, and \mathbf{G} is referred to as the corresponding Gram matrix [Dum07]. Define the block-trace operator for matrix \mathbf{G} as

$$\text{blkTr}^{(P)}(\mathbf{G}) = \sum_{m=1}^{M_0} \mathbf{G}_{m,m},\tag{3.27}$$

i.e., the summation of the $P \times P$ submatrices $\mathbf{G}_{m,m}$, for $m = 1, \dots, M_0$, on the main diagonal of matrix \mathbf{G} . By employing the block-trace operator (3.27) and the elementary Toeplitz matrices $\boldsymbol{\Theta}_m$, as introduced in (3.14), the matrix coefficients \mathbf{C}_m in (3.22) can be computed from the Gram matrix \mathbf{G} in (3.25) as

$$\mathbf{C}_m = \text{blkTr}^{(P)}((\boldsymbol{\Theta}_m \otimes \mathbf{I}_P)\mathbf{G}),\tag{3.28}$$

i.e., the summation of the $P \times P$ submatrices on the m th block-diagonal of the Gram matrix \mathbf{G} . As discussed in [Pes05], the roots of the matrix polynomial $\mathbf{F}_{\text{root-RARE}}(z)$, as given in (3.22), are the points \hat{z}_l in the complex plane where the matrix becomes rank-deficient, i.e., $\det(\mathbf{F}_{\text{root-RARE}}(\hat{z}_l)) = 0$, and the roots corresponding to the spatial frequencies of the source signals are located on the unit circle. Given the matrix polynomial coefficients \mathbf{C}_m , with $m = -M_0 + 1, \dots, M_0 - 1$, the $M - P$ roots of the matrix polynomial $\mathbf{F}_{\text{root-RARE}}(z)$ can be found, e.g., by the block companion matrix approach discussed in [Pes05]. Similar as in other subspace based methods, in practice the noise subspace \mathbf{U}_N in (3.22) is approximated by its estimate $\hat{\mathbf{U}}_N$ and the roots \hat{z}_l corresponding to the spatial frequencies of interest do not exactly lie on the unit circle. Instead, the roots closest to the unit circle are selected as the desired roots.

The case of PCAs composed of thinned linear subarrays introduced in Section 2.2.4 can be treated similar as for the root-MUSIC method for fully calibrated thinned linear arrays in (3.17), discussed in the previous section, by defining a proper selection matrix $\check{\mathbf{J}}$ and computing the matrix coefficients in (3.22) according to

$$\mathbf{C}_m = \text{blkTr}^{(P)}((\boldsymbol{\Theta}_m \otimes \mathbf{I}_P)\check{\mathbf{J}}\mathbf{G}\check{\mathbf{J}}^T).\tag{3.29}$$

The root-RARE algorithm has similar advantages over the RARE algorithm, as discussed in the previous section for the MUSIC methods: The root-RARE method provides better resolution capabilities since it searches the entire complex plane for roots, instead of just searching the unit circle. At the same time, the root-RARE method has reduced computational cost since the roots can be computed in a search-free manner. On the other hand, root-RARE requires specific structure in the subarrays and an analytical description of the subarray response is required. In this regard RARE is more flexible since it can be applied to arbitrary array topologies.

3.4 ESPRIT Algorithms

Similar to the root-MUSIC and root-RARE methods, the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) exploits special structure in the array topology [RK89]. The ESPRIT approach is based on the observation that subarrays which are identical under linear translation, exhibit signal subspaces which are identical under rotation, as will be explained in more detail in this section.

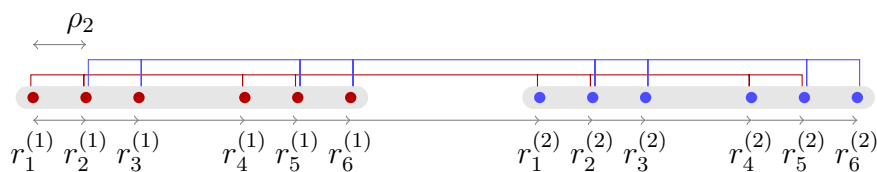


Figure 3.2: Sensor array with two overlapping shift-invariant groups

For better illustration of the ESPRIT approach, consider a partly calibrated array composed of $P = 2$ identical subarrays with $M_0 = 6$ sensors per subarray, as illustrated in Figure 3.2. Two overlapping groups of $M_1 = 8$ sensors, that are invariant under the shift ρ_2 , can be formed for the given array topology, using the corresponding selection matrices

$$\mathbf{J}_1^{(1)} = \mathbf{I}_4 \otimes \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^T \quad \text{and} \quad \mathbf{J}_1^{(2)} = \mathbf{I}_4 \otimes \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^T. \quad (3.30)$$

Based on the signal model in (2.10), the sensor group measurements are formulated as $\mathbf{Y}_1^{(1)} = \mathbf{J}_1^{(1)\top} \mathbf{Y}$ and $\mathbf{Y}_1^{(2)} = \mathbf{J}_1^{(2)\top} \mathbf{Y}$ and a compact notation to take account of the shift-invariance is introduced as

$$\begin{bmatrix} \mathbf{Y}_1^{(1)} \\ \mathbf{Y}_1^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1(\boldsymbol{\mu}) \\ \mathbf{A}_1(\boldsymbol{\mu}) \bar{\boldsymbol{\Phi}}^{\rho_2}(\boldsymbol{\mu}) \end{bmatrix} \boldsymbol{\Psi} + \begin{bmatrix} \mathbf{N}_1^{(1)} \\ \mathbf{N}_1^{(2)} \end{bmatrix}, \quad (3.31)$$

where the additive noise in $\mathbf{N}_1^{(1)}$ and $\mathbf{N}_1^{(2)}$ is defined in correspondence to the sensor group measurements $\mathbf{Y}_1^{(1)}$ and $\mathbf{Y}_1^{(2)}$, the matrix $\mathbf{A}_1(\boldsymbol{\mu}) = \mathbf{J}_1^{(1)\top} \mathbf{A}(\boldsymbol{\mu})$ is the array response for the first sensor group, and $\bar{\boldsymbol{\Phi}}(\boldsymbol{\mu}) = \text{diag}(e^{-j\pi\mu_1}, \dots, e^{-j\pi\mu_L})$ is defined

similar as in (2.34). Let the $M_1 \times L$ matrices $\mathbf{U}_{S,1}^{(1)} = \mathbf{J}_1^{(1)\top} \mathbf{U}_S$ and $\mathbf{U}_{S,1}^{(2)} = \mathbf{J}_1^{(2)\top} \mathbf{U}_S$ be obtained from the signal subspace \mathbf{U}_S of the sensor measurements \mathbf{Y} , then, under Assumption A7 that $\mathbf{A}_1(\boldsymbol{\mu})$ has full rank, there exists a non-singular matrix $\mathbf{T} \in \mathbb{C}^{L \times L}$ such that

$$\begin{bmatrix} \mathbf{U}_{S,1}^{(1)} \\ \mathbf{U}_{S,1}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1(\boldsymbol{\mu}) \\ \mathbf{A}_1(\boldsymbol{\mu}) \bar{\boldsymbol{\Phi}}^{\rho_2}(\boldsymbol{\mu}) \end{bmatrix} \mathbf{T}, \quad (3.32)$$

according to (3.5). Since \mathbf{T} is non-singular, it can be found from (3.32) that $\mathbf{A}_1(\boldsymbol{\mu}) = \mathbf{U}_{S,1}^{(1)} \mathbf{T}^{-1}$, which in turn yields

$$\begin{aligned} \mathbf{U}_{S,1}^{(2)} &= \mathbf{A}_1(\boldsymbol{\mu}) \bar{\boldsymbol{\Phi}}^{\rho_2}(\boldsymbol{\mu}) \mathbf{T} \\ &= \mathbf{U}_{S,1}^{(1)} \mathbf{T}^{-1} \bar{\boldsymbol{\Phi}}^{\rho_2}(\boldsymbol{\mu}) \mathbf{T} \\ &= \mathbf{U}_{S,1}^{(1)} \bar{\mathbf{T}}^{\rho_2}. \end{aligned} \quad (3.33)$$

Problem (3.33) constitutes a generalized eigenvalue problem and represents the ESPRIT criterion. The spatial frequencies $\boldsymbol{\mu}$ can be computed from the matrix $\bar{\boldsymbol{\Phi}}(\boldsymbol{\mu})$, which is given in the eigenvalues of $\bar{\mathbf{T}} = \mathbf{T}^{-1} \bar{\boldsymbol{\Phi}}(\boldsymbol{\mu}) \mathbf{T}$. From (3.33) it can be seen that the ESPRIT method can resolve at most $L \leq M_1$ spatial frequencies. Otherwise the linear system in (3.33) would be underdetermined, providing non-unique solutions $\bar{\mathbf{T}}$. Note that the ESPRIT formulation in (3.33) is closely related to the matrix pencil method [HS90, HW91], which similarly solves a generalized eigenvalue problem.

In the practically relevant case of a limited number of snapshots and noise-corrupted measurements, the signal subspace estimates $\hat{\mathbf{U}}_{S,1}^{(1)}$ and $\hat{\mathbf{U}}_{S,1}^{(2)}$ and the steering matrix $\mathbf{A}_1(\boldsymbol{\mu})$ do not span exactly the same range space, such that condition (3.33) cannot be fulfilled with equality for the subspace estimates. As discussed in [ROSK88, SORK92], the problem of noise corrupted measurements can be treated by matching the signal subspace estimates $\hat{\mathbf{U}}_{S,1}^{(1)} \in \mathbb{C}^{M_1 \times L}$ and $\hat{\mathbf{U}}_{S,1}^{(2)} \in \mathbb{C}^{M_1 \times L}$ to a common signal subspace estimate $\hat{\mathbf{U}}_{S,0} \in \mathbb{C}^{M \times L}$ according to the least-squares minimization problem

$$\min_{\mathbf{U}_{S,0}, \bar{\mathbf{T}}_0} \left\| \begin{bmatrix} \hat{\mathbf{U}}_{S,1}^{(1)} - \mathbf{J}_1^{(1)\top} \mathbf{U}_{S,0} \\ \hat{\mathbf{U}}_{S,1}^{(2)} - \mathbf{J}_1^{(2)\top} \mathbf{U}_{S,0} \bar{\mathbf{T}}_0^{\rho_2} \end{bmatrix} \right\|_{\text{F}}^2, \quad (3.34)$$

where the minimizer $\hat{\mathbf{U}}_{S,0} \in \mathbb{C}^{M \times L}$ models the common signal subspace estimate and $\hat{\bar{\mathbf{T}}}_0 = \hat{\mathbf{T}}^{-1} \bar{\boldsymbol{\Phi}}(\hat{\boldsymbol{\mu}}) \hat{\mathbf{T}} \in \mathbb{C}^{L \times L}$ represents the rotation matrix estimate similar as in (3.33). The problem in (3.34) is also referred to as total least squares ESPRIT, since it takes account of errors in both the signal subspace estimates $\hat{\mathbf{U}}_{S,1}^{(1)}$ and $\hat{\mathbf{U}}_{S,1}^{(2)}$, whereas a least squares approach for (3.33) would only take account of the errors either in $\hat{\mathbf{U}}_{S,1}^{(1)}$ or $\hat{\mathbf{U}}_{S,1}^{(2)}$ [OVK91].

Another advantage of the formulation in (3.34) is, that it can easily be extended to exploit multiple shift-invariances. To illustrate this, assume that the array from Figure

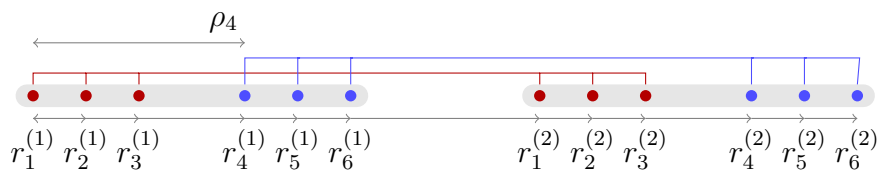


Figure 3.3: Sensor array form Figure 3.2 with alternative shift-invariant groups

3.2 exhibits an additional shift-invariance as displayed in Figure 3.3. The corresponding sensor selection matrices and signal subspace estimates are given by

$$\mathbf{J}_2^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^T \otimes \mathbf{I}_3 \quad \text{and} \quad \mathbf{J}_2^{(2)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \otimes \mathbf{I}_3, \quad (3.35)$$

and $\hat{\mathbf{U}}_{S,2}^{(p)} = \mathbf{J}_2^{(p)\top} \hat{\mathbf{U}}_S$, for $p = 1, 2$. Utilizing the additional shift-invariances in (3.34) results in the multiple invariance (MI-) ESPRIT formulation [ROSK88]

$$\min_{\mathbf{U}_{S,0}, \mathbf{T}_0} \left\| \begin{bmatrix} \hat{\mathbf{U}}_{S,1}^{(1)} - \mathbf{J}_1^{(1)\top} \mathbf{U}_{S,0} \\ \hat{\mathbf{U}}_{S,1}^{(2)} - \mathbf{J}_1^{(2)\top} \mathbf{U}_{S,0} \bar{\mathbf{T}}_0^{\rho_2} \\ \hat{\mathbf{U}}_{S,2}^{(1)} - \mathbf{J}_2^{(1)\top} \mathbf{U}_{S,0} \\ \hat{\mathbf{U}}_{S,2}^{(2)} - \mathbf{J}_2^{(2)\top} \mathbf{U}_{S,0} \bar{\mathbf{T}}_0^{\rho_4} \end{bmatrix} \right\|_{\mathbb{F}}^2. \quad (3.36)$$

Note that application of the MI-ESPRIT in (3.36) requires explicit knowledge of the linear translations ρ_2 and ρ_4 of the shift-invariant groups, i.e., MI-ESPRIT cannot directly exploit the shift-invariance of the two identical subarrays in the PCA in Figure 3.3, since it is assumed that the subarray shift $\eta = r_1^{(2)} - r_1^{(1)}$ is unknown. Another drawback of the MI-ESPRIT is that it requires nonlinear and nonconvex optimization. Gradient descent or Newton's method have been proposed to solve the MI-ESPRIT minimization problem in (3.36), where proper convergence requires initialization in a point sufficiently close to the global minimum, e.g., given by the single-invariance ESPRIT in (3.33) [ROSK88, SORK92]. In contrast to the single-invariance ESPRIT formulation discussed above, the maximum number L_{\max} of frequencies resolvable by MI-ESPRIT is more difficult to derive. As discussed in [SORK92], for the special case of a PCA of M sensors, composed of P uniform linear subarrays of M_0 sensors per subarray, the maximum number of resolvable frequencies can be bounded as

$$\min\{M_0(P-1), M-1\} \leq L_{\max} \leq M-1, \quad (3.37)$$

where it is assumed that the signal covariance matrix has full rank.

Chapter 4

Sparse Reconstruction for Fully Calibrated Arrays

The following chapter considers the application of joint sparse signal reconstruction (SSR) in fully calibrated arrays. The first section introduces the main concepts of SSR for the classical application of sparse reconstruction from a single measurement vector by means of ℓ_1 norm minimization. In the context of array processing it is more common to consider multiple snapshots for signal reconstruction, and Section 4.2 discusses the concept of joint sparse reconstruction from multiple snapshots by means of the well-known $\ell_{2,1}$ mixed-norm minimization problem [MÇW05, YL06]. Compared to more recently presented sparse methods such as SPICE [SBL11a, SBL11b, SZL14] and atomic norm minimization [YX15b, LC16, YX16a], the classical $\ell_{2,1}$ minimization problem has the general shortcoming that the number of optimization parameters in the jointly sparse signal representation grows with the number of snapshots as well as with the resolution requirement. Approaches to deal with the aforementioned problems have been presented, e.g., in [MÇW05].

The main contribution of this chapter is given in Section 4.3, by a novel reformulation of the $\ell_{2,1}$ minimization problem, which is based on a compact parameterization in which the optimization parameters represent the row-norms of the signal representation, rather than the signal matrix itself. The compact formulation is referred to as SPARse ROW-norm reconstruction (SPARROW). Given the sparse signal row-norms, the jointly sparse signal matrix is reconstructed from the measurement vectors in closed-form. Moreover, the SPARROW formulation only relies on the sample covariance matrix instead of the measurement vectors themselves. In this sense, a concentration of the optimization variables as well as the measurements is achieved, leading to a significantly reduced problem size in the case of a large number of snapshots. The resulting optimization problem can be solved by means of semidefinite programming or by a low-complexity coordinate descent implementation which is applicable to large and irregular sampling scenarios, as presented in Section 6.1.3. Furthermore, a gridless SPARROW formulation is derived for search-free parameter estimation in uniform linear arrays, which is proven to be equivalent to the atomic norm minimization problem [YX15b, LC16, YX16a]. For off-grid estimation in arbitrary array topologies, i.e., estimation of spatial frequencies which are not restricted to lie on a predefined grid of candidate frequencies, an off-grid SPARROW formulation is derived which uses linear interpolation to model the array response in between two frequency grid points. To put the SPARROW framework in context with existing work, a comparison to the recently proposed SParse Iterative Covariance-based Estimation (SPICE) method [SBL11a, SBL11b, SZL14] is performed, which extends the existing links between SPICE and $\ell_{2,1}$ mixed-norm minimization.

The chapter is concluded by a numerical analysis of the parameter estimation performance and the computational cost of the proposed SPARROW formulation. The latter results demonstrate a significant reduction in the computation time of the proposed reformulation as compared to both equivalent formulations, the classical $\ell_{2,1}$ mixed-norm [MCW05, YL06] and the atomic norm [YX15b, LC16, YX16a] problem formulations.

4.1 Sparse Reconstruction from Single Snapshots

Consider an array of M sensors and assume L point sources in the farfield region of the array, as discussed in Section 2.1 and illustrated in Figure 4.1. In relation to (2.2), a single snapshot of the noise-free signal received by the fully calibrated array is modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\mu})\boldsymbol{\psi}(t). \quad (4.1)$$

The signal model in (4.1) can equivalently be described by a sparse representation according to

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\nu})\check{\mathbf{x}}(t), \quad (4.2)$$

with $\check{\mathbf{x}}(t) \in \mathbb{C}^K$ denoting the sparse signal vector, and $\mathbf{A}(\boldsymbol{\nu}) \in \mathbb{C}^{M \times K}$ representing the overcomplete sensing matrix defined in correspondence to (2.3), where the vector $\boldsymbol{\nu} = [\nu_1, \dots, \nu_K]^\top$ is obtained by sampling the spatial frequencies in $K \gg L$ points ν_1, \dots, ν_K as illustrated in Figure 4.2. For ease of notation, the argument $\boldsymbol{\nu}$ will be dropped if the sparse reconstruction context is clear and the sensing matrix $\mathbf{A}(\boldsymbol{\nu})$ will be referred to as \mathbf{A} in the remainder of this thesis. It is assumed that the frequency grid is sufficiently fine, such that the true frequencies in $\boldsymbol{\mu}$ are contained in the frequency grid $\boldsymbol{\nu}$, i.e.,

$$\{\mu_l\}_{l=1}^L \subset \{\nu_k\}_{k=1}^K. \quad (4.3)$$

Since the true frequencies in $\boldsymbol{\mu}$ are not known in advance and the grid-size is limited in practice, the on-grid assumption (4.3) is usually not fulfilled, leading to spectral leakage effects and basis mismatch [HS10, CSPC11] in the estimation process. Methods to deal with these effects are considered in Sections 4.3.1 and 4.3.2. For ease of presentation the on-grid assumption (4.3) is assumed to hold true, elsewhere.

Under the on-grid assumption given in (4.3), the sparse signal vector $\check{\mathbf{x}}(t) = [\check{x}_1(t), \dots, \check{x}_K(t)]^\top$ in (4.2) contains only $L \ll K$ non-zero elements, according to

$$\check{x}_k(t) = \begin{cases} \psi_l(t) & \text{if } \nu_k = \mu_l \\ 0 & \text{otherwise,} \end{cases} \quad (4.4)$$

for $k = 1, \dots, K$ and $l = 1, \dots, L$, as illustrated in Figure 4.2.

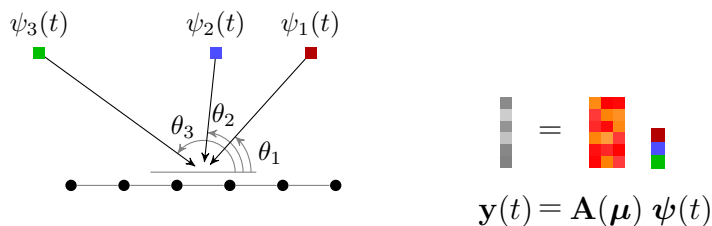


Figure 4.1: Array signal model for $M = 6$ sensors and $L = 3$ source signals

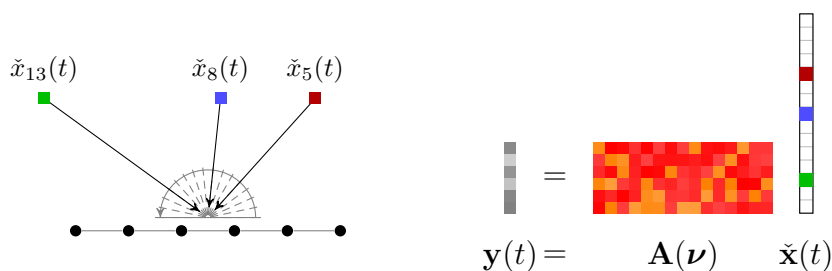


Figure 4.2: Sparse representation of the array signal model in Figure 4.1, with $K = 15$ grid points

The problem of recovering the signal vector $\tilde{\mathbf{x}}(t)$ from the measurement vector $\mathbf{y}(t)$ according to (4.2) is also referred to as linear inverse problem and occurs in similar form in various applications of signal processing, such as spectral analysis, direction of arrival (DOA) estimation, image processing, geophysics, tomography and magnetic resonance imaging, and machine learning. Accordingly, the problem has received considerable research interest in the past decades [Tib96, CDS98, DE03, CT05, Don06, CRT06a, CRT06b, CR06]. Clearly, the system in (4.2) is underdetermined and has infinitely many solutions. However, if the a-priori knowledge of a sparse solution is exploited, it was shown that this problem has a unique solution under certain conditions on the sparsity of $\tilde{\mathbf{x}}(t)$ and the sensing matrix \mathbf{A} [DE03, CT05, Don06, CRT06a, CRT06b, CR06]. The problem of recovering the signal vector $\tilde{\mathbf{x}}(t)$ in (4.2) under sparsity constraints is also referred to as sparse signal reconstruction, sparse recovery or compressed sensing.

Ideally, the sparse structure of the signal vector would be exploited by an optimization problem of the form

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{y}(t), \quad (4.5)$$

where the ℓ_0 quasi-norm of the vector $\mathbf{x} = [x_1, \dots, x_K]^T$ counts the number of its non-zero elements according to

$$\|\mathbf{x}\|_0 = |\{k \mid x_k \neq 0\}|. \quad (4.6)$$

Reconstruction guarantees for the ℓ_0 minimization problem (4.5) have been considered, e.g., in [DE03], based on the sparsity of $\tilde{\mathbf{x}}(t)$ and conditions on the sensing matrix \mathbf{A} .

In the case of noise-corrupted measurements, given by

$$\mathbf{y}(t) = \mathbf{A}\tilde{\mathbf{x}}(t) + \mathbf{n}(t), \quad (4.7)$$

the corresponding optimization problem can be formulated as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}(t)\|_2^2 \leq \beta, \quad (4.8)$$

where β is some threshold selected in relation to the noise power $E\{\|\mathbf{n}(t)\|_2^2\} = \sigma_N^2$. Given a sparse minimizer $\hat{\mathbf{x}}(t) = [\hat{x}_1(t), \dots, \hat{x}_K(t)]^\top$ for the ℓ_0 minimization problem, the DOA estimation problem reduces to identifying the support set, i.e., the indices of the non-zero elements, from which the set of estimated spatial frequencies can be obtained as

$$\{\hat{\mu}_l\}_{l=1}^{\hat{L}} = \{\nu_k \mid \hat{x}_k(t) \neq 0, k = 1, \dots, K\}, \quad (4.9)$$

where \hat{L} denotes the number of non-zero elements in $\hat{\mathbf{x}}(t)$, i.e., the estimated model order.

The ℓ_0 minimization problem is a combinatorial problem and becomes computationally intractable for large problem dimensions, cf. [FR13], such that numerous methods to approximately solve the sparse reconstruction problem have been derived in recent years. The most popular methods include greedy pursuit methods such as Orthogonal Matching Pursuit (OMP) [MZ93], threshold methods such as Iterative Hard Thresholding (IHT) [BD08], convex relaxation by means of the ℓ_1 norm [Tib96, CDS98], (re-)weighted ℓ_1 norm minimization [Zou06, CWB08], Approximate Message Passing (AMP) [DMM09], sparse Bayesian learning [Tip01, WR04], non-convex optimization and mixed-integer programming. The different methods vary in computational cost and reconstruction performance. Excellent overviews are given in [TW10, FR13]. This thesis focuses on convex relaxation approaches which show improved reconstruction performance as compared to greedy pursuit methods, but have lower computational cost than non-convex methods and mixed-integer programming.

4.1.1 ℓ_1 Norm Minimization

A popular strategy to approximately solve the ℓ_0 minimization problems (4.5) or (4.8) is to replace the ℓ_0 quasi-norm by the ℓ_1 norm. The ℓ_1 norm of a vector $\mathbf{x} = [x_1, \dots, x_K]^\top$ is computed as the absolute sum of the vector elements according to

$$\|\mathbf{x}\|_1 = \sum_k |x_k|, \quad (4.10)$$

and presents a tight convex approximation of the ℓ_0 quasi-norm. The convex approximation is considered in the sense that the ℓ_1 norm ball $\|\mathbf{x}\|_1 \leq 1$ forms the convex hull of all vectors \mathbf{x} with ℓ_∞ maximum norm and ℓ_0 quasi-norm bounded simultaneously as

$\|\mathbf{x}\|_\infty \leq 1$ and $\|\mathbf{x}\|_0 \leq 1$. In this regard, the convex relaxation of (4.5) in the noise-free case is formulated as the ℓ_1 minimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{y}(t), \quad (4.11)$$

which is commonly referred to as Basis Pursuit (BP) [CDS98]. Recovery guarantees for the ℓ_1 minimization problem have been considered, e.g., in [DE03, CT05].

In the case of noise corrupted measurements, according to (4.7), the problem in (4.11) is modified to admit an error in the reconstructed signal and the measurements, given by

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y}(t) - \mathbf{A}\mathbf{x}\|_2^2 \leq \beta, \quad (4.12)$$

where the constant $\beta \geq 0$ is selected according to the noise distribution, similar as for (4.8). Bounds on the reconstruction error $\|\tilde{\mathbf{x}}(t) - \hat{\mathbf{x}}(t)\|_2^2$, for a minimizer $\hat{\mathbf{x}}(t)$ of problem (4.12), have been considered, e.g., in [CRT06b]. Problem (4.12) is also referred to as Basis Pursuit DeNoising (BPDN) [CDS98] in the literature. Note that in contrast to the constrained ℓ_0 minimization problem in (4.8), ℓ_1 minimization according to (4.12) induces an additional bias in the magnitudes of the elements of \mathbf{x} , such that the constraint $\|\mathbf{y}(t) - \mathbf{A}\mathbf{x}\|_2^2 \leq \beta$ needs to take account of noise and bias effects, making selection of the threshold β more challenging. An alternative formulation of (4.12) is given by

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y}(t) - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (4.13)$$

where $\lambda > 0$ is a regularization parameter determining the sparsity of the minimizer $\hat{\mathbf{x}}(t)$, i.e., larger values of λ will reduce the number of non-zero elements in $\hat{\mathbf{x}}(t)$. Problem (4.13) is commonly referred to as Least Absolute Shrinkage and Selection Operator (LASSO) [Tib96] in the literature. Note that (4.12) and (4.13) are equivalent for appropriate choices of λ and β , and that both reduce to BP (4.11) in the noiseless case by letting $\lambda, \beta \rightarrow 0$. In [OPT00a] it was shown, that for

$$\lambda \geq \max_k |\mathbf{a}_k^H \mathbf{y}(t)|, \quad (4.14)$$

the LASSO (4.13) generates zeros solutions $\hat{\mathbf{x}}(t) = \mathbf{0}$. Furthermore, it was discussed in [Don95, CDS98, BTR13], that for a regularization parameter selection according to

$$\lambda = \sigma_N \sqrt{M \log M} \quad (4.15)$$

the reconstruction error $E\{\|\tilde{\mathbf{x}}(t) - \hat{\mathbf{x}}(t)\|_2^2\}$ is bounded above, proportional to the noise deviation σ_N . Another popular approach of regularization parameter selection is based on computation of the solution path

$$\hat{\mathbf{x}}(\lambda_i) = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y}(t) - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_i \|\mathbf{x}\|_1, \quad (4.16)$$

i.e., evaluation of the LASSO problem for multiple regularization parameters $\lambda_1, \dots, \lambda_i, \dots$, and selection of a solution $\hat{\mathbf{x}}(\lambda_i)$ which fulfills some predefined criteria, e.g., by cross-validation [Tib96]. Efficient computation of the solution path can be performed by the coordinate descent method [FHHT07], or, in the case of real-valued systems, by the ℓ_1 homotopy method [OPT00b] or by Least Angle Regression (LARS) [EHJT04]. The ℓ_1 homotopy method exploits the property that the LASSO solution path is piecewise linear in the case of real-valued systems, and computes the breakpoints of the LASSO path in closed form. However, in the more general case of a complex-valued system, the ℓ_1 homotopy method cannot be applied.

In [Tib96] it was discussed that the ℓ_1 penalty term $\|\mathbf{x}\|_1$ in the LASSO function can be interpreted as the negative log-likelihood joint density function of independent Laplacian priors according to

$$\pi(x_k) \sim \exp(-\lambda|x_k|), \quad (4.17)$$

such that, in the case of additive Gaussian noise, the LASSO with data matching function $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{Ax} - \mathbf{y}(t)\|_2^2$ can similarly be interpreted as a Bayesian estimator with Laplacian priors. However, other data matching functions have been investigated in literature, two of which are the Square-Root LASSO (SR-LASSO) [BCW11]

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}(t)\|_2 + \lambda_{\text{SR}}\|\mathbf{x}\|_1 \quad (4.18)$$

and the Least Absolute Deviation LASSO (LAD-LASSO) [WLJ07]

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}(t)\|_1 + \lambda_{\text{LAD}}\|\mathbf{x}\|_1. \quad (4.19)$$

Among the above listed methods, the standard LASSO (4.13) is most prevalent in literature. Compared to the standard LASSO, for which the noise is usually assumed to be Gaussian and the regularization parameter is chosen proportional to the noise deviation σ_N , e.g., according to (4.15), SR-LASSO (4.18) and LAD-LASSO (4.19) are considered to make weaker assumptions on the noise distribution and the regularization parameters can be chosen as constants that are independent of the noise level [WLJ07, BCW11].

The ℓ_1 minimization problems, presented in this section, are convex and computationally tractable. However, the ℓ_1 norm is a non-differentiable function which requires special optimization methods [Roc70, Ber99], such that problems with large dimensions might result in high computational cost. The case of large problem dimensions occurs especially in the context of parameter estimation where a high resolution is desired, resulting in sparse vectors $\check{\mathbf{x}}(t)$ with large dimensions K . Various methods for efficiently solving the ℓ_1 problem have been presented in literature, including, e.g., the homotopy method [EHJT04], the coordinate descent method [Ber99, FHHT07], the soft-thresholding and exact line search algorithm (STELA) [YP17], interior point methods [KKB07], the proximal gradient descent method [CW05, PB⁺14] and the Alternating Direction Method of Multipliers (ADMM) [BPC⁺11].

4.1.2 Atomic Norm Minimization

As mentioned in the previous section, in the context of parameter estimation it is desired to achieve a high parameter resolution, i.e., to consider a large number of sampled frequencies K . Similarly, a large number of sampled frequencies is also desired to decrease the effects of spectral leakage and basis mismatch [HS10, CSPC11] caused by violating the on-grid assumption (4.3). On the other hand, the computational cost grows with increasing number of sampled frequencies K . Additionally, the correlation of the atoms in the sensing matrix increases with the number of frequency grid points, making the optimization problem numerically instable. In this context it is highly desired to obtain gridless reconstruction methods. Two such gridless approaches have recently been presented by means of total variation norm minimization [CFG13, CFG14] and atomic norm minimization [CRPW12, BTR13, TBSR13, TBR15], which have been shown to be equivalent in [TBSR13].

The concept of Atomic Norm Minimization (ANM) was introduced in [CRPW12] as a unifying framework for different types of sparse recovery methods, such as ℓ_1 norm minimization for sparse vector reconstruction or nuclear norm minimization for low-rank matrix completion [FHB01, RFP10]. In [BTR13, TBSR13, TBR15], ANM was introduced for gridless line spectral estimation from single measurement vectors under uniform sampling. For derivation of the ANM concept, consider L source signals with spatial frequencies μ_1, \dots, μ_L , impinging on a ULA with sensor positions $r_m = (m-1)\Delta$, for $m = 1, \dots, M$. The noise-free measurement vector obtained at the array output is modeled as $\mathbf{y}(t) = \sum_{l=1}^L \mathbf{a}(\mu_l)\psi_l(t)$, where $\psi_l(t)$ denotes the signal sample transmitted by source l in time instant t . In the ANM framework, the measurement vector $\mathbf{y}(t)$ is considered as a convex combination of atoms $\mathbf{a}(\nu)b$, with $b \in \mathbb{C}$, $|b| = 1$, representing a complex phase shift and $\nu \in [-1, 1)$ representing a continuous frequency parameter. The atomic norm of $\mathbf{y}(t)$ is defined as

$$\|\mathbf{y}(t)\|_{\mathcal{A}} = \inf_{\{s_k, b_k, \nu_k\}} \left\{ \sum_k s_k : \mathbf{y}(t) = \sum_k s_k b_k \mathbf{a}(\nu_k), s_k \in \mathbb{R}_+ \right\}. \quad (4.20)$$

For the special case of ULAs, it was shown in [CRPW12, BTR13, TBSR13, TBR15] that the atomic norm in (4.20) can equivalently be computed by the semidefinite program (SDP)

$$\|\mathbf{y}(t)\|_{\mathcal{A}} = \inf_{\mathbf{w}, W} \frac{1}{2}W + \frac{1}{2M}\text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.21a)$$

$$\text{s.t.} \quad \begin{bmatrix} W & \mathbf{y}^H(t) \\ \mathbf{y}(t) & \text{HToep}(\mathbf{w}) \end{bmatrix} \succeq \mathbf{0}, \quad (4.21b)$$

where W is an auxiliary variable resulting from a Schur complement argument [VB96] and $\text{HToep}(\mathbf{w})$ denotes a Hermitian Toeplitz matrix with the vector $\mathbf{w} =$

$[w_1, \dots, w_M]^T \in \mathbb{C}^M$ as its first column, according to

$$\text{HToep}(\mathbf{w}) = \begin{bmatrix} w_1 & w_2^* & w_3^* & \cdots & w_M^* \\ w_2 & w_1 & w_2^* & \cdots & w_{M-1}^* \\ w_3 & w_2 & w_1 & \cdots & w_{M-2}^* \\ \vdots & & & \ddots & \vdots \\ w_M & & & & w_1 \end{bmatrix}. \quad (4.22)$$

Given a solution to problem (4.21), the reconstruction of the spatial frequencies ν_k and magnitudes s_k , for $k = 1, \dots, K$, is performed by means of the Vandermonde decomposition as discussed in [TBSR13], and defined as follows:

Definition 4.1 (Vandermonde decomposition [TBSR13]). *For a ULA of M sensors, the $M \times K$ sensing matrix $\mathbf{A} = [\mathbf{a}(\nu_1), \dots, \mathbf{a}(\nu_K)]$ has a Vandermonde structure, such that the convex combinations of atoms $\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)$ forms a Hermitian Toeplitz matrix according to*

$$\sum_{k=1}^K s_k \mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k) = \text{HToep}(\mathbf{w}), \quad (4.23)$$

for nonnegative weights $s_k \geq 0$. By the Caratheodory theorem [Car11, CF11, Töp11], the mapping in (4.23) is unique for any $K \leq M$ distinct frequencies ν_1, \dots, ν_K and corresponding magnitudes $s_1, \dots, s_K > 0$, such that $\text{rank}(\text{HToep}(\mathbf{w})) = K \leq M$.

In practice, the Vandermonde decomposition of a Hermitian Toeplitz matrix $\text{HToep}(\mathbf{w})$ according to (4.23) can be obtained by first recovering the frequencies ν_k , e.g., by Prony's method [dP95], the matrix pencil approach [HS90, HW91] or linear prediction methods [KT82, TK82], where the frequency recovery is performed in a gridless fashion. The corresponding signal magnitudes in $\mathbf{s} = [s_1, \dots, s_K]^T$ can be reconstructed by solving the linear system

$$\mathbf{A} \mathbf{s} = \mathbf{w}, \quad (4.24)$$

i.e., by exploiting that $[\mathbf{a}(\nu)]_1 = 1$, for all $\nu \in [-1, 1)$, and considering the first column in the representation (4.23).

Recovery conditions for problem (4.21) are given in [TBSR13], providing a lower bound on the number of sensors M required to recover a superposition of L signals with a minimum allowed frequency separation.

As proposed in [BTR13, TBR15], given a noise-corrupted measurement vector $\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\mu}) \boldsymbol{\psi}(t) + \mathbf{n}(t)$ as defined in (2.2), gridless sparse recovery can be performed by Atomic norm Soft Thresholding (AST), by using (4.20) in the form of

$$\min_{\mathbf{y}_0} \frac{1}{2} \|\mathbf{y}(t) - \mathbf{y}_0\|_2^2 + \lambda \|\mathbf{y}_0\|_{\mathcal{A}}, \quad (4.25)$$

with the noise-free estimate of the measurements in \mathbf{y}_0 . Problem (4.25) is equivalently expressed by using the SDP formulation in (4.21) as

$$\min_{\mathbf{w}, W, \mathbf{y}_0} \frac{1}{2} \|\mathbf{y}(t) - \mathbf{y}_0\|_2^2 + \frac{\lambda}{2} \left(W + \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{w})) \right) \quad (4.26a)$$

$$\text{s.t.} \quad \begin{bmatrix} W & \mathbf{y}_0^H \\ \mathbf{y}_0 & \text{HToep}(\mathbf{w}) \end{bmatrix} \succeq \mathbf{0}. \quad (4.26b)$$

Bounds on the reconstruction error of the minimizer $\hat{\mathbf{y}}$ to problem (4.25) have been considered in [BTR13, TBR15]. In the case of a thinned linear array of M sensors with an associated virtual ULA of M_0 sensors and selection matrix \mathbf{J} of dimensions $M_0 \times M$, as discussed in Section 2.1.2, the ANM formulation can be rewritten as

$$\min_{\mathbf{y}_0} \frac{1}{2} \|\mathbf{y}(t) - \mathbf{J}^T \mathbf{y}_0\|_F^2 + \lambda \|\mathbf{y}_0\|_{\mathcal{A}}, \quad (4.27)$$

where $\mathbf{y}(t)$ is the M element vector containing the sensor measurements and \mathbf{y}_0 is the M_0 element vector modeling the measurements in the virtual ULA [TBSR13]. The atomic norm framework has the advantage that it does not require discretization of the frequency space, hence the computational cost is independent of the frequency resolution, in contrast to ℓ_1 minimization where the computational cost and frequency resolution depends on the number of sampled frequencies K . On the other hand, the computational cost of the SDP formulations in (4.21) and (4.26) grows significantly with the number of sensors M , which is reflected in the dimensions of the semidefinite constraints (4.21b) and (4.26b). An efficient implementation of the SDP, based on the alternating direction method of multipliers (ADMM), has been proposed in [BTR13]. However, for large problem sizes it was proposed in [TBR13] to rather use the grid-based formulations, such as the ℓ_1 minimization problems (4.11) or (4.13) which can be solved more efficiently, rather than the SDP formulations in (4.21) and (4.26).

4.2 Joint Sparse Reconstruction from Multiple Snapshots

The sparse reconstruction approach discussed in Section 4.1 considers DOA estimation from a single measurement vector $\mathbf{y}(t)$. In the case of multiple snapshots, as common in array processing applications, the problem provides additional structure that can be exploited in the sparse reconstruction approach.

Consider a noise-free realization of the multi snapshot model in (2.5) for an array of M sensors and L sources with spatial frequencies μ_1, \dots, μ_L , given as

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\mu}) \boldsymbol{\Psi}, \quad (4.28)$$

as discussed for (2.5), where it is assumed that the spatial frequencies $\boldsymbol{\mu}$ of the source signals remain constant during the time of observation. Similar to the single snapshot

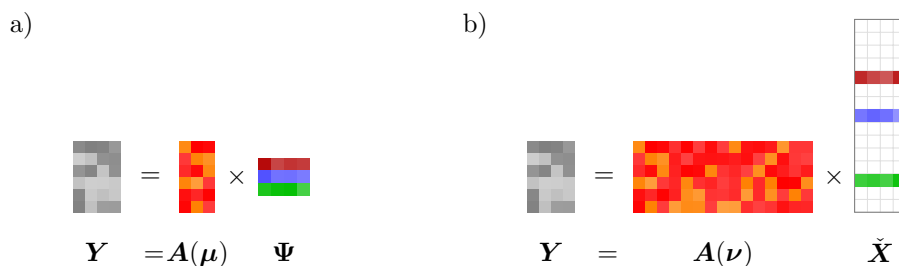


Figure 4.3: a) Signal model and b) sparse representation for $M = 6$ sensors, $L = 3$ source signals and $K = 15$ grid points

case discussed in Section 4.1, the multi snapshot model can be described by a sparse representation according to

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\nu}) \tilde{\mathbf{X}}, \quad (4.29)$$

with $\tilde{\mathbf{X}} \in \mathbb{C}^{K \times N}$ denoting a row-sparse signal matrix, and the overcomplete sensing matrix $\mathbf{A}(\boldsymbol{\nu}) \in \mathbb{C}^{M \times K}$, defined in relation to (4.2), which will be further referred to as \mathbf{A} for ease of notation. Similar as for the single snapshot case, the on-grid assumption (4.3) is expected to hold true, for ease of presentation.

The $K \times N$ sparse signal matrix $\tilde{\mathbf{X}}$ in (4.29) contains the elements

$$[\tilde{\mathbf{X}}]_{k,n} = \begin{cases} [\boldsymbol{\psi}(t_n)]_l & \text{if } \nu_k = \mu_l \\ 0 & \text{otherwise,} \end{cases} \quad (4.30)$$

for $k = 1, \dots, K$, $l = 1, \dots, L$ and $n = 1, \dots, N$, i.e., only $L \ll K$ rows are non-zero. Thus $\tilde{\mathbf{X}}$ exhibits a row-sparse structure, i.e., the elements in a row of $\tilde{\mathbf{X}}$ are either jointly zero or primarily non-zero, as illustrated in Figure 4.3.

The joint sparse reconstruction problem can be formulated by the following minimization problem

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{q,0} \quad \text{s.t. } \mathbf{Y} = \mathbf{A}\mathbf{X}, \quad (4.31)$$

where the $\ell_{q,0}$ mixed-norm of a matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]^\top$ is defined as the number of its non-zero rows \mathbf{x}_k according to

$$\|\mathbf{X}\|_{q,0} = |\{k \mid \|\mathbf{x}_k\|_q \neq 0\}|, \quad (4.32)$$

for any ℓ_q vector norm. Minimization of the $\ell_{q,0}$ mixed-norm promotes joint sparsity in the rows of \mathbf{X} , in the sense that rows are either jointly zero or primarily non-zero. Reconstruction guarantees for the joint sparse recovery problem in (4.31) have been derived, e.g., in [CH06].

In the case of noise corrupted measurements, the minimization problem can be formulated as

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{q,0} \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\mathbb{F}}^2 \leq \beta N \quad (4.33)$$

with the threshold β computed in relation to the noise power $\sigma_{\mathbb{N}}^2$. Note that for a single measurement vector, the problems (4.31) and (4.33) reduce to the standard ℓ_0 minimization problems discussed in Section 4.1.

Given a row-sparse minimizer $\hat{\mathbf{X}}$ for problems (4.31) or (4.33), the DOA estimation problem reduces to identifying the union support set, i.e., the indices of the non-zero rows, from which the set of estimated spatial frequencies can be obtained as

$$\{\hat{\mu}_l\}_{l=1}^{\hat{L}} = \{\nu_k \mid \|\hat{\mathbf{x}}_k\|_q \neq 0, k = 1, \dots, K\}, \quad (4.34)$$

where $\hat{\mathbf{x}}_k$ corresponds to the k th row of the estimated signal matrix $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_K]^{\text{T}}$ and \hat{L} denotes the number of non-zero rows in $\hat{\mathbf{X}}$, i.e., the estimated model order.

Similar to the ℓ_0 norm minimization problem, the $\ell_{q,0}$ mixed-norm minimization problem requires combinatorial optimization and becomes computationally intractable for large problem dimensions, such that approximate methods have been proposed, including greedy methods, such as Simultaneous Orthogonal Matching Pursuit (S-OMP) [TGS06], and convex relaxation to $\ell_{q,1}$ mixed-norm minimization [MÇW05, TVW05, Tro06, YL06, Kow09], the latter of which will be discussed in the following.

4.2.1 $\ell_{2,1}$ Mixed-Norm Minimization

The $\ell_{q,1}$ mixed-norm is a classical approach [MÇW05, TVW05, YL06, Tro06, Kow09, HM10] to approximate the $\ell_{q,0}$ norm in the joint sparse reconstruction problem (4.31) according to

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{q,1} \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{A}\mathbf{X}, \quad (4.35)$$

where the $\ell_{q,1}$ mixed-norm is defined as

$$\|\mathbf{X}\|_{q,1} = \sum_{k=1}^K \|\mathbf{x}_k\|_q, \quad (4.36)$$

applying an *inner* ℓ_q norm on the rows \mathbf{x}_k , for $k = 1, \dots, K$, in $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]^{\text{T}}$, and an *outer* ℓ_1 norm on the ℓ_q row-norms. The *inner* ℓ_q norm provides a nonlinear coupling among the elements in a row, which, in combination with the *outer* ℓ_1 norm, leads to the desired row-sparse structure of the signal matrix \mathbf{X} . In the case of noise-corrupted measurements, the penalized form

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_{\mathbb{F}}^2 + \lambda\sqrt{N}\|\mathbf{X}\|_{q,1}, \quad (4.37)$$

can be applied. The regularization parameter $\lambda \geq 0$ admits balancing the data fitting fidelity versus the sparsity level in \mathbf{X} , where the choice of a small λ in (4.37) tends to result in a large number of non-zero rows, whereas a large value of λ tends to result in a small number of non-zero rows. Common choices of mixed-norms are the $\ell_{2,1}$ norm [MCW05, YL06] and the $\ell_{\infty,1}$ norm [TVW05, Tro06]. In the case of a single measurement vector, the $\ell_{q,1}$ mixed-norm problem (4.37) reduces to the standard LASSO (4.13) discussed in Section 4.1. This thesis focuses on $\ell_{2,1}$ minimization in the form of

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_{\text{F}}^2 + \lambda\sqrt{N}\|\mathbf{X}\|_{2,1}. \quad (4.38)$$

Following the ideas in [OPT00a], the smallest regularization λ providing a zero solution for the problem in (4.38), is given by

$$\lambda = \max_k \|\mathbf{a}_k^{\text{H}}\mathbf{Y}\|_2/\sqrt{N}, \quad (4.39)$$

as shown in Appendix A. A novel regularization parameter selection approach for large number of sensors M and snapshots N , given as

$$\lambda = \sigma_N \sqrt{M} (\sqrt{M/N} + 1), \quad (4.40)$$

is derived in Appendix A as a contribution of this thesis. For the special case of a single snapshot $N = 1$, however, the regularization parameter selected according to (4.40) exceeds the parameter according to (4.15), given as

$$\lambda = \sigma_N \sqrt{M} \sqrt{\log(M)}, \quad (4.41)$$

which is commonly applied for single snapshot scenarios. This matter will be investigated in Section 4.4 by numerical experiments.

As discussed in [MCW05], a major drawback of the mixed-norm minimization problem in (4.38) lies in its computational cost, which is mainly determined by the size of the $K \times N$ source signal matrix \mathbf{X} . A large number of grid points K is desired to achieve high frequency resolution, while a large number of snapshots N is desired to provide good estimation performance. However, the choice of too large values K and N makes the problem computationally intractable. To reduce the computational cost of the problem (4.38) it was suggested in [MCW05] to reduce the dimension of the measurement matrix by matching only the signal subspace of \mathbf{Y} , leading to the prominent ℓ_1 -SVD method. Let the singular value decomposition (SVD) of \mathbf{Y} be given as $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$, and define the reduced $M \times r$ dimensional measurement matrix as

$$\mathbf{Y}_{\text{SV}} = \mathbf{U}\mathbf{\Sigma}\mathbf{D}_r = \mathbf{Y}\mathbf{V}\mathbf{D}_r \quad (4.42)$$

that contains most of the signal power, where $\mathbf{D}_r = [\mathbf{I}_r, \mathbf{0}]^{\text{T}}$ is of dimensions $N \times r$ with $r = \min(L, N)$. Similarly, let $\check{\mathbf{X}}_{\text{SV}} = \check{\mathbf{X}}\mathbf{V}\mathbf{D}_r$ and $\mathbf{N}_{\text{SV}} = \mathbf{N}\mathbf{V}\mathbf{D}_r$. Using this notation, the sparse representation of the signal model in (2.5) is replaced by

$$\mathbf{Y}_{\text{SV}} = \mathbf{A}\check{\mathbf{X}}_{\text{SV}} + \mathbf{N}_{\text{SV}}. \quad (4.43)$$

Based on this reduced dimensionality model, the ℓ_1 -SVD method was introduced as the constrained minimization problem

$$\min_{\mathbf{X}_{\text{SV}}} \|\mathbf{X}_{\text{SV}}\|_{2,1} \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{X}_{\text{SV}} - \mathbf{Y}_{\text{SV}}\|_{\text{F}}^2 \leq \beta, \quad (4.44)$$

leading to a reduction of the number of optimization parameters in the $K \times r$ matrix \mathbf{X}_{SV} . A drawback of the ℓ_1 -SVD approach is that a wrong estimate of the model order L or the presence of correlated source signals might impair proper computation of the signal subspace \mathbf{Y}_{SV} . Furthermore, the fact that \mathbf{N}_{SV} depends on the subspace \mathbf{V} makes the regularization parameter selection more difficult [MÇW05].

To deal with the above-mentioned problems of the ℓ_1 -SVD approach, another dimensionality reduction technique was proposed in [YLSX17], where the entire column space

$$\mathbf{Y}_{\text{RD}} = \mathbf{U}\mathbf{\Sigma}\mathbf{D}_{\min(M,N)} = \mathbf{Y}\mathbf{V}\mathbf{D}_{\min(M,N)}, \quad (4.45)$$

spanned by the measurements in \mathbf{Y} is matched. It was shown in [YLSX17] that using \mathbf{Y}_{RD} in the $\ell_{2,1}$ minimization (4.38) provides an estimate $\hat{\mathbf{X}}_{\text{RD}}$, that provides the same power spectrum as the estimate $\hat{\mathbf{X}}$ obtained by matching the original measurements in \mathbf{Y} , which can be explained by the unitary invariance of the Frobenius and $\ell_{2,1}$ mixed-norm. In contrast to the ℓ_1 -SVD approach, the computation of the column space \mathbf{Y}_{RD} does not require knowledge of the estimated model order and is robust to correlation in the source signals, however, at the expense of increased computational cost, due a larger number of optimization parameters in the corresponding $K \times \min(M, N)$ matrix \mathbf{X}_{RD} .

The above dimensionality reduction techniques deal with reducing the number of columns in the optimization variable \mathbf{X} of the $\ell_{2,1}$ minimization problem (4.38). To achieve a high frequency resolution it is furthermore desired to have a large number of sampled frequencies K , corresponding to the rows of \mathbf{X} . To reduce the computational cost associated with a large number of rows in \mathbf{X} , a heuristic approach of adaptively refining the frequency grid around estimated source locations was suggested in [MÇW05]. Another approach to achieve high frequency resolution is to apply the framework of atomic norm minimization, introduced in Section 4.1.2, to joint sparse reconstruction, as will be discussed in the following.

4.2.2 Atomic Norm Minimization

The extension of ANM to multiple measurement vectors was studied in [YX15b, LC16, YX16a], which will be briefly considered in the following. Assume L source signals with spatial frequencies μ_1, \dots, μ_L , impinging on a ULA of M sensors and consider the noise-free measurement matrix $\mathbf{Y}_0 = \sum_{l=1}^L \mathbf{a}(\mu_l)\boldsymbol{\psi}_l^{\text{T}}$ obtained at the array output, where the samples of the l th source signal are contained in the vector $\boldsymbol{\psi}_l \in \mathbb{C}^N$, for $l = 1, \dots, L$. In [YX15b, LC16, YX16a] it was proposed to model the measurement matrix \mathbf{Y}_0 as a convex combination of atoms $\mathbf{a}(\nu)\mathbf{b}^{\text{H}}$, with signal vector $\mathbf{b} \in \mathbb{C}^N$,

$\|\mathbf{b}\|_2 = 1$, and continuous spatial frequencies $\nu \in [-1, 1)$. Hence, the atomic norm of the measurement matrix \mathbf{Y}_0 is defined as

$$\|\mathbf{Y}_0\|_{\mathcal{A}} = \inf_{\{s_k, \mathbf{b}_k, \nu_k\}} \left\{ \sum_k s_k : \mathbf{Y}_0 = \sum_k s_k \mathbf{a}(\nu_k) \mathbf{b}_k^H, s_k \in \mathbb{R}_+ \right\}. \quad (4.46)$$

For the special case of ULAs, the atomic norm in (4.46) can equivalently be computed by the SDP [YX15b, LC16, YX16a]

$$\|\mathbf{Y}_0\|_{\mathcal{A}} = \inf_{\mathbf{w}, \mathbf{W}_N} \frac{1}{2} \text{Tr}(\mathbf{W}_N) + \frac{1}{2M} \text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.47a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_N & \mathbf{Y}_0^H \\ \mathbf{Y}_0 & \text{HToep}(\mathbf{w}) \end{bmatrix} \succeq \mathbf{0}, \quad (4.47b)$$

with the auxiliary variables $\mathbf{w} \in \mathbb{C}^M$ and $\mathbf{W}_N \in \mathbb{C}^{N \times N}$. Given a minimizer $\hat{\mathbf{w}}$ to problem (4.47), the reconstruction of the spatial frequencies ν_k and corresponding magnitudes s_k , for $k = 1, \dots, K$, is performed by means of the Vandermonde decomposition, as given in Definition 4.1. Recovery guarantees for problem (4.47) are given in [YX16a].

In the case of noise-corrupted measurements \mathbf{Y} , as defined in (2.5), gridless joint sparse recovery from multiple snapshots can be performed by using (4.46) in the Atomic norm Soft Thresholding (AST) problem

$$\min_{\mathbf{Y}_0} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}_0\|_{\mathbb{F}}^2 + \lambda \sqrt{N} \|\mathbf{Y}_0\|_{\mathcal{A}}, \quad (4.48)$$

or, equivalently, by using the SDP formulation in (4.47), as

$$\min_{\mathbf{w}, \mathbf{W}_N, \mathbf{Y}_0} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}_0\|_{\mathbb{F}}^2 + \frac{\lambda \sqrt{N}}{2} \left(\text{Tr}(\mathbf{W}_N) + \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{w})) \right) \quad (4.49a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_N & \mathbf{Y}_0^H \\ \mathbf{Y}_0 & \text{HToep}(\mathbf{w}) \end{bmatrix} \succeq \mathbf{0}, \quad (4.49b)$$

with auxiliary variable $\mathbf{Y}_0 \in \mathbb{C}^{M \times N}$, representing the noise-free measurement matrix estimate. Bounds on the reconstruction error of the estimated noise-free measurement matrix $\hat{\mathbf{Y}}_0$ have been presented in [LC16]. The case of a thinned linear array of M sensors, discussed in Section 2.1.2, is treated similar as for (4.27) by application of a proper selection matrix $\hat{\mathbf{J}}$ in problem (4.48).

Similar as for the $\ell_{2,1}$ mixed-norm minimization problem, in the case of a large number of snapshots N the ANM problem suffers from a large number of optimization parameters in the matrix \mathbf{Y}_0 , such that dimensionality reduction techniques similar to those discussed in Section 4.2.1 have been proposed to reduce the computational cost [YX15a]. In extension to the single snapshot case discussed in Section 4.1.2, an implementation based on ADMM has been proposed to deal with the computational cost of (4.49) in the case of large number of sensors M .

4.3 A Compact Formulation for $\ell_{2,1}$ Minimization

As seen from the previous section, the $\ell_{2,1}$ minimization problem for joint sparse reconstruction from multiple snapshots has received significant attention in SSR research in the past years. One of the main problems of the proposed approaches is the computational cost, especially in the case of a large number of snapshots, grid points or sensors. In this context, one of the main contributions of this thesis is given by the following theorem:

Theorem 4.2. *The row-sparsity inducing $\ell_{2,1}$ mixed-norm minimization problem*

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_{\text{F}}^2 + \lambda\sqrt{N} \|\mathbf{X}\|_{2,1} \quad (4.50)$$

is equivalent to the convex problem

$$\min_{\mathbf{S} \in \mathcal{D}_+^K} \text{Tr}((\mathbf{A}\mathbf{S}\mathbf{A}^{\text{H}} + \lambda\mathbf{I}_M)^{-1}\hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}), \quad (4.51)$$

with $\hat{\mathbf{R}} = \mathbf{Y}\mathbf{Y}^{\text{H}}/N$ denoting the sample covariance matrix and \mathcal{D}_+^K describing the set of nonnegative $K \times K$ diagonal matrices, in the sense that minimizers $\hat{\mathbf{X}}$ and $\hat{\mathbf{S}}$ for problems (4.50) and (4.51), respectively, are related by

$$\hat{\mathbf{X}} = \hat{\mathbf{S}}\mathbf{A}^{\text{H}}(\mathbf{A}\hat{\mathbf{S}}\mathbf{A}^{\text{H}} + \lambda\mathbf{I}_M)^{-1}\mathbf{Y}. \quad (4.52)$$

A proof of Theorem 4.2 is provided in Appendix B, while the convexity of (4.51) is proven in Appendix C by showing positive semidefiniteness of the Hessian matrix of the objective function in (4.51).

In addition to relation (4.52), it is shown in Appendix B that the matrix $\hat{\mathbf{S}} = \text{diag}(\hat{s}_1, \dots, \hat{s}_K)$ contains the row-norms of the sparse signal matrix $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_K]^{\text{T}}$ on its diagonal according to

$$\hat{s}_k = \frac{1}{\sqrt{N}} \|\hat{\mathbf{x}}_k\|_2, \quad (4.53)$$

for $k = 1, \dots, K$, such that the union support of $\hat{\mathbf{X}}$ is equivalently represented by the support of the sparse vector of row-norms $[\hat{s}_1, \dots, \hat{s}_K]$. The formulation in (4.51) is referred to as SPARse ROW-norm reconstruction (SPARROW). In this context, it should be emphasized that $\hat{\mathbf{S}}$ is not to be mistaken for a sparse representation of the source covariance matrix, i.e., in general $\hat{\mathbf{S}} \neq \text{E}\{\hat{\mathbf{X}}\hat{\mathbf{X}}^{\text{H}}\}/N$.

With regard to the problem dimension, the $\ell_{2,1}$ minimization problem in (4.50) has NK complex-valued variables in \mathbf{X} , while the SPARROW problem in (4.51) provides a reduction to only K real-valued, nonnegative variables in the diagonal matrix \mathbf{S} . Moreover, the SPARROW problem in (4.51) only depends on the sample covariance matrix $\hat{\mathbf{R}}$ instead of the actual measurement vectors in \mathbf{Y} themselves, leading to an

additional reduction in problem size, especially in the case of large number of snapshots N . Interestingly, this also indicates that the union support of the signal matrix $\hat{\mathbf{X}}$, i.e., the set of non-zero rows, represented by the diagonal matrix $\hat{\mathbf{S}}$, is fully encoded in the sample covariance $\hat{\mathbf{R}}$, rather than the instantaneous measurement vectors in \mathbf{Y} , as may be concluded from the $\ell_{2,1}$ formulation in (4.50). Similar observations were made in [YLSX17] in the context of dimensionality reduction. As seen from (4.52), the instantaneous measurement vectors in \mathbf{Y} are only required for the reconstruction of the signal matrix $\hat{\mathbf{X}}$, which, in the context of array signal processing, can be interpreted as a form of beamforming [vT02], where the row-sparse structure in $\hat{\mathbf{X}}$ is induced by premultiplication with the sparse diagonal matrix $\hat{\mathbf{S}}$.

The SPARROW formulation (4.51) can be implemented by tailored algorithms, such as the coordinate descent method presented in Section 6.1.3. For implementation of (4.51) by means of standard convex solvers, such as SeDuMi [Stu99] or MOSEK [MOS15], consider the following corollaries on the reformulation of the matrix inverse expression in (4.51) to a linear matrix inequality constraint, cf. [VB96, p.56]:

Corollary 4.3. *The SPARROW problem in (4.51) is equivalent to the semidefinite program*

$$\min_{\mathbf{S}, \mathbf{W}_N} \frac{1}{N} \text{Tr}(\mathbf{W}_N) + \text{Tr}(\mathbf{S}) \quad (4.54a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_N & \mathbf{Y}^H \\ \mathbf{Y} & \mathbf{A}\mathbf{S}\mathbf{A}^H + \lambda\mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (4.54b)$$

$$\mathbf{S} \in \mathcal{D}_+^K \quad (4.54c)$$

where \mathbf{W}_N is a Hermitian matrix of size $N \times N$.

To see the equivalence of the two problems, note that in (4.54) $\mathbf{A}\mathbf{S}\mathbf{A}^H + \lambda\mathbf{I}_M \succ \mathbf{0}$ is positive definite, since $\mathbf{S} \succeq \mathbf{0}$ and $\lambda > 0$. Further consider the Schur complement [VB96] of the constraint (4.54b),

$$\mathbf{W}_N \succeq \mathbf{Y}^H(\mathbf{A}\mathbf{S}\mathbf{A}^H + \lambda\mathbf{I}_M)^{-1}\mathbf{Y}, \quad (4.55)$$

which implies

$$\begin{aligned} \frac{1}{N} \text{Tr}(\mathbf{W}_N) &\geq \frac{1}{N} \text{Tr}(\mathbf{Y}^H(\mathbf{A}\mathbf{S}\mathbf{A}^H + \lambda\mathbf{I}_M)^{-1}\mathbf{Y}) \\ &= \text{Tr}((\mathbf{A}\mathbf{S}\mathbf{A}^H + \lambda\mathbf{I}_M)^{-1}\hat{\mathbf{R}}). \end{aligned} \quad (4.56)$$

For any optimal point $\hat{\mathbf{S}}$ of (4.51), a feasible point of (4.54) with the same objective function value can be constructed by choosing $\mathbf{W}_N = \mathbf{Y}^H(\mathbf{A}\hat{\mathbf{S}}\mathbf{A}^H + \lambda\mathbf{I}_M)^{-1}\mathbf{Y}$. Conversely, any optimal solution pair $\hat{\mathbf{W}}_N, \hat{\mathbf{S}}$ of (4.54) is also feasible for (4.51), which completes the proof.

Corollary 4.4. *The SPARROW formulation in (4.51) is equivalent to the semidefinite program*

$$\min_{\mathbf{S}, \mathbf{W}_M} \text{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \quad (4.57a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I}_M \\ \mathbf{I}_M & \mathbf{A}\mathbf{S}\mathbf{A}^H + \lambda\mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (4.57b)$$

$$\mathbf{S} \in \mathcal{D}_+^K \quad (4.57c)$$

where \mathbf{W}_M is a Hermitian matrix of size $M \times M$.

The proof of Corollary 4.4 follows the same line of arguments as in the proof of Corollary 4.3. In contrast to the constraint (4.54b), the dimension of the semidefinite constraint (4.57b) is independent of the number of snapshots N . It follows that either problem formulation (4.54) or (4.57) can be selected to solve the SPARROW problem in (4.51), depending on the number of snapshots N and the resulting dimension of the semidefinite constraint, i.e., (4.54) is preferable for $N \leq M$ and (4.57) is preferable otherwise. Note that the SDP implementations in [YX15b] have been derived using similar steps, i.e., employing the Schur complement to obtain linear matrix inequality constraints.

4.3.1 Gridless SPARROW

The SPARROW formulation in (4.51) relies on grid-based frequency estimation. As discussed for the on-grid assumption (4.3) and for atomic norm minimization in Section 4.1.2, grid-based frequency estimation requires a fine frequency grid to reduce the frequency estimation error, which, however, results in a high computational cost. The atomic norm minimization problem presented in Section 4.2.2 provides an approach for gridless frequency estimation in ULAs by means of the Vandermonde decomposition in Definition 4.1. In a similar fashion, the Vandermonde decomposition can be employed to derive a gridless version of the SPARROW formulation in (4.51).

Consider a ULA, where the sensing matrix \mathbf{A} has a Vandermonde structure and the matrix product $\mathbf{A}\mathbf{S}\mathbf{A}^H = \text{HToep}(\mathbf{w})$ forms a Hermitian Toeplitz matrix, as discussed in Section 4.1.2. Based on the uniqueness of the Vandermonde decomposition, problem (4.51) can be rewritten as the GridLess (GL-) SPARROW formulation

$$\min_{\mathbf{w}} \text{Tr}((\text{HToep}(\mathbf{w}) + \lambda\mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.58a)$$

$$\text{s.t.} \quad \text{HToep}(\mathbf{w}) \succeq \mathbf{0}, \quad (4.58b)$$

where the identity

$$\text{Tr}(\mathbf{S}) = \frac{1}{M} \text{Tr}(\mathbf{A}^H \mathbf{A} \mathbf{S}) = \frac{1}{M} \text{Tr}(\mathbf{A} \mathbf{S} \mathbf{A}^H) = \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.59)$$

was employed, with the factor $1/M$ resulting from $\|\mathbf{a}(\nu)\|_2^2 = M$, for all $\nu \in [-1, 1)$. Using the ideas of Corollaries 4.3 and 4.4, problem (4.58) is equivalent to the semidefinite program

$$\min_{\mathbf{w}, \mathbf{W}_N} \frac{1}{N} \text{Tr}(\mathbf{W}_N) + \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.60a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_N & \mathbf{Y}^H \\ \mathbf{Y} & \text{HToep}(\mathbf{w}) + \lambda \mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (4.60b)$$

$$\text{HToep}(\mathbf{w}) \succeq \mathbf{0}, \quad (4.60c)$$

or, alternatively, to

$$\min_{\mathbf{w}, \mathbf{W}_M} \text{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.61a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I}_M \\ \mathbf{I}_M & \text{HToep}(\mathbf{w}) + \lambda \mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (4.61b)$$

$$\text{HToep}(\mathbf{w}) \succeq \mathbf{0}. \quad (4.61c)$$

Given a minimizer $\hat{\mathbf{w}}$ of problem (4.60) or (4.61), the number of sources, i.e., the model order, can be estimated according to

$$\hat{L} = \text{rank}(\text{HToep}(\hat{\mathbf{w}})), \quad (4.62)$$

while the frequencies $\{\hat{\mu}_l\}_{l=1}^{\hat{L}}$ and corresponding magnitudes $\{\hat{s}_l\}_{l=1}^{\hat{L}}$ can be estimated by Vandermonde decomposition of the Hermitian Toeplitz matrix $\text{HToep}(\hat{\mathbf{w}})$ according to Definition 4.1. With the estimated frequencies in $\{\hat{\mu}_l\}_{l=1}^{\hat{L}}$ and signal magnitudes in $\{\hat{s}_l\}_{l=1}^{\hat{L}}$, the corresponding signal matrix estimate $\hat{\mathbf{X}}$ can be reconstructed by application of (4.52). As discussed in the context of the Vandermonde decomposition in Definition 4.1, different methods can be applied for frequency reconstruction from the Hermitian Toeplitz matrix $\text{HToep}(\hat{\mathbf{w}})$, such as root-MUSIC [Bar83], ESPRIT [RK89], the matrix pencil method [HS90, HW91] and others, similar to the atomic norm minimization problem in Section 4.1.2. The specific relation to the root-MUSIC estimator will be investigated in more detail in the following section. In this context the gridless SPARROW can also be considered as a signal subspace estimator, represented by the Hermitian Toeplitz matrix $\text{HToep}(\mathbf{w})$.

Unique Vandermonde decomposition requires that $\hat{L} = \text{rank}(\text{HToep}(\hat{\mathbf{w}})) < M$. The rank \hat{L} can be interpreted as the counterpart of the number of non-zero elements in the minimizer $\hat{\mathbf{S}}$ in the grid-based SPARROW formulation (4.51). Analogous to the grid-based SPARROW formulation (4.51), where the regularization parameter λ determines the sparsity level of $\hat{\mathbf{S}}$, i.e., the number of its non-zero elements, there always exists a value λ which yields a minimizer $\hat{\mathbf{w}}$ of the gridless formulations (4.60) and (4.61) which fulfills $\hat{L} = \text{rank}(\text{HToep}(\hat{\mathbf{w}})) < M$, such that a unique Vandermonde decomposition (4.59) is obtained. Appropriate regularization parameter selection schemes are given by (4.40) and (4.41).

Comparing the GL-SPARROW formulation (4.60) and the ANM problem (4.49), a similar structure in the objective functions and semidefinite constraints can be observed. In fact, both problems are equivalent as given by the following theorem:

Theorem 4.5. *The atomic norm minimization problem (4.48) and the corresponding SDP implementation (4.49), with minimizer $\hat{\mathbf{w}}_{\text{ANM}}$, are equivalent to the gridless SPARROW formulation (4.60), with minimizer $\hat{\mathbf{w}}_{\text{SPARROW}}$, in the sense that the minimizers are related by*

$$\hat{\mathbf{w}}_{\text{SPARROW}} = \hat{\mathbf{w}}_{\text{ANM}} / \sqrt{N}. \quad (4.63)$$

A proof of Theorem 4.5 is provided in Appendix D. For both problem formulations, GL-SPARROW (4.60) and ANM (4.49), the spatial frequencies $\boldsymbol{\nu}$ are encoded in the vectors $\hat{\mathbf{w}}_{\text{SPARROW}}$ and $\hat{\mathbf{w}}_{\text{ANM}}$, as found by Vandermonde decomposition (4.23), such that both formulations provide the same frequency estimates.

However, from a computational viewpoint, in contrast to the GL-SPARROW problem in (4.60), the ANM problem in (4.49) has additional MN complex-valued variables in the matrix \mathbf{Y}_0 , which need to be matched to the multi snapshot matrix \mathbf{Y} by an additional quadratic term in the objective function. The dimensionality reduction techniques for ANM, discussed in Section 4.2.2, can similarly be applied to SPARROW. Due to the reduced number of variables, the GL-SPARROW formulations (4.60) and (4.61) are less demanding in the number of computational operations and memory requirements and thus admit significantly reduced computational cost as compared to the ANM formulation (4.49).

The case of a thinned linear array of M sensors, discussed in Section 2.1.2, can be treated in the GL-SPARROW formulation by application of a proper selection matrix $\check{\mathbf{J}}$. To see this, let $\check{\mathbf{J}}$ be the $M_0 \times M$ matrix selecting the sensors in the thinned linear array from a virtual ULA of M_0 sensors. Application in the GL-SPARROW formulation (4.61) yields

$$\min_{\mathbf{w}, \mathbf{W}_M} \text{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \frac{1}{M_0} \text{Tr}(\text{HToep}(\mathbf{w})) \quad (4.64a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I}_M \\ \mathbf{I}_M & \check{\mathbf{J}}^T \text{HToep}(\mathbf{w}) \check{\mathbf{J}} + \lambda \mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (4.64b)$$

$$\text{HToep}(\mathbf{w}) \succeq \mathbf{0}, \quad (4.64c)$$

where \mathbf{w} is an M_0 element vector. A more detailed analysis of the gridless SPARROW formulation for thinned linear arrays is provided in [SSSP17b].

Frequency Estimation from the Dual Problem

The gridless SPARROW formulation in (4.58) provides a simple approach of removing the grid-based sensing matrix \mathbf{A} from the optimization problem to obtain high resolution frequency estimates. To gain further insight to the estimation problem, this

section considers an alternative approach of frequency estimation. To this end, consider the Lagrange dual problem of the SPARROW formulation (4.57), which is given by

$$\max_{\mathbf{r}_0, \mathbf{r}_1} -2 \operatorname{Re}\{\operatorname{Tr}(\mathbf{Y}_1)\} - \lambda \operatorname{Tr}(\mathbf{Y}_0) \quad (4.65a)$$

$$\text{s.t.} \quad \begin{bmatrix} \hat{\mathbf{R}} & \mathbf{r}_1 \\ \mathbf{r}_1^H & \mathbf{r}_0 \end{bmatrix} \succeq \mathbf{0} \quad (4.65b)$$

$$1 - \mathbf{a}_k^H \mathbf{Y}_0 \mathbf{a}_k \geq 0 \quad \text{for } k = 1, \dots, K, \quad (4.65c)$$

where \mathbf{Y}_0 is an $M \times M$ Hermitian, positive semidefinite matrix, while \mathbf{Y}_1 is an $M \times M$ complex-valued matrix without specific structure (see Appendix E for details). Note that the primal problem in (4.57) as well as the dual problem in (4.65) are both strictly feasible, such that strong duality applies [VB96]. The constraint (4.65c) forms the dual to the nonnegativity constraint on the elements of \mathbf{S} in the primal problem in (4.57), such that by complementary slackness any dual optimal variable $\hat{\mathbf{Y}}_0$ must fulfill

$$1 - \mathbf{a}_k^H \hat{\mathbf{Y}}_0 \mathbf{a}_k \begin{cases} = 0 & \text{if } \hat{s}_k \geq 0 \\ \geq 0 & \text{if } \hat{s}_k = 0, \end{cases} \quad (4.66)$$

for $k = 1, \dots, K$. Equation (4.66) indicates that the spatial frequency estimates $\{\hat{\mu}_l\}_{l=1}^{\hat{L}}$, as defined in (4.34), can equivalently be identified by the elements ν_k satisfying $1 - \mathbf{a}_k^H \mathbf{Y}_0 \mathbf{a}_k = 0$, i.e.,

$$\{\hat{\mu}_l\}_{l=1}^{\hat{L}} = \{\nu_k \mid 1 - \mathbf{a}_k^H \hat{\mathbf{Y}}_0 \mathbf{a}_k = 0, k = 1, \dots, K\}. \quad (4.67)$$

This leads to the conclusion that, instead of solving the primal problem (4.57) and perform frequency estimation according to equation (4.34), the dual problem (4.65) can be solved to obtain frequency estimates according (4.67). Note that similar observations have been made in the context of minimization of the total variation norm for gridless sparse recovery [CFG13, CFG14].

Consider the special case of a uniform linear array with sensor positions $r_m = (m-1) \Delta$, for $m = 1, \dots, M$, and move from a finite set of K steering vectors $\mathbf{a}(\nu_1), \dots, \mathbf{a}(\nu_K)$ to the array response vector

$$\mathbf{a}(z) = [1, z, \dots, z^{M-1}]^T \quad (4.68)$$

with $z = e^{-j\pi\mu\Delta}$, as discussed in Section 2.1.1. This transforms the finite set of constraints (4.65c) into an infinite-dimensional constraint, given as

$$1 - \mathbf{a}^H(z) \mathbf{Y}_0 \mathbf{a}(z) = 1 - \sum_{m=-M+1}^{M-1} c_m z^{-m} \geq 0. \quad (4.69)$$

Equation (4.69) constitutes a nonnegative trigonometric polynomial on the unit circle $|z| = 1$ with coefficients

$$c_m = \operatorname{Tr}(\Theta_m \mathbf{Y}_0) \quad (4.70)$$

given as the sum of the main-diagonal and off-diagonal elements of \mathbf{Y}_0 , with $\mathbf{\Theta}_m$ being the elementary Toeplitz matrix with ones on the m th diagonal and zeros elsewhere, as introduced in Section 2.1.

Define a second polynomial equation

$$1 = \sum_{m=-M+1}^{M-1} d_m z^{-m} = \mathbf{a}^H(z) \mathbf{H} \mathbf{a}(z) \quad (4.71)$$

for some Hermitian positive semidefinite matrix $\mathbf{H} \succeq \mathbf{0}$, with coefficients d_m fulfilling

$$d_m = \text{Tr}(\mathbf{\Theta}_m \mathbf{H}) = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{otherwise,} \end{cases} \quad (4.72)$$

i.e., the main-diagonal elements of \mathbf{H} must sum to 1 while the elements on the off-diagonals of \mathbf{H} must sum to 0, such that (4.71) always holds true on the unit circle $|z| = 1$. The polynomial in (4.71) is employed to provide an upper bound on the polynomial in (4.69) as

$$1 - \mathbf{a}^H(z) \mathbf{Y}_0 \mathbf{a}(z) = \mathbf{a}^H(z) (\mathbf{H} - \mathbf{Y}_0) \mathbf{a}(z) \geq 0, \quad (4.73)$$

which holds true for $\mathbf{H} - \mathbf{Y}_0 \succeq \mathbf{0}$. Equation (4.73) forms the infinite-dimensional analogue to the finite-dimensional constraint (4.65c). Using (4.73) and the conditions on \mathbf{H} in (4.72), the gridless version of the dual problem (4.65) can be formulated as

$$\max_{\mathbf{r}_0, \mathbf{Y}_1, \mathbf{H}} -2 \text{Re}\{\text{Tr}(\mathbf{Y}_1)\} - \lambda \text{Tr}(\mathbf{Y}_0) \quad (4.74a)$$

$$\text{s.t.} \quad \begin{bmatrix} \hat{\mathbf{R}} & \mathbf{Y}_1 \\ \mathbf{Y}_1^H & \mathbf{Y}_0 \end{bmatrix} \succeq \mathbf{0} \quad (4.74b)$$

$$\mathbf{H} - \mathbf{Y}_0 \succeq \mathbf{0} \quad (4.74c)$$

$$\text{Tr}(\mathbf{H}) = 1, \text{Tr}(\mathbf{\Theta}_m \mathbf{H}) = 0 \text{ for } m = 1, \dots, M. \quad (4.74d)$$

Given a minimizer $\hat{\mathbf{Y}}_0$ to the problem in (4.74), the spatial frequencies $\hat{\mu}_l$ can be estimated from the \hat{L} roots $\{\hat{z}_l\}_{l=1}^{\hat{L}}$ of the polynomial (4.69) on the unit circle, according to (4.67).

Frequency estimation from the dual problem of the SPARROW formulation shows strong links to the root-MUSIC approach discussed in Section 2.1. However, while in root-MUSIC the L roots closest to the unit circle are selected, SPARROW only considers the roots on the unit circle for frequency estimation. For a proper selection of the regularization parameter λ , the existence of roots on the unit circle is ensured by complementary slackness according to (4.66) and its continuous analogue in (4.69), i.e., roots will be located on the unit circle in the grid-based case as well as in the gridless case. This is a fundamental difference to the MUSIC approach, where no such guarantee of roots on the unit circle exists, neither for the gridless root-MUSIC nor for grid-based MUSIC method. Instead, the grid-based MUSIC approach searches the

minima on the unit circle while root-MUSIC finds the roots in the entire complex plane, as discussed in Section 3.2. This leads to a degradation in resolution performance of MUSIC as compared to root-MUSIC. As will be shown by numerical experiments in Section 4.4, no such degradation occurs for grid-based SPARROW, as compared to GL-SPARROW.

4.3.2 Off-Grid SPARROW

The gridless SPARROW formulations are an attractive approach for parameter estimation in uniform linear arrays. However, the gridless SPARROW formulation cannot be directly applied for arbitrary array topologies, such that the grid-based SPARROW formulation has to be used instead. In recent literature [HS10,ZLG11,TYN14,IRA⁺14], it has been proposed to model the basis mismatch error, resulting from the finite grid, by linear interpolation of the array response in the off-grid zones, which can similarly be applied to the grid-based SPARROW formulation to formulate an off-grid method for joint sparse recovery from multiple snapshots in arbitrary array topologies. To reach this goal, a first-order Taylor expansion of the sensing matrix \mathbf{A} , as used in the SPARROW formulation (4.51), is performed. The solution of the new SPARROW problem provides the spatial frequencies on the grid and the corresponding off-grid error to obtain continuous frequency estimates.

In this section, the on-grid assumption (4.3) is relaxed and it is assumed that

$$\{\mu_l\}_{l=1}^L \not\subseteq \{\nu_k\}_{k=1}^K, \quad (4.75)$$

such that the spatial frequency μ_l of source l admits an off-grid error e_k from its closest frequency grid point ν_k . The magnitude of the off-grid error e_k is restricted to be less than or equal to half of the grid spacing, given by $\delta/2$, where a uniform frequency grid is assumed, according to

$$\nu_k = k\delta \quad \text{for } k = 1, \dots, K. \quad (4.76)$$

Hence the spatial frequency μ_l of source l is given by

$$\mu_l = \nu_k + e_k, \quad (4.77)$$

where ν_k is the grid-frequency closest to μ_l . The off-grid errors e_k , for $k = 1, \dots, K$, are assumed to be small such that the array response can be modeled by a first-order Taylor series expansion of the array response vector according to

$$\mathbf{a}(\nu_k + e_k) = \mathbf{a}(\nu_k) + e_k \frac{\partial \mathbf{a}(\nu_k)}{\partial \nu_k} + O(e_k^2). \quad (4.78)$$

In the SPARROW formulation (4.51), the array response vectors \mathbf{a}_k are reflected in the matrix product

$$\mathbf{A}(\boldsymbol{\nu} + \mathbf{e})\mathbf{S}\mathbf{A}^H(\boldsymbol{\nu} + \mathbf{e}) = \sum_{k=1}^K s_k \mathbf{a}(\nu_k + e_k) \mathbf{a}^H(\nu_k + e_k), \quad (4.79)$$

where $\mathbf{e} = [e_1, \dots, e_K]^\top$. Application of the first-order Taylor series expansion (4.78) to the outer product of the response vectors in (4.79) yields

$$\begin{aligned} \mathbf{a}(\nu_k + e_k)\mathbf{a}^H(\nu_k + e_k) &\approx \mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k) + e_k \frac{\partial}{\partial \nu_k} (\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)) \\ &= \mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k) + e_k \mathbf{\Delta}_0 \odot (\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)), \end{aligned} \quad (4.80)$$

where \odot denotes the element-wise matrix multiplication. The $M \times M$ matrix $\mathbf{\Delta}_0$ stems from derivation of the product of array response vectors and contains the differences of the sensor positions r_1, \dots, r_M as follows

$$\mathbf{\Delta}_0 = -j\pi \begin{bmatrix} 0 & r_1 - r_2 & \cdots & r_1 - r_M \\ r_2 - r_1 & 0 & & \\ \vdots & & \ddots & \\ r_M - r_1 & & & 0 \end{bmatrix}. \quad (4.81)$$

Substituting (4.80) in (4.79) results in

$$\begin{aligned} \mathbf{A}(\boldsymbol{\nu} + \mathbf{e})\mathbf{S}\mathbf{A}^H(\boldsymbol{\nu} + \mathbf{e}) &\approx \sum_{k=1}^K s_k \left(\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k) + e_k \mathbf{\Delta}_0 \odot (\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)) \right) \\ &= \sum_{k=1}^K s_k \mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k) + \varepsilon_k \mathbf{\Delta}_0 \odot (\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)), \end{aligned} \quad (4.82)$$

where the variable

$$\varepsilon_k = s_k e_k, \quad (4.83)$$

for $k = 1, \dots, K$, is introduced with the constraint

$$|\varepsilon_k| \leq s_k \delta / 2, \quad (4.84)$$

where δ denotes the uniform grid spacing according to (4.76). As e_k can be negative, the variables ε_k , for $k = 1, \dots, K$, can also take negative values. Define the $K \times K$ real diagonal matrix $\mathbf{E} = \text{diag}(\varepsilon_1, \dots, \varepsilon_K)$ to rewrite (4.82) in compact notation as

$$\mathbf{A}(\boldsymbol{\nu} + \mathbf{e})\mathbf{S}\mathbf{A}^H(\boldsymbol{\nu} + \mathbf{e}) \approx \mathbf{A}\mathbf{S}\mathbf{A}^H + \mathbf{\Delta}_0 \odot (\mathbf{A}\mathbf{E}\mathbf{A}^H), \quad (4.85)$$

with the dictionary matrix $\mathbf{A} = [\mathbf{a}(\nu_1), \dots, \mathbf{a}(\nu_K)]$ as defined below (4.2). Applying the linear approximation (4.85) in the SPARROW formulation (4.51) results in the off-grid SPARROW problem

$$\min_{\mathbf{S} \in \mathcal{D}_+^K, \mathbf{E} \in \mathcal{D}^K} \text{Tr}((\mathbf{A}\mathbf{S}\mathbf{A}^H + \mathbf{\Delta}_0 \odot (\mathbf{A}\mathbf{E}\mathbf{A}^H) + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \quad (4.86a)$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{S}\mathbf{A}^H + \mathbf{\Delta}_0 \odot (\mathbf{A}\mathbf{E}\mathbf{A}^H) \succeq \mathbf{0} \quad (4.86b)$$

$$|\varepsilon_k| \leq s_k \delta / 2 \quad \text{for } k = 1, \dots, K, \quad (4.86c)$$

where \mathcal{D}^K denotes the set of the $K \times K$ real-valued diagonal matrices. The additional $K \times K$ real-valued diagonal matrix variable $\mathbf{E} = \text{diag}(\varepsilon_1, \dots, \varepsilon_K)$, with $\varepsilon_k = s_k e_k$,

for $k = 1, \dots, K$, as defined in (4.83), takes account of the off-grid parameter e_k , for $k = 1, \dots, K$, as introduced in (4.77), where the constraint (4.86c) follows from (4.84). The additional constraint (4.86b) results from the observation that the matrix product $\mathbf{A}(\boldsymbol{\nu} + \mathbf{e})\mathbf{S}\mathbf{A}^H(\boldsymbol{\nu} + \mathbf{e})$ and its approximation in (4.85) need to be positive semidefinite. Problem (4.86) will be referred to as OG1-SPARROW and can be implemented by semidefinite programming, as discussed for Corollary 4.3.

The off-grid frequency estimates $\hat{\mu}_l$, for $l = 1, \dots, \hat{L}$, are computed from the minimizers $\hat{\mathbf{S}} = \text{diag}(\hat{s}_1, \dots, \hat{s}_K)$ and $\hat{\mathbf{E}} = \text{diag}(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_K)$ of (4.86), according to

$$\{\hat{\mu}_l\}_{l=1}^{\hat{L}} = \{\nu_k + \hat{e}_k \mid \hat{s}_k \neq 0, \hat{e}_k = \hat{\varepsilon}_k / \hat{s}_k, k = 1, \dots, K\}, \quad (4.87)$$

where in practice only the local maxima of the spectrum $\hat{s}_1, \dots, \hat{s}_K$ are selected as frequencies of interest. As demonstrated by numerical experiments in Section 4.4.5, OG1-SPARROW provides improved estimation performance as compared to the standard grid-based SPARROW formulation (4.51). However, large errors ε_k , for $k = 1, \dots, K$, are not penalized as the parameters ε_k are not directly reflected in the objective function (4.86a). In the following, the OG1-SPARROW is modified by inserting a penalty term on the errors ε_k , for $k = 1, \dots, K$, in the objective function by means of the $\ell_{2,1}$ mixed-norm. A similar approach based on joint sparsity has been proposed in the context of covariance fitting [TYN14]. Following this approach to enforce joint sparsity in \mathbf{S} and \mathbf{E} by means of an $\ell_{2,1}$ mixed-norm, a new problem formulation is given as

$$\min_{\mathbf{S} \in \mathcal{D}_+^K, \mathbf{E} \in \mathcal{D}^K} \text{Tr}((\mathbf{A}\mathbf{S}\mathbf{A}^H + \boldsymbol{\Delta}_0 \odot (\mathbf{A}\mathbf{E}\mathbf{A}^H) + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \|\mathbf{W}_{\text{SE}}\|_{2,1} \quad (4.88a)$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{S}\mathbf{A}^H + \boldsymbol{\Delta}_0 \odot (\mathbf{A}\mathbf{E}\mathbf{A}^H) \succeq \mathbf{0} \quad (4.88b)$$

$$\mathbf{W}_{\text{SE}} = [\text{vecd}(\mathbf{S}), \text{vecd}(\mathbf{E})] \quad (4.88c)$$

$$|\varepsilon_k| \leq s_k \delta / 2, \text{ for } k = 1, \dots, K, \quad (4.88d)$$

where $\text{vecd}(\mathbf{S})$ denotes the vector containing the diagonal elements of \mathbf{S} , and the $\ell_{2,1}$ mixed-norm of the matrix $\mathbf{W}_{\text{SE}} \in \mathbb{R}^{K \times 2}$ is computed as

$$\|\mathbf{W}_{\text{SE}}\|_{2,1} = \sum_{k=1}^K \sqrt{s_k^2 + \varepsilon_k^2}. \quad (4.89)$$

Problem (4.88) is referred to as OG2-SPARROW, and can be implemented by semidefinite programming, as discussed for Corollary 4.3. As shown by simulations in Section 4.4, the introduction of the variables $\varepsilon_1, \dots, \varepsilon_K$ in the objective function (4.88) improves the estimation of the off-grid error \mathbf{e} as compared to the OG1-SPARROW approach in (4.86).

4.3.3 Related Work: SPICE

Numerous methods for SSR have been proposed in recent years and links between the various methods have been established to provide a better understanding of SSR.

The SParse Iterative Covariance-based Estimation (SPICE) method [SBL11a, SBL11b, SZL14] has been derived in the context of covariance matching and was shown to have interesting links to $\ell_{2,1}$ mixed-norm minimization. Based on the SPARROW formulation (4.51), these links can be further extended, as discussed in the following.

The SPICE method seeks to match the sample covariance matrix $\hat{\mathbf{R}} = \mathbf{Y}\mathbf{Y}^H/N$ with a sparse representation of the covariance matrix \mathbf{R}_0 . To this end, consider the covariance matrix

$$\mathbf{R} = \mathbb{E}\{\mathbf{Y}\mathbf{Y}^H\}/N = \mathbf{A}(\boldsymbol{\mu})\boldsymbol{\Pi}\mathbf{A}^H(\boldsymbol{\mu}) + \sigma^2\mathbf{I}_M, \quad (4.90)$$

as given in (3.1). In contrast to the previous considerations on SSR in this chapter, the authors in [SBL11a, SBL11b, SZL14] explicitly assume that the source signals, represented in the matrix $\boldsymbol{\Psi}$, are uncorrelated, such that the source covariance matrix

$$\boldsymbol{\Pi} = \mathbb{E}\{\boldsymbol{\Psi}\boldsymbol{\Psi}^H\}/N \quad (4.91)$$

has a diagonal structure, i.e., $\boldsymbol{\Pi} = \text{diag}(\pi_1, \dots, \pi_L)$. The sparse representation \mathbf{R}_0 of the covariance matrix in (4.90) is introduced as

$$\mathbf{R}_0 = \mathbf{A}\check{\mathbf{P}}\mathbf{A}^H + \varsigma\mathbf{I}_M, \quad (4.92)$$

where \mathbf{A} denotes the $M \times K$ sensing matrix computed for a fixed grid of frequencies ν_1, \dots, ν_K , as introduced in (4.2), $\varsigma = \sigma^2$ denotes the noise power. Under the on-grid assumption (4.3), the elements of the sparse diagonal source covariance matrix $\check{\mathbf{P}} = \text{diag}(\check{p}_1, \dots, \check{p}_K) \in \mathcal{D}_+^K$ are given as

$$\check{p}_k = \begin{cases} \pi_l & \text{if } \nu_k = \mu_l \\ 0 & \text{otherwise,} \end{cases} \quad (4.93)$$

for $k = 1, \dots, K$ and $l = 1, \dots, L$, with π_l denoting the diagonal elements of the source covariance matrix as defined (4.91).

Two types of weighted covariance matching functions have been proposed in [SBL11a, SBL11b, SZL14] for the SPICE method. The undersampled case, with $N < M$, is treated by minimization of a weighted covariance matching function according to

$$\begin{aligned} & \min_{\mathbf{P} \in \mathcal{D}_+^K, \varsigma \geq 0} \left\{ \|\mathbf{R}_0^{-1/2}(\hat{\mathbf{R}} - \mathbf{R}_0)\|_{\text{F}}^2 \quad \text{s.t. } \mathbf{R}_0 = \mathbf{A}\mathbf{P}\mathbf{A}^H + \varsigma\mathbf{I}_M \right\} \\ & = \min_{\mathbf{P} \in \mathcal{D}_+^K, \varsigma \geq 0} \left\{ \text{Tr}(\mathbf{R}_0^{-1}\hat{\mathbf{R}}^2) + \text{Tr}(\mathbf{R}_0) - 2\text{Tr}(\hat{\mathbf{R}}) \quad \text{s.t. } \mathbf{R}_0 = \mathbf{A}\mathbf{P}\mathbf{A}^H + \varsigma\mathbf{I}_M \right\}, \quad (4.94) \end{aligned}$$

where sparsity in \mathbf{P} is induced in the objective of (4.94) in form of the trace penalty term $\text{Tr}(\mathbf{R}_0)$, as can be observed from the identity

$$\text{Tr}(\mathbf{R}_0) = \varsigma M + \sum_{k=1}^K \|\mathbf{a}_k\|_2^2 \cdot p_k = M(\varsigma + \sum_{k=1}^K p_k). \quad (4.95)$$

The oversampled case, with $N \geq M$, where the sample covariance matrix $\hat{\mathbf{R}}$ is non-singular, is treated by the minimization of the weighted covariance matching function according to

$$\begin{aligned} & \min_{\mathbf{P} \in \mathcal{D}_+^K, \varsigma \geq 0} \left\{ \|\mathbf{R}_0^{-1/2} (\hat{\mathbf{R}} - \mathbf{R}_0) \hat{\mathbf{R}}^{-1/2}\|_{\text{F}}^2 \quad \text{s.t. } \mathbf{R}_0 = \mathbf{A} \mathbf{P} \mathbf{A}^{\text{H}} + \varsigma \mathbf{I}_M \right\} \\ & = \min_{\mathbf{P} \in \mathcal{D}_+^K, \varsigma \geq 0} \left\{ \text{Tr}(\mathbf{R}_0^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{R}_0 \hat{\mathbf{R}}^{-1}) - 2M \quad \text{s.t. } \mathbf{R}_0 = \mathbf{A} \mathbf{P} \mathbf{A}^{\text{H}} + \varsigma \mathbf{I}_M \right\}, \end{aligned} \quad (4.96)$$

where sparsity in \mathbf{P} is induced by summation of its diagonal elements with data dependent, nonnegative weights according to

$$\text{Tr}(\mathbf{R}_0 \hat{\mathbf{R}}^{-1}) = \varsigma \text{Tr}(\hat{\mathbf{R}}^{-1}) + \sum_{k=1}^K \mathbf{a}_k^{\text{H}} \hat{\mathbf{R}}^{-1} \mathbf{a}_k \cdot p_k. \quad (4.97)$$

Note that the SPARROW formulation in (4.51) exhibits similarities with both SPICE formulations (4.94) and (4.96). While the SPARROW formulation shares the uniformly weighted summation of its variables in $\text{Tr}(\mathbf{S})$ with the SPICE formulation in (4.94), it shares the structure of the data fitting function $\text{Tr}((\mathbf{A} \mathbf{S} \mathbf{A}^{\text{H}} + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}})$ with the SPICE formulation in (4.96). There is, however, a fundamental difference between the SPARROW formulation and the SPICE formulations in the fact that the variables in \mathbf{S} correspond to the normalized row-norms of the signal matrix, i.e., $\hat{s}_k = \frac{1}{\sqrt{N}} \|\hat{\mathbf{x}}_k\|_2$, for $k = 1, \dots, K$, as seen from (4.53), while the variables in \mathbf{P} correspond to the signal powers, i.e., $\hat{p}_k = \frac{1}{N} \text{E}\{\|\hat{\mathbf{x}}_k\|_2^2\}$, for $k = 1, \dots, K$, as seen from (4.91) and (4.93).

Moreover, the SPICE formulations make assumptions on the second-order signal statistics in form of the covariance matrix in (4.92), namely, the sparse source covariance matrix \mathbf{P} is modeled as a diagonal matrix, which involves the assumption of uncorrelated source signals. In contrast to that, the SPARROW problem in (4.51) relies on a deterministic signal model with stochastic noise and does not make any assumptions on the statistics of the signals. Regardless of these model assumptions, the SPICE method is known to show good estimation performance even in the case of correlated source signals.

In [RKH13, BS14] it has been shown, that for the case of a single snapshot $\mathbf{y}(t)$, the SPICE problem in (4.94) is equivalent to the Square-Root LASSO (SR-LASSO) [BCW11]

$$\min_{\mathbf{x}} \|\mathbf{A} \mathbf{x} - \mathbf{y}(t)\|_2 + \|\mathbf{x}\|_1, \quad (4.98)$$

as briefly introduced in Section 4.1.1, where equivalence holds in the sense that the corresponding minimizers are related by

$$\hat{\mathbf{x}} = \hat{\mathbf{P}} \mathbf{A}^{\text{H}} (\mathbf{A} \hat{\mathbf{P}} \mathbf{A}^{\text{H}} + \hat{\varsigma} \mathbf{I}_M)^{-1} \mathbf{y}(t) \quad \text{and} \quad \hat{p}_k = \frac{|\hat{x}_k| \|\mathbf{y}(t)\|_2}{\sqrt{M}}. \quad (4.99)$$

Similarly, it was shown in [SBL11b] that the SPICE formulation in (4.96) is equivalent to a weighted SR-LASSO formulation. Note that the line of arguments used in [SBL11b, RKH13, BS14] to prove the above equivalence is rather different from those used in the proof of Theorem 4.2. Furthermore, there are some significant differences between the SR-LASSO formulation (4.98) and the standard mixed-norm formulation (4.50) considered here. The latter reduces to the popular standard LASSO (4.13) in the special case of a single measurement vector. As compared to the SR-LASSO (4.98), the standard LASSO (4.13) has a squared data fitting term, such that the standard LASSO admits an interpretation as a Bayesian estimator with Laplacian priors under additive white Gaussian noise [Tib96], [PC08], as briefly discussed in Section 4.1.1. Equivalence of the standard LASSO and the SR-LASSO only holds in the noise-free case, such that in this case the SPICE formulation in (4.94) is equivalent to standard ℓ_1 norm minimization. In contrast to that, the SPARROW formulation is equivalent to the standard $\ell_{2,1}$ minimization problem (4.50) in the general and practically relevant case of noise-corrupted measurements.

Another major difference of the $\ell_{2,1}$ minimization problem in (4.50) and the SR-LASSO formulation in (4.98) lies in the absence of the regularization parameter λ in the latter approach. The mixed-norm problem (4.50) admits to obtain a solution of any desired sparsity level by tuning the regularization parameter λ , e.g., by exploiting a-priori knowledge or by applying blind techniques such as the cross-validation approach of [Tib96]. The SR-LASSO in (4.98) does not have such a regularization parameter and thus provides less flexibility in the solution. On the other hand, since the selection of the regularization parameter can be quite challenging in practice, this makes the SR-LASSO, and correspondingly the SPICE method, easily applicable in practical scenarios [SBL11a, SBL11b, SB12, SZL14].

A drawback of the SPICE method, which has not yet been reported in literature, is that the parameterization $\mathbf{R}_0 = \mathbf{A}\mathbf{P}\mathbf{A}^H + \varsigma\mathbf{I}_M$ might not be unique, i.e., depending on the array topology and the frequency grid, as reflected in the sensing matrix \mathbf{A} , there might exist a matrix $\tilde{\mathbf{P}}$ such that $\mathbf{A}\tilde{\mathbf{P}}\mathbf{A}^H = \mathbf{I}_M$ and

$$\mathbf{R}_0 = \mathbf{A}\mathbf{P}\mathbf{A}^H + \varsigma\mathbf{I}_M = \mathbf{A}(\mathbf{P} + \varsigma\tilde{\mathbf{P}})\mathbf{A}^H. \quad (4.100)$$

For example, given a ULA of M sensors, the identity matrix \mathbf{I}_M can be created by a convex combination of M equispaced atoms $\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)$, with $\nu_k = 2(k-1)/M - 1$ for $k = 1, \dots, M$, such that for $\tilde{\mathbf{P}} = 1/M\mathbf{I}_M$ the identity $\mathbf{A}\tilde{\mathbf{P}}\mathbf{A}^H = \mathbf{I}_M$ holds true. Similarly, for a thinned array of M sensors with maximum lag ΔM_0 , the identity matrix \mathbf{I}_M can be created by convex combination of M_0 equispaced atoms $\mathbf{a}(\nu_k)\mathbf{a}^H(\nu_k)$, with $\nu_k = 2(k-1)/M_0 - 1$ for $k = 1, \dots, M_0$, and $\tilde{\mathbf{P}} = 1/M_0\mathbf{I}_M$ [Fuc97]. Possible realizations of $\tilde{\mathbf{P}}$ and $\mathbf{A}(\boldsymbol{\nu})$ are not limited to the above examples, but these simple considerations reveal a general shortcoming of the SPICE method in that it cannot necessarily guarantee a unique sparse solution, since there might exist a non-sparse matrix $\tilde{\mathbf{P}}$ to any sparse solution to fulfill (4.100). A simple approach to obtain a unique solution to the SPICE method, which has not been published in previous literature, is to solve the corresponding optimization problem twice. The first run serves to compute

the estimate of the covariance matrix \mathbf{R}_0 and to find its smallest eigenvalue $\tilde{\varsigma}$. In the second run, the parameter ς is held constant in the SPICE optimization problem as $\varsigma = \tilde{\varsigma}$, leading to a sparse estimate of the source covariance \mathbf{P} . Indirectly, a similar approach is performed in the context of the gridless SPICE method, where a different parameterization of the sparse representation of the covariance matrix \mathbf{R}_0 is applied, as discussed in the following section. Note that the uniqueness of the solution for $\ell_{2,1}$ minimization (4.50) and SPARROW (4.51) depends on the choice of regularization parameter λ . For too small values of the regularization parameter λ , the solution to SPARROW might not be sparse and unique. However, such problems can be fixed by increasing the value of the regularization parameter λ .

Gridless SPICE

A gridless extension of SPICE to the GridLess Spice (GLS) method for ULAs was proposed in [YX15b], which relies on an SDP formulation of the SPICE problems (4.94) and (4.96), and Vandermonde decomposition of Hermitian Toeplitz matrices as given in Definition 4.1, similar to the gridless SPARROW and ANM problems discussed in Sections 4.3.1 and 4.1.2. In the case of ULAs, the sparse representation of the covariance matrix in (4.92) has a Hermitian Toeplitz structure $\mathbf{R}_0 = \text{HToep}(\mathbf{w})$, such that the undersampled SPICE problem (4.94) can be formulated as

$$\min_{\mathbf{w}} \left\{ \text{Tr}(\mathbf{R}_0^{-1} \hat{\mathbf{R}}^2) - 2\text{Tr}(\hat{\mathbf{R}}) \quad \text{s.t. } \mathbf{R}_0 = \text{HToep}(\mathbf{w}) \succeq \mathbf{0} \right\} \quad (4.101)$$

and, correspondingly, the oversampled SPICE problem (4.96) can be formulated as

$$\min_{\mathbf{w}} \left\{ \text{Tr}(\mathbf{R}_0^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{R}_0 \hat{\mathbf{R}}^{-1}) \quad \text{s.t. } \mathbf{R}_0 = \text{HToep}(\mathbf{w}) \succeq \mathbf{0} \right\}. \quad (4.102)$$

Given a minimizer $\hat{\mathbf{R}}_0$ to any of the two problems, the authors in [YX15b] suggest to first compute the minimum eigenvalue $\hat{\varsigma}$ of $\hat{\mathbf{R}}_0$ and then perform Vandermonde decomposition of $\hat{\mathbf{R}}_0 - \hat{\varsigma}\mathbf{I}$ to recover the frequency estimates $\{\hat{\mu}_l\}_{l=1}^L$ and the corresponding source covariances $\{\hat{p}_l\}_{l=1}^L$. In contrast to the grid-based SPICE method, the approach in [YX15b] provides a unique parameterization of $\hat{\mathbf{R}}_0$, since the Vandermonde decomposition of the rank-deficient matrix $\hat{\mathbf{R}}_0 - \hat{\varsigma}\mathbf{I}$ is unique. As shown by numerical experiments in Section 4.4, this leads to different estimates of grid-based and gridless SPICE.

In [YX16b] it has been shown that in the noise-free case, GLS can be interpreted as special versions of the ANM problem (4.47). In contrast to the results in [YX15b, YX16b], the results of equivalence between gridless SPARROW and ANM, in Section 4.3.1 of this thesis, hold in the more general case with an additional data matching term in the ANM formulation to account for noise-corrupted measurements according to (4.49).

4.4 Numerical Experiments

The parameter estimation performance of $\ell_{2,1}$ minimization, ANM and SPICE has been numerically investigated in various publications, e.g., [MÇW05, HM10, SBL11a, SBL11b, SZL14, YX15b, LC16, YX16a]. This section extends the existing simulation results by a numerical analysis of some less investigated aspects, e.g., the estimation bias, standard deviation and root-mean-square error. The experiments compare grid-based SPARROW (4.54),(4.57), gridless SPARROW (4.60),(4.61) (referred to as GL-SPARROW), the off-grid SPARROW formulations (4.86) and (4.88), under- and oversampled SPICE, i.e., (4.94) and (4.96), (referred to as SPICE-US and SPICE-OS, respectively), under- and oversampled GridLess Spice (4.101),(4.102) (referred to as GLS-US and GLS-OS, respectively), spectral MUSIC [Sch86], root-MUSIC [KV96, vT02], and the stochastic Cramér-Rao Bound (CRB) [SLG01]. Since the SPARROW formulation is equivalent to $\ell_{2,1}$ minimization and ANM, as discussed in Section 4.3, the latter two methods are not included in the performance analysis. Instead, a comparison of computation time for the equivalent approaches is provided.

The SPARROW, SPICE and MUSIC methods all make different assumptions on the availability of a-priori knowledge. While SPICE does not require any a-priori knowledge, perfect knowledge of the noise power σ_N^2 is assumed for the regularization parameter selection of SPARROW as discussed in Section 4.2.1, and perfect knowledge of the number of source signals L is assumed for the MUSIC method. Since estimation of these parameters itself might affect the estimation performance of the MUSIC and SPARROW methods, the standard assumption of perfectly known number of source signals and noise power is applied and the achievable performance under these idealized assumptions is investigated.

4.4.1 Regularization Parameter Selection

Before turning to the frequency estimation performance, this section performs a brief investigation of the regularization parameter selection. To this end the two selection approaches (4.40) and (4.41) are compared, as briefly discussed in Section 4.2.1. The experimental setup assumes two equal power source signals with source covariance matrix $E\{\boldsymbol{\psi}(t)\boldsymbol{\psi}^H(t)\} = \mathbf{I}$ and spatial frequencies $\mu_1 = 0.5$ and $\mu_2 = 0.4$, a signal-to-noise ratio (SNR), defined as $\text{SNR} = 1/\sigma^2$, of 0 dB, and considers varying number of snapshots and sensors in the receiving ULA. Figure 4.4 shows the average number of estimated source signals \hat{L} over 100 iterations per scenario.

As seen from Figure 4.4, in the gross of the investigated scenarios both approaches correctly estimate the number of source signals. In Figure 4.4 a) it is observed that approach (4.41) tends to overestimate the number of estimated signals \hat{L} for low number of snapshots N , indicating that the regularization parameter is selected too low. For low number of sensors the two closely spaced signals can not always be resolved, leading to underestimation of the number of source signals. From Figure 4.4 b) it can be

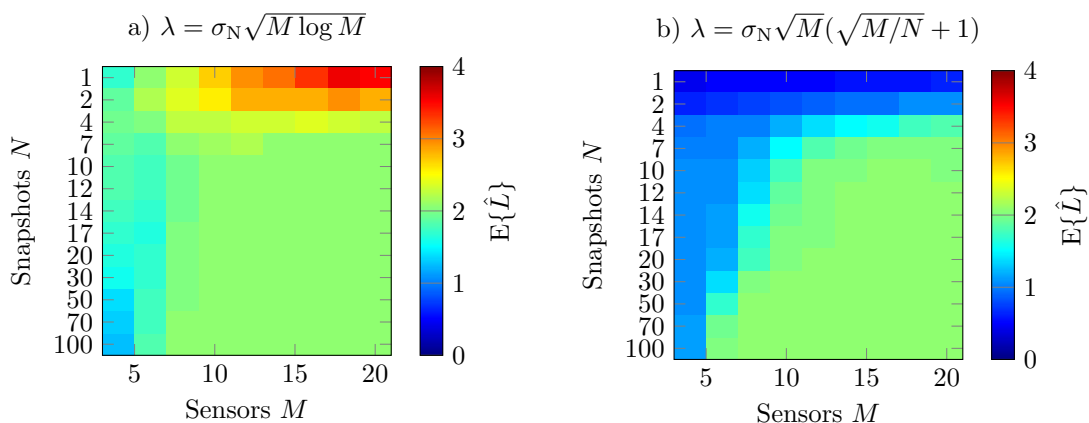


Figure 4.4: Average number \hat{L} of estimated source signals for regularization parameter selection approaches a) (4.41) and b) (4.40), over varying number of snapshots N and sensors M

noticed that approach (4.40) underestimates the number of sources for low number of snapshots, indicating that the regularization parameter is too large. In the case of a single snapshot, this regularization parameter selection approach actually estimates no signal at all. For low number of sensors, the two closely spaced sources again cannot be resolved. Since the following numerical experiments will focus on difficult scenarios with low number of sensors and snapshots, approach (4.41) will be used for the numerical simulations.

4.4.2 Resolution Performance and Estimation Bias

The setup of the following experiment consists of a uniform linear array of $M = 6$ sensors with half signal wavelength spacing, an SNR of 3 dB and a number of $N = 20$ snapshots. A number of $T = 1000$ Monte Carlo trials are performed. For each trial $L = 2$ independent complex Gaussian sources with static spatial frequencies are considered, where the first source signal has a spatial frequency of $\mu_1 = 0.5$, while the spatial frequency of the second source is varied according to $\mu_2 = \mu_1 - \Delta\mu$, with $0.01 \leq \Delta\mu \leq 1$. All grid-based estimation methods use a uniform grid of $K = 1000$ points, i.e., $\nu = [-1, -0.998, -0.996, \dots, 0.998]^T$.

Figure 4.5 shows the root-mean-square error (RMSE) of the schemes under investigation, which is computed according to

$$\text{RMSE}(\hat{\boldsymbol{\mu}}) = \sqrt{\frac{1}{LT} \sum_{t=1}^T \sum_{l=1}^L |\mu_l - \hat{\mu}_l(t)|_{\text{wa}}^2}, \quad (4.103)$$

with $\hat{\mu}_l(t)$ denoting the estimate of the l th frequency μ_l in Monte Carlo trial t and $|\hat{\mu}_1 - \hat{\mu}_2|_{\text{wa}} = \min_{i \in \mathbb{Z}} |\hat{\mu}_1 - \hat{\mu}_2 + 2i|$ denoting the wrap-around distance for frequencies

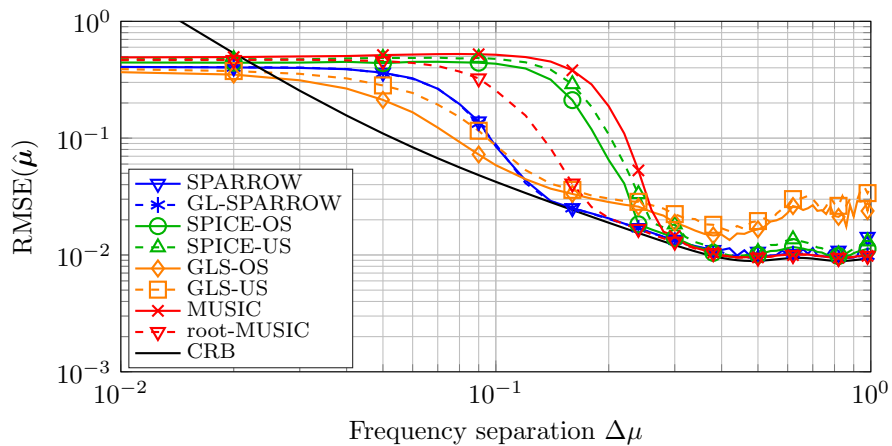


Figure 4.5: Frequency estimation performance for $L = 2$ source signals, $N = 20$ snapshots and $\text{SNR} = 3$ dB.

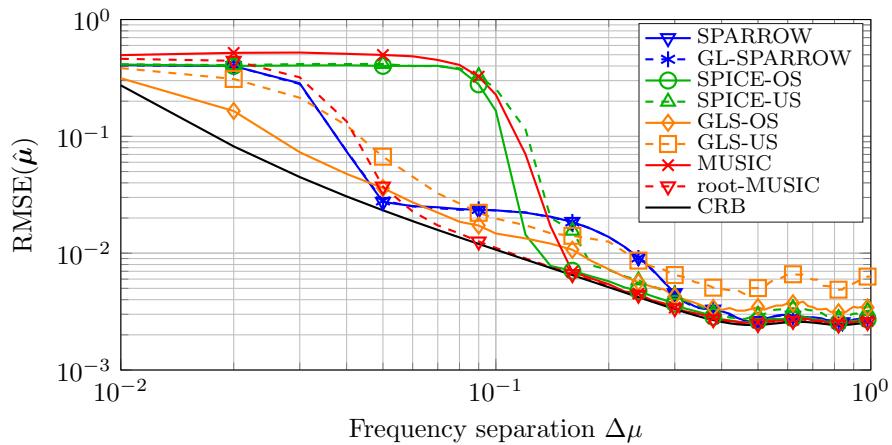


Figure 4.6: Frequency estimation performance for $L = 2$ source signals, $N = 50$ snapshots and $\text{SNR} = 10$ dB.

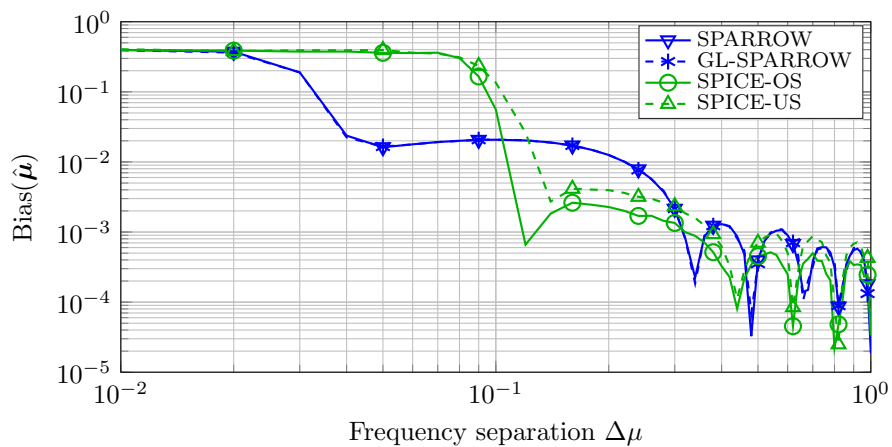


Figure 4.7: Bias of estimated frequencies for $L = 2$ source signals, $N = 50$ snapshots and $\text{SNR} = 10$ dB.

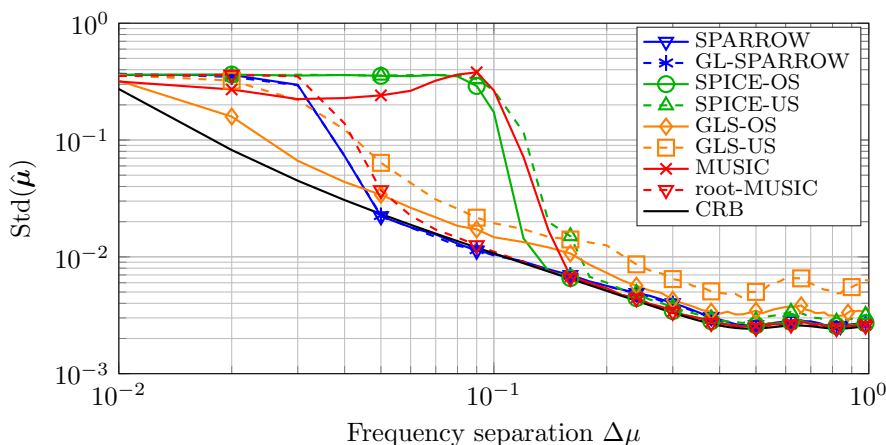


Figure 4.8: Standard deviation of estimated frequencies for $L = 2$ source signals, $N = 50$ snapshots and $\text{SNR} = 10$ dB.

$\hat{\mu}_1, \hat{\mu}_2 \in [-1, 1)$. Since the RMSE computation (4.103) requires the number of estimated source signals \hat{L} to be equal to the true number of source signals L , two special cases have to be considered: In the case of overestimation of the model order, $\hat{L} > L$, the L frequency estimates with the largest corresponding magnitudes are selected, whereas $L - \hat{L}$ additional random spatial frequencies are selected in the case of underestimation $\hat{L} < L$.

From Figure 4.5 it can be observed that the grid-based and gridless SPARROW versions show the same estimation performance. While both methods reach the CRB for high frequency separation, they also achieve a good threshold performance¹, close to that of the GLS methods and slightly better than the performance of root-MUSIC. The GLS methods show the best threshold performance but do not reach the CRB for high frequency separation. In contrast to the GLS, the grid-based SPICE methods show rather poor threshold performance, similar to that of grid-based MUSIC, and reach the CRB for high frequency separation. This difference in estimation performance of the grid-based and gridless SPICE methods can be explained by the different parameterization of \mathbf{R}_0 as discussed in Section 4.3.3. Similar differences in the estimation performance can be observed for the grid-based and gridless MUSIC methods, where the difference mainly lies in the threshold performance, with root-MUSIC showing an improved performance.

Figure 4.6 shows the RMSE performance for a less challenging scenario with $\text{SNR} = 10$ dB and $N = 50$ snapshots. As can be seen in the figure, the SPARROW methods again show good threshold performance, similar to root-MUSIC, but do not reach the CRB for frequency separations $0.05 \leq \Delta\mu \leq 0.3$. The RMSE performance of the

¹In the context of RMSE performance the threshold refers to the point where the RMSE curve deviates from the CRB, e.g., because two closely spaced sources cannot be resolved for low SNR or low number of snapshots. This region is also referred to as the non-informative region of the RMSE plot.

remaining schemes is comparable to the performance in the first scenario. The behavior of the SPARROW methods can be explained by considering the estimation bias, which is computed as

$$\text{Bias}(\hat{\boldsymbol{\mu}}) = \sqrt{\frac{1}{L} \sum_{l=1}^L (\mu_l - \text{Mean}(\hat{\mu}_l))^2}, \quad (4.104)$$

where the mean estimate for frequency μ_l is computed as

$$\text{Mean}(\hat{\mu}_l) = \frac{1}{T} \sum_{t=1}^T \hat{\mu}_l(t). \quad (4.105)$$

Furthermore, the standard deviation is computed as

$$\text{Std}(\hat{\boldsymbol{\mu}}) = \sqrt{\frac{1}{TL} \sum_{l=1}^L \sum_{t=1}^T |\text{Mean}(\hat{\mu}_l) - \hat{\mu}_l(t)|_{\text{wa}}^2}. \quad (4.106)$$

Figures 4.7 and 4.8 show the resulting estimation bias and standard deviation versus the frequency separation $\Delta\mu = |\mu_2 - \mu_1|_{\text{wa}}$. As can be observed from the figures, the proposed SPARROW method provides a relatively large bias in the case of closely spaced frequencies, with $\Delta\mu \leq 0.33$, but provides source resolution performance, i.e., threshold performance, slightly superior to that of root-MUSIC, with successful source resolution for $\Delta\mu \geq 0.05$. For frequency separation $\Delta\mu \geq 0.33$ the estimation bias reduces significantly and becomes negligible with respect to the standard deviation.

Similar to GL-SPARROW, the grid-based SPICE methods exhibit an estimation bias for closely spaced source signals with $\Delta\mu < 0.44$, however, provide a degraded source resolution performance for $\Delta < 0.14$, similar to spectral MUSIC. In contrast to that, the two GLS versions display negligible estimation bias (not shown here), while achieving a reduced estimation performance in terms of standard deviation.

Note that in the gridless SPARROW implementation for the case of ULAs, the estimation bias is inherent in the estimation method and independent of grid effects, and can be countered by bias mitigation techniques [NBA13] or by a subsequent maximum likelihood (ML) estimation step [KV96, vT02]. For instance, a combination of the SPICE and ML estimation has been proposed in [SB12] in form of the so-called LIKES method.

4.4.3 Varying Number of Snapshots

The following experiment investigates the estimation performance of the various methods for a varying number of snapshots. Consider two independent complex Gaussian sources with static spatial frequencies $\mu_1 = 0.35$ and $\mu_2 = 0.5$ and a ULA with $M = 6$ sensors. The SNR is fixed at 3 dB. Figure 4.9 shows the RMSE of the schemes under

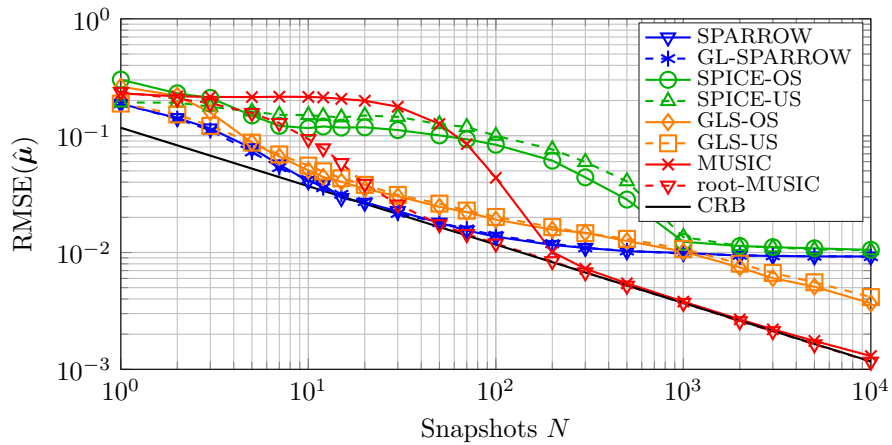


Figure 4.9: Frequency estimation performance for $L = 2$ source signals with frequency separation $\Delta\mu = 0.15$ and SNR = 3 dB

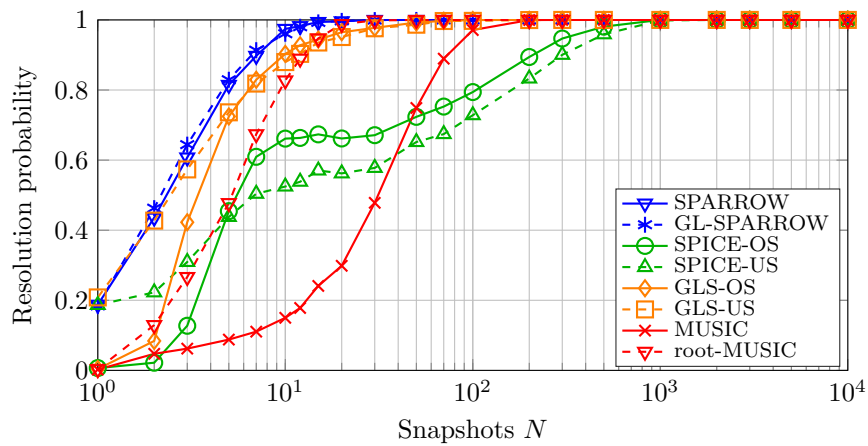


Figure 4.10: Resolution percentage for $L = 2$ source signals with frequency separation $\Delta\mu = 0.15$ and SNR = 3 dB

investigation from which it can be observed that SPARROW outperforms all other methods in terms of threshold performance. However, for larger number of snapshots, the RMSE of SPARROW saturates due to the estimation bias as discussed in Section 4.4.2. GLS shows slightly worse RMSE performance compared to SPARROW for a low number of snapshots N and also does not reach the CRB for a large number of snapshots. In contrast to that, root-MUSIC shows slightly degraded threshold performance but asymptotically reaches the CRB. The grid-based techniques MUSIC and SPICE all show poor threshold performance. While MUSIC asymptotically reaches the CRB, the SPICE techniques reach saturation, i.e. the RMSE reaches a lower bound.

To provide further insight to the resolution performance, the corresponding resolution probability is plotted in Figure 4.10, where two source signals with true frequencies μ_1, μ_2 and estimated frequencies $\hat{\mu}_1, \hat{\mu}_2$ are considered to be resolved if

$$\sum_{l=1}^2 |\mu_l - \hat{\mu}_l| \leq |\mu_1 - \mu_2|. \quad (4.107)$$

Similar as for the RMSE threshold performance, it can be observed from Figure 4.10 that GL-SPARROW outperforms the other investigated methods, providing 100% resolution percentage for $N \geq 30$ snapshots, slightly better than root-MUSIC. The GLS methods require $N \geq 100$ snapshots to provide resolution guarantee. Again, the grid-based schemes MUSIC and SPICE show significantly lower resolution performance.

4.4.4 Correlated Signals

For illustration of the estimation performance of SPARROW in the case of coherent source signals, consider a ULA with $M = 6$ sensors and $L = 2$ impinging source signals with spatial frequencies $\boldsymbol{\mu} = [0.0, 0.5]^\top$, SNR = 0 dB and varying number of snapshots N . Figure 4.11 displays the RMSE for uncorrelated source signals while Figure 4.12 shows the RMSE for coherent source signals under the same setup. As can be seen from the two figures, in the uncorrelated case all methods show similar frequency estimation performance. The MUSIC methods reach the CRB for large number of snapshots, and also SPARROW reaches the CRB since it is not affected by an estimation due to sufficient spacing between the spatial frequencies. The SPICE methods, on the other hand, do not reach the CRB for large number of snapshots. For coherent signals it can be observed that the MUSIC methods fail to resolve the two coherent source signals, as discussed in Section 3.2. The other methods, i.e., SPARROW and the SPICE methods show performance similar to the uncorrelated case, with SPARROW outperforming the SPICE methods.

4.4.5 Off-Grid SPARROW

For evaluation of the off-grid SPARROW formulations presented in Section 4.3.2, the first experiment considers a ULA of $M = 8$ sensors, with $N = 30$ snapshots and

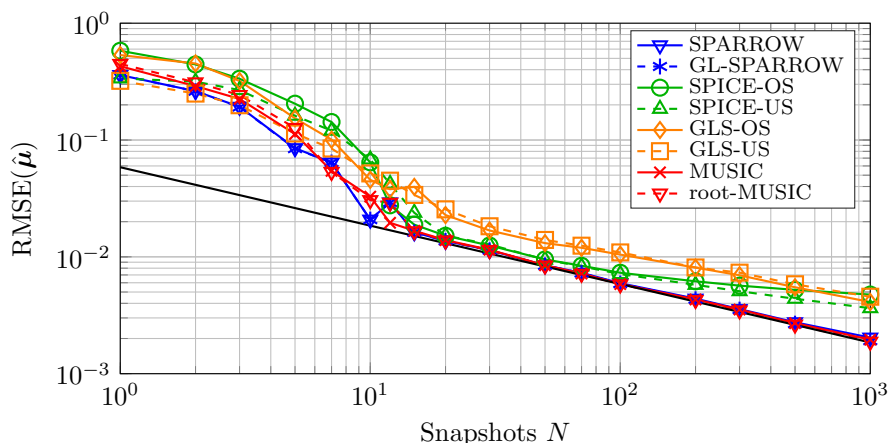


Figure 4.11: Frequency estimation performance for $L = 2$ uncorrelated source signals with frequency separation $\Delta\mu = 0.5$ and $\text{SNR} = 0$ dB

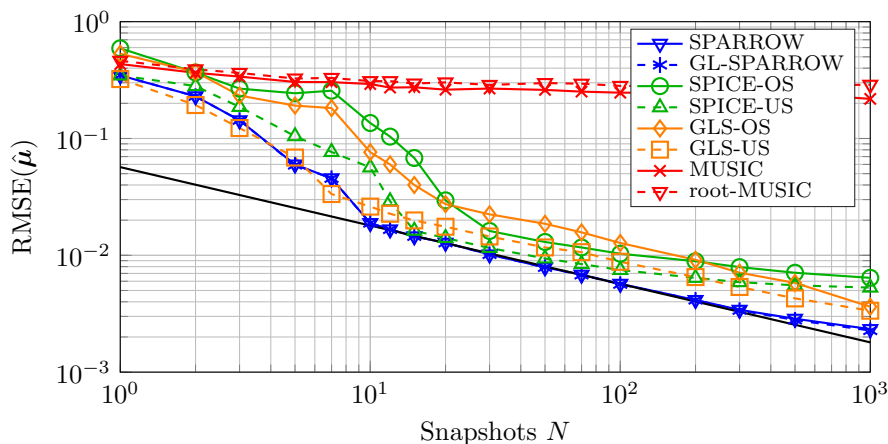


Figure 4.12: Frequency estimation performance for $L = 2$ coherent source signals with frequency separation $\Delta\mu = 0.5$ and $\text{SNR} = 0$ dB

$\text{SNR} = 20$ dB, and investigates the off-grid error estimation performance of the two formulations OG1-SPARROW in (4.86) and OG2-SPARROW in (4.88). Figure 4.13 displays the average value of the off-grid error \hat{e}_l estimated by the two off-grid schemes, for a single source with spatial frequency according to $\mu = e$, i.e., the spatial frequency is varied in the off-grid grid region around the frequency 0. An equidistant grid of $K = 200$ points is employed, corresponding to a grid spacing of $\delta = 0.01$. As can be observed from Figure 4.13, the off-grid error \hat{e} estimated by OG2-SPARROW is closer to the true off-grid error e than the estimates provided by OG1-SPARROW. Both off-grid formulations admit the additional estimation the off-grid parameter such that better DOA estimation can be obtained as compared to the grid-based SPARROW formulation (4.51).

Figure 4.14 shows the frequency estimation performance for a ULA of $M = 8$ sensors, $L = 2$ sources with spatial frequencies $\mu_1 = 0.103$ and $\mu_2 = 0.502$, $N = 20$ snapshots

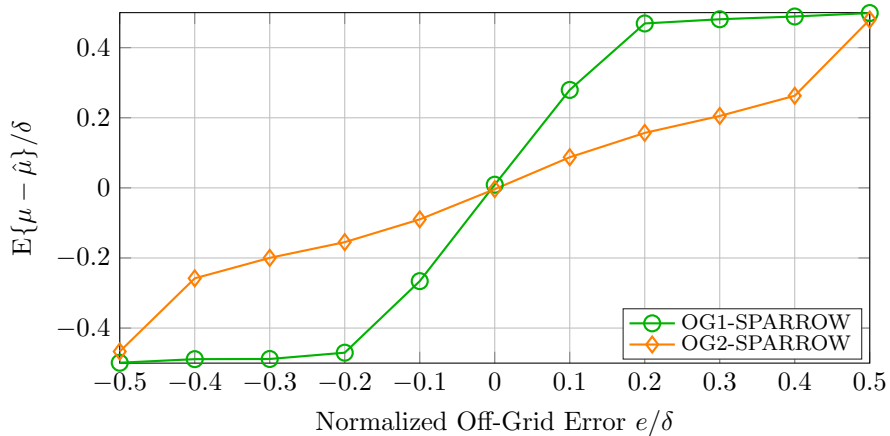


Figure 4.13: Mean of estimated off-grid error

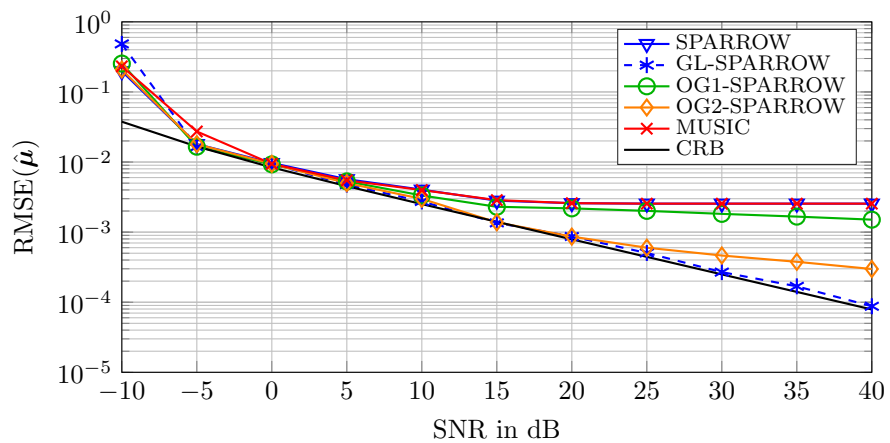


Figure 4.14: Frequency estimation performance for ULA of $M = 8$ sensors, with $L = 2$ source signals and $N = 20$ snapshots

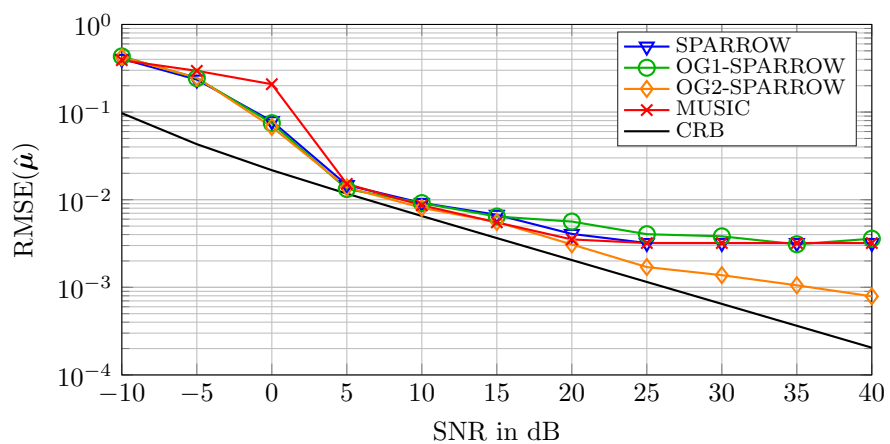


Figure 4.15: Frequency estimation performance for nonuniform linear array of $M = 8$ sensors, with $L = 2$ source signals and $N = 5$ snapshots

and varying SNR. As before, an equidistant grid of $K = 200$ points is used, corresponding to a grid spacing of $\delta = 0.01$. As seen from Figure 4.14, for high SNR the grid-based SPARROW and MUSIC methods result in a residual error caused by the finite grid. In contrast to that, GL-SPARROW achieves the CRB for high SNR and OG1- and OG2-SPARROW show a reduced error as compared to the grid-based schemes, with OG2- outperforming OG1-SPARROW.

The frequency estimation performance for a nonuniform linear array of $M = 8$ sensors with sensor positions $\mathbf{r} = [0.00, 0.57, 1.85, 3.12, 3.98, 5.84, 7.22, 8.00]^T$, and $L = 2$ source signals with spatial frequencies $\mu_1 = 0.0045$ and $\mu_2 = 0.2005$, is illustrated in Figure 4.15. The number of snapshots is fixed as $N = 5$ and the SNR is varied, while the frequency grid is selected similar to the previous setups. Figure 4.15 clearly demonstrates that OG2-SPARROW outperforms OG1-SPARROW and the grid-based schemes MUSIC and SPARROW for high SNR.

4.4.6 Computation Time of SDP Formulations

The experiments concerning the computation time of $\ell_{2,1}$ minimization, SPARROW and ANM are performed in MATLAB on a computer with an Intel Core i7-4770 CPU @ 3.40 GHz \times 8 and 16 GByte RAM. The optimization problems for the equivalent estimation problems are solved using the MOSEK framework [MOS15] with the CVX MATLAB interface [GB08, GB14].

Consider a scenario with $L = 3$ independent complex Gaussian sources with static spatial frequencies $\mu_1 = -0.1$, $\mu_2 = 0.35$ and $\mu_3 = 0.5$, and a ULA of $M = 10$ sensors. The SNR is fixed at 10 dB while the number of snapshots N is varied. Figure 4.16 shows the average CPU time of $\ell_{2,1}$ minimization (4.50), the grid-based SPARROW formulations (4.54) and (4.57), Atomic Norm Minimization (ANM) (4.49) and GL-SPARROW (4.60) and (4.61). A number of $K = 1000$ candidate frequencies is used for the grid-based methods.

Regarding the CPU time for the grid-based methods, it can be noted that the SPARROW formulation (4.54) outperforms the $\ell_{2,1}$ mixed-norm minimization (4.50) for $N < 30$ snapshots. For a larger number of snapshots the dimensions of the semidefinite constraint (4.54b) become too large, such that the computational cost is increased as compared to $\ell_{2,1}$ minimization (4.50). The SPARROW formulation (4.57) is based on the sample covariance matrix, thus, the computational cost is independent of the number of snapshots. For the gridless methods, Figure 4.16 clearly displays that the CPU time of the GL-SPARROW (4.60) formulation is significantly reduced as compared to the ANM formulation (4.47). Similar to the grid-based case, the CPU time of the covariance-based GL-SPARROW formulation (4.61) is relatively independent of the number of snapshots N and outperforms the other methods for large number of snapshots N . Independent of the number of snapshots, the gridless SPARROW formulations (4.60) and (4.61) clearly outperform their grid-based counterparts (4.54)

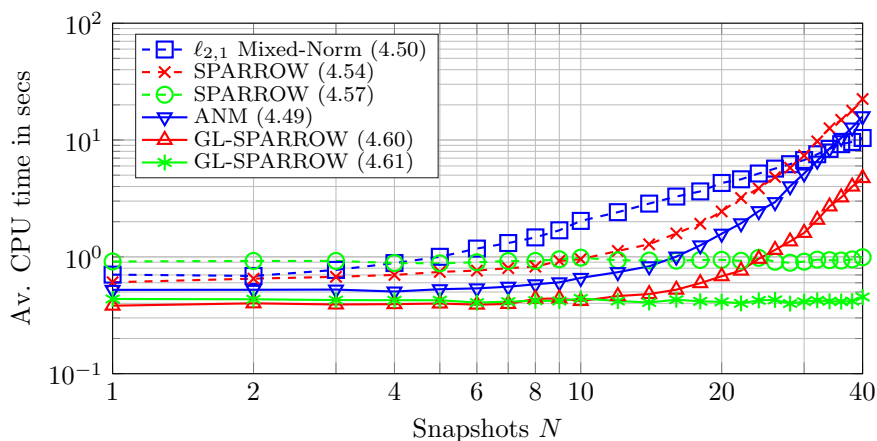


Figure 4.16: Average computation time of different SDP implementations for ULA of $M = 10$ sensors and varying number of snapshots N

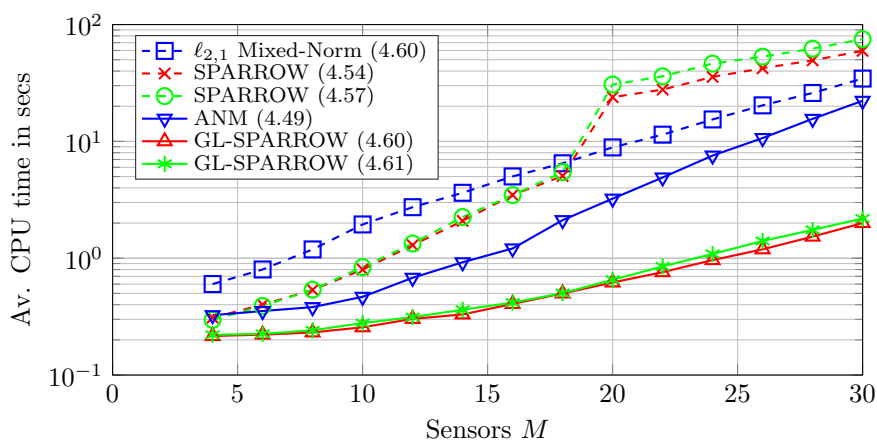


Figure 4.17: Average computation time of different SDP implementations for ULA with varying number of sensors M , and N snapshots (with dimensionality reduction)

and (4.57). The experiment also demonstrates that all the methods employing the raw measurements in \mathbf{Y} , i.e., $\ell_{2,1}$ minimization (4.50), the SPARROW formulation (4.54), the ANM formulation (4.47) and the GL-SPARROW formulation (4.60), suffer from increased computation time in the case of large number of snapshots N , demonstrating the necessity of dimensionality reduction techniques, as will be investigated in the following experiment.

For the experiment, the source signal parameters from the previous experiment are maintained and the number of snapshots is fixed as $N = 50$, while the number of sensors M in the ULA is varied. To reduce the computational burden in the methods based on the $M \times N$ raw measurement matrix \mathbf{Y} , dimensionality reduction according to [YX15a] is performed to match a matrix \mathbf{Y}_{RD} of dimensions $M \times M$ instead, as discussed in Section 4.2. Figure 4.17 displays the average CPU time for the various

equivalent methods under investigation.

Using the dimensionality reduction technique, it can be seen that both grid-based SPARROW formulations (4.54) and (4.57) require the same computational time, since the dimensions of the semidefinite constraints are identical. For $M \leq 18$ sensors the grid-based SPARROW formulations outperform the $\ell_{2,1}$ mixed-norm minimization (4.50). However, for $M > 18$ the dimensions of the semidefinite constraints in the SPARROW formulations become too large, such that the computational cost exceeds that of $\ell_{2,1}$ minimization (4.50).

Similar as for the grid-based SPARROW, the gridless SPARROW formulations (4.60) and (4.61) show identical computation time, due to the identical size of the semidefinite constraints. Both gridless SPARROW formulations clearly outperform the ANM approach (4.47), especially for large number of sensors M . This can be explained by the additional M^2 complex-valued variables in the matrix \mathbf{Y}_0 of the ANM formulation (4.47).

The experiment demonstrates that the computational cost of the grid-based SPARROW formulations exceeds that of $\ell_{2,1}$ mixed-norm minimization for large number of sensors, when the SDP formulations are used with the MOSEK solver. To deal with this problem, a low-complexity coordinate descent (CD) implementation is devised in Section 6.1.3, which exploits the special structure of the proposed SPARROW formulation. Experimental results on convergence rate of the CD implementation in Section 6.3.1 demonstrate that the SPARROW formulation also admits reduced computation time for large number of sensors.

4.5 Chapter Summary

The chapter has started with a short review of state of the art methods in terms of ℓ_1 norm and atomic norm minimization for sparse reconstruction from single snapshots, as well as $\ell_{2,1}$ mixed-norm and atomic norm minimization for joint sparse reconstruction from multiple snapshots. A major contribution of the chapter is given by a compact reformulation of the $\ell_{2,1}$ minimization problem, which is referred to as SPARROW (SPARse ROW-norm reconstruction) and has significantly reduced number of optimization parameters as compared to the original $\ell_{2,1}$ minimization problem. Furthermore, a gridless implementation of the SPARROW formulation was derived, for application in uniform and thinned linear arrays. The gridless implementation relies on Vandermonde decomposition and was shown to be equivalent to atomic norm minimization for joint sparse reconstruction, while having significantly reduced number of optimization parameters. To similarly admit off-grid estimation in arbitrary array topologies, an extension of the SPARROW formulation based on linear interpolation of the array response in the off-grid zones was presented. In context with other existing work, a detailed comparison of the SPARROW formulation and the SPICE method has been provided.

The numerical experiments in this chapter have demonstrated that sparse reconstruction by means of convex relaxation provides a viable supplement to classical subspace-based methods, such as MUSIC, especially for correlated signals and in the non-asymptotic regime of low signal-to-noise ratio and low number of snapshots. The experiments on resolution performance have demonstrated that especially the grid-based SPARROW formulation outperforms the spectral MUSIC method, indicating superior threshold performance of the SPARROW method for nonuniform array topologies, where the powerful root-MUSIC method cannot be applied. Regarding computation time, it was shown that the SPARROW formulation clearly outperforms atomic norm minimization for any size of sensor arrays. In comparison with $\ell_{2,1}$ minimization, a reduction in computation time for the SDP-based SPARROW implementations was found mainly for medium-sized sensor arrays. However, as will be shown in Section 6.1.3, improved computation time for large sensor arrays can be obtained by a tailored implementation of the SPARROW formulation in form of the coordinate descent method.

Chapter 5

Sparse Reconstruction for Partly Calibrated Arrays

In direction finding applications based on array signal processing it is desired to achieve a high angular resolution and to identify a large number of sources. This can be achieved by sensor arrays with a large aperture and a large number of sensors [KV96]. However, a large aperture size makes it difficult to achieve and maintain precise array calibration. Possible reasons for imperfect calibration are inaccuracies in the sensor positions, timing synchronization errors, or other unknown gain and phase offsets among sensors [PF97, RS87, WF91, VS94, NS96, FB01, SG04]. Standard approaches to this problem usually rely on either offline or online calibration. Offline calibration of the overall array is performed using reference sources at known positions and can easily become a challenging and time consuming task [PF97]. Alternatively, several online calibration techniques have been proposed which use calibration sources at unknown positions [RS87, WF91, VS94, NS96, FB01], but the computational cost of these techniques is often prohibitively high, and performance can be severely limited in the case of large sensor position errors [FB01]. Moreover, these techniques cannot be employed in scenarios with imperfect time synchronization of sensors or other unknown sensor gain and phase offsets.

One way to overcome the calibration problem is to partition the overall array into smaller subarrays, which are themselves comparably easy to calibrate. This type of array is referred to as partly calibrated array (PCA). Generally, direction finding approaches for this type of arrays can be classified into incoherent and coherent methods. In the incoherent case, the subarrays independently perform estimation of the directions of arrival (DOAs) or the signal covariance matrix to communicate these estimates to a central processor, where further processing is performed to achieve an improved joint estimate [WK85, SNS95, SP14]. In the coherent approach, parameter estimation is performed based on joint coherent processing of all available sensor measurements, e.g., by computing a global sample covariance matrix, and imperfect calibration among the different subarrays is taken account of in the estimation process. Coherent processing generally achieves better estimation performance, but has more stringent requirements with regard to synchronization of the subarrays in time and frequency, as compared to incoherent processing. Prominent methods for coherent processing are based on subspace separation such as the RARE method [PGW02, SG04] or ESPRIT-type methods [SORK92, SSJ01, PP11, SPZ13, SPPZ14], as discussed in Chapter 3. Similar to the case of fully calibrated arrays discussed in Chapter 4, subspace-based methods for PCAs are asymptotically optimal at tractable computational cost, but may suffer from performance degradation in the case of low number of snapshots or correlated source signals, such that SSR methods provide an attractive alternative in these difficult scenarios.

This chapter considers SSR for coherent processing in PCAs. To put the results in context with existing SSR methods for PCAs, the chapter starts with a brief review of state of the art methods for incoherent SSR in PCAs, where a block-sparse structure in the subarray signal representations is exploited [DSB⁺05, LYJ⁺15]. In the following, a new approach for coherent SSR in PCAs is derived, based on minimization of a mixed nuclear and ℓ_1 mixed-norm formulation, as presented in [SPP14], which exploits two types of sparse structures in the PCA signal model: While ℓ_1 minimization provides a block-sparse signal structure, nuclear norm minimization leads to low-rank solutions [FHB01, RFP10, CCS10, CR12]. Moreover, an equivalent compact reformulation of the $\ell_{*,1}$ mixed-norm minimization problem is derived, termed as COmpact Block- and RAnk-Sparse recovery (COBRAS) [SP18], which has a reduced number of optimization parameters as compared to the original $\ell_{*,1}$ mixed-norm minimization problem. While the original COBRAS formulation relies on grid-based implementation and is applicable to arbitrary array topologies, the special case of PCAs composed of uniform or thinned linear subarrays admits a gridless implementation of the COBRAS formulation. To put the results in context with previous work it is shown in this chapter how the gridless COBRAS formulation can equivalently be derived from the SPARROW formulation discussed in Section 4.3. Furthermore, a gridless implementation of the SPARROW formulation is derived for shift-invariant array topologies [SSSP17a]. Extensive numerical experiments demonstrate that the COBRAS approach outperforms the state of the art methods in difficult scenarios.

5.1 Incoherent Sparse Reconstruction

Sparse reconstruction for PCAs with incoherent processing has been well investigated in the literature [DSB⁺05, LYJ⁺15] and has low requirements for practical implementation in terms of subarray synchronization, as discussed in Section 2.2. Consider a linear PCA of $M = \sum_p M_p$ sensors, partitioned into P linear subarrays of M_1, \dots, M_P sensors, and with arbitrary subarray topologies, as discussed in Section 2.2. Let $\tilde{\mathbf{Y}}^{(p)} = [\mathbf{y}^{(p)}(t_1^{(p)}), \dots, \mathbf{y}^{(p)}(t_N^{(p)})] \in \mathbb{C}^{M_p \times N}$ denote the measurement matrix obtained by subarray p in sampling times $t_1^{(p)}, \dots, t_N^{(p)}$, which might vary for different subarrays $p = 1, \dots, P$. According to (2.20), the subarray measurements are modeled as

$$\tilde{\mathbf{Y}}^{(p)} = \mathbf{A}^{(p)}(\boldsymbol{\mu}) \tilde{\boldsymbol{\Psi}}^{(p)} + \tilde{\mathbf{N}}^{(p)}, \quad (5.1)$$

where $\tilde{\boldsymbol{\Psi}}^{(p)} = [\tilde{\boldsymbol{\psi}}^{(p)}(t_1^{(p)}), \dots, \tilde{\boldsymbol{\psi}}^{(p)}(t_N^{(p)})] \in \mathbb{C}^{L \times N}$ represents the source signals as observed by sensor 1 in subarray p as defined in (2.21), $\tilde{\mathbf{N}}^{(p)} \in \mathbb{C}^{M_p \times N}$ represents the sensor noise observed by subarray p , and $\mathbf{A}^{(p)}(\boldsymbol{\mu}) \in \mathbb{C}^{M_p \times L}$ denotes the steering matrix of subarray p as given in (2.22), for $p = 1, \dots, P$. A sparse representation of the subarray signal model in (5.1) is given as

$$\tilde{\mathbf{Y}}^{(p)} = \mathbf{A}^{(p)}(\boldsymbol{\nu}) \tilde{\tilde{\mathbf{X}}}^{(p)} + \tilde{\mathbf{N}}^{(p)}, \quad (5.2)$$

where $\mathbf{A}^{(p)}(\boldsymbol{\nu})$ denotes the $M_p \times K$ overcomplete subarray sensing matrix, for $p = 1, \dots, P$, which is obtained for a grid of $K \gg L$ sampled frequencies $\boldsymbol{\nu} = [\nu_1, \dots, \nu_K]^\top$, and will be referred to as $\mathbf{A}^{(p)}$ in the following. The $K \times N$ row-sparse source signal matrix $\check{\check{\mathbf{X}}}^{(p)}$ is defined similar to the FCA case in (4.30), according to

$$[\check{\check{\mathbf{X}}}^{(p)}]_{k,n} = \begin{cases} [\tilde{\boldsymbol{\psi}}^{(p)}(t_n^{(p)})]_l & \text{if } \nu_k = \mu_l \\ 0 & \text{otherwise.} \end{cases} \quad (5.3)$$

From (5.3) it can be seen that the sparse source signal representations $\check{\check{\mathbf{X}}}^{(p)}$ for subarrays $p = 1, \dots, P$ and snapshots $n = 1, \dots, N$ have the same signal support, such that the joint sparse structure can be exploited for the sparse reconstruction.

Consider the case of identical subarrays, such that $\mathbf{A}^{(0)} \triangleq \mathbf{A}^{(1)} = \mathbf{A}^{(2)} = \dots = \mathbf{A}^{(P)} \in \mathbb{C}^{M_0 \times K}$, with M_0 sensors per subarray, and define the $M_0 \times PN$ data matrix $\bar{\mathbf{Y}} = [\tilde{\mathbf{Y}}^{(1)}, \dots, \tilde{\mathbf{Y}}^{(P)}]$ and the $K \times PN$ signal matrix $\check{\check{\mathbf{X}}} = [\check{\check{\mathbf{X}}}^{(1)}, \dots, \check{\check{\mathbf{X}}}^{(P)}]$. The signal matrix $\check{\check{\mathbf{X}}}$ has a joint sparse structure, similar to the multi snapshot case for FCAs discussed in Section (4.2), which can be exploited by the sparse optimization problem [DSB⁺05, LYJ⁺15]

$$\min_{\check{\check{\mathbf{X}}}} \|\check{\check{\mathbf{X}}}\|_{q,0} \quad \text{s.t.} \quad \|\mathbf{A}^{(0)} \check{\check{\mathbf{X}}} - \bar{\mathbf{Y}}\|_{\text{F}}^2 \leq \beta, \quad (5.4)$$

for some threshold parameter β , where in the noise-free case the constraint is replaced by $\mathbf{A}^{(0)} \check{\check{\mathbf{X}}} = \bar{\mathbf{Y}}$.

For the general case of arbitrary subarrays, the subarray measurements are stacked according to $\tilde{\mathbf{Y}} = [\tilde{\mathbf{Y}}^{(1)\top}, \dots, \tilde{\mathbf{Y}}^{(P)\top}]^\top \in \mathbb{C}^{M \times N}$ and a sparse signal representation according to the model in (2.24) is employed, i.e.,

$$\tilde{\mathbf{Y}} = \mathbf{B}(\boldsymbol{\nu}) \check{\check{\mathbf{Z}}} + \check{\check{\mathbf{N}}}, \quad (5.5)$$

where the overcomplete subarray sensing block matrix $\mathbf{B}(\boldsymbol{\nu}) = [\mathbf{B}(\nu_1), \dots, \mathbf{B}(\nu_K)] \in \mathbb{C}^{M \times PK}$ is defined in correspondence with (2.16), where $\boldsymbol{\nu} = [\nu_1, \dots, \nu_K]^\top$ denotes the grid of sampled spatial frequencies, with $K \gg L$. For ease of presentation the subarray sensing block matrix $\mathbf{B}(\boldsymbol{\nu})$ will be referred to as $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_K]$ in the following. The $KP \times N$ source signal matrix $\check{\check{\mathbf{Z}}}$ in (5.5) has a block structure

$$\check{\check{\mathbf{Z}}} = [\check{\check{\mathbf{Z}}}_1^\top, \dots, \check{\check{\mathbf{Z}}}_K^\top]^\top, \quad (5.6)$$

where each block $\check{\check{\mathbf{Z}}}_k = [\check{\check{\mathbf{x}}}_k^{(1)}, \dots, \check{\check{\mathbf{x}}}_k^{(P)}]^\top \in \mathbb{C}^{P \times N}$ contains the source signals $\check{\check{\mathbf{X}}}^{(p)} = [\check{\check{\mathbf{x}}}_1^{(p)}, \dots, \check{\check{\mathbf{x}}}_K^{(p)}]^\top$ observed by the different subarrays $p = 1, \dots, P$, for sampled frequency ν_k , according to

$$[\check{\check{\mathbf{Z}}}_k]_{p,n} = \begin{cases} [\tilde{\boldsymbol{\psi}}^{(p)}(t_n^{(p)})]_l & \text{for } \nu_k = \mu_l \\ 0 & \text{otherwise.} \end{cases} \quad (5.7)$$

The matrix $\tilde{\mathbf{Z}}$ exhibits a block-sparse structure, i.e., the elements in the $P \times N$ submatrices $\tilde{\mathbf{Z}}_k$, for $k = 1, \dots, K$, are either jointly zero or primarily non-zero. This structure can be taken account of in the sparse reconstruction by the $\ell_{q,0}^{(P \times N)}$ mixed-norm according to

$$\|\tilde{\mathbf{Z}}\|_{q,0}^{(P \times N)} = |\{k \mid \|\tilde{\mathbf{Z}}_k\|_q \neq 0\}|, \quad (5.8)$$

where $\|\tilde{\mathbf{Z}}_k\|_q$ denotes an appropriate matrix norm applied to the $P \times N$ submatrix $\tilde{\mathbf{Z}}_k$, e.g., the Frobenius norm or the spectral norm, such that the $\ell_{q,0}^{(P \times N)}$ mixed-norm $\|\tilde{\mathbf{Z}}\|_{q,0}^{(P \times N)}$ counts the number of non-zero submatrices $\tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_K$, in the matrix $\tilde{\mathbf{Z}}$. Note that in the case of a single subarray $P = 1$ the block-wise mixed-norm (5.8) reduces to the row-wise mixed-norm (4.32) applied for joint sparse reconstruction in FCAs, i.e., $\|\tilde{\mathbf{Z}}\|_{q,0}^{(1 \times N)} = \|\tilde{\mathbf{Z}}\|_{q,0}$ as defined in (4.32).

Using the mixed-norm formulation (5.8), a sparse recovery approach for incoherent processing in PCAs can be formulated by the problem

$$\min_{\tilde{\mathbf{Z}}} \|\tilde{\mathbf{Z}}\|_{p,0}^{(P \times N)} \quad \text{s.t.} \quad \|\mathbf{B}\tilde{\mathbf{Z}} - \tilde{\mathbf{Y}}\|_{\text{F}}^2 \leq \beta, \quad (5.9)$$

for some proper threshold β , where in the noise-free case the constraint in (5.9) is replaced by $\mathbf{B}\tilde{\mathbf{Z}} = \tilde{\mathbf{Y}}$. Given a minimizer $\hat{\tilde{\mathbf{Z}}}$ the spatial frequencies can be identified from the block support according to

$$\{\hat{\nu}_l\}_{l=1}^{\hat{L}} = \{\nu_k \mid \|\hat{\tilde{\mathbf{Z}}}_k\|_q \neq 0\}. \quad (5.10)$$

Similar as for the sparse reconstruction problems based on ℓ_0 minimization in Section 4, the problem in (5.9) becomes intractable for large problem dimensions, such that approximate methods have to be devised.

5.1.1 $\ell_{\text{F},1}$ Mixed-Norm Minimization

As discussed in the previous section, SSR for incoherent processing with identical PCAs can be formulated by the joint sparse reconstruction problem in (5.4), which has a structure similar to that of joint sparse reconstruction for the multi snapshot model in FCAs discussed in Section 4.2. Correspondingly, the $\ell_{q,0}$ mixed-norm in (5.4) can be approximated by an $\ell_{q,1}$ mixed-norm. More specifically, consider the $\ell_{2,1}$ mixed-norm, such that the convex relaxation of the SSR problem (5.4) is formulated in the penalized form as

$$\min_{\bar{\mathbf{X}}} \frac{1}{2} \|\mathbf{A}^{(0)} \bar{\mathbf{X}} - \bar{\mathbf{Y}}\|_{\text{F}}^2 + \lambda \sqrt{PN} \|\bar{\mathbf{X}}\|_{2,1}, \quad (5.11)$$

which can be solved, e.g., by means of the SPARROW formulation discussed in Section 4.3. It follows that in the case of identical uniform linear subarrays of M_0 sensors per subarray, no more than $L < M_0$ signals can be recovered under incoherent processing,

due to the characteristics of the Vandermonde decomposition given in Definition 4.1. The regularization parameter λ can be selected according to the approaches (4.40) or (4.41) presented in Section 4.2.

In the case of arbitrary subarrays, the $\ell_{q,0}^{(P \times N)}$ mixed-norm can similarly be approximated by an $\ell_{q,1}^{(P \times N)}$ mixed-norm, defined as

$$\|\tilde{\mathbf{Z}}\|_{q,1}^{(P \times N)} = \sum_{k=1}^K \|\tilde{\mathbf{Z}}_k\|_q. \quad (5.12)$$

Consider in the following the application of the Frobenius norm on the submatrices $\tilde{\mathbf{Z}}_k$, for $k = 1, \dots, K$, such that the $\ell_{F,1}^{(P \times N)}$ mixed-norm is given as

$$\|\tilde{\mathbf{Z}}\|_{F,1}^{(P \times N)} = \sum_{k=1}^K \|\tilde{\mathbf{Z}}_k\|_F = \sum_{k=1}^K \|\text{vec}(\tilde{\mathbf{Z}}_k)\|_2. \quad (5.13)$$

From (5.13) it can be seen that the $\ell_{F,1}^{(P \times N)}$ mixed-norm can equivalently be interpreted as an $\ell_{2,1}$ mixed-norm on the vectorized submatrices $\tilde{\mathbf{Z}}_k$. Using the $\ell_{F,1}^{(P \times N)}$ mixed norm, the convex approximation of problem (5.9) yields

$$\min_{\tilde{\mathbf{Z}}} \frac{1}{2} \|\mathbf{B}\tilde{\mathbf{Z}} - \tilde{\mathbf{Y}}\|_F^2 + \lambda\sqrt{N} \|\tilde{\mathbf{Z}}\|_{F,1}^{(P \times N)}. \quad (5.14)$$

In the case of identical subarrays, as discussed above, the $\ell_{F,1}^{(P \times N)}$ minimization problem in (5.14) is equivalent to $\ell_{2,1}$ minimization problem in (5.11), as seen from the definition of the matrix blocks $\tilde{\mathbf{Z}}_k$ in (5.6) and below, and the interpretation of the $\ell_{F,1}^{(P \times N)}$ mixed-norm as an $\ell_{2,1}$ mixed-norm as discussed for (5.13). Hence, application of the $\ell_{F,1}^{(P \times N)}$ mixed-norm is a reasonable choice for incoherent processing in PCAs. Note that in the case of a single subarray $P = 1$, the SSR formulations (5.4) and (5.9) as well as the convex relaxations in (5.11) and (5.14) reduce to the formulations for joint sparse recovery in FCAs given in (4.33) and (4.37).

From above considerations it can be deduced that the gain of incoherent processing in PCAs of identical subarrays, as discussed for problem (5.11), lies in additional snapshots which can be used for estimation. However, the large array aperture and the increased number of sensors are not fully exploited. From subspace-based methods such as RARE and ESPRIT it is known that these benefits can be retained by coherent processing, which will be further investigated in the context of SSR in the following section.

5.2 Coherent Sparse Reconstruction

While incoherent processing admits simple implementation in practice, it has drawbacks in exploiting the additional structure in the PCA topology. This section considers the application of coherent processing in PCAs, i.e., synchronous sampling of the signals

received by the subarrays. To this end the signal model in (2.19) is applied, which is given as

$$\mathbf{Y} = \mathbf{B}(\boldsymbol{\mu}) \boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \boldsymbol{\Psi} + \mathbf{N}, \quad (5.15)$$

where $\mathbf{Y}, \mathbf{N} \in \mathbb{C}^{M \times N}$ represent the synchronously sampled measurements and the sensor noise of the entire sensor array, respectively, $\mathbf{B}(\boldsymbol{\mu}) \in \mathbb{C}^{M \times PL}$ contains the subarray responses, $\boldsymbol{\Phi}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \in \mathbb{C}^{LP \times L}$ accounts for the subarray shifts and $\boldsymbol{\Psi} \in \mathbb{C}^{L \times N}$ represents the source signals, as observed by sensor 1 in subarray 1.

A sparse representation of the signal model in (5.15) for the PCA case is given as

$$\mathbf{Y} = \mathbf{B}(\boldsymbol{\nu}) \boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \check{\mathbf{X}} + \mathbf{N}, \quad (5.16)$$

where the row-sparse signal matrix $\check{\mathbf{X}}$ is defined similar to the FCA case in (4.30). Furthermore, the $M \times PK$ overcomplete subarray sensing block matrix $\mathbf{B}(\boldsymbol{\nu})$ and the $PK \times K$ overcomplete subarray shift matrix $\boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ are defined in correspondence to (2.16) and (2.17), respectively, for a grid of $K \gg L$ sampled spatial frequencies $\boldsymbol{\nu} = [\nu_1, \dots, \nu_K]^\top$.

In the PCA case, the inter-subarray displacements in $\boldsymbol{\eta}$ and the subarray perturbations in $\boldsymbol{\alpha}$ are unknown and represent additional estimation variables. Hence the subarray shifts in $\boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$, which depend on the spatial frequencies in $\boldsymbol{\nu}$, the subarray displacements $\boldsymbol{\eta}$ and the subarray perturbations $\boldsymbol{\alpha}$, have to be appropriately included in the sparse reconstruction problem. To this end, a model is introduced that couples among the variables $\check{\mathbf{x}}_k$ in the rows of $\check{\mathbf{X}} = [\check{\mathbf{x}}_1, \dots, \check{\mathbf{x}}_K]^\top$ and the subarray shifts in $\boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})$, for $k = 1, \dots, K$. Define the $KP \times N$ extended signal matrix $\check{\mathbf{Z}}$ as

$$\check{\mathbf{Z}} = \boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \check{\mathbf{X}}, \quad (5.17)$$

containing the products of the subarray shifts and the signal waveforms. As compared to the total $K(N + P - 1)$ complex-valued unknowns in both the signal matrix $\check{\mathbf{X}}$ and the block-diagonal subarray shift matrix $\boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ in model (5.16), the number of unknown complex-valued signal elements in the matrix $\check{\mathbf{Z}}$, defined in (5.17), is increased to KPN . Note that the incoherent signal matrix $\check{\mathbf{Z}}$ in (5.5) has the same number of elements as the coherent signal matrix $\check{\mathbf{Z}}$ in (5.17). However, in contrast to the incoherent signal matrix $\check{\mathbf{Z}}$ in (5.5), the coherent signal matrix $\check{\mathbf{Z}} = [\check{\mathbf{Z}}_1^\top, \dots, \check{\mathbf{Z}}_K^\top]^\top$ in (5.17) enjoys a beneficial structure as it is composed of K stacked rank-one matrices

$$\check{\mathbf{Z}}_k = \boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \check{\mathbf{x}}_k^\top, \quad (5.18)$$

of dimensions $P \times N$, for $k = 1, \dots, K$, which follows from the block structure of the subarray shift matrix $\boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \text{blkdiag}(\boldsymbol{\varphi}(\nu_1, \boldsymbol{\alpha}, \boldsymbol{\eta}), \dots, \boldsymbol{\varphi}(\nu_K, \boldsymbol{\alpha}, \boldsymbol{\eta}))$, defined in correspondence to (2.17). Using the coherent signal matrix in (5.17), the sparse representation for the PCA case in (5.16) is equivalently described by

$$\mathbf{Y} = \mathbf{B}(\boldsymbol{\nu}) \check{\mathbf{Z}} + \mathbf{N}, \quad (5.19)$$

where, for ease of notation, the subarray sensing block matrix $\mathbf{B}(\boldsymbol{\nu})$ in (5.19) will be referred to as $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_K]$ in the following. An SSR approach to take account of the special structure of the coherent signal matrix $\tilde{\mathbf{Z}}$ in (5.17) can be formulated as

$$\min_{\mathbf{Z}} \sum_{k=1}^K \text{rank}(\mathbf{Z}_k) \quad \text{s.t.} \quad \|\mathbf{B}\mathbf{Z} - \mathbf{Y}\|_{\text{F}}^2 \leq \beta. \quad (5.20)$$

The formulation in (5.20) takes twofold advantage of the sparsity assumption. First, minimization of the rank-terms encourages low-rank blocks $\hat{\mathbf{Z}}_1, \dots, \hat{\mathbf{Z}}_K$ in the minimizer $\hat{\mathbf{Z}}$. Second, minimizing the sum of ranks provides a block-sparse structure of $\hat{\mathbf{Z}} = [\hat{\mathbf{Z}}_1^{\text{T}}, \dots, \hat{\mathbf{Z}}_K^{\text{T}}]^{\text{T}}$, i.e., the elements in each block $\hat{\mathbf{Z}}_k$, for $k = 1, \dots, K$, are either jointly zero or primarily non-zero. However, the minimization problem in (5.20) is generally hard to solve, since it requires minimization of the rank of a matrix [VB96], such that a convex relaxation approach will be considered in the following section.

5.2.1 $\ell_{*,1}$ Mixed-Norm Minimization

A convex approximation of the rank function is given by the nuclear norm which has been successfully applied in a variety of rank minimization problems [FHB01, CCS10, RFP10, CR12]. The definition of the nuclear norm is given as

$$\|\mathbf{Z}_k\|_* = \text{Tr}((\mathbf{Z}_k^{\text{H}}\mathbf{Z}_k)^{1/2}) = \sum_{i=1}^r \sigma_{k,i}, \quad (5.21)$$

where $r = \min(P, N)$ and $\sigma_{k,i}$ is the i th largest singular value of $\mathbf{Z}_k \in \mathbb{C}^{P \times N}$. To understand the rank minimization characteristic of the nuclear norm, observe from (5.21) that it can similarly be interpreted as an ℓ_1 norm on the singular values of \mathbf{Z}_k . Thus, minimization of the nuclear norm leads to sparsity in the singular values, corresponding to a low matrix rank. In contrast, the Frobenius norm, as applied for incoherent processing in Section 5.1.1, corresponds to the ℓ_2 norm of the singular vectors according to

$$\|\mathbf{Z}_k\|_{\text{F}} = \sqrt{\text{Tr}(\mathbf{Z}_k^{\text{H}}\mathbf{Z}_k)} = \sqrt{\sum_{i=1}^r \sigma_{k,i}^2}, \quad (5.22)$$

such that minimization of the Frobenius norm does not lead to a low-rank matrix estimate.

Regarding the rank minimization character of the nuclear norm, as a major contribution of this thesis it was proposed in [SPP14], to approximate the sparse reconstruction problem (5.20) by the following convex minimization problem in the regularized form

$$\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{B}\mathbf{Z} - \mathbf{Y}\|_{\text{F}}^2 + \lambda\sqrt{N} \|\mathbf{Z}\|_{*,1}^{(P \times N)}, \quad (5.23)$$

where $\|\mathbf{Z}\|_{*,1}$ denotes the $\ell_{*,1}^{(P \times N)}$ mixed-norm, computed as

$$\|\mathbf{Z}\|_{*,1}^{(P \times N)} = \sum_{k=1}^K \|\mathbf{Z}_k\|_*. \quad (5.24)$$

Similar to (5.20), the problem in (5.23) motivates low-rank blocks $\hat{\mathbf{Z}}_1, \dots, \hat{\mathbf{Z}}_K$ and a block-sparse structure in the minimizer $\hat{\mathbf{Z}} = [\hat{\mathbf{Z}}_1^\top, \dots, \hat{\mathbf{Z}}_K^\top]^\top$. Note that the PCA formulations for coherent processing in (5.20) and (5.23) reduce to the FCA formulations in (4.33) and (4.38), respectively, in the case of a single subarray, i.e., $P = 1$. Similarly, in the case of a single snapshot, i.e., $N = 1$, the formulations (5.20) and (5.23) reduce to the PCA formulation for incoherent processing in (5.9) and (5.14).

Performing singular value decomposition on the matrix blocks in $\hat{\mathbf{Z}}$, i.e.,

$$\hat{\mathbf{Z}}_k = \hat{\mathbf{U}}_k \hat{\Sigma}_k \hat{\mathbf{V}}_k^H, \quad \text{for } k = 1, \dots, K, \quad (5.25)$$

the signal waveform in $\hat{\mathbf{x}}_k$ and subarray shifts in $\hat{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})$ corresponding to the spatial frequency ν_k can be recovered according to

$$\hat{\mathbf{x}}_k = \hat{\sigma}_{k,1} [\hat{\mathbf{u}}_{k,1}]_1 \hat{\mathbf{v}}_{k,1} \quad \text{and} \quad \hat{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \frac{\hat{\mathbf{u}}_{k,1}}{[\hat{\mathbf{u}}_{k,1}]_1}. \quad (5.26)$$

The left and right singular vectors $\hat{\mathbf{u}}_{k,1}$ and $\hat{\mathbf{v}}_{k,1}$ in (5.26) correspond to the largest singular value $\hat{\sigma}_{k,1}$ of $\hat{\mathbf{Z}}_k$. Normalization to the first element $[\hat{\mathbf{u}}_{k,1}]_1$ of $\hat{\mathbf{u}}_{k,1}$ in (5.26) is performed to describe the subarray shifts in vectors $\hat{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})$ in reference to the first subarray, according to (2.15).

As derived in Appendix F, a novel choice of regularization parameter for the $\ell_{*,1}^{(P \times N)}$ mixed-norm minimization problem in (5.23) is given by

$$\lambda = \sigma_N \left(\sqrt{M/N} + 1 \right) \max_p \sqrt{M_p}, \quad (5.27)$$

forming another contribution of this thesis. The regularization parameter in (5.27) has been derived under similar assumptions as the regularization parameter in (4.40) for application with $\ell_{2,1}$ minimization in FCAs discussed in Section 4.2.1, i.e., for large number of sensors M and snapshots N . Ideally, in the case of a single subarray $P = 1$ and a single snapshot $N = 1$ the regularization parameter would reduce to the standard approach for ℓ_1 minimization, given in (4.15), which is clearly not the case for the approach in (5.27). To this end a heuristic approach of selecting the regularization parameter as

$$\lambda = \sigma_N \sqrt{\log(M)} \max_p \sqrt{M_p}, \quad (5.28)$$

will be considered for application with low snapshot number N , and the estimation performance will be further investigated by numerical experiments in Section 5.5.1.

For simple implementation of the SSR problem in (5.23) by standard convex solvers, such as SeDuMi [Stu99] or MOSEK [MOS15], consider the following semidefinite characterization of the nuclear norm. As discussed in [FHB01], minimization of the nuclear norm

$$\min_{\mathbf{Z}_k \in \mathcal{C}} \|\mathbf{Z}_k\|_*, \quad (5.29)$$

for $\mathbf{Z}_k \in \mathbb{C}^{P \times N}$ and some convex set \mathcal{C} , can be expressed as the SDP

$$\min_{\substack{\mathbf{Z}_k \in \mathcal{C}, \\ \mathbf{W}_{k,1}, \mathbf{W}_{k,2}}} \frac{1}{2} (\text{Tr}(\mathbf{W}_{k,1}) + \text{Tr}(\mathbf{W}_{k,2})) \quad (5.30a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_{k,1} & \mathbf{Z}_k \\ \mathbf{Z}_k^H & \mathbf{W}_{k,2} \end{bmatrix} \succeq \mathbf{0}, \quad (5.30b)$$

where $\mathbf{W}_{k,1} = \mathbf{W}_{k,1}^H$ and $\mathbf{W}_{k,2} = \mathbf{W}_{k,2}^H$ are auxiliary variables of size $P \times P$ and $N \times N$, respectively. Based on the equivalence of (5.29) and (5.30), the $\ell_{*,1}^{(P \times N)}$ mixed-norm minimization problem (5.23) can equivalently be formulated as

$$\min_{\substack{\{\mathbf{Z}_k, \mathbf{W}_{k,1}, \\ \mathbf{W}_{k,2}\}}} \|\mathbf{BZ} - \mathbf{Y}\|_F^2 + \lambda \sqrt{N} \sum_{k=1}^K \text{Tr}(\mathbf{W}_{k,1}) + \text{Tr}(\mathbf{W}_{k,2}) \quad (5.31a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_{k,1} & \mathbf{Z}_k \\ \mathbf{Z}_k^H & \mathbf{W}_{k,2} \end{bmatrix} \succeq \mathbf{0}, \text{ for } k = 1, \dots, K. \quad (5.31b)$$

Note that with the auxiliary variables in the Hermitian matrices $\mathbf{W}_{k,1}$ and $\mathbf{W}_{k,2}$, for $k = 1, \dots, K$, the problem (5.31) has $K(P + N)^2$ real-valued optimization variables as opposed to the problem formulation in (5.23) which has $2KPN$ real-valued optimization variables. For a large number of grid points K or snapshots N the SDP formulation (5.31) becomes computationally intractable and alternative implementations are required. Similar to the case of fully calibrated arrays discussed in Section 4.2.1, dimensionality reduction techniques can be applied to reduce the dimensions of the measurement matrix \mathbf{Y} , as suggested in [SSPH16], however, at the expense of potential performance degradation, e.g., in the case of correlated source signals. To overcome this difficulty, the following section provides a compact reformulation of problem (5.23).

5.3 A Compact Formulation for $\ell_{*,1}$ Minimization

Another major contribution of this thesis is formulated in the following theorem:

Theorem 5.1. *The rank- and block-sparsity inducing $\ell_{*,1}^{(P \times N)}$ mixed-norm minimization problem*

$$\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{BZ} - \mathbf{Y}\|_F^2 + \lambda \sqrt{N} \|\mathbf{Z}\|_{*,1}^{(P \times N)} \quad (5.32)$$

is equivalent to the convex problem

$$\min_{\mathbf{S} \in \mathcal{B}_{P+}^K} \text{Tr}((\mathbf{B}\mathbf{S}\mathbf{B}^H + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}), \quad (5.33)$$

with $\hat{\mathbf{R}} = \mathbf{Y}\mathbf{Y}^H/N$ denoting the sample covariance matrix and \mathcal{B}_{P+}^K representing the set of positive semidefinite block-diagonal matrices composed of K blocks of size $P \times P$. The equivalence holds in the sense that a minimizer $\hat{\mathbf{Z}}$ for problem (5.32) can be factorized as

$$\hat{\mathbf{Z}} = \hat{\mathbf{S}}\mathbf{B}^H(\mathbf{B}\hat{\mathbf{S}}\mathbf{B}^H + \lambda\mathbf{I})^{-1}\mathbf{Y}, \quad (5.34)$$

where $\hat{\mathbf{S}}$ is a minimizer for problem (5.33).

A proof of Theorem 5.1 is provided in Appendix G, while a proof of the convexity of (5.33) is provided below in Corollaries 5.2 and 5.3, by establishing equivalence to a semidefinite program.

In addition to relation (5.34), it can be shown (see Appendix G) that a minimizer $\hat{\mathbf{S}} = \text{blkdiag}(\hat{\mathbf{S}}_1, \dots, \hat{\mathbf{S}}_K)$ of (5.33) relates to the coherent signal matrix $\hat{\mathbf{Z}} = [\hat{\mathbf{Z}}_1^T, \dots, \hat{\mathbf{Z}}_K^T]^T$ according to

$$\hat{\mathbf{S}}_k = \frac{1}{\sqrt{N}}(\hat{\mathbf{Z}}_k\hat{\mathbf{Z}}_k^H)^{1/2}, \quad (5.35)$$

for $k = 1, \dots, K$, such that the block support $\{k | \hat{\mathbf{Z}}_k \neq \mathbf{0}, k = 1, \dots, K\}$ of $\hat{\mathbf{Z}}$ is equivalently represented by the block support of the matrix $[\hat{\mathbf{S}}_1^T, \dots, \hat{\mathbf{S}}_K^T]^T$. Similarly, the rank of the matrix blocks $\hat{\mathbf{Z}}_k$ is equivalently represented by the matrix blocks $\hat{\mathbf{S}}_k$, i.e., $\text{rank}(\hat{\mathbf{Z}}_k) = \text{rank}(\hat{\mathbf{S}}_k)$, for $k = 1, \dots, K$.

The problem in (5.33) only relies on the measurement matrix \mathbf{Y} through the sample covariance matrix $\hat{\mathbf{R}}$, leading to a significantly reduced problem size, especially in the case of large number of snapshots N . The formulation (5.33) contains KP^2 real-valued optimization parameters in the positive semidefinite matrix \mathbf{S} , as opposed to the $2KPN$ real-valued optimization parameters in the coherent signal matrix \mathbf{Z} in problem (5.32). Consequently, in the case of a large number of snapshots $N > P/2$, the reformulation (5.33) has reduced number of variables as compared to (5.32). In this context the formulation in (5.33) is termed as COmpact Block- and RAnk-Sparse recovery (COBRAS). Just like the $\ell_{*,1}$ minimization problem reduces to $\ell_{2,1}$ minimization in the case of a single subarray, as discussed in Section 5.2.1, the COBRAS formulation (5.33) reduces to the SPARROW formulation (4.51) in the case of a single subarray. Conversely, the COBRAS formulation can similarly be derived from the SPARROW formulation, as will be discussed in Section 5.3.1.

In order to solve the COBRAS formulation in (5.33) by means of a tractable SDP, which can be treated by standard convex solvers, consider the following corollaries on the reformulation of the matrix inverse expressions in (5.33) to linear matrix inequalities, cf. [VB96, p.56]:

Corollary 5.2. *The COBRAS formulation in (5.33) is equivalent to the convex semidefinite program*

$$\min_{\mathbf{S}, \mathbf{W}_N} \frac{1}{N} \text{Tr}(\mathbf{W}_N) + \text{Tr}(\mathbf{S}) \quad (5.36a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_N & \mathbf{Y}^H \\ \mathbf{Y} & \mathbf{B}\mathbf{S}\mathbf{B}^H + \lambda\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \quad (5.36b)$$

$$\mathbf{S} \in \mathcal{B}_{P+}^K, \quad (5.36c)$$

where \mathbf{W}_N is a Hermitian matrix of size $N \times N$.

Corollary 5.3. *The COBRAS formulation in (5.33) admits the equivalent problem formulation*

$$\min_{\mathbf{S}, \mathbf{W}_M} \text{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \quad (5.37a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I}_M \\ \mathbf{I}_M & \mathbf{B}\mathbf{S}\mathbf{B}^H + \lambda\mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (5.37b)$$

$$\mathbf{S} \in \mathcal{B}_{P+}^K \quad (5.37c)$$

where \mathbf{W}_M is a Hermitian matrix of size $M \times M$.

The proofs to Corollaries 5.2 and 5.3 follow similar arguments as in the proof of Corollary 4.3 and are therefore omitted here. In contrast to (5.36), the size of the semidefinite constraint in (5.37) is independent of the number of snapshots N . It follows that either problem formulation (5.36) or (5.37) can be selected to solve (5.33), depending on the number of snapshots N and the resulting size of the semidefinite constraint. The problems (5.36) and (5.37) have KP^2 real-valued optimization variables in \mathbf{S} and additional N^2 or M^2 real-valued parameters in \mathbf{W}_N and \mathbf{W}_M , respectively. Thus, in the undersampled case $N < M$ it is preferable to use the SDP formulation in (5.36), while in the oversampled case $N \geq M$ it is preferable to apply the SDP formulation in (5.37). Note that the dimensionality reduction techniques discussed in Section 4.2.1 can be applied to formulation (5.36), leading to a reduction in the number of optimization parameters.

5.3.1 Gridless COBRAS

While the SDP formulations in Section 5.3 are applicable to arbitrary array topologies and rely on grid-based SSR, consider in the following the special case of identical uniform linear subarrays, introduced in Section 2.2.3. As seen for the RARE and ESPRIT methods discussed in Sections 3.3 and 3.4, this type of array admits the application of gridless frequency estimation methods, and similar gridless estimation can be obtained for the COBRAS formulation as discussed in the following.

Assume a PCA composed of P identical uniform linear subarrays with M_0 sensors per subarray, and consider the matrix $\tilde{\mathbf{Q}} = \mathbf{B}\tilde{\mathbf{S}}\mathbf{B}^H = \sum_{k=1}^K \mathbf{B}_k \tilde{\mathbf{S}}_k \mathbf{B}_k^H$, as used in the

COBRAS formulation (5.33), where \mathbf{B}_k denotes the subarray response block matrix $\mathbf{B}(\nu_k)$ for sampled frequency ν_k . The matrix $\check{\mathbf{Q}}$ can be partitioned into P^2 submatrices $\check{\mathbf{Q}}_{p,q}$ of size $M_0 \times M_0$ according to

$$\check{\mathbf{Q}} = \begin{bmatrix} \check{\mathbf{Q}}_{1,1} & \check{\mathbf{Q}}_{2,1}^H & \cdots & \check{\mathbf{Q}}_{P,1}^H \\ \check{\mathbf{Q}}_{2,1} & \check{\mathbf{Q}}_{2,2} & & \vdots \\ \vdots & & \ddots & \\ \check{\mathbf{Q}}_{P,1} & \cdots & & \check{\mathbf{Q}}_{P,P} \end{bmatrix}. \quad (5.38)$$

Recall that the subarray response block matrices \mathbf{B}_k , for $k = 1, \dots, K$, have a block-diagonal structure according to $\mathbf{B}_k = \mathbf{I}_P \otimes \mathbf{a}_k^{(0)}$, as defined in (2.13), where $\mathbf{a}_k^{(0)}$ denotes the subarray response $\mathbf{a}^{(0)}(\nu_k)$ for sampled spatial frequency ν_k . According to (5.35) and (5.18), the blocks $\check{\mathbf{S}}_k$ on the diagonal of $\check{\mathbf{S}}$ can be decomposed as

$$\begin{aligned} \check{\mathbf{S}}_k &= \frac{1}{\sqrt{N}} (\check{\mathbf{Z}}_k \check{\mathbf{Z}}_k^H)^{1/2} = \frac{\|\check{\mathbf{x}}_k\|_2}{\sqrt{N} \|\boldsymbol{\varphi}_k\|_2} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^H \\ &= \check{\sigma}_k \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^H, \end{aligned} \quad (5.39)$$

where $\boldsymbol{\varphi}_k = [\varphi_{k,1}, \dots, \varphi_{k,P}]^T$ represents the subarray shift vectors $\boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})$, and $\check{\sigma}_k = \|\check{\mathbf{x}}_k\|_2 / \sqrt{N} / \|\boldsymbol{\varphi}_k\|_2$ denotes the scaled signal magnitudes, for sampled frequency ν_k , respectively, where $k = 1, \dots, K$. With above notation, rewrite matrix $\check{\mathbf{Q}}$ as

$$\begin{aligned} \check{\mathbf{Q}} &= \sum_{k=1}^K \mathbf{B}_k \check{\mathbf{S}}_k \mathbf{B}_k^H = \sum_{k=1}^K \check{\sigma}_k \mathbf{B}_k \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^H \mathbf{B}_k^H \\ &= \sum_{k=1}^K \check{\sigma}_k \begin{bmatrix} |\varphi_{k,1}|^2 \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} & \varphi_{k,1} \varphi_{k,2}^* \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} & \cdots & \varphi_{k,1} \varphi_{k,P}^* \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} \\ \varphi_{k,2} \varphi_{k,1}^* \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} & |\varphi_{k,2}|^2 \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} & & \vdots \\ \vdots & & \ddots & \\ \varphi_{k,P} \varphi_{k,1}^* \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} & \cdots & & |\varphi_{k,P}|^2 \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H} \end{bmatrix}. \end{aligned} \quad (5.40)$$

From (5.40) it can be seen that the main block-diagonal of $\check{\mathbf{Q}}$ contains a convex combination of Hermitian Toeplitz matrices of the form

$$\check{\mathbf{Q}}_{p,p} = \text{HToep}(\mathbf{w}^{(p)}) = \sum_{k=1}^K \check{\sigma}_k |\varphi_{k,p}|^2 \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H}, \quad (5.41)$$

resulting in Hermitian Toeplitz submatrices $\check{\mathbf{Q}}_{p,p}$, for $p = 1, \dots, P$, where $\text{HToep}(\mathbf{w}^{(p)})$ is defined in (4.22). Note that the elements of vector $\boldsymbol{\varphi}_k$ can have different magnitude, i.e., $|\varphi_{k,p}| = |\alpha^{(p)}|$ can be different from $|\varphi_{k,q}| = |\alpha^{(q)}|$ for $p \neq q$, corresponding to perturbations among the different subarrays. In consequence, two submatrices $\check{\mathbf{Q}}_{p,p} = \text{HToep}(\mathbf{w}^{(p)})$ and $\check{\mathbf{Q}}_{q,q} = \text{HToep}(\mathbf{w}^{(q)})$ on the main block-diagonal, for $p \neq q$, must not be identical. The off-diagonal blocks off $\check{\mathbf{Q}}$ contain a linear combination of Hermitian Toeplitz matrices of the form

$$\check{\mathbf{Q}}_{p,q} = \text{Toep}(\mathbf{w}^{(p,q)}) = \sum_{k=1}^K \check{\sigma}_k \varphi_{k,p} \varphi_{k,q}^* \mathbf{a}_k^{(0)} \mathbf{a}_k^{(0)H}, \quad (5.42)$$

resulting in Toeplitz submatrices $\check{\mathbf{Q}}_{p,q}$, for $p, q = 1, \dots, P$ with $p \neq q$, defined as

$$\text{Toep}(\mathbf{w}) = \begin{bmatrix} w_0 & w_1 & w_2 & \cdots & w_{M_0-1} \\ w_{-1} & w_0 & w_1 & \cdots & w_{M_0-2} \\ w_{-2} & w_{-1} & w_0 & \cdots & w_{M_0-3} \\ \vdots & & & \ddots & \vdots \\ w_{-M_0+1} & & & & w_0 \end{bmatrix}, \quad (5.43)$$

for a $2M_0 - 1$ element vector $\mathbf{w} = [w_{-M_0+1}, w_{-M_0+2}, \dots, w_{M_0-2}, w_{M_0-1}]^\top$.

The Toeplitz-block structure of matrix $\check{\mathbf{Q}}$ can be exploited in a gridless implementation of the COBRAS formulation (5.33) according to

$$\min_{\{\mathbf{w}^{(p)}, \mathbf{w}^{(p,q)}\}} \text{Tr}((\mathbf{Q} + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \frac{1}{M_0} \text{Tr}(\mathbf{Q}) \quad (5.44a)$$

$$\text{s.t. } \mathbf{Q}_{p,p} = \text{HToep}(\mathbf{w}^{(p)}) \succeq \mathbf{0}, \quad \text{for } p = 1, \dots, P \quad (5.44b)$$

$$\mathbf{Q}_{p,q} = \text{Toep}(\mathbf{w}^{(p,q)}), \quad \text{for } p, q = 1, \dots, P, p \neq q \quad (5.44c)$$

$$\mathbf{Q} \succeq \mathbf{0}, \quad (5.44d)$$

where additionally the identity $\text{Tr}(\mathbf{S}) = 1/M_0 \text{Tr}(\mathbf{B}\mathbf{S}\mathbf{B}^H) = 1/M_0 \text{Tr}(\mathbf{Q})$ is used. Careful observation shows that the matrix \mathbf{Q} is described by $P(2M - P)$ real-valued parameters, i.e., the variables describing the distinct Toeplitz submatrices. For implementation with standard convex solvers, the problem in (5.44) can be reformulated as an SDP, as discussed for Corollaries 5.2 and 5.3.

The case of a PCA of M sensors composed of P thinned linear subarrays can be addressed by application of a proper selection matrix $\check{\mathbf{J}}$ of size $PM_0 \times M$, selecting the active sensors from a virtual PCA of P identical subarrays of M_0 sensors, as discussed in Section 2.2.4. To this end the problem in (5.44) is modified according to

$$\min_{\{\mathbf{w}^{(p)}, \mathbf{w}^{(p,q)}\}} \text{Tr}((\check{\mathbf{J}}^\top \mathbf{Q} \check{\mathbf{J}} + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \frac{1}{M_0} \text{Tr}(\mathbf{Q}) \quad (5.45a)$$

$$\text{s.t. } \mathbf{Q}_{p,p} = \text{HToep}(\mathbf{w}^{(p)}) \succeq \mathbf{0}, \quad \text{for } p = 1, \dots, P \quad (5.45b)$$

$$\mathbf{Q}_{p,q} = \text{Toep}(\mathbf{w}^{(p,q)}), \quad \text{for } p, q = 1, \dots, P, p \neq q \quad (5.45c)$$

$$\mathbf{Q} \succeq \mathbf{0}, \quad (5.45d)$$

where \mathbf{Q} is the $PM_0 \times PM_0$ Toeplitz-block matrix corresponding to the virtual PCA of uniform linear subarrays.

Given a minimizer $\hat{\mathbf{Q}}$ to the gridless COBRAS formulation (5.44), the estimated number of signals is given by

$$\hat{L} = \text{rank}(\hat{\mathbf{Q}}), \quad (5.46)$$

and depends on the choice of the regularization parameter λ . The \hat{L} spatial frequencies $\{\mu_l\}_{l=1}^{\hat{L}}$ can be recovered from the minimizer $\hat{\mathbf{Q}}$, e.g., by the matrix pencil

method [HS90, HW91] or the subspace-based methods root-RARE [PGW02] and ESPRIT [RK89]. It should be noted that a frequency estimate $\hat{\mu}_l$ can have multiplicity greater than one, e.g., if a generalized eigenvalue obtained in the matrix pencil approach has multiplicity greater than one. For M sensors partitioned into P identical subarrays, the matrix pencil [HS90, HW91], root-RARE [PGW02] and ESPRIT [RK89] methods can identify at most

$$L \leq M - P \quad (5.47)$$

spatial frequencies, as discussed in Sections 3.3 and 3.4.

For estimation of the displacement shifts in $\varphi(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})$ from the matrix $\hat{\mathbf{Q}}$, consider the parameterization in (5.39) to formulate

$$\hat{\mathbf{Q}} = \sum_{l=1}^{\hat{L}} \hat{\sigma}_l \mathbf{B}(\hat{\mu}_l) \hat{\boldsymbol{\varphi}}_l \hat{\boldsymbol{\varphi}}_l^H \mathbf{B}^H(\hat{\mu}_l), \quad (5.48)$$

and assume without loss of generality that $[\hat{\boldsymbol{\varphi}}_l]_1 = 1$ and $[\mathbf{B}^H(\hat{\mu}_l)]_{1,1} = 1$, i.e., the first sensor of the first array is the overall phase reference. Under these assumptions, the first column $\hat{\mathbf{q}}_1$ of $\hat{\mathbf{Q}} = [\hat{\mathbf{q}}_1, \dots, \hat{\mathbf{q}}_M]$ is modeled as

$$\hat{\mathbf{q}}_1 = \sum_{l=1}^{\hat{L}} \hat{\sigma}_l \mathbf{B}(\hat{\mu}_l) \hat{\boldsymbol{\varphi}}_l \quad (5.49)$$

$$= \mathbf{B}(\hat{\boldsymbol{\mu}}) \tilde{\boldsymbol{\varphi}}, \quad (5.50)$$

where $\mathbf{B}(\hat{\boldsymbol{\mu}}) = [\mathbf{B}(\hat{\mu}_1), \dots, \mathbf{B}(\hat{\mu}_{\hat{L}})]^T$ is of dimensions $M \times \hat{L}P$ and the $\hat{L}P \times 1$ vector $\tilde{\boldsymbol{\varphi}} = [\hat{\sigma}_1 \hat{\boldsymbol{\varphi}}_1^T, \dots, \hat{\sigma}_{\hat{L}} \hat{\boldsymbol{\varphi}}_{\hat{L}}^T]^T$ contains the signal magnitudes $\hat{\sigma}_l$ and displacement shifts $\hat{\boldsymbol{\varphi}}_l$ of interest, for $l = 1, \dots, \hat{L}$. Given the frequency estimates $\hat{\mu}_1, \dots, \hat{\mu}_{\hat{L}}$, the signal magnitudes $\hat{\sigma}_l$ and the displacement shifts $\hat{\boldsymbol{\varphi}}_l$, the signal matrix $\hat{\mathbf{Z}}$ can be reconstructed according to (5.34).

Frequency Estimation from the Dual Problem

Similar to the gridless SPARROW formulation in Section 4.3.1, the gridless COBRAS can be implemented by its dual problem, which provides a new perspective on the COBRAS problem. The Lagrange dual problem of the COBRAS formulation (5.37) is given as

$$\max_{\mathbf{r}_1, \mathbf{r}_0} -2 \operatorname{Re}\{\operatorname{Tr}(\boldsymbol{\Upsilon}_1)\} - \lambda \operatorname{Tr}(\boldsymbol{\Upsilon}_0) \quad (5.51a)$$

$$\text{s.t.} \quad \begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_1 \\ \boldsymbol{\Upsilon}_1^H & \boldsymbol{\Upsilon}_0 \end{bmatrix} \succeq \mathbf{0} \quad (5.51b)$$

$$\mathbf{I}_P - \mathbf{B}^H(\nu_k) \boldsymbol{\Upsilon}_0 \mathbf{B}(\nu_k) \succeq \mathbf{0}, \quad k = 1, \dots, K, \quad (5.51c)$$

where \mathbf{Y}_0 is an $M \times M$ positive semidefinite matrix and \mathbf{Y}_1 is of size $M \times M$ and does not exhibit specific structure (see Appendix I for more details). Note that the primal problem in (5.37) as well as the dual problem in (5.51) are both strictly feasible, such that Slater's condition holds and strong duality applies [VB96]. Complementary slackness requires that

$$\text{Tr}(\mathbf{S}_k(\mathbf{I}_P - \mathbf{B}^H(\nu_k)\mathbf{Y}_0\mathbf{B}(\nu_k))) = 0, \quad (5.52)$$

for $k = 1, \dots, K$, i.e., if $\mathbf{S}_k \neq \mathbf{0}$ then $\mathbf{I}_P - \mathbf{B}^H(\nu_k)\mathbf{Y}_0\mathbf{B}(\nu_k)$ must be singular, such that

$$\det(\mathbf{I}_P - \mathbf{B}^H(\nu_k)\mathbf{Y}_0\mathbf{B}(\nu_k)) \begin{cases} = 0 & \text{if } \mathbf{S}_k \neq \mathbf{0} \\ \geq 0 & \text{if } \mathbf{S}_k = \mathbf{0}. \end{cases} \quad (5.53)$$

Condition (5.53) indicates that instead of solving the primal problem (5.33) and identifying the block support from \mathbf{S} , the dual problem (5.51) can be solved equivalently to identify the block support from the roots of (5.53).

Consider the limiting case of an infinitesimal spacing in the grid of spatial frequencies, i.e., $\nu_k - \nu_{k-1} = 2/K \rightarrow 0$ for $k = 2, \dots, K$, such that the frequency becomes a continuous parameter ν . Further consider the case of identical uniform linear subarrays, as discussed in the previous section and in Section 2.2.3, such that the subarray steering block matrix $\mathbf{B}(z)$ is given by (2.25) according to

$$\mathbf{B}(z) = \mathbf{I}_P \otimes [1 \quad z \quad \dots \quad z^{M_0-1}]^T, \quad (5.54)$$

where $z = e^{-j\pi\nu\Delta}$. Based on the definition in (5.54) it can be concluded that the matrix product $\mathbf{B}^H(z)\mathbf{Y}_0\mathbf{B}(z)$ in constraint (5.51c) constitutes a trigonometric matrix polynomial of degree $2(M_0 - 1)$, according to

$$\mathbf{F}_{\text{COBRAS}}(z) = \mathbf{B}^H(z)\mathbf{Y}_0\mathbf{B}(z) = \sum_{m=-M_0}^{M_0} \mathbf{C}_m z^{-m}, \quad (5.55)$$

with matrix coefficients \mathbf{C}_m of size $P \times P$ [Dum07]. For continuous frequency estimation, the K constraints (5.51c) are replaced by the continuous constraint

$$\mathbf{I}_q - \mathbf{B}^H(z)\mathbf{Y}_0\mathbf{B}(z) \succeq \mathbf{0}, \quad (5.56)$$

which provides an upper bound on the matrix polynomial (5.55) and can be implemented by semidefinite programming, e.g., by the problem formulation (H.10) derived in Appendix H [Dum07]. Once the continuous implementation of problem (5.51) is solved, the spatial frequencies can be recovered by finding the roots for which the left-hand side of (5.56) becomes singular, e.g., by rooting the continuous counterpart of (5.53) using the techniques discussed in [Pes05], e.g., by the block companion matrix approach.

Relation to SPARROW

The dual problem formulation (5.51) can equivalently be derived by means of the SPARROW formulation presented in Section 4.3. To do so, formulate the SPARROW problem in (4.51) by means of the PCA sensing matrix $\mathbf{A} = \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$, defined in correspondence with (2.11), as

$$\min_{\mathbf{S} \in \mathcal{D}_+^K} \text{Tr}((\mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})\mathbf{S}\mathbf{A}^H(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) + \tilde{\lambda}\mathbf{I})^{-1}\hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}), \quad (5.57)$$

where $\tilde{\lambda} > 0$ is a regularization parameter and $\mathbf{S} = \text{diag}(s_1, \dots, s_K) \succeq \mathbf{0}$ is of size $K \times K$. As shown by Corollary 4.4, problem (5.57) can be formulated as the semidefinite program

$$\min_{\mathbf{S}, \mathbf{W}_M} \text{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \quad (5.58a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I} \\ \mathbf{I} & \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})\mathbf{S}\mathbf{A}^H(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) + \tilde{\lambda}\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \quad (5.58b)$$

$$\mathbf{S} \in \mathcal{D}_+^K.$$

with the corresponding Lagrange dual problem given as

$$\max_{\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_0} -2 \text{Re}\{\text{Tr}(\tilde{\mathbf{r}}_1)\} - \tilde{\lambda} \text{Tr}(\tilde{\mathbf{r}}_0) \quad (5.59a)$$

$$\text{s.t.} \quad \begin{bmatrix} \hat{\mathbf{R}} & \tilde{\mathbf{r}}_1 \\ \tilde{\mathbf{r}}_1^H & \tilde{\mathbf{r}}_0 \end{bmatrix} \succeq \mathbf{0} \quad (5.59b)$$

$$\mathbf{a}^H(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})\tilde{\mathbf{r}}_0 \mathbf{a}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \leq 1, \quad k = 1, \dots, K. \quad (5.59c)$$

Note that the primal problem (5.58) and the dual problem (5.59) are strictly feasible, such that Slater's condition is fulfilled and strong duality applies. As discussed in Section 4.3.1, complementary slackness requires that

$$1 - \mathbf{a}^H(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})\tilde{\mathbf{r}}_0 \mathbf{a}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \begin{cases} = 0 & \text{if } s_k \geq 0 \\ \geq 0 & \text{if } s_k = 0, \end{cases} \quad (5.60)$$

such that the support of the vector $\mathbf{s} = [s_1, \dots, s_K]^T$, representing the spatial frequency estimates, can equivalently be identified by rooting the function in (5.60).

Making use of the parameterization $\mathbf{a}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \mathbf{B}(\nu_k)\boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})$, as introduced in (2.12), condition (5.60) can be rewritten as

$$\begin{aligned} 1 - \mathbf{a}^H(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})\tilde{\mathbf{r}}_0 \mathbf{a}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) &= 1 - \boldsymbol{\varphi}^H(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{B}^H(\nu_k) \tilde{\mathbf{r}}_0 \mathbf{B}(\nu_k) \boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \\ &= 1 - \tilde{\boldsymbol{\varphi}}^H(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{B}^H(\nu_k) \tilde{\mathbf{r}}_0 \mathbf{B}(\nu_k) \tilde{\boldsymbol{\varphi}}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \\ &= \tilde{\boldsymbol{\varphi}}^H(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) (\mathbf{I}_p - \mathbf{B}^H(\nu_k) \tilde{\mathbf{r}}_0 \mathbf{B}(\nu_k)) \tilde{\boldsymbol{\varphi}}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) \\ &\geq 0, \end{aligned} \quad (5.61)$$

where

$$\tilde{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta}) / \|\boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})\|_2 \quad (5.62)$$

$$\boldsymbol{\Upsilon}_0 = \|\boldsymbol{\varphi}(\nu_k, \boldsymbol{\eta})\|_2^2 \tilde{\boldsymbol{\Upsilon}}_0. \quad (5.63)$$

Condition (5.61) is fulfilled if

$$\mathbf{I}_q - \mathbf{B}^H(\nu_k) \boldsymbol{\Upsilon}_0 \mathbf{B}(\nu_k) \succeq \mathbf{0}, \quad (5.64)$$

which is identical to the constraint (5.51c) in problem (5.51). Replacing the constraint (5.59c) in problem (5.59) by the condition (5.64), and further using (5.63) and the substitutions $\lambda = \tilde{\lambda} / \|\boldsymbol{\varphi}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})\|_2^2$ and $\boldsymbol{\Upsilon}_1 = \tilde{\boldsymbol{\Upsilon}}_1$, shows that for the PCA case the dual problem of the SPARROW formulation in (5.59) can be reformulated as the dual problem of the COBRAS formulation in (5.51). As demonstrated in the previous section on frequency estimation from the dual COBRAS problem, condition (5.64) can be extended to an infinitesimal grid spacing, resulting in a matrix polynomial constraint, such that the resulting gridless estimation problem can be implemented by semidefinite programming (see Appendix H).

5.4 Coherent Sparse Reconstruction for Shift-Invariant Arrays

The considerations in Section 5.3.1 have shown that PCAs composed of identical uniform linear subarrays exhibit structure that can be exploited in the sparse reconstruction process. The case of PCAs composed of identical uniform linear subarrays falls into the more general class of shift-invariant arrays, introduced in Section 2.2.5. It is well known that PCAs composed of identical and identically oriented subarrays exhibit multiple shift-invariances that can be exploited for parameter estimation [RK89, SORK92, PPG11, PP11, PPG12, SPZ13, SPZ16]. This section investigates how the shift-invariant structure can be exploited in sparse reconstruction.

For ease of presentation, consider the example from Section 2.2.5, as displayed in Figure 5.1, consisting of a linear array of $M = PM_0$ sensors partitioned into P identical subarrays of M_0 sensors. Note that the subarrays must not be uniform linear, but can have arbitrary topology, as long as any two subarrays have identical structure. Let the sensor positions $r_m^{(p)}$ fulfill the shift-invariance property in (2.27), where the relative sensor positions ρ_m within each subarray are perfectly known while the displacements $\eta^{(p)}$ between the different subarrays are assumed to be unknown, for $m = 1, \dots, M_0$ and $p = 1, \dots, P$.

The measurements obtained at the array output are modeled by

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \boldsymbol{\Psi} + \mathbf{N}, \quad (5.65)$$

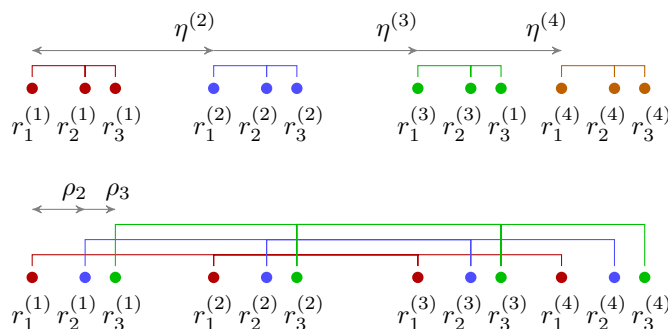


Figure 5.1: Illustration of multiple shift-invariances in array topology

as described for (2.10), where $\Psi \in \mathbb{C}^{L \times N}$ denotes the source signal matrix, $\mathbf{N} \in \mathbb{C}^{M \times N}$ represents additive sensor noise and the $M \times L$ array steering matrix $\mathbf{A}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ is defined according to (2.11). As stated by Theorem 4.2, sparse reconstruction for the model in (5.65) can be performed by application of the SPARROW problem

$$\min_{\mathbf{S} \in \mathcal{D}_+^K} \text{Tr}((\mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{S} \mathbf{A}^H(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}), \quad (5.66)$$

for some frequency grid $\boldsymbol{\nu} = [\nu_1, \dots, \nu_K]^T$. In this section no additional subarray perturbations are assumed, i.e., $\boldsymbol{\alpha} = \mathbf{1}$, such that the array response vectors defined in (2.12) have constant norm $\|\mathbf{a}(\nu_k, \boldsymbol{\alpha}, \boldsymbol{\eta})\|_2 = M$, for $k = 1, \dots, K$, and it holds that

$$\text{Tr}(\mathbf{S}) = \frac{1}{M} \text{Tr}(\mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{S} \mathbf{A}^H(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})). \quad (5.67)$$

Using (5.67), the SPARROW problem (5.66) can be further rewritten as

$$\min_{\mathbf{Q} \in \mathcal{D}_+^K} \text{Tr}((\mathbf{Q} + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \frac{1}{M} \text{Tr}(\mathbf{Q}) \quad (5.68a)$$

$$\text{s.t. } \mathbf{Q} = \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{S} \mathbf{A}^H(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}). \quad (5.68b)$$

In formulation (5.68) the objective function (5.68a) only depends on the matrix variable \mathbf{Q} , reflecting the specific structure of the sensing matrix $\mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ in the constraint (5.68b). An additional low-rank structure in the minimizer $\hat{\mathbf{Q}}$ is encouraged by the trace-term $\text{Tr}(\mathbf{Q})$ in (5.68a), since for $\mathbf{Q} \succeq \mathbf{0}$ it is equivalent to the nuclear norm of \mathbf{Q} [RFP10], i.e., $\text{Tr}(\mathbf{Q}) = \|\mathbf{Q}\|_*$.

For further investigation consider the constraint (5.68b) in the context of the shift-invariance property of the array sensing matrix $\mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$, as discussed in Section 2.2.5. Let $\mathbf{J}^{(p)}$ denote the $PM_0 \times M_0$ matrix selecting all the sensors in the p th subarray, for $p = 1, \dots, P$, and \mathbf{J}_m denote the $PM_0 \times P$ matrix selecting the m th sensor in each subarray, for $m = 1, \dots, M_0$, as defined in (2.31). By the shift-invariance of the sensor positions (2.27), the sensing matrix fulfills

$$\mathbf{J}^{(p)T} \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \mathbf{J}^{(1)T} \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \bar{\Phi}^{\eta^{(p)}}(\boldsymbol{\nu}) \quad (5.69a)$$

$$\mathbf{J}_m^T \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \mathbf{J}_1^T \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \bar{\Phi}^{\rho_m}(\boldsymbol{\nu}), \quad (5.69b)$$

as derived in (2.33), where the $L \times L$ unitary diagonal matrix

$$\bar{\Phi}(\boldsymbol{\nu}) = \text{diag}(e^{-j\pi\nu_1}, \dots, e^{-j\pi\nu_K}) \quad (5.70)$$

contains the phase shifts for the spatial frequencies in $\boldsymbol{\nu}$ on its main diagonal, in correspondence with (2.34).

Based on the conditions in (5.69a), the matrix product in (5.68b) fulfills the identity

$$\begin{aligned} \mathbf{J}^{(p)\top} \mathbf{Q} \mathbf{J}^{(p)} &= \mathbf{J}^{(p)\top} \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{S} \mathbf{A}^{\text{H}}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{J}^{(p)} \\ &= \mathbf{J}^{(1)\top} \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \bar{\Phi}^{\eta^{(p)}}(\boldsymbol{\nu}) \mathbf{S} \bar{\Phi}^{\eta^{(p)\text{H}}}(\boldsymbol{\nu}) \mathbf{A}^{\text{H}}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{J}^{(1)} \\ &= \mathbf{J}^{(1)\top} \mathbf{A}(\boldsymbol{\nu}) \mathbf{S} \mathbf{A}^{\text{H}}(\boldsymbol{\nu}) \mathbf{J}^{(1)} \\ &= \mathbf{J}^{(1)\top} \mathbf{Q} \mathbf{J}^{(1)}, \end{aligned} \quad (5.71)$$

for $p = 2, \dots, P$, where it holds that $\bar{\Phi}^{\eta^{(p)\text{H}}}(\boldsymbol{\nu}) \mathbf{S} \bar{\Phi}^{\eta^{(p)}}(\boldsymbol{\nu}) = \mathbf{S}$, due to the assumption of unit subarray perturbation $\boldsymbol{\alpha} = \mathbf{1}$. From (5.71) it can be observed that for any two shift-invariant sensor groups the corresponding submatrices in \mathbf{Q} are identical. The same observation holds true for the other groups of shift-invariances expressed in (5.69b), i.e.,

$$\mathbf{J}_m^{\top} \mathbf{Q} \mathbf{J}_m = \mathbf{J}_1^{\top} \mathbf{Q} \mathbf{J}_1, \quad (5.72)$$

for $m = 2, \dots, M_0$. As mentioned in Section 2.2.5, other types of structure, such as overlapping shift-invariant groups or centro-symmetry, may be available in the array topology, which yield similar equality constraints on \mathbf{Q} as given in (5.71) and (5.72).

Returning to the SPARROW formulation (5.68) and replacing the grid-based constraint $\mathbf{Q} = \mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta}) \mathbf{S} \mathbf{A}^{\text{H}}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ in (5.68b) by the structural constraints in (5.71) and (5.72), the gridless SPARROW problem can be formulated as

$$\min_{\mathbf{Q} \succeq \mathbf{0}} \text{Tr}((\mathbf{Q} + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \frac{1}{M} \text{Tr}(\mathbf{Q}) \quad (5.73a)$$

$$\text{s.t. } \mathbf{J}^{(1)\top} \mathbf{Q} \mathbf{J}^{(1)} = \mathbf{J}^{(p)\top} \mathbf{Q} \mathbf{J}^{(p)}, \quad p = 2, \dots, P \quad (5.73b)$$

$$\mathbf{J}_1^{\top} \mathbf{Q} \mathbf{J}_1 = \mathbf{J}_m^{\top} \mathbf{Q} \mathbf{J}_m, \quad m = 2, \dots, M_0. \quad (5.73c)$$

The program in (5.73) does not require knowledge of the overall array response in form of the sensing matrix $\mathbf{A}(\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\eta})$. In fact it does not even require knowledge of any shifts $\eta^{(p)}$, for $p = 2, \dots, P$, and ρ_m , for $m = 2, \dots, M_0$. The only information which is exploited is the composition of the shift-invariant sensor groups. The approach in (5.73) was presented in [SSSP17a] and is referred to as Shift-Invariant (SI-) SPARROW. Note that the SI-SPARROW is applicable to PCAs composed of linear subarrays with arbitrary topology, such that it generalizes the GL-COBAS approach in (5.44), which is applicable to PCAs composed of identical uniform linear subarrays. More precisely, the SI-SPARROW exploits additional structure in the assumption of unit subarray perturbations $\boldsymbol{\alpha} = \mathbf{1}$. However, in the case of arbitrary subarray perturbations in $\boldsymbol{\alpha}$,

the SI-SPARROW (5.73) can still be applied by removing constraint (5.73b) from the program.

Similar to the GL-COBRAS problem (5.44), the SI-SPARROW problem in (5.73) can be solved by special convex solvers by reformulating it as an SDP, as discussed for Corollaries 5.2 and 5.3. The case of thinned shift-invariant arrays can be treated by application of a proper selection matrix \mathbf{J} , similar to (5.45).

Given a minimizer $\hat{\mathbf{Q}}$ to problem (5.73) the underlying spatial frequencies can be recovered using different approaches, depending on the amount of shift-invariances and the exact knowledge that is available on the shifts $\eta^{(p)}$ and ρ_m , for $m = 2, \dots, M_0$ and $p = 2, \dots, P$. Assuming only a single shift structure, e.g. ρ_2 in the case of a PCA composed of 2-element subarrays, the ESPRIT method [RK89] or the matrix pencil method [HS90, HW91] can be applied to the matrix $\hat{\mathbf{Q}}$ to estimate the spatial frequencies $\hat{\boldsymbol{\mu}}$ in a search-free fashion. If knowledge of multiple shifts $\eta^{(p)}$ and ρ_m , for $m = 2, \dots, M_0$ and $p = 2, \dots, P$, are available, more sophisticated methods such as MI-ESPRIT [SORK92] can be applied to $\hat{\mathbf{Q}}$ to obtain improved frequency estimates $\hat{\boldsymbol{\mu}}$.

As discussed above, for a PCA of $M = PM_0$ sensors partitioned into P identical uniform linear subarrays of M_0 sensors with unknown subarray perturbation $\boldsymbol{\alpha}$ the SI-SPARROW coincides with the GL-COBRAS approach discussed in Section 5.3.1. Consequently, at most $L \leq M - P$ frequencies can be uniquely recovered from the estimated matrix $\hat{\mathbf{Q}}$. If multiple shift-invariances are available, as discussed for Figure 3.3, there are additional constraints for the SI-SPARROW formulation (5.73), resulting in less degrees of freedom in the optimization problem. As discussed in Section 3.4, the number of identifiable frequencies L_{\max} in the multiple-invariance case is much more difficult to obtain than in the case of a single shift-invariance.

5.5 Numerical Results

For experimental evaluation of the proposed COBRAS and SI-SPARROW methods, the estimation performance is compared to the state of the art methods for SSR under incoherent processing discussed in Section 5.1, as well as to the state of the art methods for coherent processing discussed in Section 3.3, namely spectral RARE [SG04] and root-RARE [PGW02].

All simulations consider circular complex Gaussian source signals $\boldsymbol{\psi}(t)$ with covariance matrix $\mathbb{E}\{\boldsymbol{\psi}(t)\boldsymbol{\psi}^H(t)\} = \mathbf{I}$, if not specified otherwise. Furthermore, spatio-temporal white circular complex Gaussian sensor noise $\mathbf{n}(t)$ with covariance matrix $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma_N^2 \mathbf{I}$ is assumed and the signal-to-noise ratio (SNR) is defined as $\text{SNR} = 1/\sigma_N^2$. The vector $\mathbf{r}^{(p)} = [r_1^{(p)}, \dots, r_{M_p}^{(p)}]^T$ is defined to contain the global sensor positions $r_m^{(p)}$ of subarray p , for $m = 1, \dots, M_p$, $p = 1, \dots, P$, expressed in half-wavelength, as defined in (2.8). If not stated otherwise, $T = 1000$ Monte Carlo trials

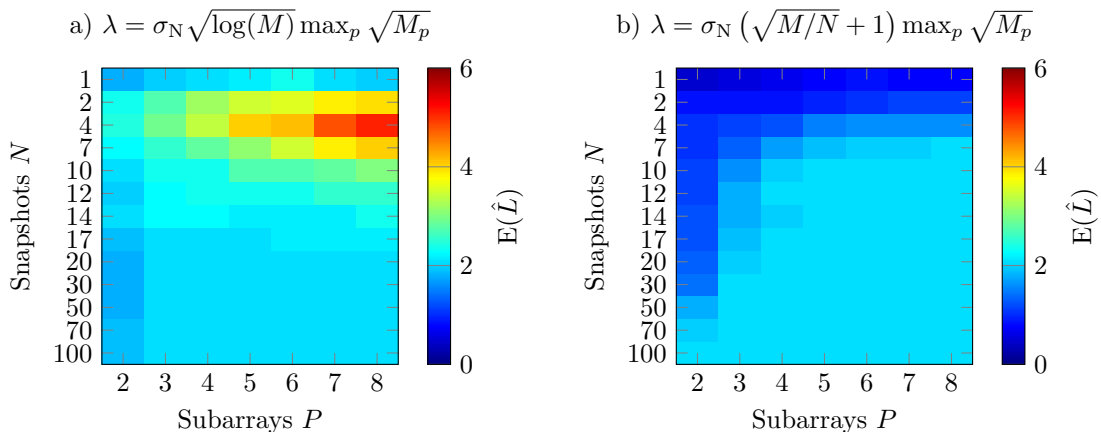


Figure 5.2: Average number of estimated source signals \hat{L} with GL-COBRAS for regularization parameter selection approaches a) (5.28) and b) (5.27), over varying number of snapshots N and subarrays P

are performed for each experimental setup to approximately compute the statistical error.

Note that the RARE and COBRAS/SI-SPARROW methods make different assumptions on the availability of a-priori knowledge. While the RARE method requires knowledge of the number of source signals, the regularization parameter selection for the COBRAS method according to (5.27) and (5.28) requires knowledge of the noise power. However, since estimation of these parameters itself might affect the frequency estimation performance of the RARE and COBRAS / SI-SPARROW methods, the standard assumption of perfectly known number of source signals and noise power is applied and the achievable performance under these idealized assumptions is investigated.

5.5.1 Regularization Parameter Selection

Similar to Section 4.4, the experiments start with a short investigation of the regularization parameter selection to evaluate the reconstruction performance for regularization parameter selection approaches (5.27) and (5.28), as discussed in Section 5.2.1. As before, two equal power source signals are assumed, with source covariance matrix $E\{\boldsymbol{\psi}(t)\boldsymbol{\psi}^H(t)\} = \mathbf{I}$ and spatial frequencies $\mu_1 = 0.5$ and $\mu_2 = 0.4$, and the signal-to-noise ratio is fixed as $\text{SNR} = 0$ dB. In the experimental setup the number of snapshots N and number of subarrays P are varied, where it is assumed that each subarray consists of $M_0 = 3$ sensors and that the overall array has a uniform linear structure. Figure 5.2 shows the average number of estimated source signals \hat{L} over 100 realizations per scenario, where in each realization the source signals and additive noise are generated from a circularly symmetric complex Gaussian random process.

Figure 5.2 illustrates that in the majority of cases, for sufficiently large number of snapshots N and subarrays P , both approaches correctly estimate the number of source signals. As seen from Figure 5.2 a), approach (5.28) tends to overestimate the number of estimated signals \hat{L} for low number of snapshots N and large number of subarrays P , indicating that the regularization parameter λ is selected too low. For a low number of sensors the two closely spaced signals can not always be resolved, leading to underestimation of the number of source signals. As observed from Figure 5.2 b), approach (5.27) tends to underestimate the number of sources for low number of snapshots N , indicating that the regularization parameter λ is selected too large. In the case of a single snapshot, this regularization parameter selection approach estimates a zero signal, i.e., no source signal at all. Similar to the other selection approach, for low number of sensors, the two closely spaced sources cannot be resolved. Since the following numerical experiments focus on difficult scenarios of low number of sensors and snapshots, approach (5.28) will be used for the following simulations.

5.5.2 Resolution Performance and Estimation Bias

For comparison of the resolution performance of the various methods, consider a uniform linear array of $M = 9$ sensors, partitioned into $P = 3$ identical, uniform linear subarrays of $M_0 = 3$ sensors each, without additional gain/phase offsets, i.e., $\boldsymbol{\alpha} = [1, 1, 1]^T$ in (2.10). For the experiment, $L = 2$ uncorrelated signals are considered where the spatial frequency of the first signal is fixed as $\mu_1 = 0.505$ while the spatial frequency of the second signal is varied according to $\mu_2 = \mu_1 - \Delta\mu$ with $10^{-2} \leq \Delta\mu \leq 1$. For all grid-based estimation methods a uniform grid of $K = 200$ points according to $\boldsymbol{\nu} = [-1, -0.99, -0.98, \dots, 0.99]^T$ is applied. The SNR and number of snapshots are fixed as SNR = 0 dB and $N = 20$.

Figure 5.3 shows the RMSE of the frequency estimates as computed according to (4.103). The class of incoherent estimation methods is represented by the root-MUSIC (MUSIC SA1) and GL-SPARROW (GL-SPARROW SA1) methods applied on the measurements of subarray 1, as well as root-MUSIC (MUSIC IP) and GL-SPARROW (GL-SPARROW IP) methods applied on the incoherent processing problem in (5.11). First, it can be observed from Figure 5.3 that the incoherent methods clearly show the worst resolution performance, where the incoherent processing of the entire array measurements (IP) still shows a small improvement over the separate processing of the measurements of subarray 1 only (SA1). Only for large frequency separation of $\Delta\mu \geq 0.6$, incoherent processing approximately achieves the performance of coherent processing. Regarding the coherent methods, it can be observed that spectral RARE performs significantly worse than root-RARE in terms of thresholding performance, i.e., spectral RARE cannot always resolve the two signals for a frequency separation of $\Delta\mu \lesssim 0.4$, while the root-RARE method can resolve the signals for $\Delta\mu \gtrsim 0.12$. The reason for this difference in resolution performance is that the root-RARE method searches the roots of the corresponding matrix polynomial in the entire complex plane, while spectral RARE only searches minima on the unit circle, as discussed in Section

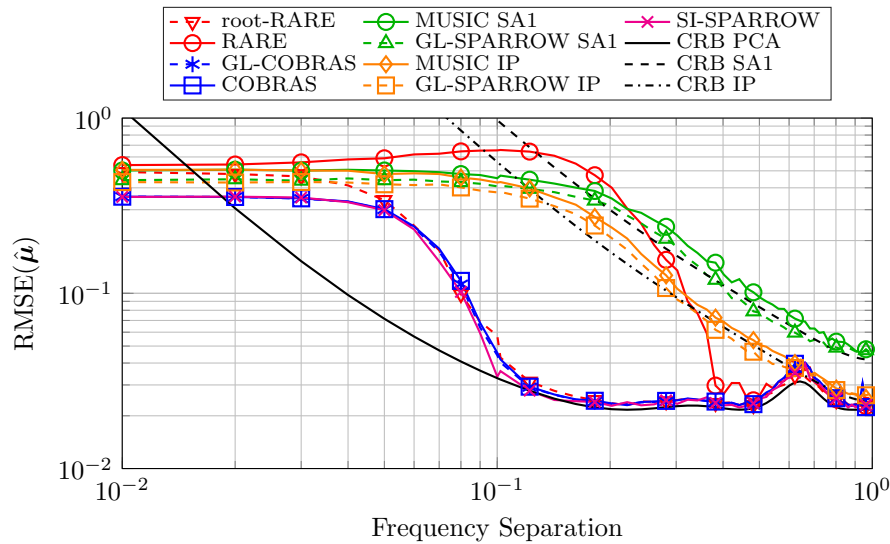


Figure 5.3: Frequency estimation performance for uniform linear PCA of $M = 9$ sensors in $P = 3$ linear subarrays, for $N = 20$ snapshots and $\text{SNR} = 0$ dB

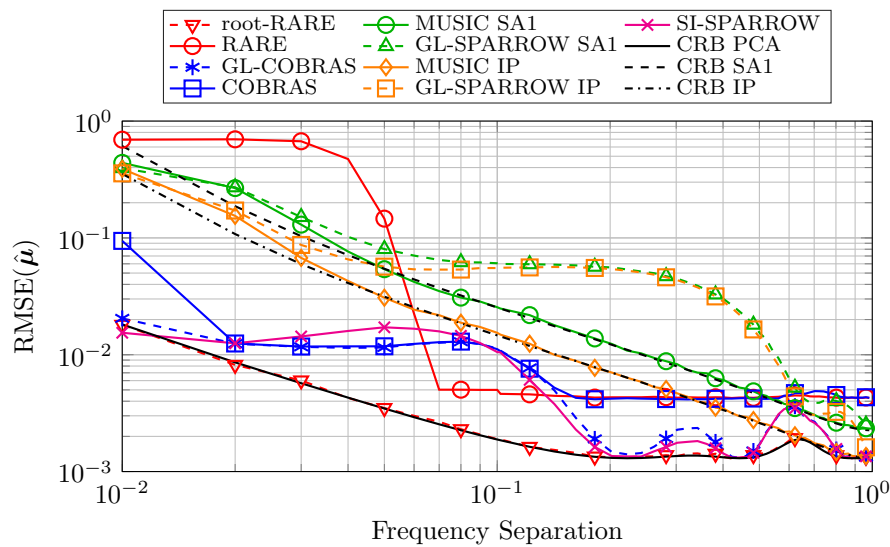


Figure 5.4: Frequency estimation performance for uniform linear PCA of $M = 9$ sensors in $P = 3$ linear subarrays, for $N = 50$ snapshots and $\text{SNR} = 20$ dB

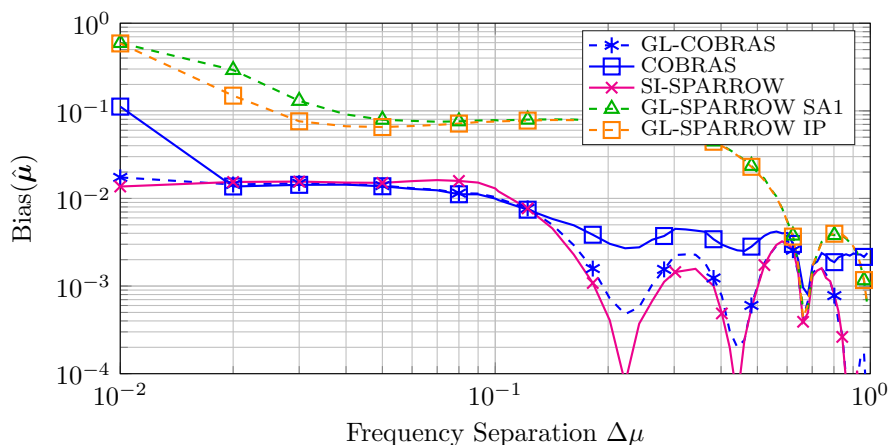


Figure 5.5: Frequency estimation bias for uniform linear PCA of $M = 9$ sensors in $P = 3$ linear subarrays, for $N = 50$ snapshots and $\text{SNR} = 20$ dB

3.3 (see also [Bar83, PGH00]). In contrast to that, the grid-based and the gridless COBRAS methods (5.33) and (5.44) both show rather similar estimation performance, comparable to that of the root-RARE method, and reach the CRB for sufficiently large frequency separation. This observation can be explained by the fact that both dual COBRAS optimization problems provide matrix polynomials with the roots of interest constrained on the unit circle, as discussed in Section 5.3.1. This explains the similar performance results of the grid-based and gridless COBRAS methods. The only difference between the grid-based and gridless COBRAS methods is, that in the first case, the roots are generated on a grid of candidate frequencies on the unit circle, while in the latter case the roots are continuously located on the unit circle. The SI-SPARROW method (5.73) shows estimation performance similar to that of the COBRAS methods, i.e., the additional information of unit subarray perturbations $\alpha = \mathbf{1}$, exploited by SI-SPARROW, does not provide improved estimation performance for the problem at hand.

In a slightly modified experiment the SNR and the number of snapshots are fixed to $\text{SNR} = 20$ dB and $N = 50$, respectively. As seen from Figure 5.4 the incoherent methods again show worst estimation performance. While the root-MUSIC methods for incoherent processing (IP) and for subarray 1 (SA1) achieve the corresponding CRB, the GL-SPARROW methods deviate from the CRB for $0.05 \leq \Delta\mu \leq 0.6$, which can be explained by the estimation bias as investigated for FCAs in Section 4.4.2. Only for a frequency separation of $\Delta\mu \geq 0.6$ do the incoherent methods obtain the estimation performance of the coherent methods. In the class of coherent methods it can be seen that the root-RARE method performs close to the CRB for the region of interest, while spectral RARE cannot always resolve the signals for $\Delta\mu \lesssim 0.06$ and reaches an estimation bias for large source separation, which is caused by the finite frequency grid. Furthermore, it can be observed that the estimation performance of the COBRAS methods deviates from that of the root-RARE method. For large frequency separation $\Delta\mu \gtrsim 0.2$ the grid-based COBRAS method reaches the grid bias, similar to spectral

RARE. However, also for low frequency separation $\Delta\mu \lesssim 0.2$ both methods do not reach the CRB. Similar to the GL-SPARROW method under incoherent processing, this behavior can be explained by an inherent frequency estimation bias for SSR methods.

For further investigation of this characteristic, the spatial frequency estimation bias is computed according to (4.104) and the results are displayed in Figure 5.5. Clearly, the incoherent processing methods GL-SPARROW SA1 and GL-SPARROW IP have the largest estimation bias, which can be explained by the small effective aperture of the utilized array. Both COBRAS methods show a relatively large estimation bias of $\text{Bias}(\hat{\boldsymbol{\mu}}) \approx 0.01$ in the case of low frequency separation $\Delta\mu \lesssim 0.2$. For larger frequency separation $\Delta\mu \gtrsim 0.2$, the bias of the grid-based COBRAS method is mainly determined by the finite grid, while the bias of the gridless COBRAS method shows to be periodic in $\Delta\mu$. The estimation bias of SI-SPARROW is approximately similar to that of GL-COBRAS. In difficult scenarios, with low SNR and low number of snapshots as in the previous setup, the estimation bias is below the CRB, such that it is negligible in the RMSE performance. The frequency estimation bias is a well known phenomenon in SSR research [MÇW05, SPP18] and bias mitigation techniques have been discussed, e.g., in [NBA13].

5.5.3 Arbitrary Array Topologies and Grid-Based Estimation

Consider a PCA with a large aperture, composed of $M = 11$ sensors which are partitioned in $P = 4$ linear subarrays with 3,2,3, and 3 sensors, respectively. The sensor positions for each subarray are $\mathbf{r}^{(1)} = [0.0, 0.6, 2.3]^T$, $\mathbf{r}^{(2)} = [12.2, 13.0]^T$, $\mathbf{r}^{(3)} = [21.5, 22.8, 23.6]^T$, and $\mathbf{r}^{(4)} = [37.6, 38.5, 41.1]^T$, in half signal wavelength, and assume no additional gain/phase offsets among the subarrays, i.e., $\boldsymbol{\alpha} = [1, 1, 1, 1]^T$ in (2.10). Furthermore, consider $L = 3$ uncorrelated Gaussian source signals with spatial frequencies $\boldsymbol{\mu} = [0.5011, 0.4672, -0.2007]^T$.

The array topology does not admit a direct implementation of the gridless COBRAS and the root-RARE methods, such that the experiments in this subsection are limited to the investigation of the grid-based COBRAS method and the spectral RARE method. For both grid-based methods a grid of $K = 400$ grid points according to $\boldsymbol{\nu} = [-1.000, -0.995, -0.990, \dots, 0.995]^T$ is used.

In the first experiment the signal-to-noise ratio is fixed as $\text{SNR} = 6$ dB, while the number of snapshots N is varied. Figure 5.6 clearly demonstrates that the proposed grid-based COBRAS technique outperforms spectral RARE for low number of signal snapshots N . While the spectral RARE method is not able to always resolve the two closely spaced signals with spatial frequencies $\mu_1 = 0.5011$ and $\mu_2 = 0.4672$ for $N \leq 500$ signal snapshots, the proposed COBRAS method resolves the signals for any $N \geq 30$ snapshots.

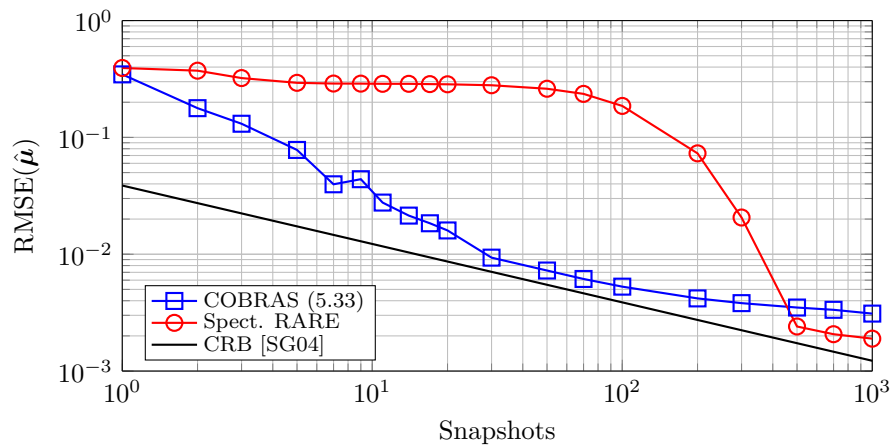


Figure 5.6: Frequency estimation performance for a PCA of $M = 11$ sensors in $P = 4$ subarrays, with $\text{SNR} = 6$ dB and varying number of snapshots N

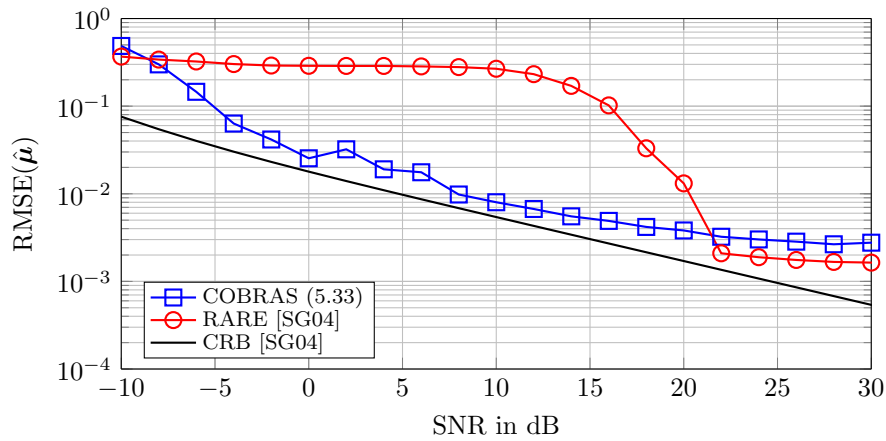


Figure 5.7: Frequency estimation performance for a PCA of $M = 11$ sensors in $P = 4$ subarrays, with $N = 20$ snapshots and varying SNR

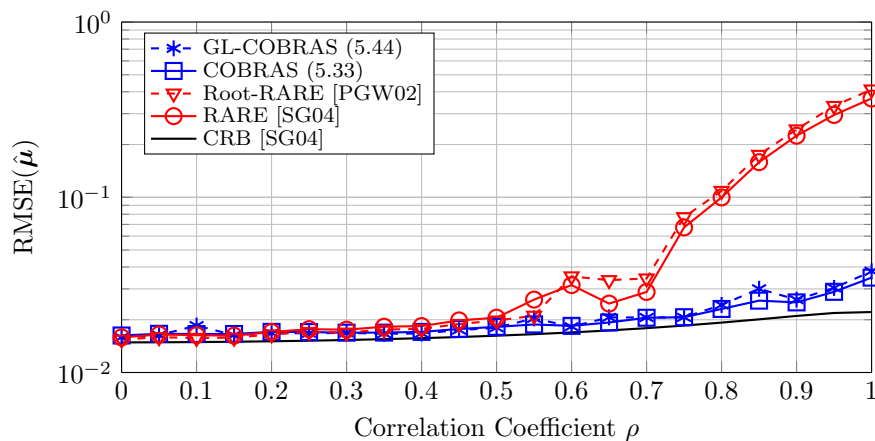


Figure 5.8: Frequency estimation performance for PCA of $M = 9$ sensors in $P = 3$ subarrays for $\text{SNR} = 0$ dB, $N = 30$ snapshot and $L = 2$ source signals with varying real-valued correlation coefficient ρ

In a second experiment the number of snapshots is fixed as $N = 20$ and the SNR is varied. As can be observed from Figure 5.7 the grid-based COBRAS method shows superior thresholding performance as compared to spectral RARE. While spectral RARE can reliably resolve the two closely spaced sources only for $\text{SNR} \geq 22$ dB, the proposed COBRAS can do so for $\text{SNR} \geq 8$ dB with an RMSE of less than 10^{-2} . For high SNR, spectral RARE reaches a bias in the RMSE which is caused mainly by the finite grid. A similar bias effect can be observed for the grid-based COBRAS method. However, for the grid-based COBRAS method the bias is larger than for the spectral RARE method and it is not only caused by the finite grid, as investigated in the previous section.

5.5.4 Correlated Signals

As discussed in the Section 5.5.2, gridless COBRAS and root-RARE show approximately equal resolution performance in difficult scenarios with low SNR, low number of snapshots and uncorrelated signals. This situation changes in the case of correlated signals, where preprocessing in form of subspace separation, as required for the RARE method, becomes difficult. For further investigation of this aspect, consider a PCA of $M = 9$ sensors partitioned into $P = 3$ subarrays of 3, 4 and 2 sensors with positions $\mathbf{r}^{(1)} = [0, 1, 3]^T$, $\mathbf{r}^{(2)} = [17.4, 18.4, 19.4, 21.4]^T$ and $\mathbf{r}^{(3)} = [24.8, 25.8]^T$, in half signal wavelength. Furthermore, consider gain/phase offsets among the subarrays according to $\boldsymbol{\alpha} = [1, 0.7 \cdot e^{j\frac{2}{3}\pi}, 1.2 \cdot e^{j\frac{1}{4}\pi}]^T$ in (2.10). The SNR and number of snapshots are selected as $\text{SNR} = 0$ dB and $N = 30$. A number of $L = 2$ source signals with spatial frequencies $\boldsymbol{\mu} = [0.505, 0.105]^T$ is assumed with a source covariance matrix given as

$$\mathbb{E}\{\boldsymbol{\psi}(t)\boldsymbol{\psi}^H(t)\} = \begin{bmatrix} 1 & \rho \\ \rho^* & 1 \end{bmatrix}, \quad (5.74)$$

where the correlation coefficient ρ is assumed to be real-valued and varied in the experiment. A grid of $K = 200$ candidate frequencies is employed for the grid-based estimation methods, defined as in the previous subsection. As seen from Figure 5.8, the spectral and root-RARE methods fail to properly estimate the spatial frequencies for high correlation ($\rho > 0.6$) while the grid-based and gridless COBRAS methods still show estimation performance close to the CRB, since these methods do not require subspace separation.

5.5.5 Array Calibration Performance

Besides estimation of the spatial frequencies, the COBRAS method also admits estimation of the subarray shifts in $\boldsymbol{\varphi}$ as defined in (2.15). Since the RARE methods do not provide direct estimation of the subarray shifts, the method presented in equation (11) of [PPG11]¹ is employed on the basis of the spatial frequency estimates obtained by the RARE methods.

¹Without the restriction that the complex phase terms must be of unit magnitude.

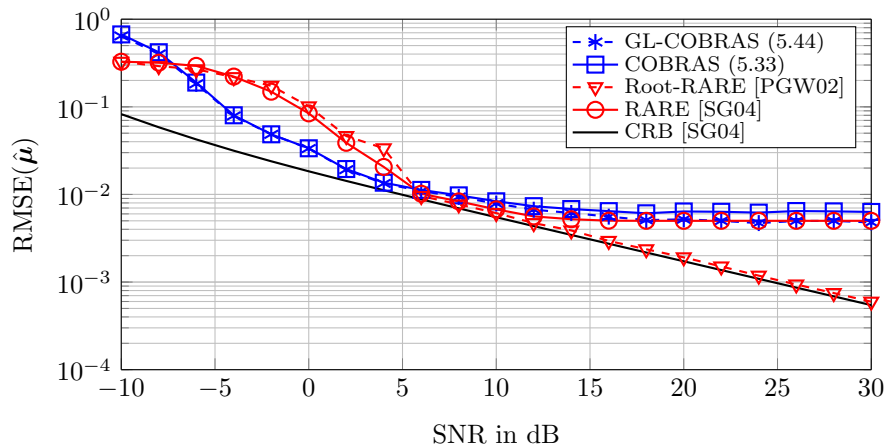


Figure 5.9: Frequency estimation performance for PCA of $M = 10$ sensors in $P = 4$ subarrays, for $N = 20$ snapshots and $L = 2$ uncorrelated source signals

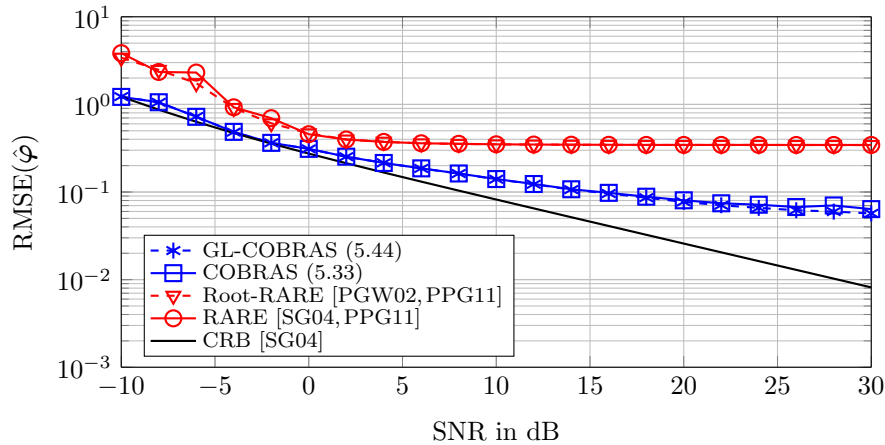


Figure 5.10: Array calibration performance for PCA of $M = 10$ sensors in $P = 4$ subarrays, for $N = 20$ snapshots and $L = 2$ uncorrelated source signals

The setup under investigation consists of a PCA of $M = 10$ sensors partitioned into $P = 4$ subarrays of 3, 2, 3 and 2 sensors at positions $\mathbf{r}^{(1)} = [0, 2, 3]^T$, $\mathbf{r}^{(2)} = [10.1, 11.1]^T$, $\mathbf{r}^{(3)} = [27.4, 28.4, 30.4]^T$ and $\mathbf{r}^{(4)} = [54.8, 56.8]^T$, in half signal wavelength. The subarray gain/phase offsets are set as $\boldsymbol{\alpha} = [1, 1.3 \cdot e^{j\frac{2}{3}\pi}, 0.7 \cdot e^{-j\frac{1}{4}\pi}, 0.9 \cdot e^{-j\frac{3}{5}\pi}]^T$ in (2.10). Furthermore, consider $L = 3$ uncorrelated source signals with spatial frequencies $\boldsymbol{\mu} = [0.605, 0.255, -0.305]^T$ and set the number of snapshots to $N = 20$.

Figure 5.9 displays the frequency estimation error of the different methods for varying SNR, where both COBRAS methods show the best thresholding performance but reach an estimation bias for $\text{SNR} \geq 15$ dB. Similarly, the spectral RARE method reaches an estimation bias which is caused by the finite grid. On the other hand, root-RARE performs asymptotically optimal and reaches the CRB for high SNR. The corresponding subarray shift estimation performance is displayed in Figure 5.10, where the root-mean-

square error is computed according to

$$\text{RMSE}(\hat{\boldsymbol{\varphi}}) = \sqrt{\frac{1}{LT(P-1)} \sum_{t=1}^T \sum_{l=1}^L \|\boldsymbol{\varphi}_l - \hat{\boldsymbol{\varphi}}_l(t)\|_2^2}, \quad (5.75)$$

with $\hat{\boldsymbol{\varphi}}_l(t)$ being the displacement phase vector estimate for signal l in Monte Carlo trial t . As can be observed from Figure 5.10, the subarray shift estimation method in [PP11], based on the frequency estimates obtained from the RARE methods, achieves a relatively large estimation bias for high SNR. In contrast to that, the grid-based and gridless COBRAS methods show a significantly reduced estimation error, which demonstrates the advantage of joint frequency and displacement phase estimation.

5.5.6 Computation Time of SDP Formulations

To investigate the computation time of the COBRAS formulation, simulations are performed in MATLAB using the SeDuMi solver [Stu99] with the CVX interface [GB08, GB14] on a machine with an Intel Core i5-760 CPU @ 2.80 GHz \times 4 and 8 GByte RAM. The scenario consists of two independent complex Gaussian sources with static spatial frequencies $\mu_1 = 0.505$ and $\mu_2 = -0.205$ and a uniform linear PCA of $M = 9$ sensors partitioned into $P = 3$ identical and uniform linear subarrays of 3 sensors. Additional subarray gain/phase offsets are neglected, i.e., $\boldsymbol{\alpha} = [1, 1, 1]^T$ in (2.10).

For the first experiment the SNR is fixed at 0 dB while the number of snapshots N is varied. Figure 5.11 shows the average computation time for $T = 100$ Monte Carlo runs of the SDP implementation of the $\ell_{*,1}$ mixed-norm minimization (5.31) and the grid-based COBRAS formulations (5.36) and (5.37) with a grid size of $K = 100$, as well as the gridless (GL-) COBRAS formulation in (5.44). The computation time is measured only for solving the corresponding optimization problem in CVX. Pre-processing steps, such as computation of the sample covariance matrix, or post-processing steps, such as peak-search or polynomial rooting, are not included into this consideration. As can be observed from Figure 5.11, for a number of $N < 5$ snapshots all grid-based methods exhibit approximately equal computation time. For $5 \leq N < 40$ the $\ell_{*,1}$ mixed-norm minimization problem has largest computation time while the COBRAS formulation (5.36) requires longest computation time for $N > 40$, due to the large dimension of the semidefinite constraint (5.36b). Regarding the computation time of the grid-based COBRAS formulation using the sample covariance matrix (5.37) it can be observed that it is relatively constant for any number of snapshots N and lower than for the other implementations especially for large number of snapshots $N > 10$. The lowest computation time is required for solving the GL-COBRAS implementation (5.44).

Figure 5.12 shows the average computation time for $T = 100$ Monte Carlo runs for a varying number of grid points K and a fixed number of $N = 9$ signal snapshots, corresponding to the case where the dimensionality reduction techniques are applied (compare [MÇW05, SSPH16]). For all grid-based methods the number of SDP constraints

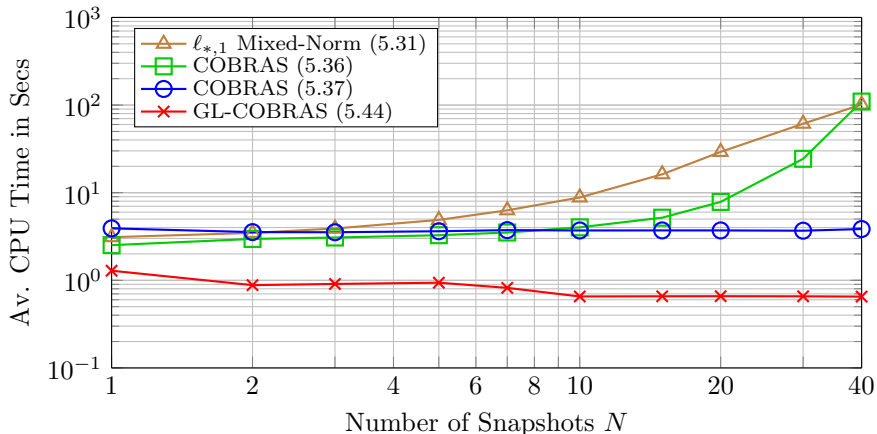


Figure 5.11: Average computation time of different SDP implementations for uniform linear PCA of $M = 9$ sensors in $P = 3$ subarrays, with $K = 100$ grid points and varying number of snapshots N

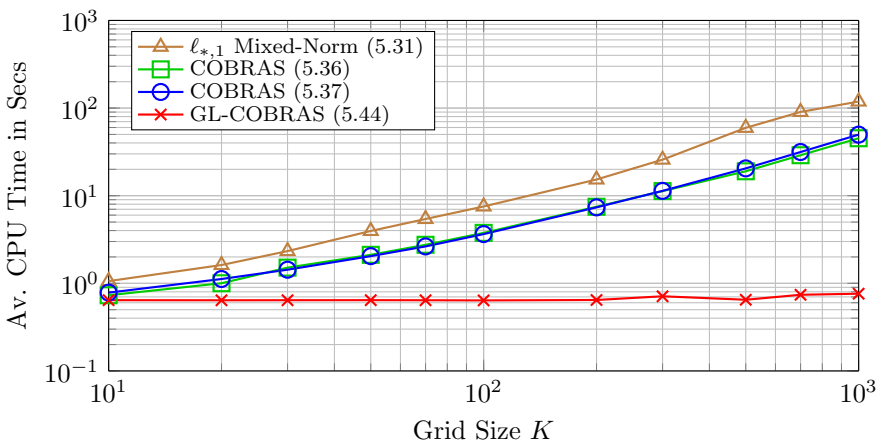


Figure 5.12: Average computation time of different SDP implementations for uniform linear PCA of $M = 9$ sensors in $P = 3$ subarrays, with $N = 9$ signal snapshots and varying grid size K

grows nearly linear with the number of grid points K . Clearly, the $\ell_{*,1}$ mixed-norm minimization approach has the largest computation time for any investigated number of grid points K . The grid-based COBRAS formulations (5.36) and (5.37) show approximately equal computation time, as both formulations have SDP constraints of identical dimensions. Since the GL-COBRAS formulation is independent of the grid size K , it is constant for all grid size numbers K in Figure 5.12 and provides the fastest computation of all methods under investigation.

5.6 Chapter Summary

The chapter has discussed sparse reconstruction for partly calibrated arrays with incoherent and coherent processing. A major contribution of this chapter is given by a novel method for sparse reconstruction in partly calibrated arrays with coherent processing. The proposed method relies on convex minimization of a mixed nuclear and ℓ_1 norm, termed as $\ell_{*,1}$ mixed-norm minimization. Another major contribution concerns a compact, equivalent formulation of the $\ell_{*,1}$ mixed-norm minimization problem, referred to as COmpact Block- and RAnk-Sparse recovery (COBRAS). The COBRAS formulation is computationally attractive, especially in the case of a large number of signal snapshots. For the special case of identical uniform linear subarrays or thinned linear subarrays, an extension to gridless estimation, referred to as gridless COBRAS (GL-COBRAS), was derived. The GL-COBRAS approach was further generalized for gridless estimation with identical subarrays of arbitrary topology, by means of the shift-invariant (SI-) SPARROW formulation.

As shown by numerical results, the proposed coherent sparse reconstruction approach clearly outperforms the existing incoherent methods. Regarding the comparison of sparse reconstruction and subspace-based methods, the grid-based COBRAS method significantly outperforms the spectral RARE method in terms of thresholding performance for closely spaced source signals. Furthermore, the COBRAS method outperforms the root-RARE method in the case of strongly correlated source signals and in the array calibration (subarray shift estimation) performance. A drawback of the $\ell_{*,1}$ mixed-norm minimization approach and the COBRAS formulation is given by an inherent estimation bias, which becomes significant in the asymptotic case of large number of snapshots or high signal-to-noise ratio. However, if higher estimation accuracy is required, the COBRAS estimates can be used to provide initial estimates, e.g., for a subsequent maximum likelihood estimator.

Chapter 6

Low-Complexity Algorithms for Sparse Reconstruction

The semidefinite programming (SDP) formulations for $\ell_{*,1}$ minimization in Section 5.2.1 and the SPARROW and COBRAS formulations in Sections 4.3 and 5.3, admit simple implementation by means of standard conic solvers such as SeDuMi [Stu99] and MOSEK [MOS15]. These solvers usually employ interior point methods for solving the corresponding SDPs, which converge to high precision in few iterations, but might suffer from high computational cost and memory requirements in the case of large dimensions of the SDP constraints [NN94, Ber99, Wri97, BTN01, BV04]. To better deal with the high computational demand required for solving SDPs of large dimensions, it was suggested in [BTR13] to apply the alternating direction method of multipliers (ADMM) for solving the SDP formulation of atomic norm minimization, which reduces memory requirements but suffers from poor convergence speed and requires eigendecomposition of matrices of possibly large dimensions in each iteration.

This chapter provides low-complexity algorithms for the sparse recovery formulations discussed in Sections 4 and 5, where the considerations are restricted to the grid-based methods. The chapter starts with a short review of the well known block coordinate descent (BCD) method for $\ell_{2,1}$ minimization [QSG13, Wri15]. The BCD relies on sequential optimization and has low computational cost and memory requirements, but might suffer from poor convergence speed, due to a possibly large number of iterations that cannot be easily parallelized. The recently proposed Soft Thresholding with Exact Line search Algorithm (STELA) [YP17] is an approach for parallel computation of ℓ_1 minimization problem, and is a first-order method similar to BCD. An extension of the STELA method for application with $\ell_{2,1}$ minimization is derived in this chapter. With regard to the SPARROW formulation, a novel implementation by means of the sequential coordinate descent (CD) method is presented, which shows superior convergence characteristics as compared to the BCD method for the original $\ell_{2,1}$ minimization problem [SPP18]. A parallel implementation for SPARROW is left for future research.

As discussed in Chapter 5, the $\ell_{*,1}$ minimization problem shares many characteristics with the $\ell_{2,1}$ minimization problem, such that many of the concepts of $\ell_{2,1}$ minimization can be transferred to $\ell_{*,1}$ minimization. In this context, the sequential BCD method and the parallel STELA method for $\ell_{2,1}$ minimization, are extended for application with $\ell_{*,1}$ minimization, which admit simple implementation and have low memory requirements. Similarly, a novel CD implementation for the COBRAS formulation is presented, which obtains improved convergence rate as compared to the BCD implementation for $\ell_{*,1}$ minimization, but has slightly higher computational cost. A reduction of the computational cost for the CD implementation and a parallel implementation for COBRAS are left for future research.

The chapter concludes with a short numerical analysis of the convergence speed of the different proposed methods, highlighting benefits and drawbacks of the specific algorithms.

6.1 Algorithms for Sparse Reconstruction in FCAs

6.1.1 Coordinate Descent Method for $\ell_{2,1}$ Minimization

A prominent approach in sparse reconstruction is based on the coordinate descent (CD) method [Ber99, FHHT07], which, for application in joint sparse reconstruction, has been modified to the so-called block coordinate descent (BCD) method [QSG13, Wri15]. In the BCD method, coordinate-wise unconstrained minimization according to

$$\mathbf{x}_k^{(\tau)} = \arg \min_{\mathbf{x}_k \in \mathbb{C}^N} f\left(\mathbf{x}_1^{(\tau)}, \dots, \mathbf{x}_{k-1}^{(\tau)}, \mathbf{x}_k, \mathbf{x}_{k+1}^{(\tau-1)}, \dots, \mathbf{x}_K^{(\tau-1)}\right) \quad (6.1)$$

is performed in an iterative manner, where k and τ determine the coordinate and iteration index, respectively. The BCD method is particularly suitable in the case that the subproblems have analytic solutions and the desired minimizer is sparse, since zero-valued coordinates can be excluded from the iterative computation [FHHT07]. As will be discussed later, both these properties are fulfilled for $\ell_{2,1}$ minimization.

For ease of presentation, in this chapter it is assumed, without loss of generality, that the columns \mathbf{a}_k in the sensing matrix $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_K] \in \mathbb{C}^{M \times K}$ are normalized, i.e., $\|\mathbf{a}_k\|_2 = 1$, for $k = 1, \dots, K$.

Let the approximate solution of the signal matrix in iteration τ and coordinate k be given as

$$\mathbf{X}_k^{(\tau)} = [\mathbf{x}_1^{(\tau)}, \dots, \mathbf{x}_{k-1}^{(\tau)}, \mathbf{x}_k^{(\tau)}, \mathbf{x}_{k+1}^{(\tau-1)}, \dots, \mathbf{x}_K^{(\tau-1)}]^\top \in \mathbb{C}^{K \times N}. \quad (6.2)$$

Using (6.2), the $\ell_{2,1}$ minimization problem given in (4.38) is reformulated, in terms of minimization with respect to the k th row \mathbf{x}_k , as

$$\mathbf{x}_k^{(\tau)} = \arg \min_{\mathbf{x}_k} \frac{1}{2} \|\mathbf{Y} - \mathbf{A} \mathbf{X}_k^{(\tau)}\|_{\mathbb{F}}^2 + \lambda \sqrt{N} \|\mathbf{X}_k^{(\tau)}\|_{2,1} \quad (6.3)$$

$$= \arg \min_{\mathbf{x}_k} \frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^\top\|_{\mathbb{F}}^2 + \lambda \sqrt{N} \|\mathbf{x}_k\|_2, \quad (6.4)$$

where the residual measurement matrix $\mathbf{Y}_{-k}^{(\tau)} \in \mathbb{C}^{M \times N}$ for coordinate k in iteration τ is given as

$$\mathbf{Y}_{-k}^{(\tau)} = \mathbf{Y} - \mathbf{A}_{-k} \mathbf{X}_{-k}^{(\tau)}, \quad (6.5)$$

with

$$\mathbf{A}_{-k} = [\mathbf{a}_1, \dots, \mathbf{a}_{k-1}, \mathbf{a}_{k+1}, \dots, \mathbf{a}_K] \in \mathbb{C}^{M \times K-1} \quad (6.6)$$

denoting the reduced sensing matrix with column \mathbf{a}_k excluded from the representation, and

$$\mathbf{X}_{-k}^{(\tau)} = [\mathbf{x}_1^{(\tau)}, \dots, \mathbf{x}_{k-1}^{(\tau)}, \mathbf{x}_{k+1}^{(\tau-1)}, \dots, \mathbf{x}_K^{(\tau-1)}]^\top \in \mathbb{C}^{K-1 \times N} \quad (6.7)$$

denoting the approximate solution of the signal matrix in iteration τ , with coordinate \mathbf{x}_k excluded from the representation. Note that $\mathbf{Y}_{-k}^{(\tau)}$ in (6.5) can also be computed iteratively according to

$$\mathbf{Y}_{-k}^{(\tau)} = \mathbf{Y}_{-(k-1)}^{(\tau)} - \mathbf{a}_{k-1} \mathbf{x}_{k-1}^{(\tau)\top} + \mathbf{a}_k \mathbf{x}_k^{(\tau-1)\top}, \quad (6.8)$$

which reduces the computational overhead.

Consider the subdifferential [Ber99] of the objective function in (6.4) with respect to \mathbf{x}_k , given as

$$\partial_{\mathbf{x}_k} \left(\frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^\top\|_F^2 + \lambda \sqrt{N} \|\mathbf{x}_k\|_2 \right) = -\mathbf{a}_k^H \left(\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^\top \right) + \lambda \sqrt{N} \mathbf{h}_k^\top. \quad (6.9)$$

In (6.9), \mathbf{h}_k is a subgradient of the penalty term $\|\mathbf{x}_k\|_2$. As discussed in [Ber99], a subgradient \mathbf{h}_k is an element of the subdifferential of the ℓ_2 -norm, defined as

$$\mathbf{h}_k \in \partial \|\mathbf{x}_k\|_2 = \begin{cases} \mathbf{x}_k / \|\mathbf{x}_k\|_2 & \text{if } \mathbf{x}_k \neq \mathbf{0} \\ \{\tilde{\mathbf{h}}_k \mid \|\tilde{\mathbf{h}}_k\|_2 \leq 1\} & \text{if } \mathbf{x}_k = \mathbf{0}. \end{cases} \quad (6.10)$$

For optimality, the subdifferential in (6.9) must contain the $\mathbf{0}$ vector. Consequently, assuming a non-zero minimizer $\mathbf{x}_k \neq \mathbf{0}$, application of (6.10) in (6.9) results in the optimality condition

$$(1 + \lambda \sqrt{N} / \|\mathbf{x}_k\|_2) \mathbf{x}_k^\top = \mathbf{a}_k^H \mathbf{Y}_{-k}^{(\tau)}. \quad (6.11)$$

Computing the ℓ_2 norm of (6.11) and solving for $\|\mathbf{x}_k\|_2$ yields

$$\|\mathbf{x}_k\|_2 = \|\mathbf{a}_k^H \mathbf{Y}_{-k}^{(\tau)}\|_2 - \lambda \sqrt{N}, \quad (6.12)$$

which is feasible for $\|\mathbf{a}_k^H \mathbf{Y}_{-k}^{(\tau)}\|_2 \geq \lambda \sqrt{N}$ and provides an optimality condition on the ℓ_2 norm of \mathbf{x}_k . From (6.11) it can further be concluded that for $\mathbf{x}_k \neq \mathbf{0}$, the direction of \mathbf{x}_k only depends on the direction of $\mathbf{a}_k^H \mathbf{Y}_{-k}^{(\tau)}$. Define the soft-thresholding operator as

$$\mathcal{S}_\varepsilon(\mathbf{w}) = \max(\|\mathbf{w}\|_2 - \varepsilon, 0) \cdot \text{sign}(\mathbf{w}), \quad (6.13)$$

where

$$\text{sign}(\mathbf{w}) = \begin{cases} \mathbf{w} / \|\mathbf{w}\|_2 & \text{for } \mathbf{w} \neq \mathbf{0} \\ \mathbf{0} & \text{for } \mathbf{w} = \mathbf{0}. \end{cases} \quad (6.14)$$

Using (6.13) and based on above considerations on the optimal ℓ_2 norm and direction of \mathbf{x}_k , the minimizer for (6.4) is computed as

$$\mathbf{x}_k^{(\tau)} = \mathcal{S}_{\lambda \sqrt{N}} \left(\mathbf{a}_k^H \mathbf{Y}_{-k}^{(\tau)} \right). \quad (6.15)$$

Algorithm 1 $\ell_{2,1}$ -CD

-
- 1: Initialize approximate solution $\mathbf{X}_1^{(1)} \leftarrow \mathbf{0}$
 - 2: Initialize residual matrix $\mathbf{Y}_{-1}^{(1)} \leftarrow \mathbf{Y}$
 - 3: Initialize iteration index $\tau \leftarrow 1$
 - 4: **repeat**
 - 5: **for** $k \leftarrow 1, \dots, K$ **do**
 - 6: Compute residual matrix $\mathbf{Y}_{-k}^{(\tau)}$ by Eq. (6.8)
 - 7: Update approximate solution $\mathbf{x}_k^{(\tau)}$ by Eq. (6.15)
 - 8: **end for**
 - 9: Update iteration index $\tau \leftarrow \tau + 1$
 - 10: **until** convergence
-

Note that the minimizer in (6.15) is unique, since the objective function in (6.4) is strictly convex in \mathbf{x}_k , for $k = 1, \dots, K$. The expression in (6.15) is computed sequentially over the coordinates $k = 1, \dots, K$ and iteration index $\tau = 1, \dots, \tau_{\max}$, until some convergence criterion is fulfilled. Each coordinate update of the BCD method requires $2MN$ complex-valued additions and multiplications for updating the residual matrix in (6.8) and $(M + 2)N$ complex-valued additions and multiplications for computing the soft-thresholding operator in (6.15), resulting in a total of $(3M + 2)N$ complex-valued additions and multiplications per coordinate. For sake of numerical stability, i.e., to prevent error propagation, it is advisable to fully compute the residual matrix according to (6.5) after a fixed number of iterations, instead of only performing the low-complexity updates according to (6.8), slightly increasing the required number of operations. The computation time can further be reduced if zero coordinates $\mathbf{x}_k^{(\tau)} = \mathbf{0}$ are excluded from future updates. The BCD method is summarized in Algorithm 1, and is contained in the class of non-differentiable minimization with certain separability and regulatory constraints, for which convergence of the BCD iterates has been established in [Tse01].

6.1.2 STELA Method for $\ell_{2,1}$ Minimization

The Soft-Thresholding with Exact Line search Algorithm (STELA) is a special realization of the iterative convex approximation framework proposed in [YP17]. The framework considers optimization problems of the form

$$\min_{\mathbf{X}} f(\mathbf{X}) + g(\mathbf{X}) \quad \text{s.t. } \mathbf{X} \in \mathcal{C}, \quad (6.16)$$

where \mathcal{C} is a closed convex set and $f(\mathbf{X}) : \mathbb{C}^{K \times N} \rightarrow \mathbb{R}$ is a proper and differentiable function with a continuous gradient. The function $g(\mathbf{X})$ is assumed to be convex and may be nondifferentiable. Let $\tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)})$ be an approximate function of $f(\mathbf{X})$ around the point $\mathbf{X}^{(\tau)}$ in iteration τ of the iterative method. The approximate function $\tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)})$ is assumed to be

- pseudo-convex in \mathbf{X} for any $\mathbf{X}^{(\tau)} \in \mathcal{C}$,
- continuously differentiable in \mathbf{X} for any given $\mathbf{X}^{(\tau)} \in \mathcal{C}$, and
- the gradient of $\tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)})$ and the gradient of $f(\mathbf{X})$ are identical at $\mathbf{X} = \mathbf{X}^{(\tau)}$ for any $\mathbf{X}^{(\tau)} \in \mathcal{C}$,

as discussed in [YP17]. The approximate problem of the original problem in (6.16) is formulated as

$$\min_{\mathbf{X}} \tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)}) + g(\mathbf{X}) \quad \text{s.t. } \mathbf{X} \in \mathcal{C} \quad (6.17)$$

and its optimal point is denoted as the best response matrix $\mathbb{B}\mathbf{X}^{(\tau)} \in \mathbb{C}^{K \times N}$, computed as

$$\mathbb{B}\mathbf{X}^{(\tau)} \in \arg \min_{\mathbf{X} \in \mathcal{C}} \tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)}) + g(\mathbf{X}). \quad (6.18)$$

If $\mathbf{X}^{(\tau)}$ is not an optimal point, the update $\mathbf{X}^{(\tau+1)}$ in iteration $\tau + 1$ is defined as

$$\mathbf{X}^{(\tau+1)} = \mathbf{X}^{(\tau)} + \gamma^{(\tau)}(\mathbb{B}\mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)}), \quad (6.19)$$

where $\mathbb{B}\mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)}$ denotes a descent direction and $\gamma^{(\tau)} \in [0, 1]$ is an appropriate stepsize that can be determined by either exact line search or successive line search. If the problem admits exact line search, the optimal stepsize is computed according to

$$\gamma^{(\tau)} = \arg \min_{0 \leq \gamma \leq 1} f(\mathbf{X}^{(\tau)} + \gamma^{(\tau)}(\mathbb{B}\mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)})) + \gamma(g(\mathbb{B}\mathbf{X}^{(\tau)}) - g(\mathbf{X}^{(\tau)})). \quad (6.20)$$

In the case of $\ell_{2,1}$ minimization according to (4.37), the functions $f(\cdot)$ and $g(\cdot)$ in (6.16) are given as

$$f(\mathbf{X}) \triangleq \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_{\mathbb{F}}^2 \quad \text{and} \quad g(\mathbf{X}) \triangleq \lambda \sqrt{N} \|\mathbf{X}\|_{2,1}, \quad (6.21)$$

with optimization variable $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]^{\top}$. The problem is convex in \mathbf{X} , but its objective function is non-differentiable and does not have a closed-form solution. Let $\mathbf{X}^{(\tau)}$ denote the approximate solution in iteration τ . An approximate function for $f(\mathbf{X})$, fulfilling the technical conditions given in [YP17], is given by

$$\begin{aligned} \tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)}) &= \sum_{k=1}^K \tilde{f}(\mathbf{x}_k, \mathbf{X}_{-k}^{(\tau)}) \\ &= \sum_{k=1}^K \frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^{\top}\|_{\mathbb{F}}^2, \end{aligned} \quad (6.22)$$

where $\tilde{f}(\mathbf{x}_k, \mathbf{X}_{-k}^{(\tau)}) = \frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^{\top}\|_{\mathbb{F}}^2$ and $\mathbf{Y}_{-k}^{(\tau)} = \mathbf{Y} - \mathbf{A}_{-k} \mathbf{X}_{-k}^{(\tau)}$, with \mathbf{A}_{-k} defined in (6.6) and $\mathbf{X}_{-k}^{(\tau)} = [\mathbf{x}_1^{(\tau)}, \dots, \mathbf{x}_{k-1}^{(\tau)}, \mathbf{x}_{k+1}^{(\tau)}, \dots, \mathbf{x}_K^{(\tau)}]^{\top}$. Using (6.22) and (6.21) in (6.18), the best response is computed as

$$\mathbb{B}\mathbf{X}^{(\tau)} = \arg \min_{\mathbf{X}} \sum_{k=1}^K \frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^{\top}\|_{\mathbb{F}}^2 + \lambda \sqrt{N} \|\mathbf{X}\|_{2,1}. \quad (6.23)$$

Algorithm 2 $\ell_{2,1}$ -STELA

-
- 1: Initialize approximate solution $\mathbf{X}^{(1)} \leftarrow \mathbf{0}$,
 - 2: Initialize iteration index $\tau \leftarrow 1$
 - 3: **repeat**
 - 4: Compute the best response vectors $\mathbb{B}_k \mathbf{X}^{(\tau)}$ in (6.24), for $k = 1, \dots, K$
 - 5: Compute the update stepsize $\gamma^{(\tau)}$ in (6.25)
 - 6: Compute the variable update $\mathbf{X}^{(\tau+1)}$ in (6.26)
 - 7: Update iteration index $\tau \leftarrow \tau + 1$
 - 8: **until** convergence
-

Note that the minimizer $\mathbb{B} \mathbf{X}^{(\tau)}$ (6.23) is unique, since the approximate function in (6.23) is strictly convex. Similar to the approximate function $\tilde{f}(\mathbf{X}; \mathbf{X}^{(\tau)})$, the function $g(\mathbf{X}) = \lambda\sqrt{N} \|\mathbf{X}\|_{2,1} = \lambda\sqrt{N} \sum_{k=1}^K \|\mathbf{x}_k\|_2$ can be decomposed among the rows \mathbf{x}_k , such that the matrix problem in (6.23) reduces to K independent subproblems

$$\begin{aligned}
\mathbb{B}_k \mathbf{X}^{(\tau)} &= \arg \min_{\mathbf{x}_k} \bar{f}(\mathbf{x}_k, \mathbf{X}_{-k}^{(\tau)}) + g(\mathbf{x}_k) \\
&= \arg \min_{\mathbf{x}_k} \frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{a}_k \mathbf{x}_k^\top\|_{\mathbb{F}}^2 + \lambda\sqrt{N} \|\mathbf{x}_k\|_2 \\
&= \mathcal{S}_{\lambda\sqrt{N}} \left(\mathbf{a}_k^\mathbf{H} \mathbf{Y}_{-k}^{(\tau)} \right),
\end{aligned} \tag{6.24}$$

where $\mathcal{S}_{\lambda\sqrt{N}}(\cdot)$ denotes the soft-thresholding operator defined in (6.13). Note that in contrast to first or second order approximations of the objective function in (4.37), the approximation in (6.24) relies on the concept of best-response, i.e., optimization for one row vector \mathbf{x}_k while fixing the others. Based on (6.24) the best response is given as $\mathbb{B} \mathbf{X}^{(\tau)} = [\mathbb{B}_1^\top \mathbf{X}^{(\tau)}, \dots, \mathbb{B}_K^\top \mathbf{X}^{(\tau)}]^\top$. The optimal step size according to (6.20) can be computed in closed form as

$$\begin{aligned}
\gamma^{(\tau)} &= \arg \min_{0 \leq \gamma \leq 1} \frac{1}{2} \|\mathbf{A}(\mathbf{X}^{(\tau)} + \gamma(\mathbb{B} \mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)})) - \mathbf{Y}\|_{\mathbb{F}}^2 + \gamma\lambda\sqrt{N} (\|\mathbb{B} \mathbf{X}^{(\tau)}\|_{2,1} - \|\mathbf{X}^{(\tau)}\|_{2,1}) \\
&= \left[-\frac{\operatorname{Re}\{\operatorname{Tr}((\mathbf{A} \mathbf{X}^{(\tau)} - \mathbf{Y})^\mathbf{H} \mathbf{A}(\mathbb{B} \mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)}))\} + \lambda\sqrt{N} (\|\mathbb{B} \mathbf{X}^{(\tau)}\|_{2,1} - \|\mathbf{X}^{(\tau)}\|_{2,1})}{\|\mathbf{A}(\mathbb{B} \mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)})\|_{\mathbb{F}}^2} \right]_0^1.
\end{aligned} \tag{6.25}$$

Given the best response $\mathbb{B} \mathbf{X}^{(\tau)}$ and update stepsize $\gamma^{(\tau)}$, the variable update is performed as

$$\mathbf{X}^{(\tau+1)} = \mathbf{X}^{(\tau)} + \gamma^{(\tau)} (\mathbb{B} \mathbf{X}^{(\tau)} - \mathbf{X}^{(\tau)}), \tag{6.26}$$

in correspondence with (6.19).

The proposed update has several desirable features that make it appealing in practice. Firstly, in each iteration, all elements $\mathbb{B}_k \mathbf{X}^{(\tau)}$ of the best response $\mathbb{B} \mathbf{X}^{(\tau)}$ can be updated in parallel, based on the independent subproblems in (6.24). Secondly, the proposed exact line search (6.25) not only yields notable progress in each iteration, but also enjoys

an easy implementation given the closed-form expression. In fact, the main advantage of STELA over other descent direction methods is the update in the form of (6.26), which admits closed-form solutions, in contrast to other methods where successive line search has to be applied. The implementation of the STELA method for $\ell_{2,1}$ minimization is summarized in Algorithm 2. Convergence of the STELA iterates to a stationary point is guaranteed by Theorem 2 in [YP17].

6.1.3 Coordinate Descent Method for SPARROW

The computational cost of the BCD method for the conventional $\ell_{2,1}$ minimization problem discussed in Section 6.1.1 increases with the number of snapshots N . In contrast, neglecting the comparably small overhead required for computing the sample covariance matrix $\hat{\mathbf{R}}$, the computational cost of the SPARROW formulation in (4.51) is independent of the number of snapshots N and, as will be shown in this section, a simple CD implementation also exists for the SPARROW formulation which does not involve an explicit matrix inversion per CD iteration, as proposed in [SPP18].

Consider a function $f(\mathbf{S})$ which is jointly convex in the variables s_1, \dots, s_K . To be consistent with previous notation, the variables are summarized in the diagonal matrix $\mathbf{S} = \text{diag}(s_1, \dots, s_K)$. Furthermore, consider uncoupled constraints of the form $s_k \geq 0$, for $k = 1, \dots, K$. As discussed in Section 6.1.1, the CD method provides sequential and iterative coordinate updates, where in contrast to the definition (6.1) for the BCD method, in this section it is assumed that coordinate $s_k^{(\tau)}$ in iteration τ is updated with an optimal stepsize $\hat{d}_k^{(\tau)}$, computed as

$$\hat{d}_k^{(\tau)} = \arg \min_d f(\mathbf{S}_k^{(\tau)} + d \mathbf{E}_k) \quad (6.27a)$$

$$\text{s.t. } s_k^{(\tau)} + d \geq 0. \quad (6.27b)$$

In (6.27), the diagonal matrix

$$\mathbf{S}_k^{(\tau)} = \text{diag}\left(s_1^{(\tau+1)}, \dots, s_{k-1}^{(\tau+1)}, s_k^{(\tau)}, \dots, s_K^{(\tau)}\right) \quad (6.28)$$

denotes the approximate solution for the minimizer of $f(\mathbf{S})$ in iteration τ , before updating coordinate k , and matrix \mathbf{E}_k with elements

$$[\mathbf{E}_k]_{m,n} = \begin{cases} 1 & \text{if } m = n = k \\ 0 & \text{otherwise} \end{cases} \quad (6.29)$$

denotes a selection matrix. Given the update stepsize $\hat{d}_k^{(\tau)}$, the coordinate update is performed according to

$$\mathbf{S}_k^{(\tau+1)} = \mathbf{S}_k^{(\tau)} + \hat{d}_k^{(\tau)} \mathbf{E}_k. \quad (6.30)$$

Regarding the SPARROW problem in (4.51), the objective function of the subproblem in (6.27) is given as

$$f(\mathbf{S}_k^{(\tau)} + d \mathbf{E}_k) = \text{Tr}((\mathbf{Q}_{k,\tau} + d \mathbf{a}_k \mathbf{a}_k^H)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}_k^{(\tau)}) + d, \quad (6.31)$$

with $\mathbf{Q}_{k,\tau} = \mathbf{A} \mathbf{S}_k^{(\tau)} \mathbf{A}^H + \lambda \mathbf{I}_M \in \mathbb{C}^{M \times M}$. Upon application of the matrix inversion lemma [Hag89]

$$(\mathbf{Q}_{k,\tau} + d \mathbf{a}_k \mathbf{a}_k^H)^{-1} = \mathbf{Q}_{k,\tau}^{-1} - \frac{d \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k \mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1}}{1 + d \mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k} \quad (6.32)$$

and by exploiting the cyclic property of the trace operator, Equation (6.31) is rewritten as

$$f(\mathbf{S}_k^{(\tau)} + d \mathbf{E}_k) = \text{Tr}(\mathbf{Q}_{k,\tau}^{-1} \hat{\mathbf{R}}) - \frac{d \mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \hat{\mathbf{R}} \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k}{1 + d \mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k} + \text{Tr}(\mathbf{S}_k^{(\tau)}) + d. \quad (6.33)$$

The function $f(\mathbf{S}_k^{(\tau)} + d \mathbf{E}_k)$ in (6.33) behaves asymptotically linear in d and has stationary points in

$$\tilde{d}_{1,2} = \frac{\pm \sqrt{\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \hat{\mathbf{R}} \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k - 1}}{\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k}, \quad (6.34)$$

symmetrically located around the simple pole in

$$\tilde{d}_0 = -\frac{1}{\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k} = -\frac{1 + s_k^{(\tau)} \mathbf{a}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{a}_k}{\mathbf{a}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{a}_k}, \quad (6.35)$$

where the last identity in (6.35) follows from the matrix inversion lemma [Hag89] applied to $\mathbf{Q}_{k,\tau}^{-1} = (\mathbf{Q}_{-k,\tau} + s_k^{(\tau)} \mathbf{a}_k \mathbf{a}_k^H)^{-1}$, with $\mathbf{Q}_{-k,\tau} = \mathbf{A}_{-k} \mathbf{S}_{-k}^{(\tau)} \mathbf{A}_{-k}^H + \lambda \mathbf{I}_M$, where

$$\mathbf{S}_{-k}^{(\tau)} = \text{diag}(s_1^{(\tau+1)}, \dots, s_{k-1}^{(\tau+1)}, s_{k+1}^{(\tau)}, \dots, s_K^{(\tau)}) \quad (6.36)$$

and \mathbf{A}_{-k} defined in (6.6). By taking account of the constraint $s_k^{(\tau)} + d \geq 0$ in (6.27b), it can easily be verified that the optimal stepsize must fulfill $\hat{d}_k^{(\tau)} \geq -s_k^{(\tau)} > \tilde{d}_0$, i.e., it must be located on the right hand side of the pole \tilde{d}_0 , such that the optimal stepsize according to (6.27) is computed as

$$\hat{d}_k^{(\tau)} = \max \left(\frac{\sqrt{\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \hat{\mathbf{R}} \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k - 1}}{\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k}, -s_k^{(\tau)} \right). \quad (6.37)$$

Given the stepsize $\hat{d}_k^{(\tau)}$, the variable update is performed according to (6.30). The matrix inverse $\mathbf{Q}_{k+1,\tau}^{-1}$, including the updated coordinate $s_k^{(\tau+1)} = s_k^{(\tau)} + \hat{d}_k^{(\tau)}$ as required for updating the next coordinate $s_{k+1}^{(\tau)}$, can be computed iteratively by the matrix inversion lemma as shown in (6.32), such that computationally expensive explicit matrix inversion can be avoided. To reduce the effect of numerical error propagation, it is advisable to explicitly compute the matrix inverse $\mathbf{Q}_{k,\tau}^{-1} = (\mathbf{A} \mathbf{S}_k^{(\tau)} \mathbf{A}^H + \lambda \mathbf{I}_M)^{-1}$ instead

Algorithm 3 SPARROW-CD

-
- 1: Initialize approximate solution $\mathbf{S}_{1,1} \leftarrow \mathbf{0}$,
 - 2: Initialize matrix inverse $\mathbf{Q}_{1,1}^{-1} \leftarrow \frac{1}{\lambda} \mathbf{I}$
 - 3: Initialize iteration index $\tau \leftarrow 1$
 - 4: **repeat**
 - 5: **for** $k \leftarrow 1, \dots, K$ **do**
 - 6: Compute stepsize $\hat{d}_k^{(\tau+1)}$ by Eq. (6.37)
 - 7: Update approximate solution $\mathbf{S}_k^{(\tau+1)}$ by Eq. (6.30)
 - 8: Update matrix inverse $\mathbf{Q}_{k+1,\tau}^{-1}$ by Eq. (6.32)
 - 9: **end for**
 - 10: Update matrix inverse $\mathbf{Q}_{1,\tau+1}^{-1} \leftarrow \mathbf{Q}_{K+1,\tau}^{-1}$
 - 11: Update iteration index $\tau \leftarrow \tau + 1$
 - 12: **until** convergence
-

of using the rank-one updates, after a number of iterations, depending on the variable precision and desired accuracy of the solution. From experiments in MATLAB with double precision floating-point numbers it was found that a closed form computation after every 100 CD iterations achieves good reconstruction performance. The overall procedures of the proposed CD method are summarized in Algorithm 3. Note that in a practical implementation only the $M \times M$ Hermitian matrix $\mathbf{Q}_{k+1,\tau}^{-1}$ as well as the diagonal elements in $\mathbf{S}_k^{(\tau)}$ need to be stored and updated over coordinates k and iterations τ , and that the computation time of the CD method can be drastically reduced if the sparsity in $\mathbf{S}_k^{(\tau)}$ is exploited, by excluding zero elements in $\mathbf{S}_k^{(\tau)}$ from the computation in future iterations. The proposed CD implementation of SPARROW can be implemented with about $(3M + 2)M$ complex-valued multiplications and additions per coordinate and iteration, which is in the same order as for the BCD method discussed in Section 6.1.1. In the undersampled case, with $N < M$, the number of operations can be further reduced by replacing

$$\sqrt{\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \hat{\mathbf{R}} \mathbf{Q}_{k,\tau}^{-1} \mathbf{a}_k} = \|\mathbf{a}_k^H \mathbf{Q}_{k,\tau}^{-1} \mathbf{Y}\|_2 / \sqrt{N} \quad (6.38)$$

in the update stepsize computation (6.37). As shown by numerical experiments in Section 4.4, the CD implementation of the SPARROW formulation provides superior convergence rate over the BCD method for $\ell_{2,1}$ minimization.

Regarding the convergence of the SPARROW-CD method, it is shown in Appendix C that the objective function in the SPARROW formulation is continuously differentiable and strictly convex in its single components s_1, \dots, s_K , such that Proposition 2.7.1 in [Ber99] applies and convergence of the SPARROW-CD method to a stationary point is guaranteed.

6.2 Algorithms for Sparse Reconstruction in PCAs

6.2.1 Coordinate Descent Method for $\ell_{*,1}$ Minimization

For application of the BCD method to the $\ell_{*,1}$ minimization problem given in (5.23), define

$$\bar{\mathbf{Z}}_k^{(\tau)} = [\mathbf{Z}_1^{(\tau)}, \dots, \mathbf{Z}_{k-1}^{(\tau)}, \mathbf{Z}_k^{(\tau)}, \mathbf{Z}_{k+1}^{(\tau-1)}, \dots, \mathbf{Z}_K^{(\tau-1)}]^\top \quad (6.39)$$

as the approximate solution of the signal matrix for coordinate k in iteration τ , and assume, without loss of generality, that the blocks \mathbf{B}_k of the subarray sensing block matrix $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_K]$ are normalized according to $\mathbf{B}_k^H \mathbf{B}_k = \mathbf{I}_P$, for $k = 1, \dots, K$. For coordinatewise optimization, the problem in (5.23) can be rewritten as

$$\mathbf{Z}_k^{(\tau)} = \arg \min_{\mathbf{Z}_k} \frac{1}{2} \|\mathbf{Y} - \mathbf{B} \bar{\mathbf{Z}}_k^{(\tau)}\|_F^2 + \lambda \sqrt{N} \|\bar{\mathbf{Z}}_k^{(\tau)}\|_{*,1} \quad (6.40)$$

$$= \arg \min_{\mathbf{Z}_k} \frac{1}{2} \|\mathbf{Y}_{-k}^{(\tau)} - \mathbf{B}_k \mathbf{Z}_k\|_F^2 + \lambda \sqrt{N} \|\mathbf{Z}_k\|_*, \quad (6.41)$$

where the $M \times N$ residual matrix for coordinate k in iteration τ is given by

$$\mathbf{Y}_{-k}^{(\tau)} = \mathbf{Y} - \mathbf{B}_{-k} \mathbf{Z}_{-k}^{(\tau)} \in \mathbb{C}^{M \times N}, \quad (6.42)$$

with

$$\mathbf{B}_{-k} = [\mathbf{B}_1, \dots, \mathbf{B}_{k-1}, \mathbf{B}_{k+1}, \dots, \mathbf{B}_K] \in \mathbb{C}^{M \times P(K-1)} \quad (6.43)$$

denoting the subarray sensing block matrix and

$$\mathbf{Z}_{-k}^{(\tau)} = [\mathbf{Z}_1^{(\tau)\top}, \dots, \mathbf{Z}_{k-1}^{(\tau)\top}, \mathbf{Z}_{k+1}^{(\tau-1)\top}, \dots, \mathbf{Z}_K^{(\tau-1)\top}]^\top \in \mathbb{C}^{P(K-1) \times N} \quad (6.44)$$

denoting an approximation of the signal representation in iteration τ , with the k th submatrix removed, respectively. Similar as in (6.8), the residual matrix $\mathbf{Y}_{-k}^{(\tau)}$ can also be computed iteratively according to

$$\mathbf{Y}_{-k}^{(\tau)} = \mathbf{Y}_{-(k-1)}^{(\tau)} - \mathbf{B}_{k-1} \mathbf{Z}_{k-1}^{(\tau)} + \mathbf{B}_k \mathbf{Z}_k^{(\tau-1)}, \quad (6.45)$$

which reduces the computational overhead. Under the assumption of normalized subarray sensing block matrices \mathbf{B}_k , it can easily be verified that problem (6.41) is equivalent to

$$\mathbf{Z}_k^{(\tau)} = \arg \min_{\mathbf{Z}_k} \frac{1}{2} \|\tilde{\mathbf{Z}}_k^{(\tau)} - \mathbf{Z}_k\|_F^2 + \lambda \sqrt{N} \|\mathbf{Z}_k\|_*, \quad (6.46)$$

where

$$\tilde{\mathbf{Z}}_k^{(\tau)} = \mathbf{B}_k^H \mathbf{Y}_{-k}^{(\tau)} \in \mathbb{C}^{P \times N} \quad (6.47)$$

denotes the least-squares estimate of $\mathbf{Z}_k^{(\tau)}$. Note that the minimizer $\mathbf{Z}_k^{(\tau)}$ in (6.46) is unique, since the objective function is strictly convex in \mathbf{Z}_k . It has been shown that

Algorithm 4 $\ell_{*,1}$ -CD

-
- 1: Initialize approximate solution $\mathbf{Z}_1^{(1)} \leftarrow \mathbf{0}$
 - 2: Initialize residual matrix $\mathbf{Y}_{-1}^{(1)} \leftarrow \mathbf{Y}$
 - 3: Initialize iteration index $\tau \leftarrow 1$
 - 4: **repeat**
 - 5: **for** $k \leftarrow 1, \dots, K$ **do**
 - 6: Compute residual matrix $\mathbf{Y}_{-k}^{(\tau)}$ by Eq. (6.45)
 - 7: Update approximate solution $\mathbf{Z}_k^{(\tau)}$ by SVT in Eq. (6.48)
 - 8: **end for**
 - 9: Update iteration index $\tau \leftarrow \tau + 1$
 - 10: **until** convergence
-

the problem in (6.46) can be solved in closed form by the singular value thresholding (SVT) method [CCS10], which is performed according to

$$\mathbf{Z}_k^{(\tau)} = \mathbf{U}_k^{(\tau)} \left(\tilde{\Sigma}_k^{(\tau)} - \lambda\sqrt{N}\mathbf{I}_r \right)_+ \mathbf{V}_k^{(\tau)\text{H}} \quad (6.48)$$

where $[(\mathbf{W})_+]_{ij} = \max([\mathbf{W}]_{ij}, 0)$, and the $r \times r$ matrix $\tilde{\Sigma}_k^{(\tau)}$, the $K \times r$ matrix $\mathbf{U}_k^{(\tau)}$, and the $N \times r$ matrix $\mathbf{V}_k^{(\tau)}$, for $r = \min(P, N)$, are obtained from compact singular value decomposition (SVD) of the least-squares estimate $\tilde{\mathbf{Z}}_k^{(\tau)}$ in (6.47), i.e.,

$$\tilde{\mathbf{Z}}_k^{(\tau)} = \mathbf{U}_k^{(\tau)} \tilde{\Sigma}_k^{(\tau)} \mathbf{V}_k^{(\tau)\text{H}}. \quad (6.49)$$

The steps of the BCD method for $\ell_{*,1}$ minimization are given in Algorithm 4, where the BCD iterations are performed until some convergence criterion is met [QSG13, Wri15]. The major computational cost for the proposed BCD method is given by the SVT step in (6.48), which requires SVD of the $P \times N$ least-squares estimate $\tilde{\mathbf{Z}}_k^{(\tau)}$ with $\mathcal{O}(PN^2)$ operations, for $P \leq N$ [TBI97]. The BCD method in Algorithm 4 falls within the class of non-differentiable minimization with certain separability and regulatory constraints, for which convergence has been established in [Tse01].

6.2.2 STELA Method for $\ell_{*,1}$ Minimization

Similar to the discussion in Section 6.1.2, the key idea for application of STELA for $\ell_{*,1}$ minimization is to solve a sequence of approximate problems instead of the original minimization problem given in (5.23). Analogous to (6.23), in iteration τ of the STELA method the problem in (5.23) is approximated by K independent subproblems of the form

$$\begin{aligned} \mathbb{B}_k \mathbf{Z}^{(\tau)} &= \arg \min_{\mathbf{Z}_k} \frac{1}{2} \left\| \mathbf{Y} - \mathbf{B}_{-k} \mathbf{Z}_{-k}^{(\tau)} - \mathbf{B}_k \mathbf{Z}_k \right\|_{\text{F}}^2 + \lambda\sqrt{N} (\| \mathbf{Z}_{-k}^{(\tau)} \|_{*,1} + \| \mathbf{Z}_k \|_*) \\ &= \arg \min_{\mathbf{Z}_k} \frac{1}{2} \left\| \mathbf{Y}_{-k}^{(\tau)} - \mathbf{B}_k \mathbf{Z}_k \right\|_{\text{F}}^2 + \lambda\sqrt{N} \| \mathbf{Z}_k \|_*, \end{aligned} \quad (6.50)$$

for $k = 1, \dots, K$, with the residual matrix

$$\mathbf{Y}_{-k}^{(\tau)} = \mathbf{Y} - \mathbf{B}_{-k} \mathbf{Z}_{-k}^{(\tau)}, \quad (6.51)$$

where $\mathbf{B}_{-k} \in \mathbb{C}^{M \times P(K-1)}$, as defined in (6.43), and

$$\mathbf{Z}_{-k}^{(\tau)} = [\mathbf{Z}_1^{(\tau)\top}, \dots, \mathbf{Z}_{k-1}^{(\tau)\top}, \mathbf{Z}_{k+1}^{(\tau)\top}, \dots, \mathbf{Z}_K^{(\tau)\top}]^\top \in \mathbb{C}^{P(K-1) \times N} \quad (6.52)$$

denote the subarray sensing block matrix and an approximation of the signal representation in iteration τ , with the k th submatrix removed, respectively. The matrix $\mathbb{B}_k \mathbf{Z}^{(\tau)}$ in (6.50) denotes the k th submatrix of the best response matrix $\mathbb{B} \mathbf{Z}^{(\tau)} = [\mathbb{B}_1 \mathbf{Z}^{(\tau)\top}, \dots, \mathbb{B}_K \mathbf{Z}^{(\tau)\top}]^\top$ in iteration τ and is the unique minimizer since (6.50) is strictly convex in \mathbf{Z}_k . As discussed for (6.46), in the case of unitary submatrices with $\mathbf{B}_k^\mathbf{H} \mathbf{B}_k = \mathbf{I}_P$ as assumed here, the problem (6.50) can be solved in closed form by singular value thresholding [CCS10] according to

$$\mathbb{B}_k \mathbf{Z}^{(\tau)} = \mathbf{U}_k^{(\tau)} \left(\tilde{\boldsymbol{\Sigma}}_k^{(\tau)} - \lambda \sqrt{N} \mathbf{I}_r \right)_+ \mathbf{V}_k^{(\tau)\mathbf{H}}, \quad (6.53)$$

based on the singular value decomposition $\tilde{\mathbf{Z}}_n^{(\tau)} = \mathbf{U}_k^{(\tau)} \tilde{\boldsymbol{\Sigma}}_k^{(\tau)} \mathbf{V}_k^{(\tau)\mathbf{H}}$ of the least-squares estimate $\tilde{\mathbf{Z}}_k^{(\tau)} = \mathbf{B}_k^\mathbf{H} \mathbf{Y}_{-k}^{(\tau)}$, as given in (6.49) and (6.47).

As discussed for (6.19) the matrix $\mathbb{B} \mathbf{Z}^{(\tau)} - \mathbf{Z}^{(\tau)}$ is a descent direction of the objective function in (5.23). Therefore a variable update is performed according to

$$\mathbf{Z}^{(\tau+1)} = \mathbf{Z}^{(\tau)} + \gamma^{(\tau)} (\mathbb{B} \mathbf{Z}^{(\tau)} - \mathbf{Z}^{(\tau)}), \quad (6.54)$$

with stepsize $\gamma^{(\tau)}$. Convergence of the sequence $\{\mathbf{Z}^{(\tau)}\}_{\tau=1}^\infty$ to a stationary point strongly depends on proper selection of the stepsize parameter $\gamma^{(\tau)}$, e.g., by successive or exact line search methods. Following the ideas in [YP17], exact line search is performed according to

$$\begin{aligned} \gamma^{(\tau)} &= \arg \min_{0 \leq \gamma \leq 1} \frac{1}{2} \left\| \mathbf{Y} - \mathbf{B} (\mathbf{Z}^{(\tau)} + \gamma (\mathbb{B} \mathbf{Z}^{(\tau)} - \mathbf{Z}^{(\tau)})) \right\|_{\mathbb{F}}^2 + \gamma \lambda \sqrt{N} (\|\mathbb{B} \mathbf{Z}^{(\tau)}\|_{*,1} - \|\mathbf{Z}^{(\tau)}\|_{*,1}) \\ &= \left[\frac{\operatorname{Re} \{ \operatorname{Tr} ((\mathbf{B} \mathbf{Z}^{(\tau)} - \mathbf{Y})^\mathbf{H} \mathbf{B} (\mathbb{B} \mathbf{Z}^{(\tau)} - \mathbf{Z}^{(\tau)})) \} + \lambda \sqrt{N} (\|\mathbb{B} \mathbf{Z}^{(\tau)}\|_{*,1} - \|\mathbf{Z}^{(\tau)}\|_{*,1})}{\|\mathbf{B} (\mathbb{B} \mathbf{Z}^{(\tau)} - \mathbf{Z}^{(\tau)})\|_{\mathbb{F}}^2} \right]_0^1. \end{aligned} \quad (6.55)$$

In contrast to standard exact line search approaches, which generally have to be evaluated numerically, for problem (5.23) the stepsize parameter proposed in (6.55) can be computed in closed-form which significantly reduces the computational cost. Similarly, as seen from (6.53), all submatrices $\mathbb{B}_k \mathbf{Z}^{(\tau)}$, for $k = 1, \dots, K$, admit closed form expressions and can be computed independently and, thus, in parallel. It can be verified that the problem in (5.23) and the approximate subproblems in (6.50) fulfill all assumptions specified by Theorem 2 in [YP17], such that the sequence $\{\mathbf{Z}^{(\tau)}\}_{t=1}^\infty$ converges to the global optimum $\hat{\mathbf{Z}}$ of problem (5.23). Furthermore, due to the exact line search, parallel updates and closed form expressions, the convergence speed of STELA is generally much faster than that of block coordinate descent methods or gradient-based methods, as will be shown in the numerical results later in this chapter. The corresponding steps of the STELA implementation are summarized in Algorithm 5.

Algorithm 5 $\ell_{*,1}$ -STELA

-
- 1: Initialize approximate solution $\mathbf{Z}^{(1)} \leftarrow \mathbf{0}$
 - 2: Initialize iteration index $\tau \leftarrow 1$
 - 3: **repeat**
 - 4: Compute the best response submatrices $\mathbb{B}_k \mathbf{Z}^{(\tau)}$ in (6.53), for $k = 1, \dots, K$
 - 5: Compute the update stepsize $\gamma^{(\tau)}$ in (6.55)
 - 6: Compute the variable update $\mathbf{Z}^{(\tau+1)}$ in (6.54)
 - 7: Update iteration index $\tau \leftarrow \tau + 1$
 - 8: **until** convergence
-

6.2.3 Coordinate Descent Method for COBRAS

As indicated in Section 6.1.3, the CD implementation of SPARROW offers superior convergence as compared to the BCD implementation for $\ell_{2,1}$ minimization. Inspired by these improvements, a CD implementation for the COBRAS formulation is derived in this section.

Define the block-diagonal matrix

$$\bar{\mathbf{S}}_k^{(\tau)} = \text{blkdiag}(\mathbf{S}_1^{(\tau)}, \dots, \mathbf{S}_{k-1}^{(\tau)}, \mathbf{S}_k, \mathbf{S}_{k+1}^{(\tau-1)}, \dots, \mathbf{S}_K^{(\tau-1)}) \quad (6.56)$$

as the approximate solution of the COBRAS-CD method in iteration τ and coordinate k , and consider the coordinate-wise optimization of the COBRAS problem in (5.33) as

$$\mathbf{S}_k^{(\tau)} = \arg \min_{\mathbf{S}_k \succeq \mathbf{0}} \text{Tr}((\mathbf{B} \bar{\mathbf{S}}_k^{(\tau)} \mathbf{B}^H)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\bar{\mathbf{S}}_k^{(\tau)}) \quad (6.57)$$

$$= \arg \min_{\mathbf{S}_k \succeq \mathbf{0}} \text{Tr}((\mathbf{Q}_{\tau,-k} + \mathbf{B}_k \mathbf{S}_k \mathbf{B}_k^H)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}_k), \quad (6.58)$$

where $\mathbf{Q}_{\tau,-k} = \mathbf{B}_{-k} \bar{\mathbf{S}}_{-k}^{(\tau)} \mathbf{B}_{-k} + \lambda \mathbf{I}_M \in \mathbb{C}^{M \times M}$ is computed from the approximate solution with the k th block removed, according to

$$\bar{\mathbf{S}}_{-k}^{(\tau)} = \text{blkdiag}(\mathbf{S}_1^{(\tau)}, \dots, \mathbf{S}_{k-1}^{(\tau)}, \mathbf{S}_{k+1}^{(\tau-1)}, \dots, \mathbf{S}_K^{(\tau-1)}), \quad (6.59)$$

and \mathbf{B}_{-k} is defined in (6.43). Neglecting the constraint $\mathbf{S}_k \succeq \mathbf{0}$, the unconstrained version of problem (6.58) is given by

$$\tilde{\mathbf{S}}_k^{(\tau)} = \arg \min_{\tilde{\mathbf{S}}_k} \text{Tr}((\mathbf{Q}_{\tau,-k} + \mathbf{B}_k \tilde{\mathbf{S}}_k \mathbf{B}_k^H)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\tilde{\mathbf{S}}_k). \quad (6.60)$$

Optimality of (6.60) requires that

$$\begin{aligned} & \frac{\partial}{\partial \tilde{\mathbf{S}}_k} \text{Tr}((\mathbf{Q}_{-k,\tau} + \mathbf{B}_k \tilde{\mathbf{S}}_k \mathbf{B}_k^H)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\tilde{\mathbf{S}}_k) \\ &= \mathbf{I}_P - \mathbf{B}_k^H (\mathbf{Q}_{-k,\tau} + \mathbf{B}_k \tilde{\mathbf{S}}_k \mathbf{B}_k^H)^{-1} \hat{\mathbf{R}} (\mathbf{Q}_{-k,\tau} + \mathbf{B}_k \tilde{\mathbf{S}}_k \mathbf{B}_k^H)^{-1} \mathbf{B}_k = \mathbf{0}. \end{aligned} \quad (6.61)$$

Algorithm 6 COBRAS-CD

-
- 1: Initialize eigenvalues and -vectors in $\Sigma_k^{(1)} \leftarrow \mathbf{0}$ and $\mathbf{U}_k^{(1)} \leftarrow \mathbf{I}$, for $k = 1, \dots, K$
 - 2: Initialize matrix inverse $\mathbf{Q}_{1,1}^{-1} \leftarrow \frac{1}{\lambda} \mathbf{I}$
 - 3: Initialize iteration index $\tau \leftarrow 1$
 - 4: **repeat**
 - 5: **for** $k \leftarrow 1, \dots, K$ **do**
 - 6: Successively compute matrix inverse $\mathbf{Q}_{-k,\tau}^{-1}$ from $\mathbf{Q}_{k,\tau}^{-1}$ by Eq. (6.65)
 - 7: Update eigenvalues and -vectors in $\Sigma_k^{(\tau)}$ and $\mathbf{U}_k^{(\tau)}$ by Eqs. (6.63) and (6.64)
 - 8: Successively compute matrix inverse $\mathbf{Q}_{k+1,\tau}^{-1}$ from $\mathbf{Q}_{-k,\tau}^{-1}$ by Eq. (6.66)
 - 9: **end for**
 - 10: Update matrix inverse $\mathbf{Q}_{1,\tau+1}^{-1} \leftarrow \mathbf{Q}_{K+1,\tau}^{-1}$
 - 11: Update iteration index $\tau \leftarrow \tau + 1$
 - 12: **until** convergence
-

Performing basic transformations of (6.61), including application of matrix inverse identities, yields

$$\left((\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k)^{-1} + \tilde{\mathbf{S}}_k \right)^2 = (\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k)^{-1} \mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \hat{\mathbf{R}} \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k (\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k)^{-1}. \quad (6.62)$$

By taking the positive-semidefinite matrix square-root and solving for $\tilde{\mathbf{S}}_k$, the minimizer of the unconstrained problem (6.60) results to

$$\tilde{\mathbf{S}}_k^{(\tau)} = \left((\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k)^{-1} (\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \hat{\mathbf{R}} \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k) (\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k)^{-1} \right)^{1/2} - (\mathbf{B}_k^H \mathbf{Q}_{-k,\tau}^{-1} \mathbf{B}_k)^{-1}. \quad (6.63)$$

Let the eigendecomposition of $\tilde{\mathbf{S}}_k^{(\tau)}$ be given as $\tilde{\mathbf{S}}_k^{(\tau)} = \mathbf{U}_k^{(\tau)} \tilde{\Sigma}_k^{(\tau)} \mathbf{U}_k^{(\tau)H}$. A minimizer for (6.58), fulfilling $\mathbf{S}_k^{(\tau)} \succeq \mathbf{0}$, is computed from $\tilde{\mathbf{S}}_k^{(\tau)}$ in (6.63) by projection to the set of positive semidefinite matrices, according to

$$\mathbf{S}_k^{(\tau)} = \mathbf{U}_k^{(\tau)} (\tilde{\Sigma}_k^{(\tau)})_+ \mathbf{U}_k^{(\tau)H}, \quad (6.64)$$

where $[(\mathbf{W})_+]_{i,j} = \max([\mathbf{W}]_{i,j}, 0)$. Note that the computation of $\mathbf{S}_k^{(\tau)}$ has comparably high computational cost, since it involves computation of the matrix inverse and the matrix square-root of $P \times P$ matrices in (6.63), as well as the eigendecomposition of a $P \times P$ matrix in (6.64). It is left for future research to find a computationally more efficient implementation of (6.63).

To reduce the computational burden, the inverse of the $M \times M$ matrix $\mathbf{Q}_{-k,\tau} = \mathbf{B}_{-k} \bar{\mathbf{S}}_{-k}^{(\tau)} \mathbf{B}_{-k} + \lambda \mathbf{I}_M$, in (6.63), can be computed efficiently from the matrix $\mathbf{Q}_{k,\tau}^{-1} = (\mathbf{B} \bar{\mathbf{S}}_k^{(\tau)} \mathbf{B} + \lambda \mathbf{I}_M)^{-1}$. To see this, observe that the submatrix $\mathbf{S}_k^{(\tau-1)}$, obtained in the previous iteration, has a low-rank structure, as its rank is minimized in the COBRAS optimization. Hence, by performing successive rank-one updates of

$$\mathbf{Q}_{-k,\tau}^{-1} = (\mathbf{Q}_{k,\tau} - \mathbf{U}_k^{(\tau-1)} \Sigma_k^{(\tau-1)} \mathbf{U}_k^{(\tau-1)H})^{-1}, \quad (6.65)$$

based on the low-rank eigendecomposition of $\mathbf{S}_k^{(\tau-1)} = \mathbf{U}_k^{(\tau-1)} \boldsymbol{\Sigma}_k^{(\tau-1)} \mathbf{U}_k^{(\tau-1)\text{H}}$, direct matrix inversion of $\mathbf{Q}_{-k,\tau}^{-1}$ can be avoided, similar to the SPARROW-CD method discussed in Section 6.1.3. After updating $\mathbf{S}_k^{(\tau)}$ and the corresponding eigenvectors and eigenvalues in $\mathbf{U}_k^{(\tau)}$ and $\boldsymbol{\Sigma}_k^{(\tau)}$, by computing (6.63) and (6.64), the matrix inverse

$$\mathbf{Q}_{k+1,\tau}^{-1} = (\mathbf{Q}_{-k,\tau} + \mathbf{U}_k^{(\tau)} \boldsymbol{\Sigma}_k^{(\tau)} \mathbf{U}_k^{(\tau)\text{H}})^{-1}, \quad (6.66)$$

for computation of the next coordinate $k+1$ can be computed by successive rank-one updates of $\mathbf{Q}_{-k,\tau}^{-1}$. Hence, it suffices to track the eigenvectors and eigenvalues in $\mathbf{U}_k^{(\tau)}$ and $\boldsymbol{\Sigma}_k^{(\tau)}$ as well as the matrix inverse $\mathbf{Q}_{k,\tau}^{-1}$ for cyclic computation of (6.63). The proposed COBRAS-CD approach is summarized in Algorithm 6.

6.3 Numerical Experiments

For evaluation of the CD and STELA implementations discussed in the previous sections, basic MATLAB implementations are performed on a computer with an Intel Core i7-4770 CPU @ 3.40 GHz \times 8 and 16 GByte RAM. Numerical experiments show that the computation times of the CD and STELA methods not only depend on the number of measurements and variables, represented by the number of sensors M , the number of snapshots N and the number of grid points K . Furthermore, the computation time also depends on signal characteristics, such as the number of source signals L , the frequency spacing and the SNR, where, e.g., a larger number of source signals L and a higher SNR can result in higher computation time. For sake of brevity the experiments in this section will be limited to one scenario for the FCA and the PCA case, respectively, which demonstrates the main features of the different CD and STELA implementations.

6.3.1 Sparse Reconstruction in FCAs

For investigation of the computational cost of the $\ell_{2,1}$ -CD implementation in Algorithm 1, the $\ell_{2,1}$ -STELA implementation in Algorithm 2 and the SPARROW-CD implementation in Algorithm 3, consider a fully calibrated uniform linear array of $M = 100$ sensors. Furthermore, assume $L = 9$ equal power complex Gaussian source signals with spatial frequencies $\boldsymbol{\mu} = [-0.50, -0.48, -0.30, -0.28, -0.10, -0.08, 0.22, 0.24, 0.27]^\text{T}$ and consider $N = 100$ snapshots with an SNR of 0 dB. A grid of $K = 1000$ points is used for all CD and STELA implementations, resulting in a number of $KN = 200000$ real-valued parameters for the $\ell_{2,1}$ minimization problem and $K = 1000$ nonnegative optimization variables for the SPARROW problem.

To illustrate the convergence speed, Figures 6.1-6.3 consider the distance to optimality $\|\mathbf{S}_k^{(\tau)} - \hat{\mathbf{S}}\|_\text{F}$, which is given as the distance of the optimal solution $\hat{\mathbf{S}}$ to the approximate solution $\mathbf{S}_k^{(\tau)}$ in iteration τ for coordinate k .

The first Figure 6.1 displays the distance to optimality over the iteration index τ , where one iteration corresponds to one sweep over all coordinates. Clearly, in this representation the SPARROW-CD method provides the best convergence characteristics, obtaining a distance to optimality of 10^{-5} after roughly $\tau = 130$ iterations. In comparison, the $\ell_{2,1}$ -CD method and the $\ell_{2,1}$ -STELA method require about $\tau = 720$ iterations and $\tau = 1500$, respectively, to obtain the same distance to optimality. However, in terms of computation time the representation over the number of iterations τ is not fair since the CD methods suffer from the fact that the coordinate updates can only be performed in a sequential manner, while for the STELA methods all the coordinate updates for one iteration can be computed in parallel.

To provide better comparison of the CD and STELA methods, Figures 6.2-6.3 display the distance to optimality over the number of computation time slots, where one computation time slot for the CD methods corresponds to one sequential coordinate update while one computation time slot for STELA corresponds to the parallel update of all coordinates. In a naive implementation of the CD methods, the coordinate updates would be performed in a cyclic fashion over all available coordinates. The resulting computation time is displayed in Figure 6.2, showing that the STELA method clearly outperforms the two CD implementations.

A simple way to exploit sparsity in the CD implementations is based on the observation that coordinates with zero entries will usually not change in future iterations, and thus can be skipped in future iterations, as discussed in [FHHT07]. This simple approach reduces the number of sequential computations and, consequently, the computation time. However, it cannot be guaranteed that a coordinate entry will remain a zero value in future iterations. Hence, it is advisable to recompute all coordinates after a fixed number of iterations $\Delta\tau$. For the experiments it was found that a computation of all coordinates after every 10 iterations provides a good compromise of speed and accuracy. Figure 6.3 displays the resulting computation time for the sparsity exploiting CD implementations. Although the CD implementations cannot reach the convergence rate of the STELA method, they show drastically improved convergence rate as compared to the naive implementations in Figure 6.2. The re-computation of all coordinates in the CD implementations becomes visible from the step-like character of the corresponding error functions in Figure 6.3.

The MATLAB implementations resulted in a computation time of 15 seconds for the $\ell_{2,1}$ -CD implementation, 12 seconds for the $\ell_{2,1}$ -STELA implementation and 4 seconds for the SPARROW-CD implementation. This reduction in CPU time as compared to the experimental results for the SDP implementations with MOSEK in Section 4.4.6 can be explained by the low-complexity coordinate updates. Still, it should be noted that these computation times were achieved for very basic MATLAB implementations and might possibly be improved for more sophisticated implementations, e.g., in C/C++.

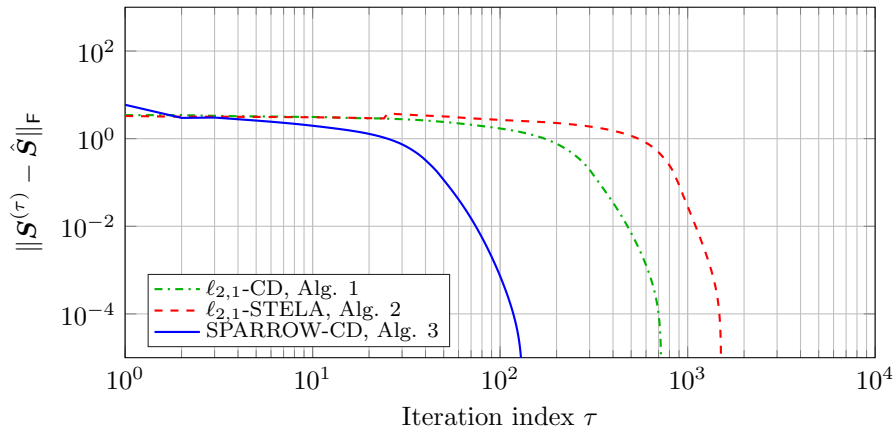


Figure 6.1: Convergence rate of FCA algorithms: Distance to optimality over iteration index

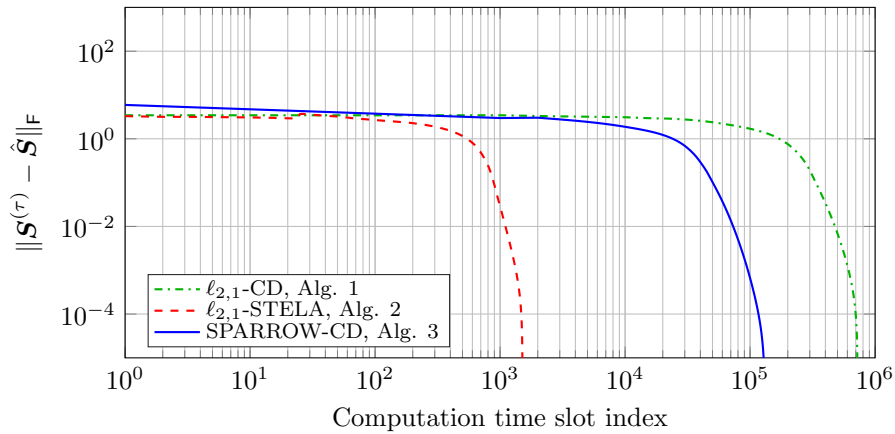


Figure 6.2: Convergence rate of FCA algorithms: Distance to optimality over computation time slots with naive CD implementations

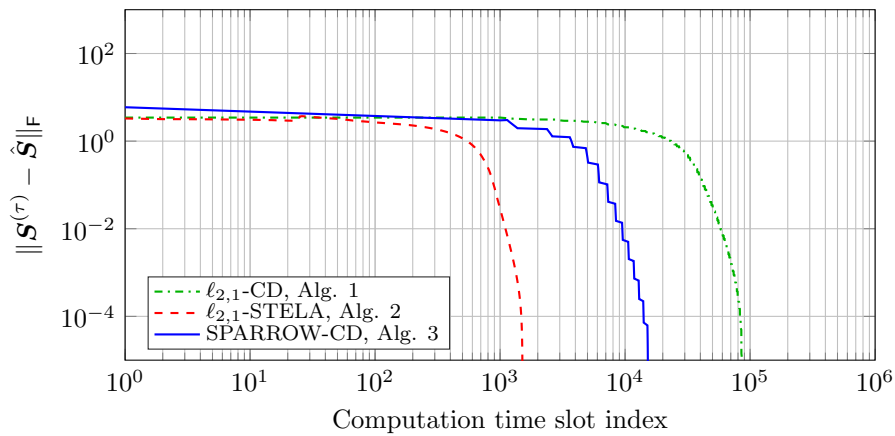


Figure 6.3: Convergence rate of FCA algorithms: Distance to optimality over computation time slots with sparse CD implementations

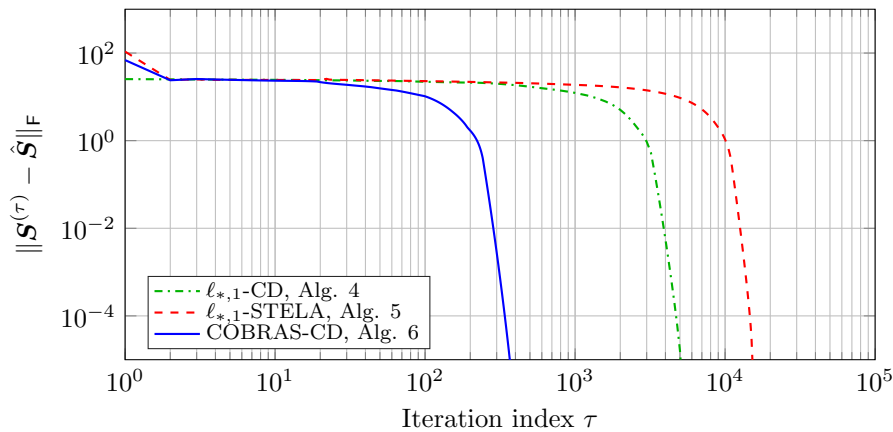


Figure 6.4: Convergence rate of PCA algorithms: Distance to optimality over iteration index

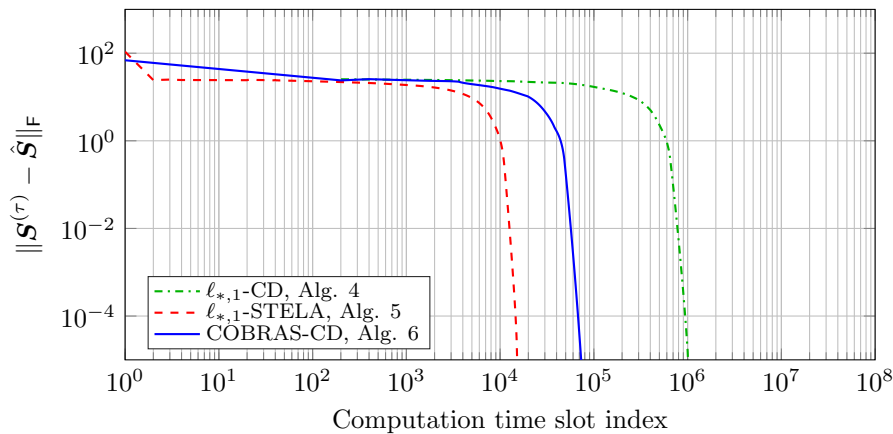


Figure 6.5: Convergence rate of PCA algorithms: Distance to optimality over computation time slots with naive CD implementations

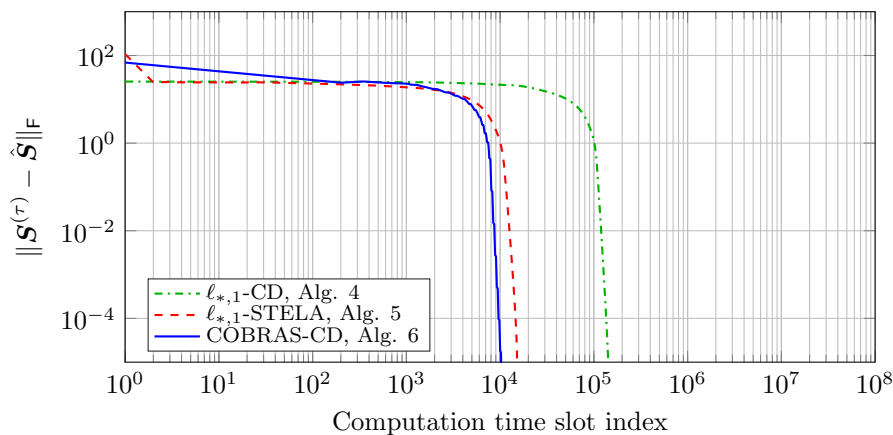


Figure 6.6: Convergence rate of PCA algorithms: Distance to optimality over computation time slots with sparse CD implementations

6.3.2 Sparse Reconstruction in PCAs

For further evaluation of the $\ell_{*,1}$ -CD method in Algorithm 4, the $\ell_{*,1}$ -STELA method in Algorithm 5 and the COBRAS-CD method in Algorithm 6, in this experiment a uniform linear array of $M = 100$ sensors is considered, partitioned in $P = 10$ identical uniform linear subarrays of $M_0 = 10$ sensors. A total of $N = 100$ snapshots is observed from $L = 8$ equal power, complex Gaussian sources signals with spatial frequencies $\boldsymbol{\mu} = [-0.50, -0.46, -0.3, -0.25, -0.08, 0.22, 0.27, 0.33]^\top$ and an SNR of 10 dB. A frequency grid of $K = 200$ points is employed, resulting in a total of $KPN = 400000$ real-valued parameters for the $\ell_{*,1}$ minimization problem and $KP^2 = 20000$ real-valued optimization parameters for the COBRAS formulation.

Figures 6.4-6.6 display the distance to optimality $\|\mathbf{S}_k^{(\tau)} - \hat{\mathbf{S}}\|_F$, as introduced in the previous section. In Figure 6.4 the distance to optimality over the iteration index τ is displayed, i.e., one iteration corresponds to one sweep over all coordinates. Clearly, the COBRAS-CD method provides the best convergence characteristics, obtaining a distance to optimality of 10^{-5} after roughly $\tau = 360$ iterations. In contrast, the $\ell_{*,1}$ -CD and STELA methods require about $\tau = 5000$ and $\tau = 15000$ iterations to obtain the same distance to optimality.

For a fair comparison in terms of convergence over computation time slots, Figures 6.5-6.6 compare the STELA implementation with naive and sparsity exploiting CD implementations, similar to the discussion in the previous section. From Figure 6.5 it can be observed that the $\ell_{*,1}$ -STELA implementation clearly outperforms the naive CD implementations. However, if sparsity is exploited in the CD implementations, as discussed in the previous section, the convergence rate of the CD implementations clearly improves, as displayed in Figure 6.6. In fact, in this representation the COBRAS-CD method even outperforms the STELA method with parallel implementation.

The computation time of the MATLAB implementations was measured as 10 minutes and 12 seconds for the $\ell_{*,1}$ -STELA method, where parallelization has not been exploited, 2 minutes and 26 seconds for the $\ell_{*,1}$ -CD method and 14 seconds for the COBRAS-CD method. Again, it should be emphasized that the computation time strongly depends on the implementation and the characteristics of the source signals and might change for other setups.

6.4 Chapter Summary

The chapter has considered a number of low-complexity methods for joint sparse signal reconstruction by means of convex relaxation. A short introductory review of the state of the art approach in terms of the block coordinate descent method in Section 6.1.1 was followed by a basic extension of the STELA method [YP17] for $\ell_{2,1}$ minimization in Section 6.1.2. The considerations on $\ell_{2,1}$ minimization were concluded by a novel

coordinate descent method for the equivalent SPARROW formulation in Section 6.1.3, which forms one of the major contributions of this chapter.

As another major contribution of this chapter, the methods presented for $\ell_{2,1}$ minimization and the SPARROW formulation were modified for application with $\ell_{*,1}$ minimization and the COBRAS formulation. In Sections 6.2.1 and 6.2.2 two novel low-complexity methods for $\ell_{*,1}$ mixed-norm minimization were proposed, that are based on the coordinate descent and the STELA method. A novel coordinate descent implementation of the COBRAS problem was presented in Section 6.2.3.

Numerical experiments for scenarios with large sensor arrays and large number of snapshots have demonstrated the benefits of the different proposed methods.

Chapter 7

Conclusion and Outlook

This thesis has studied the application of joint sparse signal reconstruction methods in fully and partly calibrated sensor arrays. A novel method for coherent sparse reconstruction in partly calibrated arrays and compact formulations for convex sparse reconstruction have been derived, that admit efficient implementation and provide new perspectives on sparse signal reconstruction by means of convex relaxation.

After introducing the signal model and giving a short overview of the state of the art in Chapters 2 and 3, joint sparse signal reconstruction in fully calibrated arrays has been considered in Chapter 4. The major contribution of Chapter 4 is given by a compact formulation, termed as SPARROW (SPARse-ROW norm reconstruction), for joint sparse signal reconstruction in fully calibrated arrays. In context with existing methods, it was shown that the SPARROW formulation is equivalent to the classical $\ell_{2,1}$ mixed-norm minimization problem and the atomic norm minimization problem for joint sparse signal reconstruction from multiple measurement vectors. At the same time, the SPARROW formulation provides a significant reduction in optimization parameters, as compared to $\ell_{2,1}$ and atomic norm minimization, and only relies on the sample covariance matrix instead of the instantaneous measurement vectors, leading to a reduction in computational cost and memory requirements. While the original SPARROW formulation considers grid-based signal reconstruction in arbitrary array topologies, a gridless as well as an off-grid version of the SPARROW formulation for application in uniform linear and arbitrary array topologies have been derived in Sections 4.3.1 and 4.3.2, respectively.

Chapter 5 has considered the application of joint sparse signal reconstruction in partly calibrated arrays. One of the major contributions of this chapter is given in Section 5.2, by a novel method for joint sparse signal reconstruction in partly calibrated arrays with coherent processing. The novel method is based on minimization of a mixed nuclear and ℓ_1 norm and is referred to as $\ell_{*,1}$ mixed-norm minimization. While, existing methods for sparse reconstruction in partly calibrated arrays under incoherent processing exploit a block-sparse structure in the incoherent signal model, the proposed $\ell_{*,1}$ minimization exploits an additional low-rank structure in the coherent signal model, leading to improved estimation performance.

Another major contribution of Chapter 5 is given in Section 5.3, by a compact, equivalent formulation of the $\ell_{*,1}$ mixed-norm minimization problem, referred to as COBRAS (COmpact Block- and RAnk-Sparse recovery). The COBRAS formulation is attractive especially in the case of a large number of signal snapshots. For the special case of identical uniform linear subarrays, a gridless version of the COBRAS method (GL-COBRAS) has been presented in Section 5.3.1. A generalization of the GL-COBRAS method was proposed in Section 5.4. The corresponding method is applicable in

PCAs composed of identical subarrays with arbitrary topology and termed as Shift-Invariant (SI-) SPARROW. Although being introduced for application in PCAs, the SI-SPARROW method can similarly be applied to any type of arrays exhibiting shift-invariances, including, e.g., centro-symmetric arrays considered in Section 2.2.5.

As discussed in Sections 5.1 and 5.2, the sparse reconstruction problems for fully calibrated arrays and partly calibrated arrays with incoherent and coherent processing are strongly related, and coincide in the special cases of a single subarray or a single snapshot. Similarly, the SPARROW and COBRAS formulations have strong links and one can be derived from the other, as shown in Section 5.3.1.

The numerical experiments for partly calibrated arrays in Section 5.5.2 have clearly demonstrated that the coherent processing methods outperform their incoherent counterparts. Regarding the comparison with state of the art subspace-based methods, the numerical experiments for fully and partly calibrated arrays in Sections 4.4 and 5.5 have illustrated that sparse reconstruction methods by means of convex relaxation provide a viable supplement to classical subspace-based methods, such as MUSIC, RARE and ESPRIT. Gains in estimation performance can be obtained especially for correlated signals as well as in the non-asymptotic regime of low signal-to-noise ratio and low number of snapshots. In these scenarios, especially the grid-based sparse reconstruction methods show significantly improved resolution performance over the grid-based versions of the subspace-based methods. These benefits of sparse reconstruction in estimation performance are slightly diminished by an inherent parameter estimation bias, which can however be compensated by appropriate post-processing methods, as discussed in Section 4.4.2.

For efficient implementation of the sparse reconstruction methods discussed in this thesis, Chapter 6 has presented various novel low-complexity methods based on the coordinate descent and the STELA method [YP17], which admit sequential and parallel computation, respectively. As demonstrated by numerical experiments, the coordinate descent implementations for SPARROW and COBRAS significantly outperform their counterparts for $\ell_{2,1}$ and $\ell_{*,1}$ minimization. It is left for future research to transfer this reduction in computation time to parallel implementations based on the STELA framework.

While reconstruction guarantees for joint sparse signal reconstruction for fully calibrated arrays are available, e.g., in [CH06], the derivation of such guarantees for sparse reconstruction in partly calibrated arrays is still an open problem. In this regard, the original problems for block- and rank-sparse reconstruction in (5.20) and for block-sparse reconstruction in (5.9) can serve as a starting point to derive such reconstruction guarantees. Related results for further research are given in [EM09, EKB10, WWX13, APJ16]. Another point of interest in this context is, under what conditions the convex relaxations in (5.23) and (5.14) can solve the original block- and rank-sparse reconstruction problem (5.20) and the block-sparse reconstruction problem (5.9). In this context, the nullspace property for nuclear norm minimization [DF10] might be of interest for further research.

A promising extension of the sparse reconstruction methods proposed in this thesis is given by reweighted minimization, to enhance sparsity and parameter estimation performance. Reweighted minimization was originally proposed for ℓ_1 minimization [CWB08] and has similarly been applied for nuclear norm minimization [FHB03, MF10] and atomic norm minimization [YX15a], where it was found to provide improved signal reconstruction performance.

Considering the strong links between the COBRAS, SPARROW and SPICE methods, as discussed in Sections 5.3 and 4.3.3, the structural constraints of GL-COBRAS (5.44) and SI-SPARROW (5.73) could likewise be implemented in the gridless SPICE methods (4.101) and (4.102). This approach would make the SPICE methods applicable for PCAs, admitting hyperparameter-free estimation.

Regarding related applications, it was shown in [SYP16] that the $\ell_{*,1}$ minimization approach can similarly be applied to multidimensional parameter estimation. In this context it would be of interest to further investigate the estimation performance of the COBRAS and SI-SPARROW formulations for multidimensional estimation, and to put the results in relation with more recent publications on sparse recovery for multidimensional estimation [YXS16, TZW17].

Appendix

A Regularization Parameter for $\ell_{2,1}$ Minimization

Based on the ideas in [OPT00a] for ℓ_1 minimization, consider the subdifferential of the $\ell_{2,1}$ minimization problem in (4.50), given as

$$\partial_{\mathbf{X}} \left(\frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\mathbb{F}}^2 + \lambda\sqrt{N} \|\mathbf{X}\|_{2,1} \right) = -\mathbf{A}^{\text{H}} (\mathbf{Y} - \mathbf{A}\mathbf{X}) + \lambda\sqrt{N}\mathbf{H}, \quad (\text{A.1})$$

where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^{\text{T}}$ is a subgradient of the penalty term $\|\mathbf{X}\|_{2,1} = \sum_{k=1}^K \|\mathbf{x}_k\|_2$, with rows $\mathbf{h}_k \in \partial \|\mathbf{x}_k\|_2$ being elements of the subdifferentials of the ℓ_2 row-norms $\|\mathbf{x}_k\|_2$ of the matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]^{\text{T}}$. The subdifferential [Roc70, Ber99] of the ℓ_2 norm is defined as

$$\begin{aligned} \partial \|\mathbf{x}_k\|_2 &= \{ \mathbf{h}_k \mid \|\tilde{\mathbf{x}}_k\|_2 \geq \|\mathbf{x}_k\|_2 + \text{Re}\{(\tilde{\mathbf{x}}_k - \mathbf{x}_k)^{\text{H}}\mathbf{h}_k\} \} \\ &= \begin{cases} \mathbf{x}_k / \|\mathbf{x}_k\|_2 & \text{for } \mathbf{x}_k \neq \mathbf{0} \\ \{ \mathbf{h}_k \mid \|\mathbf{h}_k\|_2 \leq 1 \} & \text{for } \mathbf{x}_k = \mathbf{0} \end{cases} \end{aligned}$$

with $\mathbf{h}_k, \tilde{\mathbf{x}}_k \in \mathbb{C}^N$, for $k = 1, \dots, K$. From this it can be concluded that

$$\|\mathbf{h}_k\|_2 \begin{cases} = 1 & \text{for } \mathbf{x}_k \neq \mathbf{0} \\ \leq 1 & \text{for } \mathbf{x}_k = \mathbf{0}. \end{cases} \quad (\text{A.2})$$

Hence, the row support of \mathbf{X} can similarly be identified from a subgradient \mathbf{H} by identifying the rows \mathbf{h}_k with ℓ_2 norm equal to 1.

Optimality conditions require that the subdifferential (A.1) contains the $\mathbf{0}$ matrix, i.e., there must exist matrices $\hat{\mathbf{X}}$ and $\hat{\mathbf{H}}$ that fulfill

$$-\mathbf{A}^{\text{H}} (\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}}) + \lambda\sqrt{N}\hat{\mathbf{H}} = \mathbf{0}, \quad (\text{A.3})$$

and equivalently the row-vectors must obey

$$\lambda\hat{\mathbf{h}}_k = \mathbf{a}_k^{\text{H}} (\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}}) / \sqrt{N}. \quad (\text{A.4})$$

Since $\|\hat{\mathbf{h}}_k\|_2 \leq 1$, it follows from (A.4) that

$$\lambda \geq \|\mathbf{a}_k^{\text{H}} (\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}})\|_2 / \sqrt{N}, \quad (\text{A.5})$$

which shows that for a regularization parameter selection according to

$$\lambda \geq \|\mathbf{a}_k^{\text{H}} \mathbf{Y}\|_2 / \sqrt{N} \quad (\text{A.6})$$

a zero solution $\hat{\mathbf{X}} = \mathbf{0}$ is obtained for the $\ell_{2,1}$ minimization problem (4.50), independent of the underlying sparse signal $\check{\mathbf{X}}$ in (4.29). Under the assumption that there are no

source signals present in the measurements, i.e. $\hat{\mathbf{X}} = \mathbf{0}$ such that $\mathbf{Y} = \mathbf{N}$, condition (A.6) reduces to

$$\lambda \geq \|\mathbf{a}_k^H \mathbf{N}\|_2 / \sqrt{N}. \quad (\text{A.7})$$

Thus, to obtain a proper zero estimate $\hat{\mathbf{X}} = \mathbf{0}$ for the $\ell_{2,1}$ minimization problem (4.50) with $\mathbf{Y} = \mathbf{N}$, the regularization parameter λ must fulfill condition (A.7), for $k = 1, \dots, K$. In terms of statistical expectation, an upper bound on the right hand side in (A.7) is given as

$$\begin{aligned} \mathbb{E}\left\{\max_k \|\mathbf{a}_k^H \mathbf{N}\|_2 / \sqrt{N}\right\} &\leq \mathbb{E}\left\{\max_k \|\mathbf{a}_k\|_2 \|\mathbf{N}\|_2 / \sqrt{N}\right\} \\ &= \max_k \|\mathbf{a}_k\|_2 \mathbb{E}\{\|\mathbf{N}\|_2\} / \sqrt{N} \\ &\leq \max_k \sigma_N \|\mathbf{a}_k\|_2 (\sqrt{M} + \sqrt{N}) / \sqrt{N} \\ &= \sigma_N \sqrt{M} (\sqrt{M/N} + 1), \end{aligned} \quad (\text{A.8})$$

where $\|\mathbf{a}_k\|_2 = \sqrt{M}$, for $k = 1, \dots, K$, and the expectation of the spectral norm of an $M \times N$ complex Gaussian matrix, with entries according to $[\mathbf{N}]_{m,n} \sim \mathcal{N}_C(0, \sigma_N^2)$, is bounded as [RV10, T⁺15, PE10]

$$\mathbb{E}\{\|\mathbf{N}\|_2\} \leq \sigma(\sqrt{M} + \sqrt{N}). \quad (\text{A.9})$$

From (A.8) it can be concluded that a regularization parameter according to

$$\lambda = \sigma_N \sqrt{M} (\sqrt{M/N} + 1) \quad (\text{A.10})$$

fulfills condition (A.7) and will suppress the noise in the signal estimate, retaining only signal components in the sparse estimate $\hat{\mathbf{X}}$. Note that the bound on the spectral norm in (A.9) is tight for large values of sensors M and snapshots N . In the non-asymptotic case of small values for M or N , the bound in (A.9) might be loose and the regularization parameter might be too large, leading to suppression not only of noise, but also of signal components in the estimate $\hat{\mathbf{X}}$.

B Equivalence of SPARROW and $\ell_{2,1}$ Minimization

Proof of Theorem 4.2. A key component in establishing the equivalence in Equations (4.50) and (4.51) is the observation that the ℓ_2 norm of a vector $\mathbf{x}_k \in \mathbb{C}^N$ can be rewritten as

$$\|\mathbf{x}_k\|_2 = \min_{\gamma_k, \mathbf{h}_k} \frac{1}{2} (|\gamma_k|^2 + \|\mathbf{h}_k\|_2^2) \quad (\text{B.1a})$$

$$\text{s.t. } \gamma_k \mathbf{h}_k = \mathbf{x}_k, \quad (\text{B.1b})$$

where γ_k is a complex-valued scalar and \mathbf{h}_k is a complex-valued vector of N elements. For the optimal solution of (B.1), it holds that

$$\|\mathbf{x}_k\|_2 = |\gamma_k|^2 = \|\mathbf{h}_k\|_2^2. \quad (\text{B.2})$$

To see this, consider that any feasible solution must fulfill

$$\|\mathbf{x}_k\|_2 = \sqrt{|\gamma_k|^2 \|\mathbf{h}_k\|_2^2} \leq \frac{1}{2}(|\gamma_k|^2 + \|\mathbf{h}_k\|_2^2), \quad (\text{B.3})$$

based on the inequality of arithmetic and geometric means, with equality holding if and only if $|\gamma_k| = \|\mathbf{h}_k\|_2$.

The idea in (B.1) can be extended to the $\ell_{2,1}$ mixed-norm of the source signal matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]^\top$ according to

$$\|\mathbf{X}\|_{2,1} = \sum_{k=1}^K \|\mathbf{x}_k\|_2 = \min_{\mathbf{\Gamma} \in \mathcal{D}^K, \mathbf{H} \in \mathbb{C}^{K \times N}} \frac{1}{2} (\|\mathbf{\Gamma}\|_{\text{F}}^2 + \|\mathbf{H}\|_{\text{F}}^2) \quad (\text{B.4a})$$

$$\text{s.t.} \quad \mathbf{X} = \mathbf{\Gamma}\mathbf{H}, \quad (\text{B.4b})$$

where $\mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_K)$ is taken from the set \mathcal{D}^K of $K \times K$ complex-valued diagonal matrices and $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^\top$ is a $K \times N$ complex-valued matrix with rows \mathbf{h}_k , for $k = 1, \dots, K$. Upon substitution of (B.4) into the $\ell_{2,1}$ minimization problem in (4.50), the following minimization problem can be formulated

$$\min_{\mathbf{\Gamma} \in \mathcal{D}^K, \mathbf{H} \in \mathbb{C}^{K \times N}} \frac{1}{2} \|\mathbf{A}\mathbf{\Gamma}\mathbf{H} - \mathbf{Y}\|_{\text{F}}^2 + \frac{\lambda\sqrt{N}}{2} (\|\mathbf{\Gamma}\|_{\text{F}}^2 + \|\mathbf{H}\|_{\text{F}}^2). \quad (\text{B.5})$$

For a fixed matrix $\mathbf{\Gamma}$, the minimizer $\hat{\mathbf{H}}$ of problem (B.5) admits the closed form expression

$$\begin{aligned} \hat{\mathbf{H}} &= (\mathbf{\Gamma}^\text{H} \mathbf{A}^\text{H} \mathbf{A} \mathbf{\Gamma} + \lambda\sqrt{N} \mathbf{I}_K)^{-1} \mathbf{\Gamma}^\text{H} \mathbf{A}^\text{H} \mathbf{Y} \\ &= \mathbf{\Gamma}^\text{H} \mathbf{A}^\text{H} (\mathbf{A} \mathbf{\Gamma} \mathbf{\Gamma}^\text{H} \mathbf{A}^\text{H} + \lambda\sqrt{N} \mathbf{I}_M)^{-1} \mathbf{Y}, \end{aligned} \quad (\text{B.6})$$

where the last identity is derived from the matrix inversion lemma [Hag89]. Reinserting the optimal matrix $\hat{\mathbf{H}}$ into Equation (B.5) and performing basic reformulations of the objective function results in the compact minimization problem

$$\min_{\mathbf{\Gamma} \in \mathcal{D}^K} \frac{\lambda\sqrt{N}}{2} \left(\text{Tr}((\mathbf{A} \mathbf{\Gamma} \mathbf{\Gamma}^\text{H} \mathbf{A}^\text{H} + \lambda\sqrt{N} \mathbf{I}_M)^{-1} \mathbf{Y} \mathbf{Y}^\text{H}) + \text{Tr}(\mathbf{\Gamma} \mathbf{\Gamma}^\text{H}) \right). \quad (\text{B.7})$$

Upon substituting $\mathbf{Y} \mathbf{Y}^\text{H} = N \hat{\mathbf{R}}$ and defining the nonnegative diagonal matrix

$$\mathbf{S} = \mathbf{\Gamma} \mathbf{\Gamma}^\text{H} / \sqrt{N} \in \mathcal{D}_+^K, \quad (\text{B.8})$$

the problem in (B.7) can be rewritten as

$$\min_{\mathbf{S} \in \mathcal{D}_+^K} \frac{\lambda N}{2} \left(\text{Tr}((\mathbf{A} \mathbf{S} \mathbf{A}^\text{H} + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \right). \quad (\text{B.9})$$

Ignoring the factor $\lambda N/2$ in (B.9), results in formulation (4.51). From Equation (B.2) and the definition of $\mathbf{S} = \text{diag}(s_1, \dots, s_K)$ in (B.8) it can furthermore be concluded that

$$s_k = \frac{1}{\sqrt{N}} \|\mathbf{x}_k\|_2, \quad (\text{B.10})$$

for $k = 1, \dots, K$, as given by (4.53). Making further use of the factorization in (B.4b) yields

$$\begin{aligned}\hat{\mathbf{X}} &= \hat{\Gamma} \hat{\mathbf{H}} \\ &= \hat{\Gamma} \hat{\Gamma}^H \mathbf{A}^H (\mathbf{A} \hat{\Gamma} \hat{\Gamma}^H \mathbf{A}^H + \lambda \sqrt{N} \mathbf{I}_M)^{-1} \mathbf{Y} \\ &= \hat{\mathbf{S}} \mathbf{A}^H (\mathbf{A} \hat{\mathbf{S}} \mathbf{A}^H + \lambda \mathbf{I}_M)^{-1} \mathbf{Y}\end{aligned}\tag{B.11}$$

which is (4.52). \square

C Convexity of the SPARROW Problem

One approach of proving convexity of a function is by showing positive semidefiniteness of its Hessian matrix. To this end, consider the objective function of the SPARROW formulation in (4.50)

$$f(\mathbf{s}) = \text{Tr}(\mathbf{Q}^{-1} \hat{\mathbf{R}}) + \mathbf{1}^T \mathbf{s},\tag{C.1}$$

where $\mathbf{Q} = \mathbf{A} \text{diag}(\mathbf{s}) \mathbf{A}^H + \lambda \mathbf{I}$, the vector $\mathbf{s} = [s_1, \dots, s_K]^T$ contains the nonnegative elements $s_1, \dots, s_K \geq 0$ and $\mathbf{1}$ is a vector of ones. Using the elementwise derivatives

$$\frac{\partial \mathbf{Q}}{\partial s_k} = \frac{\partial}{\partial s_k} \sum_{k=1}^K s_k \mathbf{a}_k \mathbf{a}_k^H + \lambda \mathbf{I} = \mathbf{a}_k \mathbf{a}_k^H\tag{C.2}$$

$$\frac{\partial \mathbf{Q}^{-1}}{\partial s_k} = -\mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial s_k} \mathbf{Q}^{-1} = -\mathbf{Q}^{-1} \mathbf{a}_k \mathbf{a}_k^H \mathbf{Q}^{-1}\tag{C.3}$$

the gradient of (C.1) is given as

$$\frac{\partial f(\mathbf{s})}{\partial \mathbf{s}} = \mathbf{1} - \text{vecd}(\mathbf{A}^H \mathbf{Q}^{-1} \hat{\mathbf{R}} \mathbf{Q}^{-1} \mathbf{A})\tag{C.4}$$

where $\text{vecd}(\mathbf{X})$ denotes the vector containing the elements on the main diagonal of matrix \mathbf{X} . The Hessian matrix of (C.1) is computed as

$$\frac{\partial^2 f(\mathbf{s})}{\partial \mathbf{s} \partial \mathbf{s}^T} = 2 \text{Re}\{(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A})^T \odot (\mathbf{A}^H \mathbf{Q}^{-1} \hat{\mathbf{R}} \mathbf{Q}^{-1} \mathbf{A})\},\tag{C.5}$$

with \odot denoting the Hadamard product, i.e., elementwise multiplication. From the Schur product theorem [HJ90] it can be concluded that the Hessian matrix in (C.5) is positive semidefinite, since for $s_1, \dots, s_K \geq 0$ it holds that $\mathbf{Q} \succeq \mathbf{0}$. In other words, the SPARROW formulation in (C.1) is convex for nonnegative $s_1, \dots, s_K \geq 0$. Note also that the SPARROW function is strictly convex in its single components, since

$$\frac{\partial^2 f(\mathbf{s})}{\partial s_k^2} = 2 \text{Re}\{(\mathbf{a}_k^H \mathbf{Q}^{-1} \mathbf{a}_k) \cdot (\mathbf{a}_k^H \mathbf{Q}^{-1} \hat{\mathbf{R}} \mathbf{Q}^{-1} \mathbf{a}_k)\} > 0.\tag{C.6}$$

D Equivalence of SPARROW and AST

Proof of Theorem 4.5. Consider the GL-SPARROW formulation

$$\min_{\mathbf{u}, \mathbf{U}} \frac{\lambda}{2} \text{Tr}(\mathbf{U}) + \frac{\lambda N}{2M} \text{Tr}(\text{HToep}(\mathbf{u})) \quad (\text{D.1a})$$

$$\text{s.t.} \begin{bmatrix} \mathbf{U}/\sqrt{N} & \mathbf{Y}^{\text{H}} \\ \mathbf{Y} & \sqrt{N} \text{HToep}(\mathbf{u}) + \lambda\sqrt{N} \mathbf{I}_M \end{bmatrix} \succeq \mathbf{0} \quad (\text{D.1b})$$

$$\text{HToep}(\mathbf{u}) \succeq \mathbf{0}, \quad (\text{D.1c})$$

where the objective function is scaled by the factor $\lambda N/2$ as compared to (4.60) and the Schur complement of constraint (D.1b) is identical to that of (4.60b). Furthermore, consider the ANM formulation (4.49)

$$\min_{\mathbf{v}, \mathbf{V}, \mathbf{Y}_0} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}_0\|_{\text{F}}^2 + \frac{\lambda\sqrt{N}}{2} \left(\text{Tr}(\mathbf{V}) + \frac{1}{M} \text{Tr}(\text{HToep}(\mathbf{v})) \right) \quad (\text{D.2a})$$

$$\text{s.t.} \begin{bmatrix} \mathbf{V} & \mathbf{Y}_0^{\text{H}} \\ \mathbf{Y}_0 & \text{HToep}(\mathbf{v}) \end{bmatrix} \succeq \mathbf{0}. \quad (\text{D.2b})$$

Both problems are equivalent in the sense that the corresponding minimizers are related by

$$\hat{\mathbf{v}} = \sqrt{N} \hat{\mathbf{u}} \quad (\text{D.3})$$

$$\hat{\mathbf{Y}}_0 = \text{HToep}(\hat{\mathbf{u}}) (\text{HToep}(\hat{\mathbf{u}}) + \lambda \mathbf{I}_M)^{-1} \mathbf{Y} \quad (\text{D.4})$$

$$\hat{\mathbf{U}} = \sqrt{N} \hat{\mathbf{V}} + \frac{1}{\lambda} (\mathbf{Y} - \hat{\mathbf{Y}}_0)^{\text{H}} (\mathbf{Y} - \hat{\mathbf{Y}}_0). \quad (\text{D.5})$$

Inserting (D.3) and (D.5) into the objective functions (D.1a) and (D.2a), it can easily be verified that both problems achieve the same minimum value. It remains to show that the optimal point $(\hat{\mathbf{U}}, \hat{\mathbf{u}})$ of the GL-SPARROW formulation (D.1) is feasible for the ANM formulation (D.2), and, conversely, that the optimal point $(\hat{\mathbf{V}}, \hat{\mathbf{v}}, \hat{\mathbf{Y}}_0)$ of the ANM formulation (D.2) is feasible for the GL-SPARROW formulation (D.1).

First it will be shown that the optimal point $(\hat{\mathbf{V}}, \hat{\mathbf{v}}, \hat{\mathbf{Y}}_0)$ of the ANM formulation (D.2) is feasible for the GL-SPARROW formulation (D.1). Defining $\mathbf{W} = \mathbf{Y} - \hat{\mathbf{Y}}_0$ and inserting (D.3) and (D.5) into the GL-SPARROW constraint (D.1b) results in

$$\begin{aligned} & \begin{bmatrix} \hat{\mathbf{U}}/\sqrt{N} & \mathbf{Y}^{\text{H}} \\ \mathbf{Y} & \sqrt{N} \text{HToep}(\hat{\mathbf{u}}) + \lambda\sqrt{N} \mathbf{I}_M \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{V}} + \frac{1}{\lambda\sqrt{N}} \mathbf{W}^{\text{H}} \mathbf{W} & \mathbf{Y}^{\text{H}} \\ \mathbf{Y} & \text{HToep}(\hat{\mathbf{v}}) + \lambda\sqrt{N} \mathbf{I}_M \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{V}} & \mathbf{Y}_0^{\text{H}} \\ \mathbf{Y}_0 & \text{HToep}(\hat{\mathbf{v}}) \end{bmatrix} + \begin{bmatrix} \frac{1}{\lambda\sqrt{N}} \mathbf{W}^{\text{H}} \mathbf{W} & \mathbf{W}^{\text{H}} \\ \mathbf{W} & \lambda\sqrt{N} \mathbf{I}_M \end{bmatrix} \\ & \succeq \mathbf{0}. \end{aligned} \quad (\text{D.6})$$

From (D.6) it can be seen that any minimizers $(\hat{\mathbf{V}}, \hat{\mathbf{v}}, \hat{\mathbf{Y}}_0)$ of the ANM problem (D.2), fulfilling constraint (D.2b), are feasible for the GL-SPARROW problem (D.1), since

$$\begin{bmatrix} \frac{1}{\lambda\sqrt{N}} \mathbf{W}^H \mathbf{W} & \mathbf{W}^H \\ \mathbf{W} & \lambda\sqrt{N} \mathbf{I}_M \end{bmatrix} = \lambda\sqrt{N} \begin{bmatrix} \frac{1}{\lambda\sqrt{N}} \mathbf{W}^H \\ \mathbf{I}_M \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda\sqrt{N}} \mathbf{W}^H \\ \mathbf{I}_M \end{bmatrix}^H \succeq \mathbf{0}. \quad (\text{D.7})$$

In the second step it is proven that an optimal point $(\hat{\mathbf{u}}, \hat{\mathbf{U}})$ of (D.1) is feasible for the ANM problem (D.2). According to Corollary 4.3 it can be assumed w.l.o.g. that

$$\hat{\mathbf{U}} = \mathbf{Y}^H (\text{HToep}(\hat{\mathbf{u}}) + \lambda \mathbf{I}_M)^{-1} \mathbf{Y} \quad (\text{D.8})$$

is optimal for (D.1). Using (D.4) and (D.8) in (D.5) and solving for $\hat{\mathbf{V}}$ results in

$$\hat{\mathbf{V}} = \frac{1}{\sqrt{N}} \mathbf{Y}^H (\text{HToep}(\hat{\mathbf{u}}) + \lambda \mathbf{I}_M)^{-1} \text{HToep}(\hat{\mathbf{u}}) (\text{HToep}(\hat{\mathbf{u}}) + \lambda \mathbf{I}_M)^{-1} \mathbf{Y}. \quad (\text{D.9})$$

Considering constraint (D.2b) of the ANM problem and inserting (D.4) and (D.9) it can be seen that

$$\begin{aligned} & \begin{bmatrix} \mathbf{V} & \mathbf{Y}_0^H \\ \mathbf{Y}_0 & \text{HToep}(\mathbf{v}) \end{bmatrix} \\ &= \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{Y}^H (\text{HToep}(\hat{\mathbf{u}}) + \lambda \mathbf{I}_M)^{-1} \\ \sqrt{N} \mathbf{I} \end{bmatrix} \text{HToep}(\hat{\mathbf{u}}) \begin{bmatrix} \mathbf{Y}^H (\text{HToep}(\hat{\mathbf{u}}) + \lambda \mathbf{I}_M)^{-1} \\ \sqrt{N} \mathbf{I} \end{bmatrix}^H \\ &\succeq \mathbf{0}, \end{aligned} \quad (\text{D.10})$$

i.e., a feasible and optimal point $(\hat{\mathbf{V}}, \hat{\mathbf{v}}, \hat{\mathbf{Y}}_0)$ for the ANM formulation can be constructed from the minimizers $(\hat{\mathbf{u}}, \hat{\mathbf{U}})$ of the GL-SPARROW problem, which concludes the proof. \square

E Dual Problem of the SPARROW Formulation

Consider the SDP form (4.57) of the SPARROW formulation, given as

$$\min_{\mathbf{S}, \mathbf{W}_M} \text{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \quad (\text{E.1a})$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I}_M \\ \mathbf{I}_M & \mathbf{A} \mathbf{S} \mathbf{A}^H + \lambda \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \quad (\text{E.1b})$$

$$\mathbf{S} \in \mathcal{D}_+^K, \mathbf{W}_M \succeq \mathbf{0}, \quad (\text{E.1c})$$

where $\mathbf{S} = \text{diag}(s_1, \dots, s_K)$. Upon introducing the matrix

$$\tilde{\mathbf{W}} = \begin{bmatrix} \tilde{\mathbf{W}}_{11} & \tilde{\mathbf{W}}_{12} \\ \tilde{\mathbf{W}}_{21} & \tilde{\mathbf{W}}_{22} \end{bmatrix} \quad (\text{E.2})$$

and summarizing $\mathbf{s} = [s_1, \dots, s_K]^T$, the SDP in (E.1) can be rewritten as

$$\min_{\mathbf{s}, \tilde{\mathbf{W}}} \text{Tr}(\tilde{\mathbf{W}}_{11} \hat{\mathbf{R}}) + \sum_{k=1}^K s_k \quad (\text{E.3a})$$

$$\text{s.t. } s_k \geq 0, \text{ for } k = 1, \dots, K \quad (\text{E.3b})$$

$$\tilde{\mathbf{W}} \succeq \mathbf{0} \quad (\text{E.3c})$$

$$\tilde{\mathbf{W}}_{12} = \mathbf{I}, \tilde{\mathbf{W}}_{21} = \mathbf{I} \quad (\text{E.3d})$$

$$\tilde{\mathbf{W}}_{22} = \sum_{k=1}^K s_k \mathbf{a}_k \mathbf{a}_k^H + \lambda \mathbf{I}. \quad (\text{E.3e})$$

The Lagrangian function [VB96] for the problem in (E.3) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{s}, \tilde{\mathbf{W}}, \boldsymbol{\omega}, \boldsymbol{\Gamma}, \boldsymbol{\Upsilon}) = & \text{Tr}(\tilde{\mathbf{W}}_{11} \hat{\mathbf{R}}) + \sum_{k=1}^K s_k \\ & - \sum_{k=1}^K \omega_k s_k - \text{Re}\{\text{Tr}(\boldsymbol{\Gamma} \tilde{\mathbf{W}})\} \\ & + \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{21}(\tilde{\mathbf{W}}_{12} - \mathbf{I}))\} + \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{12}(\tilde{\mathbf{W}}_{21} - \mathbf{I}))\} \\ & + \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{22}(\tilde{\mathbf{W}}_{22} - \sum_{k=1}^K s_k \mathbf{a}_k \mathbf{a}_k^H - \lambda \mathbf{I}))\}, \end{aligned} \quad (\text{E.4})$$

with $\omega_1, \dots, \omega_K \geq 0$, summarized in $\boldsymbol{\omega} = [\omega_1, \dots, \omega_K]^T$, being the Lagrangian multipliers accounting for the nonnegative constraint on s_1, \dots, s_K in (E.3b), and $\boldsymbol{\Gamma} \succeq \mathbf{0}$ being the Lagrangian multipliers accounting for the positive semidefinite constraint with regard to $\tilde{\mathbf{W}}$ in (E.3c). Furthermore, the matrices $\boldsymbol{\Upsilon}_{12}, \boldsymbol{\Upsilon}_{21}, \boldsymbol{\Upsilon}_{22}$, summarized in

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{\Upsilon}_{11} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix}, \quad (\text{E.5})$$

are the Lagrangian multipliers accounting for the equality constraints on submatrices $\tilde{\mathbf{W}}_{12}, \tilde{\mathbf{W}}_{21}, \tilde{\mathbf{W}}_{22}$ in (E.3d) and (E.3e). In a more compact notation, the Lagrangian function (E.4) can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{s}, \tilde{\mathbf{W}}, \boldsymbol{\omega}, \boldsymbol{\Gamma}, \boldsymbol{\Upsilon}) = & \text{Re}\left\{\text{Tr}\left(\left(\begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} - \boldsymbol{\Gamma}\right) \begin{bmatrix} \tilde{\mathbf{W}}_{11} & \tilde{\mathbf{W}}_{12} \\ \tilde{\mathbf{W}}_{21} & \tilde{\mathbf{W}}_{22} \end{bmatrix}\right)\right\} \\ & + \sum_{k=1}^K s_k ((1 - \omega_k) - \mathbf{a}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{a}_k) \\ & - \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{12} + \boldsymbol{\Upsilon}_{21} + \lambda \boldsymbol{\Upsilon}_{22})\}. \end{aligned} \quad (\text{E.6})$$

Minimization of (E.6) with respect to \mathbf{s} and $\tilde{\mathbf{W}}$ yields

$$\begin{aligned} & \min_{\mathbf{s}, \tilde{\mathbf{W}}} \mathcal{L}(\mathbf{s}, \tilde{\mathbf{W}}, \boldsymbol{\omega}, \boldsymbol{\Gamma}, \boldsymbol{\Upsilon}) \\ & = \begin{cases} -\operatorname{Re} \operatorname{Tr}(\boldsymbol{\Upsilon}_{12} + \boldsymbol{\Upsilon}_{21} + \lambda \boldsymbol{\Upsilon}_{22}) & \text{if } \begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} - \boldsymbol{\Gamma} = \mathbf{0}, \\ & (1 - \omega_k) - \mathbf{a}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{a}_k = 0, \text{ for } k = 1, \dots, K \\ -\infty & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{E.7})$$

Since $\boldsymbol{\Gamma} \succeq \mathbf{0}$, the first constraint in (E.7) can be relaxed to

$$\begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} \succeq \mathbf{0}. \quad (\text{E.8})$$

Moreover, with $\omega_k \geq 0$, the second constraint can be relaxed to

$$1 - \mathbf{a}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{a}_k \geq 0 \text{ for } k = 1, \dots, K, \quad (\text{E.9})$$

such that the dual problem of (E.3) can be formulated as

$$\max_{\boldsymbol{\Upsilon}} -\operatorname{Re}\{\operatorname{Tr}(\boldsymbol{\Upsilon}_{12} + \boldsymbol{\Upsilon}_{21} + \lambda \boldsymbol{\Upsilon}_{22})\} \quad (\text{E.10a})$$

$$\text{s.t. } \begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} \succeq \mathbf{0} \quad (\text{E.10b})$$

$$1 - \mathbf{a}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{a}_k \geq 0, \text{ for } k = 1, \dots, K, \quad (\text{E.10c})$$

as given in (4.65).

F Regularization Parameter for $\ell_{*,1}$ Minimization

Similar to the considerations in [OPT00a] and Appendix A of this thesis, consider the subdifferential of the $\ell_{*,1}$ minimization problem in (5.32), given as

$$\partial_{\mathbf{Z}} \left(\frac{1}{2} \|\mathbf{Y} - \mathbf{B}\mathbf{Z}\|_{\mathbb{F}}^2 + \lambda \sqrt{N} \|\mathbf{Z}\|_{*,1}^{(P \times N)} \right) = -\mathbf{B}^H (\mathbf{Y} - \mathbf{B}\mathbf{Z}) + \lambda \sqrt{N} \mathbf{H}. \quad (\text{F.1})$$

In (F.1), $\mathbf{H} = [\mathbf{H}_1^T, \dots, \mathbf{H}_K^T]^T$ is a subgradient of the penalty term $\|\mathbf{Z}\|_{*,1}^{(P \times N)} = \sum_{k=1}^K \|\mathbf{Z}_k\|_*$, where the blocks $\mathbf{H}_k \in \partial \|\mathbf{Z}_k\|_*$ are elements of the subdifferential of the nuclear norm $\|\mathbf{Z}_k\|_*$ of the submatrices of $\mathbf{Z} = [\mathbf{Z}_1^T, \dots, \mathbf{Z}_K^T]^T$. Taking the definition from [Wat92], the subdifferential of the nuclear norm is defined as

$$\partial \|\mathbf{Z}_k\|_* = \left\{ \mathbf{H}_k \mid \|\tilde{\mathbf{Z}}_k\|_* \geq \|\mathbf{Z}_k\|_* + \operatorname{Re}\{\operatorname{Tr}((\tilde{\mathbf{Z}}_k - \mathbf{Z}_k)^H \mathbf{H}_k)\} \right\}, \quad (\text{F.2})$$

for $\mathbf{H}_k, \tilde{\mathbf{Z}}_k \in \mathbb{C}^{P \times N}$. Assuming r non-zero singular values in the singular value decomposition of $\mathbf{Z}_k = \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^H$, and partitioning

$$\mathbf{U}_k = [\mathbf{U}_{k,1}, \mathbf{U}_{k,0}] \quad \text{and} \quad \mathbf{V}_k = [\mathbf{V}_{k,1}, \mathbf{V}_{k,0}], \quad (\text{F.3})$$

where $\mathbf{U}_{k,1} \in \mathbb{C}^{P \times r}$ and $\mathbf{V}_{k,1} \in \mathbb{C}^{N \times r}$ contain the left and right singular vectors corresponding to the non-zero singular values, the subdifferential is given by

$$\partial \|\mathbf{Z}_k\|_* = \left\{ \mathbf{U}_{k,1} \mathbf{V}_{k,1}^H + \tilde{\mathbf{Z}}_k \mid \tilde{\mathbf{Z}}_k \in \mathbb{C}^{P \times N}, \mathbf{U}_{k,1}^H \tilde{\mathbf{Z}}_k = \mathbf{0}, \tilde{\mathbf{Z}}_k \mathbf{V}_{k,1} = \mathbf{0}, \|\tilde{\mathbf{Z}}_k\|_2 \leq 1 \right\}, \quad (\text{F.4})$$

where $\|\tilde{\mathbf{Z}}_k\|_2$ denotes the spectral norm, i.e., the maximum singular value of $\tilde{\mathbf{Z}}_k$. From (F.4) it can be concluded that

$$\|\hat{\mathbf{H}}_k\|_2 \begin{cases} = 1 & \text{for } \mathbf{Z}_k \neq \mathbf{0} \\ \leq 1 & \text{for } \mathbf{Z}_k = \mathbf{0}. \end{cases} \quad (\text{F.5})$$

Hence, the block support of \mathbf{Z} can similarly be identified from a subgradient \mathbf{H} by identifying the submatrices \mathbf{H}_k with spectral norm 1.

Under optimality, the subdifferential in (F.1) must contain the matrix of zeros $\mathbf{0}$, such that there must exist matrices $\hat{\mathbf{Z}}$ and $\hat{\mathbf{H}}$ fulfilling

$$-\mathbf{B}^H (\mathbf{Y} - \mathbf{B}\hat{\mathbf{Z}}) + \lambda \sqrt{N} \hat{\mathbf{H}} = \mathbf{0}. \quad (\text{F.6})$$

From (F.6) it can be concluded that the submatrices $\hat{\mathbf{H}}_k$ of a subgradient $\hat{\mathbf{H}} = [\hat{\mathbf{H}}_1^T, \dots, \hat{\mathbf{H}}_K^T]^T$ fulfill

$$\lambda \hat{\mathbf{H}}_k = \mathbf{B}_k^H (\mathbf{Y} - \mathbf{B}\hat{\mathbf{Z}}) / \sqrt{N}, \quad (\text{F.7})$$

and based on (F.5) it can be observed that for a regularization parameter selection according to

$$\lambda \geq \|\mathbf{B}_k^H \mathbf{Y}\|_2 / \sqrt{N} \quad (\text{F.8})$$

the $\ell_{*,1}$ minimization problem in (5.32) provides a zero solution $\hat{\mathbf{Z}} = \mathbf{0}$. Assume the sensor measurements only contain noise according to $\mathbf{Y} = \mathbf{N}$, i.e., $\tilde{\mathbf{Z}} = \mathbf{0}$ in the sparse representation (5.19), such that condition (F.8) reduces to

$$\lambda \geq \|\mathbf{B}_k^H \mathbf{N}\|_2 / \sqrt{N}. \quad (\text{F.9})$$

Ideally, in this case the $\ell_{*,1}$ minimization problem (5.32) should provide a zero signal matrix $\hat{\mathbf{Z}} = \mathbf{0}$. Thus, by finding an upper bound on the spectral norm of the matrix products $\mathbf{B}_k^H \mathbf{N} / \sqrt{N}$, for $k = 1, \dots, K$, a regularization parameter λ can be found such

that $\hat{\mathbf{Z}} = \mathbf{0}$. In terms of statistical expectation, an upper bound on the right hand side in (F.9) is given as

$$\begin{aligned} \mathbb{E}\left\{\max_k \|\mathbf{B}_k^H \mathbf{N}\|_2 / \sqrt{N}\right\} &\leq \mathbb{E}\left\{\max_k \|\mathbf{B}_k\|_2 \|\mathbf{N}\|_2 / \sqrt{N}\right\} \\ &= \max_k \|\mathbf{B}_k\|_2 \mathbb{E}\{\|\mathbf{N}\|_2\} / \sqrt{N} \\ &\leq \max_k \sigma_N \|\mathbf{B}_k\|_2 (\sqrt{M} + \sqrt{N}) / \sqrt{N} \\ &= \max_k \sigma_N \|\mathbf{B}_k\|_2 (\sqrt{M/N} + 1), \end{aligned} \quad (\text{F.10})$$

where the expectation of the spectral norm of an $M \times N$ complex Gaussian matrix, with entries according to $[\mathbf{N}]_{m,n} \sim \mathcal{N}_C(0, \sigma_N^2)$, is bounded as $\mathbb{E}\{\|\mathbf{N}\|_2\} \leq \sigma(\sqrt{M} + \sqrt{N})$ [RV10, T⁺15, PE10], as discussed for (A.8).

In the case of partly calibrated arrays, the sensing submatrices are defined as

$$\mathbf{B}_k = \text{blkdiag}(\mathbf{b}^{(1)}(\mu_k), \dots, \mathbf{b}^{(P)}(\mu_k)) \quad (\text{F.11})$$

and the spectral norm can be computed from the square-root of the eigenvalues of the matrix products

$$\begin{aligned} \mathbf{B}_k^H \mathbf{B}_k &= \text{diag}(\|\mathbf{b}^{(1)}(\mu_k)\|_2^2, \dots, \|\mathbf{b}^{(P)}(\mu_k)\|_2^2) \\ &= \text{diag}(M_1, \dots, M_P), \end{aligned} \quad (\text{F.12})$$

which is determined by the largest element of the resulting diagonal matrix such that

$$\|\mathbf{B}_k\|_2 = \max_p \sqrt{M_p}. \quad (\text{F.13})$$

Using (F.13), the bound in (F.10) can further be rewritten as

$$\begin{aligned} \mathbb{E}\left\{\max_k \|\mathbf{B}_k^H \mathbf{N}\|_2 / \sqrt{N}\right\} &\leq \max_k \sigma_N \|\mathbf{B}_k\|_2 (\sqrt{M/N} + 1) \\ &= \max_p \sigma_N \sqrt{M_p} (\sqrt{M/N} + 1). \end{aligned} \quad (\text{F.14})$$

From (F.9) and (F.14), it can be concluded that a regularization parameter selection according to

$$\lambda = \max_p \sigma_N \sqrt{M_p} (\sqrt{M/N} + 1) \quad (\text{F.15})$$

fulfills condition (F.9) and will suppress the noise in the signal estimate, retaining only signal components in the sparse estimate $\hat{\mathbf{Z}}$. Similar to the discussion in Appendix A, the upper bound on the spectral norm used in (F.10) is tight for large values of sensors M and snapshots N . In the non-asymptotic case of small values for M or N , the bound used in (F.10) might be loose and the regularization parameter in (F.15) might be too large, leading to suppression not only of noise, but also of signal components in the estimate $\hat{\mathbf{Z}}$.

G Equivalence of COBRAS and $\ell_{*,1}$ Minimization

The proof of Theorem 5.1 relies on the following lemma, see [SS05, RFP10]:

Lemma G.1. *The nuclear norm of the $P \times N$ matrix \mathbf{Z}_k is equivalently computed by the minimization problem*

$$\|\mathbf{Z}_k\|_* = \min_{\mathbf{\Gamma}_k, \mathbf{H}_k} \left\{ \frac{1}{2} (\|\mathbf{\Gamma}_k\|_{\mathbb{F}}^2 + \|\mathbf{H}_k\|_{\mathbb{F}}^2) : \mathbf{\Gamma}_k \mathbf{H}_k = \mathbf{Z}_k \right\}, \quad (\text{G.1})$$

where $\mathbf{\Gamma}_k$ and \mathbf{H}_k are complex matrices of dimensions $P \times r$ and $r \times N$, respectively, with $r = \min(N, P)$.

Proof of Lemma G.1. Define the compact singular value decomposition of matrix \mathbf{Z}_k as

$$\mathbf{Z}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^{\text{H}}, \quad (\text{G.2})$$

such that the factorization terms of $\mathbf{Z}_k = \mathbf{\Gamma}_k \mathbf{H}_k$ can be expressed as

$$\mathbf{\Gamma}_k = \mathbf{U}_k \tilde{\mathbf{\Sigma}}_{k,1} \mathbf{W}_k^{\text{H}} \quad \text{and} \quad \mathbf{H}_k = \mathbf{W}_k \tilde{\mathbf{\Sigma}}_{k,2} \mathbf{V}_k^{\text{H}}, \quad (\text{G.3})$$

for $\mathbf{\Sigma}_k = \tilde{\mathbf{\Sigma}}_{k,1} \tilde{\mathbf{\Sigma}}_{k,2}$ of size $r \times r$, and some $r \times r$ arbitrary unitary matrix \mathbf{W}_k , i.e., $\mathbf{W}_k^{\text{H}} \mathbf{W}_k = \mathbf{I}_r$. Based on (G.3) it holds that

$$\begin{aligned} \|\mathbf{Z}_k\|_* &= \|\mathbf{\Sigma}_k\|_* = \|\tilde{\mathbf{\Sigma}}_{k,1} \tilde{\mathbf{\Sigma}}_{k,2}\|_* \\ &\leq \|\tilde{\mathbf{\Sigma}}_{k,1}\|_{\mathbb{F}} \|\tilde{\mathbf{\Sigma}}_{k,2}\|_{\mathbb{F}} \\ &= \|\mathbf{\Gamma}_k\|_{\mathbb{F}} \|\mathbf{H}_k\|_{\mathbb{F}}, \end{aligned} \quad (\text{G.4})$$

where the inequality stems from the Cauchy-Schwartz inequality and is fulfilled with equality if and only if $\tilde{\mathbf{\Sigma}}_{k,1} = \tilde{\mathbf{\Sigma}}_{k,2} = \mathbf{\Sigma}_k^{1/2}$. In this case, the matrix factors in (G.3) are given as

$$\mathbf{\Gamma}_k = \mathbf{U}_k \mathbf{\Sigma}_k^{1/2} \mathbf{W}_k^{\text{H}} \quad \text{and} \quad \mathbf{H}_k = \mathbf{W}_k \mathbf{\Sigma}_k^{1/2} \mathbf{V}_k^{\text{H}}. \quad (\text{G.5})$$

Furthermore, by the arithmetic-geometric-mean inequality it follows that

$$\|\mathbf{\Gamma}_k\|_{\mathbb{F}} \|\mathbf{H}_k\|_{\mathbb{F}} \leq \frac{1}{2} (\|\mathbf{\Gamma}_k\|_{\mathbb{F}}^2 + \|\mathbf{H}_k\|_{\mathbb{F}}^2), \quad (\text{G.6})$$

where equality holds if $\|\mathbf{\Gamma}_k\|_{\mathbb{F}} = \|\mathbf{H}_k\|_{\mathbb{F}}$, such that the minimum of (G.1) is given by $\|\mathbf{Z}_k\|_* = \frac{1}{2} (\|\mathbf{\Gamma}_k\|_{\mathbb{F}}^2 + \|\mathbf{H}_k\|_{\mathbb{F}}^2)$, with $\mathbf{\Gamma}_k$ and \mathbf{H}_k given in (G.5). \square

Proof of Theorem 5.1. Based on Lemma G.1, the $\ell_{*,1}$ mixed-norm of the source signal matrix $\mathbf{Z} = [\mathbf{Z}_1^{\text{T}}, \dots, \mathbf{Z}_K^{\text{T}}]^{\text{T}}$, as defined in (5.24), is equivalently computed by

$$\begin{aligned}
\|\mathbf{Z}\|_{*,1}^{(P \times N)} &= \sum_{k=1}^K \|\mathbf{Z}_k\|_* \\
&= \min_{\{\boldsymbol{\Gamma}_k \in \mathbb{C}^{P \times r}, \mathbf{H}_k \in \mathbb{C}^{r \times N}\}} \left\{ \frac{1}{2} \sum_{k=1}^K (\|\boldsymbol{\Gamma}_k\|_F^2 + \|\mathbf{H}_k\|_F^2) : \boldsymbol{\Gamma}_k \mathbf{H}_k = \mathbf{Z}_k \right\} \\
&= \min_{\boldsymbol{\Gamma} \in \mathcal{B}_{P \times r}^K, \mathbf{H} \in \mathbb{C}^{Kr \times N}} \left\{ \frac{1}{2} (\|\boldsymbol{\Gamma}\|_F^2 + \|\mathbf{H}\|_F^2) : \mathbf{Z} = \boldsymbol{\Gamma} \mathbf{H} \right\}, \tag{G.7}
\end{aligned}$$

where $r = \min(N, P)$, $\boldsymbol{\Gamma} = \text{blkdiag}(\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_K)$ is taken from the set $\mathcal{B}_{P \times r}^K$ of block-diagonal matrices composed of K blocks of size $P \times r$ on the main diagonal, and $\mathbf{H} = [\mathbf{H}_1^\top, \dots, \mathbf{H}_K^\top]^\top$ is a $Kr \times N$ complex matrix composed of blocks \mathbf{H}_k , for $k = 1, \dots, K$. Inserting Equation (G.7) into the $\ell_{*,1}$ minimization problem in (5.32), the following minimization problem can be formulated

$$\min_{\boldsymbol{\Gamma} \in \mathcal{B}_{P \times r}^K, \mathbf{H} \in \mathbb{C}^{Kr \times N}} \frac{1}{2} \|\mathbf{B} \boldsymbol{\Gamma} \mathbf{H} - \mathbf{Y}\|_F^2 + \frac{\lambda \sqrt{N}}{2} (\|\boldsymbol{\Gamma}\|_F^2 + \|\mathbf{H}\|_F^2). \tag{G.8}$$

For a fixed matrix $\boldsymbol{\Gamma}$, the minimizer $\hat{\mathbf{H}}$ of problem (G.7) has the closed form expression

$$\begin{aligned}
\hat{\mathbf{H}} &= (\boldsymbol{\Gamma}^\mathbf{H} \mathbf{B}^\mathbf{H} \mathbf{B} \boldsymbol{\Gamma} + \lambda \sqrt{N} \mathbf{I})^{-1} \boldsymbol{\Gamma}^\mathbf{H} \mathbf{B}^\mathbf{H} \mathbf{Y} \\
&= \boldsymbol{\Gamma}^\mathbf{H} \mathbf{B}^\mathbf{H} (\mathbf{B} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\mathbf{H} \mathbf{B}^\mathbf{H} + \lambda \sqrt{N} \mathbf{I})^{-1} \mathbf{Y}, \tag{G.9}
\end{aligned}$$

where the last equation is derived from the Woodbury matrix identity [Sea82, p.151]. Reinserting the optimal matrix $\hat{\mathbf{H}}$ into Equation (G.8) and performing basic reformulations of the objective function results in the concentrated minimization problem

$$\min_{\boldsymbol{\Gamma} \in \mathcal{B}_{P \times r}^K} \frac{\lambda \sqrt{N}}{2} \left(\text{Tr}((\mathbf{B} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\mathbf{H} \mathbf{B}^\mathbf{H} + \lambda \sqrt{N} \mathbf{I})^{-1} \mathbf{Y} \mathbf{Y}^\mathbf{H}) + \text{Tr}(\boldsymbol{\Gamma} \boldsymbol{\Gamma}^\mathbf{H}) \right). \tag{G.10}$$

Upon summarizing $\mathbf{Y} \mathbf{Y}^\mathbf{H} / N = \hat{\mathbf{R}}$ and defining the positive semidefinite block-diagonal matrix

$$\mathbf{S} = \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\mathbf{H} / \sqrt{N} \in \mathcal{B}_{P+}^K, \tag{G.11}$$

problem (G.10) can be rewritten as

$$\min_{\mathbf{S} \in \mathcal{B}_{P+}^K} \frac{\lambda N}{2} \left(\text{Tr}((\mathbf{B} \mathbf{S} \mathbf{B}^\mathbf{H} + \lambda \mathbf{I})^{-1} \hat{\mathbf{R}}) + \text{Tr}(\mathbf{S}) \right). \tag{G.12}$$

Neglecting the factor $\lambda N / 2$ in (G.12), results in formulation (5.33). Using Equations (G.2), (G.5) and the definition of $\hat{\mathbf{S}} = \text{blkdiag}(\hat{\mathbf{S}}_1, \dots, \hat{\mathbf{S}}_K)$ in (G.11) it can be concluded that

$$\hat{\mathbf{S}}_k = \frac{1}{\sqrt{N}} \hat{\boldsymbol{\Gamma}}_k \hat{\boldsymbol{\Gamma}}_k^\mathbf{H} = \frac{1}{\sqrt{N}} (\hat{\mathbf{Z}}_k \hat{\mathbf{Z}}_k^\mathbf{H})^{1/2}, \tag{G.13}$$

as given in (5.35). Making further use of (G.9) and the factorization in (G.7) yields

$$\begin{aligned}
\hat{\mathbf{Z}} &= \hat{\boldsymbol{\Gamma}} \hat{\mathbf{H}} \\
&= \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Gamma}}^\mathbf{H} \mathbf{B}^\mathbf{H} (\mathbf{B} \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Gamma}}^\mathbf{H} \mathbf{B}^\mathbf{H} + \lambda \sqrt{N} \mathbf{I})^{-1} \mathbf{Y} \\
&= \hat{\mathbf{S}} \mathbf{B}^\mathbf{H} (\mathbf{B} \hat{\mathbf{S}} \mathbf{B}^\mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{Y}, \tag{G.14}
\end{aligned}$$

which corresponds to relation (5.34). \square

H SDP Form of the Matrix Polynomial Constraint

As discussed in Section 5.3.1 for Equation (5.55), the matrix $\mathbf{F}_{\text{COBRAS}}(z)$ represents a matrix polynomial of degree $2(M_0 - 1)$ in the variable z [Dum07]. To see the relation between matrix \mathbf{Y}_0 and the matrix coefficients \mathbf{C}_m , consider the $M \times P$ matrix

$$\mathbf{\Omega}(z) = [\mathbf{I}_P \quad z\mathbf{I}_P \quad z^2\mathbf{I}_P \quad \dots \quad z^D\mathbf{I}_P]^\top, \quad (\text{H.1})$$

as defined in (3.23), and introduce the $M \times M$ permutation matrix \mathbf{J} such that the subarray steering block matrix can be expressed as

$$\mathbf{B}(z) = \mathbf{J}\mathbf{\Omega}(z), \quad (\text{H.2})$$

as discussed for (3.24). Similar to (3.25), inserting (H.2) in (5.55) yields

$$\begin{aligned} \mathbf{F}_{\text{COBRAS}}(z) &= \mathbf{B}^\text{H}(z)\mathbf{Y}_0\mathbf{B}(z) \\ &= \mathbf{\Omega}^\text{H}(z)\mathbf{J}^\text{H}\mathbf{Y}_0\mathbf{J}\mathbf{\Omega}(z) \\ &= \mathbf{\Omega}^\text{H}(z)\mathbf{G}\mathbf{\Omega}(z), \end{aligned} \quad (\text{H.3})$$

where $\mathbf{G} = \mathbf{J}^\text{H}\mathbf{Y}_0\mathbf{J}$ is of size $M \times M$ and is composed of the $P \times P$ blocks $\mathbf{G}_{i,j}$, for $i, j = 1, \dots, M_0$, as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \cdots & \mathbf{G}_{1,M_0} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{M_0,1} & \cdots & \mathbf{G}_{M_0,M_0} \end{bmatrix}. \quad (\text{H.4})$$

Equation (H.3) is also referred to as the Gram matrix representation of the polynomial $\mathbf{F}_{\text{COBRAS}}(z)$, and \mathbf{G} is referred to as the corresponding Gram matrix [Dum07].

Using the block trace operator defined in (3.27) and the elementary Toeplitz matrices $\mathbf{\Theta}_m$ introduced in (3.28), the matrix coefficients \mathbf{C}_m in (5.55) can be computed from the Gram matrix \mathbf{G} in (H.3) as

$$\mathbf{C}_m = \text{blkTr}^{(P)}((\mathbf{\Theta}_m \otimes \mathbf{I}_P)\mathbf{G}), \quad (\text{H.5})$$

i.e., the summation of the $P \times P$ submatrices on the m th block-diagonal of the Gram matrix \mathbf{G} .

Define a second matrix polynomial with constant value \mathbf{I}_P as

$$\mathbf{\Omega}^\text{H}(z)\mathbf{H}\mathbf{\Omega}(z) = \mathbf{I}_P \quad (\text{H.6})$$

such that the corresponding Gram matrix \mathbf{H} of size $M \times M$ fulfills

$$\text{blkTr}^{(P)}(\mathbf{H}) = \mathbf{I}_P, \quad (\text{H.7a})$$

$$\text{blkTr}^{(P)}((\mathbf{\Theta}_m \otimes \mathbf{I}_P)\mathbf{H}) = \mathbf{0} \text{ for } m \neq 0. \quad (\text{H.7b})$$

By using (H.3), (H.6) and (H.7), the matrix polynomial constraint (5.56) can be expressed as

$$\mathbf{I}_P - \mathbf{B}(z)^H \boldsymbol{\Upsilon}_0 \mathbf{B}(z) = \boldsymbol{\Omega}^H (\mathbf{H} - \mathbf{J}^H \boldsymbol{\Upsilon}_0 \mathbf{J}) \boldsymbol{\Omega} \succeq \mathbf{0}, \quad (\text{H.8})$$

which is fulfilled for

$$\mathbf{H} - \mathbf{J}^H \boldsymbol{\Upsilon}_0 \mathbf{J} \succeq \mathbf{0}. \quad (\text{H.9})$$

Applying (H.7) and (H.9) in problem (5.51), the gridless frequency estimation problem

$$\max_{\mathbf{r}_1, \boldsymbol{\Upsilon}_0, \mathbf{H}} -2 \operatorname{Re} \operatorname{Tr}(\boldsymbol{\Upsilon}_1) - \lambda \operatorname{Tr}(\boldsymbol{\Upsilon}_0) \quad (\text{H.10a})$$

$$\text{s.t.} \quad \begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_1 \\ \boldsymbol{\Upsilon}_1^H & \boldsymbol{\Upsilon}_0 \end{bmatrix} \succeq \mathbf{0} \quad (\text{H.10b})$$

$$\mathbf{H} - \mathbf{J}^H \boldsymbol{\Upsilon}_0 \mathbf{J} \succeq \mathbf{0} \quad (\text{H.10c})$$

$$\operatorname{blkTr}^{(P)}(\mathbf{H}) = \mathbf{I}_P \quad (\text{H.10d})$$

$$\operatorname{blkTr}^{(P)}((\boldsymbol{\Theta}_m \otimes \mathbf{I}_P) \mathbf{H}) = \mathbf{0} \text{ for } m \neq 0 \quad (\text{H.10e})$$

can be defined. Given a minimizer $\hat{\boldsymbol{\Upsilon}}_0$ to problem (H.10), the frequency estimation problem reduces to finding roots for which the constraint (5.56) becomes singular, as discussed in Section 5.3.1.

I Dual Problem of the COBRAS Formulation

Consider the SDP formulation (5.37) of the COBRAS problem, given as

$$\min_{\mathbf{S}, \mathbf{W}_M} \operatorname{Tr}(\mathbf{W}_M \hat{\mathbf{R}}) + \operatorname{Tr}(\mathbf{S}) \quad (\text{I.1a})$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{W}_M & \mathbf{I}_M \\ \mathbf{I}_M & \mathbf{B} \mathbf{S} \mathbf{B}^H + \lambda \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \quad (\text{I.1b})$$

$$\mathbf{S} \in \mathcal{B}_{P+}^K, \mathbf{W}_M \succeq \mathbf{0}, \quad (\text{I.1c})$$

where $\mathbf{S} = \operatorname{blkdiag}(\mathbf{S}_1, \dots, \mathbf{S}_K)$. Upon introducing the matrix

$$\tilde{\mathbf{W}} = \begin{bmatrix} \tilde{\mathbf{W}}_{11} & \tilde{\mathbf{W}}_{12} \\ \tilde{\mathbf{W}}_{21} & \tilde{\mathbf{W}}_{22} \end{bmatrix} \quad (\text{I.2})$$

the SDP in (I.1) can be reformulated as

$$\min_{\mathbf{S}, \tilde{\mathbf{W}}} \operatorname{Tr}(\tilde{\mathbf{W}}_{11} \hat{\mathbf{R}}) + \sum_{k=1}^K \operatorname{Tr}(\mathbf{S}_k) \quad (\text{I.3a})$$

$$\text{s.t.} \quad \mathbf{S}_k \succeq \mathbf{0}, \text{ for } k = 1, \dots, K \quad (\text{I.3b})$$

$$\tilde{\mathbf{W}} \succeq \mathbf{0} \quad (\text{I.3c})$$

$$\tilde{\mathbf{W}}_{12} = \mathbf{I}, \tilde{\mathbf{W}}_{21} = \mathbf{I} \quad (\text{I.3d})$$

$$\tilde{\mathbf{W}}_{22} = \sum_{k=1}^K \mathbf{B}_k \mathbf{S}_k \mathbf{B}_k^H + \lambda \mathbf{I}. \quad (\text{I.3e})$$

The Lagrangian function of (I.3) is given by

$$\begin{aligned}
\mathcal{L}(\mathbf{S}, \tilde{\mathbf{W}}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Upsilon}) = & \text{Tr}(\tilde{\mathbf{W}}_{11} \hat{\mathbf{R}}) + \sum_{k=1}^K \text{Tr}(\mathbf{S}_k) \\
& - \sum_{k=1}^K \text{Tr}(\boldsymbol{\Omega}_k \mathbf{S}_k) - \text{Tr}(\boldsymbol{\Gamma} \tilde{\mathbf{W}}) \\
& + \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{21}(\tilde{\mathbf{W}}_{12} - \mathbf{I}))\} + \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{12}(\tilde{\mathbf{W}}_{21} - \mathbf{I}))\} \\
& + \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{22}(\tilde{\mathbf{W}}_{22} - \sum_{k=1}^K \mathbf{B}_k \mathbf{S}_k \mathbf{B}_k^H - \lambda \mathbf{I}))\}, \tag{I.4}
\end{aligned}$$

with $\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_K \succeq \mathbf{0}$, summarized in $\boldsymbol{\Omega} = \text{blkdiag}(\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_K)$, being the Lagrangian multipliers accounting for the positive semidefinite constraints on $\mathbf{S}_1, \dots, \mathbf{S}_K$ in (I.3b), and $\boldsymbol{\Gamma} \succeq \mathbf{0}$ being the Lagrangian multipliers accounting for the positive semidefinite constraint with regard to $\tilde{\mathbf{W}}$ in (I.3c). The matrices $\boldsymbol{\Upsilon}_{12}, \boldsymbol{\Upsilon}_{21}, \boldsymbol{\Upsilon}_{22}$, summarized in

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{\Upsilon}_{11} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix}, \tag{I.5}$$

represent the Lagrangian multipliers accounting for the equality constraints on submatrices $\tilde{\mathbf{W}}_{12}, \tilde{\mathbf{W}}_{21}, \tilde{\mathbf{W}}_{22}$ in (I.3d) and (I.3e). In a more compact notation, the Lagrangian function (I.4) can be written as

$$\begin{aligned}
\mathcal{L}(\mathbf{S}, \tilde{\mathbf{W}}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Upsilon}) = & \text{Tr} \left(\left(\begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} - \boldsymbol{\Gamma} \right) \begin{bmatrix} \tilde{\mathbf{W}}_{11} & \tilde{\mathbf{W}}_{12} \\ \tilde{\mathbf{W}}_{21} & \tilde{\mathbf{W}}_{22} \end{bmatrix} \right) \\
& + \sum_{k=1}^K \text{Tr}(\mathbf{S}_k(\mathbf{I} - \boldsymbol{\Omega}_k) - \mathbf{S}_k \mathbf{B}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{B}_k) \\
& - \text{Re}\{\text{Tr}(\boldsymbol{\Upsilon}_{12} + \boldsymbol{\Upsilon}_{21} + \lambda \boldsymbol{\Upsilon}_{22})\}. \tag{I.6}
\end{aligned}$$

Minimization of (I.6) with respect to \mathbf{S} and $\tilde{\mathbf{W}}$ yields

$$\begin{aligned}
& \min_{\mathbf{S}, \tilde{\mathbf{W}}} \mathcal{L}(\mathbf{S}, \tilde{\mathbf{W}}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Upsilon}) \\
= & \begin{cases} -\text{Tr}(\boldsymbol{\Upsilon}_{12} + \boldsymbol{\Upsilon}_{21} + \lambda \boldsymbol{\Upsilon}_{22}) & \text{if } \begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} - \boldsymbol{\Gamma} = \mathbf{0}, \\ & \mathbf{I} - \boldsymbol{\Omega}_k - \mathbf{B}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{B}_k = \mathbf{0}, \text{ for } k = 1, \dots, K \\ -\infty & \text{otherwise.} \end{cases} \tag{I.7}
\end{aligned}$$

Since $\boldsymbol{\Gamma} \succeq \mathbf{0}$, the first constraint in (I.7) can be relaxed to

$$\begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} \succeq \mathbf{0}. \tag{I.8}$$

Moreover, with $\boldsymbol{\Omega}_k \succeq \mathbf{0}$, the second constraint can be relaxed to

$$\mathbf{I} - \mathbf{B}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{B}_k \succeq \mathbf{0} \text{ for } k = 1, \dots, K, \tag{I.9}$$

such that the dual problem of (I.3) can be formulated as

$$\max_{\boldsymbol{\Upsilon}} -\operatorname{Re}\{\operatorname{Tr}(\boldsymbol{\Upsilon}_{12} + \boldsymbol{\Upsilon}_{21} + \lambda\boldsymbol{\Upsilon}_{22})\} \quad (\text{I.10a})$$

$$\text{s.t. } \begin{bmatrix} \hat{\mathbf{R}} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{bmatrix} \succeq \mathbf{0} \quad (\text{I.10b})$$

$$\mathbf{I} - \mathbf{B}_k^H \boldsymbol{\Upsilon}_{22} \mathbf{B}_k \succeq \mathbf{0}, \quad \text{for } k = 1, \dots, K, \quad (\text{I.10c})$$

as given in (5.51).

List of Acronyms

ADMM	Alternating direction method of multipliers
AMP	Approximate message passing
ANM	Atomic norm minimization
BCD	Block coordinate descent
BP	Basis pursuit
BPDN	Basis pursuit denoising
CD	Coordinate descent
COBRAS	Compact block- and rank-sparse reconstruction
DOA	Direction of arrival
ESPRIT	Estimation of signal parameters via rotational invariance techniques
FCA	Fully calibrated array
GLS	Gridless SPICE
GL-COBRAS	Gridless COBRAS
GL-SPARROW	Gridless SPARROW
IHT	Iterative hard thresholding
LAD-LASSO	Least absolute deviation LASSO
LASSO	Least absolute shrinkage and selection operator
LARS	Least angle regression
MI-ESPRIT	Multiple-invariance ESPRIT
ML	Maximum likelihood
MUSIC	Multiple signal classification
OMP	Orthogonal matching pursuit
PCA	Partly calibrated array
RDS	Rekonstruktion dünnbesetzter Signale
RMSE	Root-mean-square error

SDP	Semidefinite program
SNR	Signal-to-noise ratio
SPARROW	Sparse row-norm reconstruction
SPICE	Sparse iterative covariance-based estimation
SR-LASSO	Square-root LASSO
SSR	Sparse signal reconstruction
STELA	Soft-thresholding and exact line search algorithm
SVD	Singular value decomposition
SVT	Singular value thresholding
TLA	Thinned linear array
ULA	Uniform linear array

List of Symbols

Constants and Sets

j	Imaginary unit $\sqrt{-1}$
$\mathbf{0}$	Matrix of zeros of conformable dimensions
$\mathbf{1}$	Matrix of ones of conformable dimensions
\mathbf{I}_M	Identity matrix of dimensions $M \times M$
\mathbf{J}	Selection and permutation matrix
Θ_m	Elementary Toeplitz matrix with ones on m th diagonal and zeros elsewhere
\mathbb{R}	Real numbers
\mathbb{R}^M	Real $M \times 1$ vectors
$\mathbb{R}^{M \times N}$	Real $M \times N$ matrices
\mathbb{C}	Complex numbers
\mathbb{C}^M	Complex $M \times 1$ vectors
$\mathbb{C}^{M \times N}$	Complex $M \times N$ matrices
\mathcal{D}^K	$K \times K$ diagonal matrices
\mathcal{D}_+^K	Nonnegative $K \times K$ diagonal matrices
$\mathcal{B}_{P \times r}^K$	Block-diagonal matrices composed of K blocks of size $P \times r$ on the main diagonal
\mathcal{B}_{P+}^K	Positive semidefinite block-diagonal matrices composed of K blocks of size $P \times P$ on the main diagonal

Operations and Characterizations

x^*	Complex conjugation of x
$\text{Re}\{x\}$	Real part of x
$\text{E}\{x\}$	Statistical expectation of a random variable x
\mathbf{X}^\top	Transpose of matrix \mathbf{X}
\mathbf{X}^H	Hermitian, i.e., conjugate transpose, of matrix \mathbf{X}
$\mathbf{X} \succeq \mathbf{0}$	Hermitian, positive semidefinite matrix
$\mathbf{X} \succ \mathbf{0}$	Hermitian, positive definite matrix
\otimes	Kronecker product
\odot	Hadamard product, i.e., element-wise matrix multiplication

$\text{diag}(x_1, \dots, x_K)$	Diagonal matrix containing the elements x_1, \dots, x_K on its main diagonal
$\text{blkdiag}(\mathbf{X}_1, \dots, \mathbf{X}_K)$	Block-diagonal matrix containing the submatrices $\mathbf{X}_1, \dots, \mathbf{X}_K$ on its main diagonal
$\text{vec}(\mathbf{X})$	Vectorization of the matrix \mathbf{X}
$\text{vecd}(\mathbf{X})$	Vectorization of the diagonal elements of square matrix \mathbf{X}
$\text{Tr}(\mathbf{X})$	Trace of square matrix \mathbf{X}
$\text{blkTr}^{(P)}(\mathbf{X})$	Block-trace of \mathbf{X} , i.e., sum of $P \times P$ submatrices on main diagonal of matrix \mathbf{X}
$\text{rank}(\mathbf{X})$	Rank of \mathbf{X}
$\text{Toep}(\mathbf{x})$	Toeplitz matrix with elements of \mathbf{x} in its first row and column
$\text{HToep}(\mathbf{x})$	Hermitian Toeplitz matrix with elements of \mathbf{x} in its first column
$ x $	absolute value of x
$\ \mathbf{x}\ _0$	ℓ_0 quasi-norm of vector \mathbf{x}
$\ \mathbf{x}\ _1$	ℓ_1 norm of vector \mathbf{x}
$\ \mathbf{x}\ _2$	ℓ_2 norm of vector \mathbf{x}
$\ \mathbf{x}\ _\infty$	ℓ_∞ norm of vector \mathbf{x} , i.e., largest absolute element of vector \mathbf{x}
$\ \mathbf{X}\ _{2,1}$	$\ell_{2,1}$ mixed-norm of matrix \mathbf{X}
$\ \mathbf{X}\ _2$	Spectral norm, i.e., largest singular value, of matrix \mathbf{X}
$\ \mathbf{X}\ _*$	Nuclear norm, i.e., sum of singular values, of matrix \mathbf{X}
$\ \mathbf{X}\ _F$	Frobenius norm of matrix \mathbf{X}
$\ \mathbf{X}\ _{*,1}^{(P \times N)}$	Mixed nuclear and ℓ_1 norm on $P \times N$ submatrices of \mathbf{X}
$\ \mathbf{X}\ _{F,1}^{(P \times N)}$	Mixed Frobenius and ℓ_1 norm on $P \times N$ submatrices of \mathbf{X}

Variables

$\alpha^{(p)}$	Perturbation coefficient for subarray p
β	Threshold parameter for sparse signal reconstruction
δ	Uniform frequency grid spacing
Δ	Common sensor array baseline
$\eta^{(p)}$	Position of subarray p relative to subarray 1
θ_l	Angular direction of source signal l
λ	Regularization parameter for sparse signal reconstruction
μ_l	Spatial frequency corresponding to source direction θ_l
ν_k	Sampled spatial frequency with index k

ξ_m	m th eigenvalue of the signal covariance matrix \mathbf{R}
Ξ	Diagonal matrix containing the eigenvalues of the signal covariance matrix \mathbf{R}
Π	Source covariance matrix
$\rho_m^{(p)}$	Position of sensor m in subarray p relative to sensor 1 in subarray p
σ_N^2	Noise variance
τ	Iteration index for iterative optimization methods
$\phi(\mu, \alpha, \eta)$	Subarray displacement shift vector
$\Phi(\mu, \alpha, \eta)$	Block-diagonal matrix containing the subarray displacement shifts
$\bar{\Phi}(\mu)$	Diagonal matrix containing direction dependent complex phase shifts
$\psi(t)$	Source signal vector in time instant t
Ψ	Source signal matrix under coherent processing
$\tilde{\Psi}^{(p)}$	Source signal matrix observed by subarray p under incoherent processing
$\hat{\Psi}$	Source signal matrix observed by all subarrays under incoherent processing
$\mathbf{a}(\mu)$	Array response vector of entire array
$\mathbf{a}^{(p)}(\mu)$	Array response vector of subarray p
$\mathbf{B}(\mu)$	Subarray response block matrix p
L	Number of source signals
\hat{L}	Estimated number of source signal
K	Number of sampled frequencies
M	Number of sensors in entire array
M_p	Number of sensors in subarray p
N	Number of signal snapshots
$\mathbf{n}(t)$	Sensor noise vector in entire array in time instant t
$\mathbf{n}^{(p)}(t)$	Sensor noise vector in subarray p and time instant t
\mathbf{N}	Sensor noise matrix in entire array
$\mathbf{N}^{(p)}$	Sensor noise matrix in subarray p
$\tilde{\mathbf{N}}^{(p)}$	Sensor noise matrix in subarray p under incoherent processing
P	Number of subarrays
r_m	Position of sensor m in FCA
$r_m^{(p)}$	Position of sensor m in subarray p in PCA

\mathbf{R}	Signal covariance matrix
\mathbf{R}_0	Sparse representation of signal covariance matrix
$\hat{\mathbf{R}}$	Sample covariance matrix
$\hat{\mathbf{S}}$	Estimated (block-) diagonal signal magnitude matrix
t_n	Time at sample index n
$t_n^{(p)}$	Time at sample index n in subarray p
\mathbf{U}_S	Signal subspace
\mathbf{U}_N	Noise subspace
\mathbf{T}	Subspace / steering matrix rotation operator
$\tilde{\mathbf{x}}(t)$	True sparse representation of signal vector $\boldsymbol{\psi}(t)$
$\hat{\mathbf{x}}(t)$	Estimated sparse representation of signal vector $\boldsymbol{\psi}(t)$
$\tilde{\mathbf{X}}$	True row-sparse representation of signal matrix $\boldsymbol{\Psi}$
$\hat{\mathbf{X}}$	Estimated row-sparse representation of signal matrix $\boldsymbol{\Psi}$
$\mathbf{y}(t)$	Sensor measurement vector in time instant t
$\mathbf{y}^{(p)}(t)$	Sensor measurement vector in subarray p and time instant t
\mathbf{Y}	Sensor measurement matrix
$\mathbf{Y}^{(p)}$	Sensor measurement matrix in subarray p
$\tilde{\mathbf{Y}}^{(p)}$	Sensor measurement matrix in subarray p under incoherent processing
$\tilde{\mathbf{Y}}$	Sensor measurement matrix under incoherent processing
$\tilde{\mathbf{Z}}$	True block- and rank-sparse extended signal matrix for coherent processing in PCA
$\hat{\mathbf{Z}}$	Estimated block- and rank-sparse extended signal matrix for coherent processing in PCA
$\tilde{\tilde{\mathbf{Z}}}$	True block-sparse extended signal matrix for incoherent processing in PCA
$\hat{\tilde{\mathbf{Z}}}$	Estimated block-sparse extended signal matrix for incoherent processing in PCA

Bibliography

- [APJ16] F. Afdideh, R. Phlypo, and C. Jutten, “Recovery guarantees for mixed norm ℓ_{p_1, p_2} block sparse representations,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Aug. 2016, pp. 378–382.
- [Bar48] M. S. Bartlett, “Smoothing periodograms from time-series with continuous spectra,” *Nature*, vol. 161, no. 4096, pp. 686–687, May 1948.
- [Bar83] A. Barabell, “Improving the resolution performance of eigenstructure-based direction-finding algorithms,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 8, Apr. 1983, pp. 336–339.
- [BCW11] A. Belloni, V. Chernozhukov, and L. Wang, “Square-root LASSO: Pivotal recovery of sparse signals via conic programming,” *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.
- [BD08] T. Blumensath and M. E. Davies, “Iterative thresholding for sparse approximations,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, Dec. 2008.
- [Ber99] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont: Athena Scientific, 1999.
- [Böh84] J. Böhme, “Estimation of source parameters by maximum likelihood and nonlinear regression,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 9, Mar. 1984, pp. 271–274.
- [Böh86] J. F. Böhme, “Estimation of spectral parameters of correlated signals in wavefields,” *Signal Processing*, vol. 11, no. 4, pp. 329–337, 1986.
- [BPC⁺11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “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.
- [BS14] P. Babu and P. Stoica, “Connection between SPICE and square-root LASSO for sparse parameter estimation,” *Signal Processing*, vol. 95, pp. 10–14, 2014.
- [BTN01] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2001.
- [BTR13] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
- [BV04] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

- [Cap69] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [Car11] C. Carathéodory, “Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen,” *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 193—217, 1911.
- [CCS10] J. Cai, E. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, pp. 1956–1982, 2010.
- [CDS98] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal On Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [CF11] C. Carathéodory and L. Fejér, “Über den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landauschen Satz,” *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 218—239, 1911.
- [CFG13] E. J. Candès and C. Fernandez-Granda, “Super-resolution from noisy data,” *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.
- [CFG14] ———, “Towards a mathematical theory of super-resolution,” *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [CH06] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
- [CR06] E. J. Candès and J. Romberg, “Quantitative robust uncertainty principles and optimally sparse decompositions,” vol. 6, no. 2, pp. 227–254, 2006.
- [CR12] E. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Communications of the ACM*, vol. 55, no. 6, pp. 111–119, 2012.
- [CRPW12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [CRT06a] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [CRT06b] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

- [CSPC11] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.
- [CT05] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [CW05] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [CWB08] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier analysis and applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [DE03] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization,” in *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, 2003, pp. 2197–2202.
- [DF10] K. Dvijotham and M. Fazel, “A nullspace analysis of the nuclear norm heuristic for rank minimization,” in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 3586–3589.
- [DMM09] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [Don92] D. L. Donoho, “Superresolution via sparsity constraints,” *SIAM Journal on Mathematical Analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.
- [Don95] ———, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [Don06] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [dP95] G. de Prony, “Essai expérimental et analytique: sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l’eau et de la vapeur de l’alcool à différentes températures,” *Journal de l’École Polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.
- [DSB⁺05] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, “Distributed compressed sensing of jointly sparse signals,” in *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, Oct. 2005, pp. 1537–1541.
- [Dum07] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*. Berlin: Springer, 2007.
- [EHJT04] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, pp. 407–499, 2004.

- [EKB10] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [EM09] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [FB01] B. P. Flanagan and K. L. Bell, “Array self-calibration with large sensor position errors,” *Signal Processing*, vol. 81, no. 10, pp. 2201–2214, 2001.
- [FHB01] M. Fazel, H. Hindi, and S. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in *Proceedings of the American Control Conference (ACC)*, vol. 6, 2001, pp. 4734–4739.
- [FHB03] M. Fazel, H. Hindi, and S. P. Boyd, “Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices,” in *Proceedings of the American Control Conference (ACC)*, vol. 3, 2003, pp. 2156–2162.
- [FHHT07] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, “Pathwise coordinate optimization,” *The Annals of Applied Statistics*, vol. 1, pp. 302–332, 2007.
- [FR13] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Basel: Birkhäuser, 2013.
- [Fuc97] J.-J. Fuchs, “Extension of the Pisarenko method to sparse linear arrays,” *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2413–2421, 1997.
- [GB08] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.
- [GB14] ———, “CVX: MATLAB software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, Mar. 2014.
- [GE95] A. B. Gershman and V. T. Ermolaev, “Optimal subarray size for spatial smoothing,” *IEEE Signal Processing Letters*, vol. 2, no. 2, pp. 28–30, Feb. 1995.
- [GRP10] A. B. Gershman, M. Rübsamen, and M. Pesavento, “One-and two-dimensional direction-of-arrival estimation: An overview of search-free techniques,” *Signal Processing*, vol. 90, no. 5, pp. 1338–1349, 2010.
- [Hag89] W. W. Hager, “Updating the inverse of a matrix,” *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.
- [HJ90] J. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

- [HM10] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.
- [HPREK14] M. Haardt, M. Pesavento, F. Römer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Academic Press Library in Signal Processing*. Elsevier, 2014, vol. 3, pp. 651–717.
- [HS90] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.
- [HS10] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal of Selected topics in signal processing*, vol. 4, no. 2, pp. 342–349, 2010.
- [HW91] B. Himed and D. D. Weiner, "Application of the matrix pencil approach to direction finding," DTIC Document, Tech. Rep., 1991.
- [IRA⁺14] M. Ibrahim, F. Römer, R. Alieiev, G. Del Galdo, and R. S. Thoma, "On the estimation of grid offsets in CS-based direction-of-arrival estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6776–6780.
- [KKB07] K. Koh, S.-J. Kim, and S. Boyd, "An interior-point method for large-scale ℓ_1 -regularized logistic regression," *Journal of Machine learning research*, vol. 8, no. Jul., pp. 1519–1555, 2007.
- [Kow09] M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 2009.
- [KT82] R. Kumaresan and D. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 6, pp. 833–840, Dec. 1982.
- [KV96] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [LC16] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.
- [LYJ⁺15] Z. Lu, R. Ying, S. Jiang, P. Liu, and W. Yu, "Distributed compressed sensing off the grid," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 105–109, Jan. 2015.
- [MÇW05] D. Malioutov, M. Çetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

- [MF10] K. Mohan and M. Fazel, “Reweighted nuclear norm minimization with application to system identification,” in *Proceedings of the American Control Conference (ACC)*, 2010, pp. 2953–2959.
- [Mof68] A. Moffet, “Minimum-redundancy linear arrays,” *IEEE Transactions on antennas and propagation*, vol. 16, no. 2, pp. 172–175, 1968.
- [MOS15] *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28)*, MOSEK ApS, 2015. [Online]. Available: <http://docs.mosek.com/7.1/toolbox/index.html>
- [MZ93] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [NBA13] E. T. Northardt, I. Bilik, and Y. I. Abramovich, “Spatial compressive sensing for direction-of-arrival estimation with bias mitigation via expected likelihood,” *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1183–1195, Mar. 2013.
- [NN94] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1994.
- [NS96] B. C. Ng and C. M. S. See, “Sensor-array calibration using a maximum-likelihood approach,” *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 6, pp. 827–835, Jun. 1996.
- [OPT00a] M. R. Osborne, B. Presnell, and B. A. Turlach, “On the LASSO and its dual,” *Journal of Computational and Graphical statistics*, vol. 9, no. 2, pp. 319–337, 2000.
- [OPT00b] M. Osborne, B. Presnell, and B. Turlach, “A new approach to variable selection in least squares problems,” vol. 20, no. 3, pp. 389–403, 2000.
- [OVK91] B. Ottersten, M. Viberg, and T. Kailath, “Performance analysis of the total least squares ESPRIT algorithm,” *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.
- [OVSN93] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, “Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing,” in *Radar array processing*. Springer, 1993, pp. 99–151.
- [Par12] P. Parvazi, “Sensor array processing in difficult and non-idealistic conditions,” Ph.D. dissertation, Technische Universität Darmstadt, 2012.
- [PB⁺14] N. Parikh, S. Boyd *et al.*, “Proximal algorithms,” *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [PC08] T. Park and G. Casella, “The Bayesian LASSO,” *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

- [PE10] D. P. Palomar and Y. C. Eldar, Eds., *Convex optimization in signal processing and communications*. Cambridge University Press, 2010.
- [Pes05] M. Pesavento, “Fast algorithms for multidimensional harmonic retrieval,” Ph.D. dissertation, Ruhr-Universität Bochum, 2005.
- [PF97] B. Porat and B. Friedlander, “Accuracy requirements in off-line array calibration,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 2, pp. 545–556, Apr. 1997.
- [PGH00] M. Pesavento, A. B. Gershman, and M. Haardt, “Unitary root-MUSIC with a real-valued eigendecomposition: A theoretical and experimental performance study,” *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1306–1314, May 2000.
- [PGW02] M. Pesavento, A. Gershman, and K. M. Wong, “Direction finding in partly calibrated sensor arrays composed of multiple subarrays,” *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2103–2115, 2002.
- [PK89] S. U. Pillai and B. H. Kwon, “Forward/backward spatial smoothing techniques for coherent signal identification,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 8–15, Jan. 1989.
- [PP11] P. Parvazi and M. Pesavento, “A new direction-of-arrival estimation and calibration method for arrays composed of multiple identical subarrays,” in *Proceedings of the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2011, pp. 171–175.
- [PPG11] P. Parvazi, M. Pesavento, and A. Gershman, “Direction-of-arrival estimation and array calibration for partly-calibrated arrays,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 2552–2555.
- [PPG12] P. Parvazi, M. Pesavento, and A. B. Gershman, “Rooting-based harmonic retrieval using multiple shift-invariances: The complete and the incomplete sample cases,” *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1556–1570, 2012.
- [PV10] P. Pal and P. Vaidyanathan, “Nested arrays: A novel approach to array processing with enhanced degrees of freedom,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.
- [QSG13] Z. Qin, K. Scheinberg, and D. Goldfarb, “Efficient block-coordinate descent algorithms for the group LASSO,” *Mathematical Programming Computation*, vol. 5, no. 2, pp. 143–169, 2013.
- [RFP10] 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.

- [RH89] B. D. Rao and K. V. S. Hari, “Performance analysis of root-MUSIC,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.
- [RK89] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [RKH13] C. Rojas, D. Katselis, and H. Hjalmarsson, “A note on the SPICE method,” *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4545–4551, Sep. 2013.
- [Roc70] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
- [ROSK88] R. Roy, B. Ottersten, L. Swindlehurst, and T. Kailath, “Multiple invariance ESPRIT,” in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1988, pp. 583–587.
- [RS87] Y. Rockah and P. Schultheiss, “Array shape calibration using sources in unknown locations—part I: Far-field sources,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 3, pp. 286–299, Mar. 1987.
- [RV10] M. Rudelson and R. Vershynin, “Non-asymptotic theory of random matrices: Extreme singular values,” in *Proceedings of the International Congress of Mathematicians*, vol. 3, Mar. 2010, pp. 1576–1602.
- [SA89] P. Stoica and N. Arye, “MUSIC, maximum likelihood, and Cramér-Rao bound,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.
- [SB12] 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.
- [SBL11a] P. Stoica, P. Babu, and J. Li, “New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data,” *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35–47, Jan. 2011.
- [SBL11b] —, “SPICE: A sparse covariance-based estimation method for array processing,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb. 2011.
- [Sch86] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [Sea82] S. Searle, *Matrix algebra useful for statistics*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1982.

- [SG04] C. See and A. Gershman, "Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 329–338, 2004.
- [SLG01] P. Stoica, G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *Signal Processing Letters, IEEE*, vol. 8, no. 5, pp. 148–150, 2001.
- [SNS95] P. Stoica, A. Nehorai, and T. Söderström, "Decentralized array processing using the MODE algorithm," *Circuits, Systems, and Signal Processing*, vol. 14, no. 1, pp. 17–38, 1995.
- [SORK92] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 867–881, Apr. 1992.
- [SP14] W. Suleiman and P. Parvazi, "Search-free decentralized direction-of-arrival estimation using common roots for non-coherent partly calibrated arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2292–2296.
- [SP18] C. Steffens and M. Pesavento, "Block- and rank-sparse recovery for direction finding in partly calibrated arrays," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 384–399, Jan. 2018.
- [SPP14] C. Steffens, P. Parvazi, and M. Pesavento, "Direction finding and array calibration based on sparse reconstruction in partly calibrated arrays," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 21–24.
- [SPP18] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1483–1497, Mar. 2018.
- [SPPZ14] W. Suleiman, P. Parvazi, M. Pesavento, and A. Zoubir, "Decentralized direction finding using Lanczos method," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 9–12.
- [SPZ13] W. Suleiman, M. Pesavento, and A. Zoubir, "Decentralized direction finding using partly calibrated arrays," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Sep. 2013, pp. 1–5.
- [SPZ16] W. Suleiman, M. Pesavento, and A. M. Zoubir, "Performance analysis of the decentralized eigendecomposition and ESPRIT algorithm," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2375–2386, May 2016.
- [SS05] N. Srebro and A. Shraibman, "Rank, trace-norm and max-norm," in *Proceedings of the Annual Conference on Learning Theory*. Springer-Verlag, 2005, pp. 545–560.

- [SSJ01] A. Swindlehurst, P. Stoica, and M. Jansson, “Exploiting arrays with multiple invariances using MUSIC and MODE,” *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2511–2521, Nov. 2001.
- [SSPH16] J. Steinwandt, C. Steffens, M. Pesavento, and M. Haardt, “Sparsity-aware direction finding for strictly non-circular sources based on rank minimization,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.
- [SSSP17a] C. Steffens, W. Suleiman, A. Sorg, and M. Pesavento, “Gridless compressed sensing under shift-invariant sampling,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 4735–4739.
- [SSSP17b] W. Suleiman, C. Steffens, A. Sorg, and M. Pesavento, “Gridless compressed sensing for fully augmentable arrays,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Kos Island, Greece, Aug. 2017, pp. 1986–1990.
- [Stu99] J. Sturm, “Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones,” *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.
- [SWK85] T.-J. Shan, M. Wax, and T. Kailath, “On spatial smoothing for direction-of-arrival estimation of coherent signals,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 806–811, Aug. 1985.
- [SYP16] C. Steffens, Y. Yang, and M. Pesavento, “Multidimensional sparse recovery for MIMO channel parameter estimation,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, Sep. 2016, pp. 66–70.
- [SZL14] P. Stoica, D. Zachariah, and J. Li, “Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation,” *Digital Signal Processing*, vol. 33, pp. 1–12, 2014.
- [T⁺15] J. A. Tropp *et al.*, “An introduction to matrix concentration inequalities,” *Foundations and Trends® in Machine Learning*, vol. 8, no. 1-2, pp. 1–230, 2015.
- [TBI97] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 1997.
- [TBR13] G. Tang, B. Bhaskar, and B. Recht, “Sparse recovery over continuous dictionaries—just discretize,” in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, Nov. 2013, pp. 1043–1047.
- [TBR15] G. Tang, B. N. Bhaskar, and B. Recht, “Near minimax line spectral estimation,” *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

- [TBSR13] G. Tang, B. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
- [TF09] T. Tuncer and B. Friedlander, Eds., *Classical and Modern Direction-of-Arrival Estimation*, 1st ed. Elsevier Science, 2009.
- [TGS06] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit,” *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.
- [Tib96] R. Tibshirani, “Regression shrinkage and selection via the LASSO,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, pp. 267–288, 1996.
- [Tip01] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *Journal of machine learning research*, vol. 1, pp. 211–244, Jun. 2001.
- [TK82] D. W. Tufts and R. Kumaresan, “Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood,” *Proceedings of the IEEE*, vol. 70, no. 9, pp. 975–989, Sep. 1982.
- [TLD⁺10] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, “Beyond Nyquist: Efficient sampling of sparse bandlimited signals,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.
- [Töp11] O. Töplitz, “Zur Theorie der quadratischen und bilinearen Formen von unendlich vielen Veränderlichen,” *Mathematische Annalen*, vol. 70, no. 3, pp. 351–376, 1911.
- [Tro06] J. A. Tropp, “Algorithms for simultaneous sparse approximation. Part II: Convex relaxation,” *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.
- [Tse01] P. Tseng, “Convergence of a block coordinate descent method for non-differentiable minimization,” *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.
- [TVW05] B. A. Turlach, W. N. Venables, and S. J. Wright, “Simultaneous variable selection,” *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.
- [TW10] J. A. Tropp and S. J. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.
- [TYN14] Z. Tan, P. Yang, and A. Nehorai, “Joint sparse recovery method for compressed sensing with structured dictionary mismatches,” *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4997–5008, 2014.
- [TZW17] Z. Tian, Z. Zhang, and Y. Wang, “Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization,” in *Proceedings of the IEEE International Conference on*

- Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 3071–3075.
- [VB96] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.
- [VP11] P. P. Vaidyanathan and P. Pal, “Sparse sensing with co-prime samplers and arrays,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.
- [VS94] M. Viberg and A. Swindlehurst, “A Bayesian approach to auto-calibration for parametric array signal processing,” *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3495–3507, Dec. 1994.
- [vT02] H. L. van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York: John Wiley & Sons, Inc., 2002.
- [Wat92] G. A. Watson, “Characterization of the subdifferential of some matrix norms,” *Linear algebra and its applications*, vol. 170, pp. 33–45, 1992.
- [Wax85] M. Wax, “Detection and estimation of superimposed signals,” Ph.D. dissertation, Stanford University, 1985.
- [Wax91] ———, “Detection and localization of multiple sources via the stochastic signals model,” *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2450–2456, Nov. 1991.
- [WF91] A. J. Weiss and B. Friedlander, “Self-calibration in high-resolution array processing,” in *Advances in Spectrum Estimation and Array Processing*, S. Haykin, Ed. Prentice-Hall, 1991, vol. II.
- [WK85] M. Wax and T. Kailath, “Decentralized processing in sensor arrays,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 1123–1129, 1985.
- [WLJ07] H. Wang, G. Li, and G. Jiang, “Robust regression shrinkage and consistent variable selection through the LAD-LASSO,” *Journal of Business & Economic Statistics*, vol. 25, no. 3, pp. 347–355, 2007.
- [WR04] D. P. Wipf and B. D. Rao, “Sparse Bayesian learning for basis selection,” *IEEE Transactions on Signal processing*, vol. 52, no. 8, pp. 2153–2164, 2004.
- [Wri97] S. J. Wright, *Primal-dual interior-point methods*. Philadelphia: Society for Industrial and Applied Mathematics, 1997.
- [Wri15] ———, “Coordinate descent algorithms,” *Mathematical Programming*, vol. 151, no. 1, pp. 3–34, 2015.
- [WWX13] Y. Wang, J. Wang, and Z. Xu, “On recovery of block-sparse signals via mixed ℓ_2/ℓ_q ($0 < q \leq 1$) norm minimization,” *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–17, 2013.

- [YL06] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [YLSX17] Z. Yang, J. Li, P. Stoica, and L. Xie, “Sparse methods for direction-of-arrival estimation,” in *Academic Press Library in Signal Processing - Array, Radar and Communications Engineering*, 1st ed., S. Theodoridis and R. Chellappa, Eds. Academic Press, Oct. 2017, vol. 7, ch. 11.
- [YP17] Y. Yang and M. Pesavento, “A unified successive pseudoconvex approximation framework,” *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3313–3328, Jul. 2017.
- [YX15a] Z. Yang and L. Xie, “Enhancing sparsity and resolution via reweighted atomic norm minimization,” *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 995–1006, Feb. 2015.
- [YX15b] —, “On gridless sparse methods for line spectral estimation from complete and incomplete data,” *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.
- [YX16a] —, “Exact joint sparse frequency recovery via optimization methods,” *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, Oct. 2016.
- [YX16b] —, “On gridless sparse methods for multi-snapshot DOA estimation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 3236–3240.
- [YXS16] Z. Yang, L. Xie, and P. Stoica, “Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.
- [ZLG11] H. Zhu, G. Leus, and G. Giannakis, “Sparsity-cognizant total least-squares for perturbed compressive sampling,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, May 2011.
- [Zou06] H. Zou, “The adaptive LASSO and its oracle properties,” *Journal of the American statistical association*, vol. 101, no. 476, pp. 1418–1429, 2006.

Lebenslauf

Name: Christian Steffens
Geburtsdatum: 24.05.1981
Geburtsort: Norden, Deutschland

Studium

10/2002 - 06/2010 Diplom in Elektro- und Informationstechnik,
Universität Bremen

Schulausbildung

06/1994 - 06/2000 Allgemeine Hochschulreife,
Gymnasium Ulricianum, Aurich

Berufstätigkeit

seit 06/2017 Senior Systems Engineer bei Telespazio VEGA Deutschland,
Darmstadt
11/2010 - 10/2016 Wissenschaftlicher Mitarbeiter am Fachgebiet Nachrichtentechnische Systeme, Technische Universität Darmstadt
08/2008 - 10/2008 Fachpraktikum bei der MIMOon GmbH, Duisburg
04/2007 - 04/2008 Studentische Hilfskraft am Arbeitsbereich Nachrichtentechnik, Universität Bremen

Wettbewerbe und Auszeichnungen

06/2014 "Student best paper award" (dritter Preis) für den Beitrag "Direction Finding and Array Calibration Based on Sparse Reconstruction in Partly Calibrated Arrays", bei dem IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM) 2014, A Coruna, Spanien (Koautoren Pouyan Parvazi und Marius Pesavento)
06/2009 Teilnahme an der Finalrunde des Rohde & Schwarz Fallstudienwettbewerbs "Ready for take off - die Zukunft der Flugsicherung liegt in Ihrer Hand!", München