Sensor calibration for off-the-grid spectral estimation

Yonina C. Eldar [a], Wenjing Liao [b], Sui Tang [c]

[a] *Department of EE Technion, Israel Institute of Technology, Haifa, Israel*
[b] *School of Mathematics, Georgia Institute of Technology, United States of America*
[c] *Department of Mathematics, Johns Hopkins University, United States of America*

A R T I C L E   I N F O

A B S T R A C T

This paper studies sensor calibration in spectral estimation where the true frequencies are located on a continuous domain. We consider a uniform array of sensors that collects measurements whose spectrum is composed of a finite number of frequencies, where each sensor has an unknown calibration parameter. Our goal is to recover the spectrum and the calibration parameters simultaneously from multiple snapshots of the measurements. In the noiseless case with an infinite number of snapshots, we prove uniqueness of this problem up to certain trivial, inevitable ambiguities based on an algebraic method, as long as there are more sensors than frequencies. We then analyze the sensitivity of this algebraic technique with respect to the number of snapshots and noise.

We next propose an optimization approach that makes full use of the measurements by minimizing a non-convex objective which is non-negative and continuously differentiable over all calibration parameters and Toeplitz matrices. We prove that, in the case of infinite snapshots and noiseless measurements, the objective vanishes only at equivalent solutions to the true calibration parameters and the measurement covariance matrix. The objective is minimized using Wirtinger gradient descent which is proven to converge to a critical point. We show empirically that this critical point provides a good approximation of the true calibration parameters and the underlying frequencies.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

High-performance systems in signal processing often require precise calibration of sensors. However, such advanced sensors can be very expensive and difficult to build. It is, therefore, beneficial to use the measurements themselves to adjust calibration parameters and perform signal recovery simultaneously. We treat sensor calibration in spectral estimation where the frequencies of interest are located on a continuous domain.

---

* Corresponding author.

E-mail addresses: yonina@ee.technion.ac.il (Y.C. Eldar), wenjing.liao@math.gatech.edu (W. Liao), stang@math.jhu.edu (S. Tang).

A uniform array of N sensors collects measurements whose spectrum is composed of s spikes located at $\mathcal{S} := \{\omega_j \in [0,1)\}_{j=1}^{s}$ with amplitudes $x(t) := \{x_j(t)\}_{j=1}^{s} \in \mathbb{C}^{s}$ at time t. Each sensor has an unknown calibration parameter g_n, $n = 0, \ldots, N-1$. The measurement vector at the array output can then be written as

$$y_e(t) = GAx(t) + e(t), \tag{1}$$

where $y_e(t) = \{y_{e,n}(t)\}_{n=0}^{N-1} \in \mathbb{C}^{N}$ and $y_{e,n}(t)$ is the noisy measurement collected by the n-th sensor at time t, $G = \mathrm{diag}(g) \in \mathbb{C}^{N \times N}$ with $g = \{g_n\}_{n=0}^{N-1} \in \mathbb{C}^{N}$ is the calibration matrix, $e(t) \in \mathbb{C}^{N}$ is the noise vector at time t, and $A \in \mathbb{C}^{N \times s}$ is the sensing matrix with elements

$$A_{n,j} = \frac{1}{\sqrt{N}} e^{2\pi i n \omega_j} \tag{2}$$

where $n = 0, \ldots, N-1$ and $j = 1, \ldots, s$. Our goal is to recover the spectrum $\mathcal{S}$ and the calibration parameters g simultaneously from noisy measurements $\{y_e(t), t \in \Gamma\}$ where $L := \#\Gamma$ is the number of snapshots.

Spectral estimation modeled by (1) is a fundamental problem in imaging and signal processing. It is widely used in speech analysis, direction of arrival (DOA) estimation, array imaging and remote sensing. For example, in array imaging [14,16,30,31], we assume there are s sources located at $\{\omega_j\}_{j=1}^{s}$ with amplitudes $\{x_j(t)\}_{j=1}^{s}$, and use a uniform array of sensors to collect measurements $\{y_e(t), \ t \in \Gamma\}$. Our goal is to recover the source locations and the sensor calibration parameters simultaneously from the measurements.

When all sensors are perfectly calibrated (g is known), many methods have been proposed to recover frequencies located on a continuous domain, such as Prony's method [28], MUSIC [30,31], ESPRIT [29], ℓ_1 minimization [5,36], and greedy algorithms [10,13]. We refer the reader to [11,33] for a comprehensive review. In this paper, the problem is complicated by the fact that each sensor has an unknown calibration parameter.

The calibration problem modeled by (1) has been considered in [27,42,43,15] with the assumptions that the underlying frequencies have random and uncorrelated amplitudes, frequencies and noises are independent, and noises at different sensors are uncorrelated. In [27], Paulraj and Kailath investigated DOA estimation using a uniform linear array in the presence of unknown calibration parameters. By exploiting the Toeplitz structure of the measurement covariance in the case of perfect sensors (which we refer to as the case in which all calibration parameters are equal to 1), they proposed an algebraic method relying on a least squares solution of two linear systems of equations for the calibration amplitudes and phases, respectively. However, the issue of phase wrapping in the set of equations for the calibration phase estimation is not taken into account and can degrade the DOA estimator performance (see Section 3.1). This method is called the full algebraic method in our paper. A similar approach is followed in [43]. In [21] Li and Er showed that, the bias in the full algebraic method with finite snapshots of measurements is nonzero, and if the problem of phase wrapping is resolved, then the variance for the calibration phases is $O(1/\sqrt{L})$ where L is the number of snapshots. However, it remains unclear how to resolve phase wrapping in the full algebraic method. Furthermore, sensitivity to noise and sensitivity in spectral estimation are not treated in [21]. In [42], Wylie, Roy, and Schmitt proposed a partial algebraic technique which successfully avoids the problem of phase wrapping by removing redundancy in the linear system. A shortcoming of their approach is that a large part of the measurements are not used in the recovery process.

In [15], an alternating algorithm is proposed by Friedlander and Weiss for the same calibration problem modeled by (1). This algorithm is based on a two-step procedure. First, one assumes that the calibration parameters are known, and estimates the frequencies through the MUSIC algorithm. Then one solves an optimization problem to obtain the best calibration parameters based on the recovered frequencies. However, no performance guarantee for this approach is provided. In addition, this algorithm does not perform as well as the partial algebraic method in the presence of noise.

1.1. Our contributions

We begin by studying uniqueness of the calibration problem given by (1) and show that there are certain inevitable ambiguities in this problem. We characterize all trivial ambiguities, and prove that, both the spectrum and the calibration parameters are uniquely determined from infinite snapshots of noiseless measurements up to a trivial ambiguity, as long as there are more sensors than frequencies. Our proof is based on the algebraic methods proposed in [27,42]. We then present a sensitivity analysis of the partial algebraic method [42] with respect to the number of snapshots L and noise level σ. In particular, Theorem 2 shows that, if the underlying frequencies are separated by $1/N$ (the standard resolution in spectral estimation), then the reconstruction error of the calibration parameters in the partial algebraic method is $O\left(L^{-\frac{1}{2}}(C_1 + C_2 \max(\sigma, \sigma^2))\right)$ for some constants $C_1, C_2 > 0$. This rate is verified by numerical experiments. As for frequency localization, we prove that, the noise-space correlation function whose s smallest local minima correspond to the recovered frequencies in the MUSIC algorithm, is perturbed by at most $O\left(L^{-\frac{1}{2}}(C_3 + C_4 \max(\sigma, \sigma^2))\right)$ for some constants $C_3, C_4 > 0$.

The partial algebraic method in [42] only exploits partial entries of the measurement covariance matrix and while full algebraic method in [27] is affected by phase wrapping. We therefore propose an optimization approach to make full use of the measurements. We consider an objective function composed of two terms: one is a quadratic loss and the other is a penalty which prevents calibration parameters going to ∞ and frequency amplitudes decreasing to 0 (or vice versa). This objective is continuously differentiable but non-convex. We propose to minimize it over all possible calibration parameters and Toeplitz matrices by Wirtinger gradient descent [6]. We prove that, Wirtinger gradient descent converges to a critical point, and show empirically that this critical point provides a good approximation to the true calibration parameters and the underlying frequencies.

Finally we perform a systematic numerical study to compare the partial algebraic method [42], the alternating algorithm [15], and our optimization approach. With respect to the stability to L and σ, our algorithm has the best performance.

In summary, the main contributions of this paper are (i) characterizing all trivial ambiguities of the calibration problem modeled by (1) and proving uniqueness up to a trivial ambiguity; (ii) presenting a sensitivity analysis of the partial algebraic method; (iii) proposing an optimization approach with superior numerical performance over previous methods.

1.2. Related work

Recently, many works addressed the following single-snapshot calibration problem:

$$y = \text{diag}(g)\Phi x_0 + e \tag{3}$$

where $g \in \mathbb{C}^N, x_0 \in \mathbb{C}^M$ are the unknown calibration parameters and signal of interest respectively, $\Phi \in \mathbb{C}^{N \times M}$ is a given sensing matrix, $e \in \mathbb{C}^N$ represents noise, and $y \in \mathbb{C}^N$ is the measurement vector [2,25,20]. The goal is to recover g and x_0 from y. Without additional assumptions, solutions to this problem are not unique since there are more unknowns than equations. In [2], Ahmed, Recht, and Romberg assumed that g lies in a known subspace, and used a lifting technique to transform the problem into that of recovering a rank-one matrix from an underdetermined system of linear equations. They proved explicit conditions under which nuclear norm minimization is guaranteed to recover the original solution in the case where Φ is a random Gaussian matrix. In [25], Ling and Strohmer extended the framework in [2] to allow for sparse signals, and a random Gaussian matrix or a random partial Discrete Fourier Transform (DFT) matrix. The latter is closely related to spectral estimation assuming a discretized frequency grid with spacing equal to $1/N$. It is well known that when the underlying frequencies are on a continuous domain, this discretization

may cause a large gridding error [8,10,12,13]. Here we do not discretize the frequencies in order to avoid gridding errors.

The lifting technique has been applied on a wide range of blind deconvolution and sensor calibration models [4,1,9,7,44], among which [7,44] are mostly related with our model. In [7], Chi considered a slightly different single-snapshot model:

$$y = \mathrm{diag}(g)Ax \tag{4}$$

where $g \in \mathbb{C}^N$ contains unknown calibration parameters, A is the same as in (2), $x \in \mathbb{C}^s$ is an unknown amplitude vector, and $y \in \mathbb{C}^N$ is the measurement vector. The goal is to recover g, x, and the frequencies $\{\omega_j\}_{j=1}^s$ from y. This problem is the same as our model in (1) with a single snapshot. Chi solved (4) using a lifting technique and atomic norm minimization, and proved that, in the noiseless case where g lies in a random subspace of dimension K with coherence parameter μ, and if the underlying frequencies are separated by $4/N$, then exact recovery is guaranteed with high probability as long as $N \geq C\mu K^2 s^2$ up to a log factor. Chi's result is generalized by Yang et al. in [44] to the model

$$y(n) = \sum_{j=1}^s x_j e^{2\pi niw_j} g_j(n), \ n = 0, \cdots, N-1, \tag{5}$$

where $x_j \in \mathbb{C}$, $w_j \in [0,1)$ and $\{g_j(n)\}$ are the unknown amplitude, location, and samples of the waveform associated with the jth complex exponential. It is assumed that (i) all g_j live in a common random subspace of dimension K with coherence parameter μ; (ii) the ω_j's are separated by $4/N$. Yang et al. proved exact recovery with high probability as long as $N \geq C\mu s K$ up to a log factor. Noise was not considered in [7,44].

Since the lifting technique greatly increases computational complexity, many non-convex optimization approaches have been proposed to address problems in signal processing, such as phase retrieval [6,3,39,34], dictionary recovery [35], blind deconvolution [20], and low-rank matrix estimation [45]. In [20], Li et al. formulated a non-convex optimization problem for the calibration problem modeled by (3), and solved it with a two-step scheme composed of a good initialization and gradient descent [20]. Performance guarantees were proved when g lies in a known subspace and Φ is a random Gaussian matrix. This theory does not apply to spectral estimation since our sensing matrix is not random Gaussian.

After sensors are built, it is usually cheap to take multiple snapshots of measurements. This paper studies the calibration problem modeled by (1) with multiple snapshots. In comparison with the works in [2,7,25,20], we remove the assumption that g lies in a known subspace. Instead we utilize the Fourier structure and take advantage of the multiple snapshots. In addition, we only require more sensors than frequencies, namely, $N > s$.

To consider multiple snapshots of measurements, the works in [22,23] addressed the following calibration model:

$$Y = \mathrm{diag}(g)\Phi X_0 + E \tag{6}$$

where $g \in \mathbb{C}^N$, $X_0 \in \mathbb{C}^{M \times L}$ are the unknown calibration parameters and signals in L snapshots respectively, $\Phi \in \mathbb{C}^{N \times M}$ is a given sensing matrix, $E \in \mathbb{C}^{N \times L}$ represents noise, and $Y \in \mathbb{C}^{N \times L}$ is the measurement matrix in L snapshots. In [22], Li, Lee and Bresler proved uniqueness up to a scaling ambiguity for generic signals X_0 and a generic sensing matrix Φ provided that $N > M$ and $\frac{N-1}{N-M} \leq L \leq M$. In the case where X_0 has s non-zero rows, uniqueness was proved for generic signals with s non-zero rows and a generic sensing matrix Φ provided that $N > 2s$ and $\frac{N-1}{N-2s} \leq L \leq s$. These conditions are optimal in terms of sample complexity. In [23], the authors formulated this calibration problem as an eigenvalue/eigenvector problem, and solved it via power iterations, or when X is sparse or jointly sparse, via truncated power iterations. In [17], the

same kind of power method was applied to solve a multichannel blind deconvolution problem. However, the theory and algorithms in [17,22,23] do not apply to our case for the following reasons: (i) our sensing matrix defined in (2) is unknown, since it depends on the unknown frequencies; (ii) even if we discretize the frequency domain and approximate every frequency by the nearest grid point to fit the model (6), our sensing matrix is not generic. In comparison, our uniqueness results consider the Fourier structure of the measurements but assume infinite snapshots. It is also interesting to study the optimal sample complexity in our case; we leave this problem for future research.

1.3. Organization and notation

This paper is organized as follows. Uniqueness results are described in Section 2.1. The partial algebraic method and its sensitivity analysis are presented in Section 3. Section 4 considers our optimization approach and its convergence to a critical point. Numerical simulations are presented in Section 5. We conclude and discuss future research directions in Section 6. Most of the proofs are relegated to the Appendices.

Throughout the paper we use s, N, L to denote the number of frequencies, sensors, and snapshots respectively. The expression $C = (A\ B)$ horizontally concatenates matrices A and B, while $C = (A; B)$ concatenates them vertically. For $x \in \mathbb{C}^N$, $\sum x := \sum_{j=1}^{N} x_j$, and $\operatorname{diag}(x)$ is the $N \times N$ diagonal matrix whose diagonal is x. We use $|x| \in \mathbb{R}^N$ and $\angle x \in \mathbb{R}^N$ to denote the magnitude and phase vectors of x respectively such that $|x|_j = |x_j|$ and $(\angle x)_j = \angle x_j, j = 1, \ldots, N$. The dynamic range of x is the ratio between the largest amplitude and the smallest amplitude of x, denoted by $\mathrm{DR}_x := \max_i |x_i| / \min_i |x_i|$. For $X \in \mathbb{C}^{N \times N}$, we use $\operatorname{diag}(X)$ to denote the main diagonal of X, $\operatorname{diag}(X, k), k > 0$ to denote the kth diagonal of X above the main diagonal, and $\operatorname{diag}(X, k), k < 0$ to denote the kth diagonal of X below the main diagonal. The notation $\|X\|$ and $\|X\|_F$ represent the spectral norm and the Frobenius norm of X, respectively. The expression $A \preceq B$ for two square matrices A and B of the same size means $B - A$ is positive semidefinite. The expression $a \lesssim b$ for two scalars a and b means $a \leq Cb$ with a constant C independent of a, b. We use $\mathbf{0}$ to denote the null vector or the null matrix.

2. Uniqueness results

2.1. Trivial ambiguity and uniqueness

In the case of a single snapshot, uniqueness does not exist since there are fewer measurements (N) than unknowns $(N + 2s)$. After sensors are built, it is often cheap to take multiple snapshots. Therefore, for uniqueness we study the setting where infinite snapshots of noiseless measurements are taken. Clearly certain trivial ambiguities between the spectrum and the calibration parameters are inevitable. For example, one can add a gain to $x(t)$ and then divide it out in g, so that there is always a gain ambiguity. Similarly, there is always a shift ambiguity in the frequencies since we can shift all frequencies by a constant and all this will do is add a phase modulation to g. We define trivial ambiguities in the calibration problem as follows:

Definition 1 *(Trivial ambiguity).* Let $\{g, \mathcal{S}, x(t)\}$ be a solution to the calibration problem modeled by (1). Then $\{\tilde{g}, \tilde{\mathcal{S}}, \tilde{x}(t)\}$ is called equivalent to $\{g, \mathcal{S}, x(t)\}$ up to a trivial ambiguity if there exist $c_0 > 0, c_1, c_2 \in \mathbb{R}$ such that

$$\tilde{g} = \{\tilde{g}_n = c_0 e^{i(c_1 + n c_2)} g_n\}_{n=0}^{N-1}$$
$$\tilde{\mathcal{S}} = \{\tilde{\omega}_j : \tilde{\omega}_j = \omega_j - c_2/(2\pi)\}_{j=1}^{s}$$
$$\tilde{x}(t) = x(t) c_0^{-1} e^{-i c_1}.$$

To proceed we make the following assumptions on the statistics of the frequencies (or sources in array imaging) and noises:

A1 Calibration parameters do not vanish: $|g_n| \neq 0, \ n = 0, \ldots, N - 1$.
A2 Sources and noises have zero mean: $\mathbb{E}x(t) = \mathbf{0}$ and $\mathbb{E}e(t) = \mathbf{0}$.
A3 Sources are uncorrelated such that $R^x := \mathbb{E}x(t)x^*(t) = \mathrm{diag}(\{\gamma_j^2\}_{j=1}^s)$.
A4 Sources and noises are independent, i.e., $\mathbb{E}x(t)e^*(t) = \mathbf{0}$.
A5 Noises at different sensors are uncorrelated so that $\mathbb{E}e(t)e^*(t) = \sigma^2 I_N$ where σ represents the noise level.

Define

$$f_n = \frac{1}{N} \sum_{j=1}^{s} \gamma_j^2 e^{2\pi i n \omega_j}, \ n = 0, \ldots, N - 1. \tag{7}$$

The values $\{f_n\}_{n=0}^{N-1}$ contain sufficient information to recover all frequencies by standard spectral estimation. Notice that $f_0 = \sum_{j=1}^{s} \gamma_j^2 > 0$. In the case of infinite snapshots, uniqueness of the calibration problem exists up to a trivial ambiguity as long as $|f_1| > 0$. This is a sufficient condition to guarantee that the sub-diagonal entries in the covariance measurement do not vanish. If the amplitudes $\{\gamma_j^2\}_{j=1}^s$ are generic, then we have $|f_1| > 0$ almost surely.

Theorem 1. *Suppose $|f_1| > 0$, $N \geq s + 1$, and the assumptions A1–A5 hold. Let $\{g, \mathcal{S}, x(t)\}$ be a solution to the calibration problem modeled by* (1). *If there is another solution $\{\tilde{g}, \tilde{\mathcal{S}}, \tilde{x}(t)\}$, then $\{\tilde{g}, \tilde{\mathcal{S}}, \tilde{x}(t)\}$ is equivalent to $\{g, \mathcal{S}, x(t)\}$ up to a trivial ambiguity.*

2.2. Proof of uniqueness

We prove Theorem 1 based on the algebraic technique proposed in [27,42]. We begin by considering the covariance of y_e:

$$R_e^y := \mathbb{E}y_e(t)y_e^*(t) = GA\underbrace{\mathbb{E}x(t)x^*(t)}_{R^x}A^*G^* + GA\underbrace{\mathbb{E}x(t)e^*(t)}_{R^{xe}} + \underbrace{\mathbb{E}e(t)x^*(t)}_{R^{ex}}A^*G^* + \underbrace{\mathbb{E}e(t)e^*(t)}_{R^e}. \tag{8}$$

Denote the noiseless data by $y(t) := GAx(t)$ and its covariance by

$$R^y := \mathbb{E}y(t)y^*(t) = GAR^xA^*G^*. \tag{9}$$

Under Assumptions (A1–A5), we have

$$R_e^y = R^y + R^e = GAR^xA^*G^* + \sigma^2 I_N. \tag{10}$$

In the noiseless case, $\sigma = 0$ and $R_e^y = R^y$. If infinite snapshots are collected, then we can assume R^y is known.

Define the Toeplitz operator which maps a sequence to a Toeplitz matrix:

$$\mathcal{T} : \mathbb{R} \times \mathbb{C}^{N-1} \to \mathbb{C}^{N \times N} : \mathcal{T}(f) := \begin{bmatrix} f_0 & \bar{f}_1 & \ddots & \bar{f}_{N-2} & \bar{f}_{N-1} \\ f_1 & f_0 & \ddots & \ddots & \bar{f}_{N-2} \\ \ddots & \ddots & f_0 & \ddots & \ddots \\ f_{N-2} & \ddots & \ddots & \ddots & \bar{f}_1 \\ f_{N-1} & f_{N-2} & \ddots & f_1 & f_0 \end{bmatrix}.$$

With this notation

$$R^y = \text{diag}(g)\mathcal{T}(f)\text{diag}(\bar{g}), \tag{11}$$

where f is the sequence defined in (7). The (m,n)th entry of R^y is

$$R^y_{m,n} = \frac{g_m \bar{g}_n}{N} \sum_{j=1}^{s} \gamma_j^2 e^{2\pi i \omega_j (m-n)} = g_m \bar{g}_n f_{m-n}. \tag{12}$$

Throughout the paper we write $g_n = \alpha_n e^{i\beta_n}$, where α_n is the calibration amplitude and β_n is the calibration phase at the nth sensor. By (12), all calibration amplitudes can be uniquely determined from the diagonal entries of R^y up to a scaling, and the calibration phases can be determined from the subdiagonal of R^y up to a trivial ambiguity as long as f_1 does not vanish.

Lemma 1. *Suppose $|f_1| > 0$ and the assumptions A1–A5 hold. If there is another set $\{\tilde{g} \in \mathbb{C}^N, \tilde{f} \in \mathbb{R} \times \mathbb{C}^{N-1}\}$ satisfying $\text{diag}(\tilde{g})\mathcal{T}(\tilde{f})\text{diag}(\bar{\tilde{g}}) = \text{diag}(g)\mathcal{T}(f)\text{diag}(\bar{g})$, then there exist $c_0 > 0$ and $c_1, c_2 \in \mathbb{R}$ such that*

$$\tilde{g}_n = c_0 e^{i(c_1 + nc_2)} g_n, \quad \tilde{f}_n = c_0^{-2} e^{-inc_2} f_n, \quad n = 0, \ldots, N-1.$$

Proof. We write $g_n = \alpha_n e^{i\beta_n}$ where α_n is the calibration amplitude and β_n is the calibration phase at the nth sensor. Then $\alpha = |g|$ and $\beta = \angle g$. Similarly, let $\tilde{\alpha} = |\tilde{g}|$ and $\tilde{\beta} = \angle \tilde{g}$. Observe that the diagonal entries of R^y are

$$R^y_{n,n} = \alpha_n^2 f_0, \ n = 0, \ldots, N-1.$$

If f_0 is given, then $\tilde{\alpha}_n = \sqrt{R^y_{n,n}/f_0}$; otherwise, the unknown f_0 leads to a scaling ambiguity such that $\tilde{\alpha} = c_0 \alpha$ for some constant $c_0 > 0$.

The sub-diagonal entries of R^y are

$$R^y_{n,n-1} = \alpha_n \alpha_{n-1} e^{i(\beta_n - \beta_{n-1})} f_1 \neq 0, \ n = 1, \ldots, N-1.$$

This gives rise to $N-2$ equations regarding the β_n's:

$$e^{i(\beta_{n+1} - 2\beta_n + \beta_{n-1})} = \frac{\alpha_{n-1} R^y_{n+1,n}}{\alpha_{n+1} R^y_{n,n-1}},$$

which are equivalent to

$$\beta_{n+1} - 2\beta_n + \beta_{n-1} = \angle \frac{R^y_{n+1,n}}{R^y_{n,n-1}} + 2\pi k_n, \ n = 1, \ldots, N-2, \ k_n \in \mathbb{Z}. \tag{13}$$

The linear system for β given by (13) has $N-2$ independent equations and N variables. Solving (13) results in

$$\tilde{\beta} = \beta + c_1 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ N-2 \\ N-1 \end{bmatrix} \quad \text{mod } 2\pi.$$

Therefore, $\tilde{\beta}_n = \beta_n + c_1 + nc_2 \mod 2\pi$. Combined with $\tilde{\alpha} = c_0 \alpha$, we have $\tilde{g}_n = c_0 e^{i(c_1 + nc_2)} g_n$. As for $\tilde{f}$, since $R^y_{m,n} = g_m \bar{g}_n f_{m-n} = \tilde{g}_m \bar{\tilde{g}}_n \tilde{f}_{m-n}$, we obtain $\tilde{f}_{m-n} = c_0^{-2} e^{-i(m-n)c_2} f_{m-n}$, which concludes the proof. $\quad\square$

After obtaining the calibration parameters $\{\tilde{g}_n = c_0 e^{i(c_1+nc_2)} g_n\}_{n=0}^{N-1}$, we simply divide $\tilde{g}$ out of R^y to obtain

$$\tilde{F} = \mathrm{diag}(\tilde{g})^{-1} R^y \mathrm{diag}(\bar{\tilde{g}})^{-1} = c_0^{-2} D_{c_2} A R^x A^* D_{c_2}^* = c_0^{-2} \tilde{A} R^x \tilde{A}^*,$$

where $D_{c_2} = \mathrm{diag}\left(\{e^{-inc_2}\}_{n=0}^{N-1}\right)$ and

$$\tilde{A}_{n,j} \in \mathbb{C}^{N\times s}: \quad \tilde{A}_{n,j} = \frac{1}{\sqrt{N}} e^{2\pi i n(\omega_j - \frac{c_2}{2\pi})}. \tag{14}$$

We can then perform standard spectral estimation on $\tilde{F}$ using the MUltiple Signal Classification (MUSIC) algorithm proposed by Schmidt [30,31] to retrieve the frequencies. MUSIC is introduced in Section 3. For now, we use the result that MUSIC guarantees exact recovery of frequencies in the noiseless case as long as $N \geq s + 1$ (see Proposition 1). Combining this with Lemma 1 gives rise to Theorem 1.

2.3. A general condition to guarantee uniqueness

Theorem 1 guarantees uniqueness when $f_1 \neq 0$. This condition can be generalized as follows. For $k = 1, \ldots N-1$, we have $R^y_{l+k,l} = \alpha_{l+k}\alpha_l e^{i(\beta_{l+k}-\beta_l)} f_k$ where $l = 0, \ldots, N-k-1$. As long as $|f_k| \neq 0$, we can compute $R^y_{l+k+1,l+1}/R^y_{l+k,l}$ for $l = 0, \ldots, N-k-2$ and obtain the following system of linear equations

$$\beta_{l+k+1} - \beta_{l+k} - \beta_{l+1} + \beta_l \equiv \angle R^y_{l+k+1,l+1}/R^y_{l+k,l} \mod 2\pi \tag{15}$$

for $l = 0, \ldots, N-k-2$. Here $a \equiv b \mod c$ means a is equal to b modulo c. We may write these $N-k-1$ linear equations (15) in matrix form as

$$\Phi_k \beta \equiv \delta_k \mod 2\pi,$$

where

$$\Phi_1 = \begin{bmatrix} 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ & & & \cdots & & & & \\ & & & \cdots & & & & \\ & & & \cdots & & & & \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{bmatrix} \in \mathbb{C}^{(N-2)\times N},$$

$$\Phi_k = \begin{bmatrix} 1 & -1 & \underbrace{0 \cdots 0}_{k-2} & -1 & 1 & & & \\ & 1 & -1 & \underbrace{0 \cdots 0}_{k-2} & -1 & 1 & & \\ & & \ddots & & \ddots & \ddots & \ddots & \ddots \\ & & & 1 & -1 & \underbrace{0 \cdots 0}_{k-2} & -1 & 1 \end{bmatrix} \in \mathbb{C}^{(N-k-1)\times N}$$

for $k = 2, \ldots N-2$ and

$$\delta_k = [\angle R^y_{k+1,1}/R^y_{k,0} \quad \angle R^y_{k+2,2}/R^y_{1+k,1} \quad \cdots \quad \angle R^y_{N-1,N-k-1}/R^y_{N-2,N-k-1}]^T \in \mathbb{C}^{N-k-1}.$$

Let $\Lambda = \{k : |f_k| \neq 0, k = 1, \cdots, N-2\}$. Concatenating the matrices Φ_k for $k \in \Lambda$ vertically yields the following matrix

$$\Phi_\Lambda = \begin{bmatrix} \vdots \\ \Phi_k \\ \vdots \end{bmatrix}_{k \in \Lambda} .$$

By exploiting all entries in the covariance matrix, we can generalize the condition $|f_1| > 0$ in Theorem 1 to the condition that $\mathrm{rank}(\Phi_\Lambda) = N-2$. Thus, uniqueness in Theorem 1 holds under the more general condition: $\mathrm{rank}(\Phi_\Lambda) = N - 2$. Observe that $|f_1| > 0$ is sufficient but not necessary to guarantee $\mathrm{rank}(\Phi_\Lambda) = N - 2$.

3. Algebraic methods

3.1. Full algebraic method and phase wrapping

In [27], Paulraj and Kailath proposed the first method for sensor calibration in DOA estimation. By exploiting the Toeplitz structure of $\mathcal{T}(f)$, they obtained two linear systems for calibration amplitudes and phases, respectively.

Consider the case with infinite snapshots of noiseless measurements. When k varies from 1 to $N - 1$, the k-th sub-diagonal entries of R^y satisfy

$$\alpha_{l+k}\alpha_l e^{i(\beta_{l+k}-\beta_l)}f_k = R^y_{l+k,l}, \ \ l = 0, \ldots, N - k - 1.$$

For $k = 1, \ldots, N - 1$, as long as $|f_k| \neq 0$, one can obtain the following system of equations for $\ln \alpha$

$$\ln \alpha_{l+k+1} + \ln \alpha_{l+1} - \ln \alpha_{l+k} - \ln \alpha_l = \ln \frac{|R^y_{l+k+1,l+1}|}{|R^y_{l+k,l}|} \tag{16}$$

where $l = 0, \ldots, N - k - 2$, as well as (15) for calibration phases β. Paulraj and Kailath proposed to substitute $\equiv$ (equal modulo 2π) with $=$ in (15) and solve these linear systems by least squares. However, one has to consider phase wrapping in (15) so that (15) is equivalent to

$$\beta_{l+k+1} - \beta_{l+k} - \beta_{l+1} + \beta_l = \angle R^y_{l+k+1,l+1}/R^y_{l+k,l} + 2\pi p_{k,l} \text{ where } p_{k,l} \in \mathbb{Z} \tag{17}$$

where $l = 0, \ldots, N - k - 2$ and $k = 1, \ldots N - 1$. Importantly, the $p_{k,l}$'s in (17) are not independent. For example, the parameters $p_{1,0}, p_{1,1}, p_{2,0}$ need to satisfy

$$\angle R^y_{2,1}/R^y_{1,0} + 2\pi p_{1,0} + \angle R^y_{3,2}/R^y_{2,1} + 2\pi p_{1,1} = \angle R^y_{3,1}/R^y_{2,0} + 2\pi p_{2,0}. \tag{18}$$

The $p_{k,l}$'s are constrained by many more equations like (18). Solving (17) with the constraints involves phase synchronization, which itself is highly nontrivial.

3.2. Partial algebraic method

In [42], Wylie, Roy and Schmitt proposed a partial algebraic method by using the system in (13), corresponding to the set of equations in (15) with $k = 1$, to recover the calibration phases. These linear equations are independent so that there is no problem of phase wrapping. This partial algebraic method is summarized in Algorithm 1.

In practice, we take L snapshots of independent measurements, i.e., $\{y_e(t), t \in \Gamma, \#\Gamma = L\}$, and approximate R^y_e by the empirical covariance matrix $\widetilde{R}^y_e$ in (19). The noise level σ can be estimated from the smallest $N - s$ eigenvalues of $\widetilde{R}^y_e$ and the noise component can be subtracted from $\widetilde{R}^y_e$ to yield $\widehat{R}^y$ as an approximation of R^y (see Step 4 in Algorithm 1). We then identify the calibration amplitudes $|g|$ from the

---

Algorithm 1 Partial algebraic method.

Input: Measurements $\{y_e(t), t \in \Gamma\}$ and sparsity s.

Output: Calibration parameters $\widehat{g} = \{\widehat{g}_n := \widehat{\alpha}_n e^{i\widehat{\beta}_n}\}_{n=0}^{N-1}$ and recovered spectrum $\{\widehat{\omega}_j\}_{j=1}^{s}$

1: Form the empirical covariance matrix

$$\widetilde{R}_e^y = \frac{1}{L} \sum_{t \in \Gamma} y_e(t) y_e^*(t). \tag{19}$$

2: Compute the eigenvalue decomposition:

$$\widetilde{R}_e^y = U \Sigma U^*$$

where $\Sigma = \mathrm{diag}(\lambda_0(\widetilde{R}_e^y), \ldots, \lambda_{N-1}(\widetilde{R}_e^y))$, $\lambda_0(\widetilde{R}_e^y) \geq \lambda_1(\widetilde{R}_e^y) \geq \ldots$.

3: Estimate the noise level $\widehat{\sigma} = \sqrt{\sum_{l=s}^{N-1} \lambda_l(\widetilde{R}_e^y)/(N-s)}$.

4: Subtract the noise component: $\widehat{R}^y \leftarrow \widetilde{R}_e^y - \widehat{\sigma}^2 I_N$.

5: Estimate calibration amplitudes: $\widehat{\alpha}_n = \sqrt{\widehat{R}_{n,n}^y}$, $n = 0, \ldots, N-1$.

6: Solve the following linear system $\Phi\widehat{\beta} = \widehat{b}$ to obtain calibration phases $\widehat{\beta}$

$$\underbrace{\begin{bmatrix} \frac{1}{N^2} & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ & & & \ldots & & & & \\ & & & \ldots & & & & \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & \frac{1}{N^2} \end{bmatrix}}_{\Phi \in \mathbb{R}^{N \times N}} \underbrace{\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_{N-3} \\ \widehat{\beta}_{N-2} \\ \widehat{\beta}_{N-1} \end{bmatrix}}_{\widehat{\beta} \in \mathbb{R}^N} = \underbrace{\begin{bmatrix} 0 \\ \angle(\widehat{R}_{2,1}^y / \widehat{R}_{1,0}^y) \\ \angle(\widehat{R}_{3,2}^y / \widehat{R}_{2,1}^y) \\ \vdots \\ \vdots \\ \angle(\widehat{R}_{N-1,N-2}^y / \widehat{R}_{N-2,N-1}^y) \\ 0 \end{bmatrix}}_{\widehat{b} \in \mathbb{R}^N}. \tag{20}$$

7: Compute the matrix $\widehat{F} = \widehat{G}^{-1} \widehat{R}^y \overline{\widehat{G}}^{-1}$ where $\widehat{G} = \mathrm{diag}(\widehat{g})$ and $\widehat{g}_n = \widehat{\alpha}_n e^{i\widehat{\beta}_n}$, $n = 0, \ldots, N-1$.

8: Apply the MUSIC algorithm on $\widehat{F}$:
   i) Compute the eigenvalue decomposition: $\widehat{F} = [V_1 \ V_2]\mathrm{diag}(\lambda_1(\widehat{F}), \ldots, \lambda_s(\widehat{F}), \ldots)[V_1 \ V_2]^*$ where $V_1 \in \mathbb{C}^{N \times s}$, and $\lambda_1(\widehat{F}) \geq \lambda_2(\widehat{F}) \geq \ldots$.
   ii) Compute the imaging function

$$\widehat{\mathcal{J}}(\omega) = \frac{\|\phi(\omega)\|}{\|V_2^* \phi(\omega)\|}$$

   where $\phi(\omega) = [1 \ e^{2\pi i\omega} \ \ldots e^{2\pi i(N-1)\omega}]^T$.
   iii) Return the spectrum $\{\widehat{\omega}_j\}_{j=1}^{s}$ corresponding to the s largest local maxima of $\widehat{\mathcal{J}}(\omega)$.

---

diagonal entries of $\widehat{R}^y$ and calibration phases $\angle g$ from the sub-diagonal entries of $\widehat{R}^y$ by solving (20) which gives a specific solution to (13) with $\beta_0 = \beta_{N-1} = 0$. After all calibration parameters are recovered, MUSIC is applied for the usual spectral estimation, which guarantees exact recovery with exact data.

Proposition 1. *Suppose s is known and the input of MUSIC is exact: $F = AR^xA^*$. If $N \geq s+1$, then*

$$\omega \in \mathcal{S} \Longleftrightarrow \mathcal{R}(\omega) = 0 \Longleftrightarrow \mathcal{J}(\omega) = \infty$$

where $\mathcal{J}(\omega)$ is the imaging function defined in Step 8.ii in Algorithm 1 and $\mathcal{R}(\omega) := 1/\mathcal{J}(\omega)$ is called the noise-space correlation function.

In the noiseless case, one can extract the s zeros of the noise-space correlation function $\mathcal{R}(\omega)$, or the s largest local maxima of the imaging function $\mathcal{J}(\omega)$ as an estimate of $\mathcal{S}$. In the presence of noise, suppose the input of MUSIC is approximate: $\widehat{F} = AR^xA^* + E$, and the noise-space correlation function is perturbed

from $\mathcal{R}$ to $\widehat{\mathcal{R}}$. Stability of MUSIC depends on the perturbation of the noise-space correlation function which can be estimated in the following lemma, thanks to classical perturbation theory of singular subspaces by Wedin [40,32,19, Theorem 3.4].

Proposition 2. *Let $N \geq s + 1$. Suppose s is known and the input of MUSIC is approximate: $\widehat{F} = F + E = AR^x A^* + E$. Let $\lambda_1(F) \geq \ldots \geq \lambda_s(F)$ be the nonzero eigenvalues of F. Suppose $\mathcal{R}(\omega)$ and $\widehat{\mathcal{R}}(\omega)$ are the noise-space correlation functions when MUSIC is applied on F and $\widehat{F}$ respectively. If $2\|E\| < \lambda_s(F)$, then*

$$\sup_{\omega \in [0,1)} |\widehat{\mathcal{R}}(\omega) - \mathcal{R}(\omega)| \leq \frac{2}{\lambda_s(F)} \|E\|.$$

3.3. Sensitivity of the partial algebraic method

Theorem 1 guarantees exact recovery up to a trivial ambiguity with infinite snapshots of noiseless measurements. In practice, only finite snapshots of noisy measurements are taken. We next present a sensitivity analysis of the partial algebraic method in Algorithm 1 with respect to the number of snapshots L and noise level σ. In particular, we prove that, there exist $C_1, C_2 > 0$ such that

$$\text{Reconstruction error of calibration parameters} \leq O\left(\frac{C_1 + C_2 \max(\sigma, \sigma^2)}{\sqrt{L}}\right)$$

by the partial algebraic method when the true frequencies are separated by $1/N$. As for frequency localization in the MUSIC algorithm, the recovered frequencies correspond to the s smallest local minima of the noise-space correlation function. Here we prove that, the noise-space correlation function is perturbed when using finite snapshots and in the presence of noise by at most $O\left(L^{-\frac{1}{2}}(C_3 + C_4 \max(\sigma, \sigma^2))\right)$ for some constants $C_3, C_4 > 0$. The constants C_1, C_2, C_3 and C_4 depend on the number of sources s, the number of sensors N, and dynamic ranges of the calibration parameters and source amplitudes. We will make these dependencies explicit in Remark 2. In the theorem below, let $\gamma_{\max} = \max_j \gamma_j, \gamma_{\min} = \min_j \gamma_j$, and $\alpha_{\max} = \max_n |g_n|, \alpha_{\min} = \min_n |g_n|$.

Theorem 2. *In addition to the assumptions A1–A5, assume $N \geq s + 1$, $|f_1| > 0$ and the source and noise amplitudes $\|x(t)\|$ and $\|e(t)\|$ are almost surely bounded. Let $\widehat{R}^y$ be the outcome in Step 4, $\widehat{g}$ be the recovered calibration parameters, and $\widehat{F}$ be the outcome in Step 7 of the partial algebraic method in Algorithm 1. Define*

$$\Delta R^y := 2\alpha_{\max}^2 \sigma_{\max}^2(A)\left(\frac{\gamma_{\max} \max_{t \in \Gamma} \|x(t)\| \sqrt{2 \log 4s}}{\sqrt{L}} + \frac{\gamma_{\max}^2 + \max_{t \in \Gamma} \|x(t)\|^2}{3L} \log 4s\right)$$

$$+ 4\alpha_{\max} \sigma_{\max}(A)\left(\frac{\sigma \gamma_{\max} \sqrt{2N \log(N + s)}}{\sqrt{L}} + \frac{\max_{t \in \Gamma} \|x(t)\| \|e(t)\|}{3L} \log(N + s)\right)$$

$$+ 2\left(\frac{\sigma \max_{t \in \Gamma} \|e(t)\| \sqrt{2 \log 2N}}{\sqrt{L}} + \frac{\sigma^2 + \max_{t \in \Gamma} \|e(t)\|^2}{3L} \log 2N\right). \tag{21}$$

Then

$$\mathbb{E}\|R^y - \widehat{R}^y\| \leq \Delta R^y.$$

Let $\mathcal{R}(\omega)$ and $\widehat{\mathcal{R}}(\omega)$ be the noise-space correlation functions in MUSIC with the input data $F = AR^x A^*$ and $\widehat{F}$ respectively. Then

$$\mathbb{E}\min_{c_0>0,c_1,c_2\in\mathbb{R}}\max_n\|c_0\widehat{g}_n - e^{i(c_1+nc_2)}g_n\|_\infty \le \left(\frac{3(\|g\|^2+N\alpha_{\max}^2)}{2\alpha_{\min}\|g\|^2 f_0}+144N^2\frac{\alpha_{\max}^5}{\alpha_{\min}^6|f_1|}\right)\Delta R^y \tag{22}$$

$$\mathbb{E}\min_{c_2\in\mathbb{R}}\sup_{\omega\in[0,1)}\left|\widehat{\mathcal{R}}(\omega)-\mathcal{R}\left(\omega-\frac{c_2}{2\pi}\right)\right| \le \frac{2}{\lambda_s(F)}\Delta F \tag{23}$$

where

$$\Delta F = \left[\frac{9}{\alpha_{\min}^2}+\frac{12\alpha_{\max}^2\gamma_{\max}^2\sigma_{\max}^2(A)}{\alpha_{\min}^3}\left(\frac{3(\|g\|^2+N\alpha_{\max}^2)}{2\alpha_{\min}\|g\|^2 f_0}+\frac{144N^2\alpha_{\max}^5}{\alpha_{\min}^6|f_1|}\right)\right]\Delta R^y.$$

Remark 1. The expectations in (22) and (23) are taken over random source amplitudes and random noises. Our estimates suggest that, the partial algebraic method is more stable in the cases where 1) the noise level σ is small and the number of snapshots L large; 2) $\|g\|, f_0, |f_1|$ are large; 3) the calibration parameters g have a small dynamic range such that $\alpha_{\max}/\alpha_{\min}\approx 1$; 4) the minimal calibration amplitude $\alpha_{\min}$ is large; 5) source amplitudes have a small dynamic range such that $\gamma_{\max}/\gamma_{\min}\approx 1$; 6) frequencies are well separated such that $\sigma_{\max}(A)/\sigma_{\min}(A)\approx 1$.

Remark 2. Notice that $\mathbb{E}\|x(t)\| = \sqrt{\sum_{j=1}^s\gamma_j^2}$ and $\mathbb{E}\|e(t)\| = \sigma\sqrt{N}$. Suppose $\|x(t)\|$ and $\|e(t)\|$ concentrate around $\mathbb{E}\|x(t)\|$ and $\mathbb{E}\|e(t)\|$ respectively. In the case that the true frequencies in $\mathcal{S}$ are separated by $q > 1/N$, discrete Ingham inequalities [24, Theorem 2] guarantee that $r_1(q,N)\le\sigma_{\min}(A)\le\sigma_{\max}(A)\le r_2(q,N)$ for some positive constants r_1,r_2 depending on q,N, which implies $\lambda_1(F)\le\gamma_{\max}^2 r_2^2$ and $\lambda_s(F)\ge\gamma_{\min}^2 r_1^2$. Then, when L is sufficiently large, we have

$$\Delta R^y \le O\left(\frac{B_1+B_2\max(\sigma,\sigma^2)}{\sqrt{L}}\right) \tag{24}$$

for some positive constants B_1,B_2 depending on $g,\{\gamma_j\}_{j=1}^s,\mathcal{S},N,s$. Therefore

$$\mathbb{E}\min_{c_0>0,c_1,c_2\in\mathbb{R}}\max_n\|c_0\widehat{g}_n - e^{i(c_1+nc_2)}g_n\|_\infty \le O\left(\frac{C_1+C_2\max(\sigma,\sigma^2)}{\sqrt{L}}\right)$$

$$\mathbb{E}\min_{c_2\in\mathbb{R}}\sup_{\omega\in[0,1)}\left|\widehat{\mathcal{R}}(\omega)-\mathcal{R}\left(\omega-\frac{c_2}{2\pi}\right)\right| \le O\left(\frac{C_3+C_4\max(\sigma,\sigma^2)}{\sqrt{L}}\right)$$

for some positive constants C_1,C_2,C_3,C_4 depending on $g,\{\gamma_j\}_{j=1}^s,\mathcal{S},N,s$. In particular, if $\alpha_{\max},\alpha_{\min},\gamma_{\max},\gamma_{\min}\approx 1$, and the frequencies are separated above $1/N$, then $B_1\sim\sqrt{s\log 4s}$, $B_2\sim\sqrt{N\log(2N)}$, $C_1,C_3\sim N^2\sqrt{s\log 4s}$ and $C_2,C_4\sim N^2\sqrt{N\log(2N)}$.

Remark 3. In Theorem 2, the expression in (21) may appear intimidating. However, it simply results from Bernstein inequalities [37], based on which we estimate the deviation of the sampled covariance matrix $\widetilde{R}_e^y$ from the covariance matrix R_e^y. Notice that

$$\|R_e^y-\widetilde{R}_e^y\| \le \|GA(R^x-\widetilde{R}^x)A^*G^*+GA(R^{xe}-\widetilde{R}^{xe})+(R^{ex}-\widetilde{R}^{ex})A^*G^*+R^e-\widetilde{R}^e\|$$

$$\le \sigma_{\max}^2(G)\sigma_{\max}^2(A)\|R^x-\widetilde{R}^x\|+\sigma_{\max}(G)\sigma_{\max}(A)\|R^{xe}-\widetilde{R}^{xe}\| \tag{25}$$

$$+\sigma_{\max}(G)\sigma_{\max}(A)\|R^{ex}-\widetilde{R}^{ex}\|+\|R^e-\widetilde{R}^e\|. \tag{26}$$

Applying Bernstein inequalities gives rise to (21) where the three terms correspond to upper bounds of (25) and (26).

Remark 4. By using Bernstein inequalities, we require that $\|x(t)\|$ and $\|e(t)\|$ are almost surely bounded and $\max_{t \in \Gamma} \|x(t)\|$ and $\max_{t \in \Gamma} \|e(t)\|$ appear in the upper bound. This result can be generalized to the case where the entries in $x(t)$ and $e(t)$ are independent sub-gaussian random variables (so we can drop the boundedness condition) by using theorem 4.7.1 in [38]. Then (21) becomes

$$\Delta R^y := C \left[\frac{\alpha_{\max}^2 \sigma_{\max}^2(A) \gamma_{\max}^3}{\gamma_{\min}^2} \left(\sqrt{\frac{s}{L}} + \frac{s}{L} \right) + 4\alpha_{\max} \sigma_{\max}(A) \gamma_{\max} N s \sqrt{Ns} \sigma \left(\frac{1}{\sqrt{L}} + \frac{1}{L} \right) \right.$$
$$\left. + 2\sigma \left(\sqrt{\frac{N}{L}} + \frac{N}{L} \right) \right], \tag{27}$$

and other results hold similarly.

A sensitivity analysis of the full algebraic method to the number of snapshots can be found in [21]. Assuming the problem of phase wrapping in the full algebraic method is resolved, Li and Er [21] split the reconstruction errors of the calibration amplitudes and phases to a bias term and a variance term. They claim that the bias is nonzero, and the variance of the calibration phases is $O(1/\sqrt{L})$ where L is the number of snapshots. Below we point out some differences between Theorem 2 and the analysis in [21].

1. In [21] the authors did not give an explicit bound on the bias but claimed it is non-zero. In this case the total reconstruction error for the calibration phases does not approach 0 as $L \to \infty$. In comparison, we show that the reconstruction error of the calibration parameters and the perturbation of the noise-space correlation function in MUSIC converge to 0 as $L \to \infty$ in Theorem 2.
2. We present a sensitivity analysis of the partial algebraic method to both the number of snapshots and noise, while the sensitivity to noise is not addressed in [21].
3. The upper bounds in Theorem 2 are explicitly given in terms of g, $\{\gamma_j\}_{j=1}^s$, N, s, $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$. When the underlying frequencies are separated by $q > 1/N$, we can further bound $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ in terms of q and N by discrete Ingham inequalities [24]. In comparison, all bounds in [21] are implicit in the sense that the bias is defined but not estimated, and the variance is expressed in terms of the trace of certain matrices that are not explicitly given.
4. One needs to perform standard spectral estimation after calibration parameters are recovered. Theorem 2 includes a sensitivity analysis of the MUSIC algorithm, which is not addressed in [21].

4. Optimization approach

As discussed in Section 3.1, it is nontrivial to make use of all entries in the covariance matrix in algebraic methods. Instead we now propose an optimization approach which takes advantage of all measurements.

Suppose $\widehat{R}^y$ is an estimate of R^y. According to Lemma 1, we can recover exact calibration parameters g and the vector f defined in (7), by solving the following optimization problem:

$$\min_{\mathbf{g}, \mathbf{f} \in \mathbb{C}^N} \mathcal{L}(\mathbf{g}, \mathbf{f}) := \left\| \mathrm{diag}(\mathbf{g}) \mathcal{T}(\mathbf{f}) \mathrm{diag}(\bar{\mathbf{g}}) - \widehat{R}^y \right\|_F^2. \tag{28}$$

Here we use boldface letters $\mathbf{g}, \mathbf{f}$ to denote variables in optimization and g, f to denote the ground truth. With infinite snapshots of noiseless measurements, the covariance matrix is exactly known so that $\widehat{R}^y = R^y$, and Lemma 1 implies that the global minimizer of (28) is the ground truth up to a trivial ambiguity. If finite snapshots of noisy measurements are taken, then we run Steps 1–4 in Algorithm 1 to obtain $\widehat{R}^y$ as an approximation to R^y.

As pointed out in Lemma 1, if $(\mathbf{g}, \mathbf{f})$ is a solution to (28), then so is $(c_0 \mathbf{g}, c_0^{-2} \mathbf{f})$ for any $c_0 \neq 0$. In order to guarantee numerical stability, we avoid the case that $\|\mathbf{g}\| \to 0$ and $\|\mathbf{f}\| \to \infty$ (or vice versa) by adding a

penalty to the objective function. Let $n_0 := \|g\|^2 \|f\|$ which can be estimated from $\widehat{R}^y$ based on the following lemma (see Appendix B for the proof):

Lemma 2. *Let R^y be defined in* (9). *Then*

$$n_0 = \left(\sum_{n=0}^{N-1} R^y_{n,n} \right) \sqrt{1 + \frac{1}{N-k} \sum_{k=1}^{N-1} \sum_{n=0}^{N-k-1} \frac{|R^y_{n+k,n}|^2}{R^y_{n+k,n+k} R^y_{n,n}}}. \tag{29}$$

Let $\widehat{n}_0$ be an estimate of n_0 from (30). Theorem 2 shows that $\|\widehat{R}^y - R^y\| \le \Delta R^y$ with ΔR^y given by (21). When the true frequencies are separated by $1/N$, (24) implies that $\Delta R^y \le O(L^{-\frac{1}{2}})$ and $\widehat{n}_0 \approx n_0$ when L is sufficiently large.

Algorithm 2 Optimization approach.

Input: Measurements $\{y_e(t), \ t \in \Gamma\}$ and sparsity s.
Output: Calibration parameters $\widehat{\mathbf{g}}$ and recovered spectrum $\{\widehat{\omega}_j\}_{j=1}^s$.
1: Run Step 1–7 in Algorithm 1 and compute

$$\widehat{n}_0 = \sum_{n=0}^{N-1} \widehat{R}^y_{n,n} \sqrt{1 + \frac{1}{N-k} \sum_{k=1}^{N-1} \sum_{n=0}^{N-k-1} \frac{|\widehat{R}^y_{n+k,n}|^2}{\widehat{R}^y_{n+k,n+k} \widehat{R}^y_{n,n}}}. \tag{30}$$

2: **Initialization:**
   (i) Let $\mathbf{g}^0 \leftarrow \widehat{g}$ and $F^0 \leftarrow \widehat{F}$ where $\widehat{g}$ and $\widehat{F}$ are from Steps 1–7 of Algorithm 1.
   (ii) Let $\mathbf{f}^0 \in \mathbb{C}^N$ such that $\mathbf{f}^0_k = \frac{1}{N-k} \sum \text{diag}(F^0, -k), k = 0, \dots, N-1$.
   (iii) Normalization: $\mathbf{g}^0 \leftarrow \sqrt[4]{\widehat{n}_0} \frac{\mathbf{g}^0}{\|\mathbf{g}^0\|}$ and $\mathbf{f}^0 \leftarrow \sqrt{\widehat{n}_0} \frac{\mathbf{f}^0}{\|\mathbf{f}^0\|}$.
3: **for** $k = 1, 2, \dots$, **do**
4:    $\mathbf{g}^k = \mathbf{g}^{k-1} - \eta^k \nabla_{\mathbf{g}} \tilde{\mathcal{L}}(\mathbf{g}^{k-1}, \mathbf{f}^{k-1})$.
5:    $\mathbf{f}^k = \mathbf{f}^{k-1} - \eta^k \nabla_{\mathbf{f}} \tilde{\mathcal{L}}(\mathbf{g}^{k-1}, \mathbf{f}^{k-1})$.
6: **end for**
7: Output of gradient descent: $\widehat{\mathbf{g}}$ and $\widehat{\mathbf{f}}$.
8: Apply MUSIC to $\mathcal{T}(\widehat{\mathbf{f}})$ to obtain the spectrum $\{\widehat{\omega}_j\}_{j=1}^s$.

Consider the following bounded set:

$$\mathcal{N}_{\widehat{n}_0} = \{(\mathbf{g}, \mathbf{f}) : \|\mathbf{g}\|^2 \le 2\sqrt{\widehat{n}_0}, \|\mathbf{f}\| \le 2\sqrt{\widehat{n}_0}\}. \tag{31}$$

We pick an initial point satisfying

$$(\mathbf{g}^0, \mathbf{f}^0) : \|\mathbf{g}^0\|^2 \le \sqrt{2\widehat{n}_0}, \|\mathbf{f}^0\| \le \sqrt{2\widehat{n}_0}, \tag{32}$$

through the partial algebraic method. The solution from the partial algebraic method has a scaling ambiguity, so we simply normalize it to guarantee (32). In order to ensure all the iterates remain in $\mathcal{N}_{\widehat{n}_0}$, we minimize the following regularized function:

$$\min_{\mathbf{g}, \mathbf{f} \in \mathbb{C}^N} \tilde{\mathcal{L}}(\mathbf{g}, \mathbf{f}) := \mathcal{L}(\mathbf{g}, \mathbf{f}) + \mathcal{G}(\mathbf{g}, \mathbf{f}) \tag{33}$$

where $\mathcal{L}(\mathbf{g}, \mathbf{f})$ is defined in (28) and $\mathcal{G}(\mathbf{g}, \mathbf{f})$ is a penalty function of the form

$$\mathcal{G}(\mathbf{g}, \mathbf{f}) = \rho \left[\mathcal{G}_0 \left(\frac{\|\mathbf{f}\|^2}{2\widehat{n}_0} \right) + \mathcal{G}_0 \left(\frac{\|\mathbf{g}\|^2}{\sqrt{2\widehat{n}_0}} \right) \right]$$

where $\mathcal{G}_0(z) = (\max(z-1, 0))^2$ and $\rho \ge (\sqrt{2}-1)^{-2} \left(3\widehat{n}_0 + \|R^y - \widehat{R}^y\|_F \right)$. When the exact frequencies are separated by $1/N$, we have $\|\widehat{R}^y - R^y\| = O[L^{-\frac{1}{2}}(B_1 + B_2 \max(\sigma, \sigma^2))]$. It follows that $\|\widehat{R}^y - R^y\|_F \le$

$\sqrt{N}\|\widehat{R}^y - R^y\| \leq \sqrt{N}\Delta R^y \to 0$ as $L \to \infty$. We therefore take $\rho \geq 3(\sqrt{2} - 1)^{-2}\widehat{n}_0$ when L is sufficiently large.

The objective function in (33) is continuously differentiable but non-convex. We choose an initial point satisfying (32) by the partial algebraic method and solve (33) by gradient descent where the derivative can be interpreted as a Wirtinger derivative.[1] The Wirtinger gradient of $\tilde{\mathcal{L}}$ is given by

$$\nabla\tilde{\mathcal{L}} = \begin{bmatrix} \nabla_{\mathbf{g}}\tilde{\mathcal{L}} & \nabla_{\mathbf{f}}\tilde{\mathcal{L}} \end{bmatrix}^T = \begin{bmatrix} \nabla_{\mathbf{g}}\mathcal{L} + \nabla_{\mathbf{g}}\mathcal{G} & \nabla_{\mathbf{f}}\mathcal{L} + \nabla_{\mathbf{f}}\mathcal{G} \end{bmatrix}^T$$

with

$$\nabla_{\mathbf{g}}\mathcal{L} = 2\mathrm{diag}\Big[\overline{\mathcal{T}(\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}})\Big(\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}}) - \widehat{R}^y\Big)}\Big], \tag{34}$$

$$\nabla_{\mathbf{f}}\mathcal{L} = \mathcal{T}^a\Big[\overline{\mathrm{diag}(\bar{\mathbf{g}})\Big(\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}}) - \widehat{R}^y\Big)\mathrm{diag}(\mathbf{g})}\Big], \tag{35}$$

$$\nabla_{\mathbf{g}}\mathcal{G} = \frac{\rho}{\sqrt{2\widehat{n}_0}}\mathcal{G}_0'\left(\frac{\|\mathbf{g}\|^2}{\sqrt{2\widehat{n}_0}}\right)\mathbf{g},$$

$$\nabla_{\mathbf{f}}\mathcal{G} = \frac{\rho}{2\widehat{n}_0}\mathcal{G}_0'\left(\frac{\|\mathbf{f}\|^2}{2\widehat{n}_0}\right)\mathbf{f},$$

and $\mathcal{G}_0'(z) = 2\max(z - 1, 0)$. The operator $\mathcal{T}^a : \mathbb{C}^{N \times N} \to \mathbb{C}^N$ is defined as

$$\mathcal{T}^a : \mathbb{C}^{N \times N} \to \mathbb{C}^N : \ \mathcal{T}^a(X) = \begin{bmatrix} \sum\big(\mathrm{diag}(X) + \mathrm{diag}(\bar{X})\big) \\ \sum\big(\mathrm{diag}(X, 1) + \mathrm{diag}(\bar{X}, -1)\big) \\ \vdots \\ \sum\big(\mathrm{diag}(X, N - 1) + \mathrm{diag}(\bar{X}, -(N - 1))\big) \end{bmatrix}.$$

One can verify that $\mathcal{T}^a$ satisfies

$$\frac{\partial}{\partial\mathbf{f}}\Big(\langle\mathcal{T}(\mathbf{f}), X\rangle + \langle X, \mathcal{T}(\mathbf{f})\rangle\Big) = \mathcal{T}^a(X), \ \forall X \in \mathbb{C}^{N \times N},$$

and therefore

$$\frac{\partial\mathcal{L}}{\partial\mathbf{f}} = \frac{\partial}{\partial\mathbf{f}}\Big(\Big\langle\mathcal{T}(\mathbf{f}), \mathrm{diag}(\bar{\mathbf{g}})\Big(\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}}) - \widehat{R}^y\Big)\mathrm{diag}(\mathbf{g})\Big\rangle$$

$$+ \Big\langle\mathrm{diag}(\bar{\mathbf{g}})\Big(\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}}) - \widehat{R}^y\Big)\mathrm{diag}(\mathbf{g}), \mathcal{T}(\mathbf{f})\Big\rangle\Big)$$

$$= \mathcal{T}^a\Big[\mathrm{diag}(\bar{\mathbf{g}})\Big(\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}}) - \widehat{R}^y\Big)\mathrm{diag}(\mathbf{g})\Big]$$

which gives rise to (35).

Our optimization approach for sensor calibration is summarized in Algorithm 2. In the next theorem we prove that the gradient descent in Steps 3–6 of Algorithm 2 converges to a critical point of (33).

---

[1] Let $z = x + iy$ and $h(z) = h(x, y) = u(x, y) + iv(x, y)$. The Wirtinger derivatives and gradient of h are

$$\frac{\partial h}{\partial z} := \frac{1}{2}\left(\frac{\partial h}{\partial x} - i\frac{\partial h}{\partial y}\right), \quad \nabla_z h := \frac{\partial h}{\partial\bar{z}} := \frac{1}{2}\left(\frac{\partial h}{\partial x} + i\frac{\partial h}{\partial y}\right).$$

Theorem 3. *Let (g, f) be the ground truth. Let $\widehat{R}^y$ be an estimate of R^y. Assume that the initial point $(\mathbf{g}^0, \mathbf{f}^0)$ satisfies $\|\mathbf{g}^0\| \leq \sqrt[4]{2\widehat{n}_0}$ and $\|\mathbf{f}^0\| \leq \sqrt{2\widehat{n}_0}$, and $\rho \geq (\sqrt{2}-1)^{-2}(3\widehat{n}_0 + \|R^y - \widehat{R}^y\|_F)$. Then running Algorithm 2 with step size*

$$\eta^k \leq 2/C_{\text{Lip}} \tag{36}$$

where

$$C_{\text{Lip}} \leq 166\widehat{n}_0 \max(\sqrt{\widehat{n}_0}, \sqrt[4]{\widehat{n}_0}) + 8\widehat{n}_0 + 16\max(\sqrt{\widehat{n}_0}, \sqrt[4]{\widehat{n}_0})\|R^y - \widehat{R}^y\|_F + \frac{12\rho}{\min(\widehat{n}_0, \sqrt{\widehat{n}_0})}$$

gives rise to a sequence $(\mathbf{g}^k, \mathbf{f}^k) \in \mathcal{N}_{\widehat{n}_0}$, and

$$\|\nabla\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)\| \to 0, \ as \ k \to \infty.$$

Theorem 3 (see Appendix C for the proof) shows that Wirtinger gradient descent converges to a critical point of (33). Our numerical experiments suggest that this point indeed provides a good approximation of the ground truth up to a trivial ambiguity.

5. Numerical experiments

We perform systematic numerical simulations to compare the performance of existing methods for the sensor calibration problem modeled by (1). In our simulations, $\mathcal{S}$ contains s frequencies located on the continuous domain $[0, 1)$. Theorem 2 shows that the problem is more challenging when the dynamic ranges of x and g increase. We denote the dynamic range of γ and g by $\text{DR}_\gamma > 0$ and $\text{DR}_g > 0$ respectively, and let $\gamma_i = (\mathbb{E}x_i^2(t))^{\frac{1}{2}} \in [1, \text{DR}_\gamma], i = 1, \ldots, s$ and $|g_i| \in [1, \text{DR}_g], i = 0, \ldots, N-1$. The phases of $x_j(t)$ are randomly chosen from $[0, 2\pi)$ to guarantee that $\mathbb{E}x(t)x^*(t) = \text{diag}(\{\gamma_i^2\}_{i=1}^s)$. We add i.i.d. Gaussian noise to the measurements such that $y_e(t) = y(t) + e(t)$ with $e(t) \sim \mathcal{N}(0, \sigma^2 I_N)$.

Suppose we take L snapshots of independent measurements, i.e., $\{y_e(t) : t = 1, \ldots, L\}$, and form the empirical covariance matrix $\widetilde{R}_e^y$. We assume s is known and denote the support of the recovered frequencies by $\widehat{\mathcal{S}} = \{\widehat{\omega}_j\}_{j=1}^s$. Due to the discrete set-up of sensors, we assume periodicity of the frequency domain $[0, 1)$ on which the distance between two frequencies $d(\omega_j, \omega_l)$ is understood as the wrap-around distance on the torus. Frequency support error is measured by the Hausdorff distance between $\mathcal{S}$ and $\widehat{\mathcal{S}}$ up to a translation:

$$\text{SuppError} = d(\mathcal{S}, \widehat{\mathcal{S}}) := \min_{c_2 \in [0, 2\pi)} \max\left(\max_{\widehat{\omega} \in \widehat{\mathcal{S}}} \min_{\omega \in \mathcal{S}} d\left(\widehat{\omega} + \frac{c_2}{2\pi}, \omega\right), \max_{\omega \in \mathcal{S}} \min_{\widehat{\omega} \in \widehat{\mathcal{S}}} d\left(\widehat{\omega} + \frac{c_2}{2\pi}, \omega\right)\right). \tag{37}$$

Let c_2^* be the minimizer in (37). In the noiseless case, we expect the recovered calibration parameters to be of the form $\widehat{g}_n = c_0 e^{ic_1} e^{inc_2^*} g_n$ for some $c_0 > 0$ and $c_1 \in [0, 2\pi)$. Let $\tilde{g}_n = \widehat{g}_n e^{-inc_2^*}$ and $C^* = \text{argmin}_C \sum_{n=0}^{N-1} |\tilde{g}_n - Cg_n|^2$. We measure the relative calibration error for the n-th sensor and the average relative calibration error as

$$\text{CalError}_n = \frac{|\tilde{g}_n - C^* g_n|}{|g_n|}, \quad \text{CalError} = \frac{1}{N}\sum_{n=0}^{N-1} \text{CalError}_n.$$

We test the following methods:

- the partial algebraic method in Algorithm 1;

- the optimization approach in Algorithm 2: In practice we choose the step length η^k according to the backtracking line search in Algorithm 3 [26, Algorithm 3.1]. This backtracking approach ensures that the selected step length η_k is short enough to guarantee a sufficient decrease of $\tilde{\mathcal{L}}$ but not too short. The latter claim holds since the accepted step length η_k is within a factor θ of the previous trial value η^k/θ, which was rejected for violating the sufficient decrease of $\tilde{\mathcal{L}}$, that is, for being too large. In our implementation, we set $\bar{\eta} = \tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)/\|\nabla\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)\|$, $\theta = 0.5, c = 0.5$ and terminate gradient descent while $\eta^k < 10^{-4}$;

Algorithm 3 Backtracking line search.

1: Choose $\bar{\eta} > 0, \theta \in (0,1), c \in (0,1)$; Set $\eta \leftarrow \bar{\eta}$; Denote $\mathbf{z}^k = (\mathbf{g}^k, \mathbf{f}^k)$ and $\mathbf{p}_k = -\nabla\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)$.
2: **repeat**
3:     $\eta \leftarrow \theta\eta$
4: **until** $\mathcal{L}(\mathbf{z}_k + \eta\mathbf{p}_k) \leq \mathcal{L}(\mathbf{z}_k) - c\eta\|\mathbf{p}_k\|^2$.
5: **return** $\eta^k = \eta$.

- an alternating algorithm proposed by Friedlander and Weiss [15]. This algorithm is based on a two-step procedure. First, one assumes that the calibration parameters are known, and estimates the frequencies with the MUSIC algorithm. Given the recovered frequencies, one minimizes the squared sum of the noise-space correlation functions evaluated at the recovered frequencies over all calibration parameters. We choose an initial point using the partial algebraic method and terminate the iterations when the squared sum of the noise-space correlation functions evaluated at the recovered frequencies decreases by 10^{-4} or less.

5.1. Partial algebraic method and optimization approach

We expect the optimization approach to outperform the partial algebraic method in almost all cases since all measurements are used. To illustrate this, we perform reconstructions on 20 frequencies separated by $2/N$. We set $\mathrm{DR}_\gamma = \mathrm{DR}_g = 2$ and $\sigma = 0.5$. We apply both methods to the same set of measurements. In Fig. 1, the imaging functions in the MUSIC algorithm are displayed for the partial algebraic method and the optimization approach, respectively. Both techniques succeed as imaging functions peak around the true frequencies. However, the optimization approach yields peaks that are higher and sharper, and the support error is smaller.

5.2. Sensitivity to the number of snapshots

The performance of all algorithms improves as the number of snapshots L increases. We prove in Theorem 2 that, for the partial algebraic method, when the underlying frequencies are separated by $1/N$ or above, the reconstruction error of calibration parameters decays like $O(1/\sqrt{L})$. In order to verify this result, we perform reconstructions on 20 frequencies separated by $2/N$ when L increases from 30 to 10^4. We set $\mathrm{DR}_\gamma = \mathrm{DR}_g = 2$, and let the noise level be $\sigma = 0, 0.5, 1, 2$. Fig. 2 displays the relative reconstruction error of calibration parameters and the success probability of support recovery in 100 independent experiments versus L in a logarithmic scale. The frequency support is successfully recovered if $d(\mathcal{S}, \widehat{\mathcal{S}}) \leq 0.2/N$. We observe that, (1) the reconstruction errors of calibration parameters for the partial algebraic method and the optimization approach decay like $O(1/\sqrt{L})$ since the slopes in Fig. 2 (a) are roughly -0.5; (2) in terms of stability to the number of snapshots, the alternating algorithm in [15] works the best when $\sigma = 0$, but its performance degrades dramatically when noise exists. In the presence of noise, our optimization approach has the best performance, and the partial algebraic method is the second best performer.
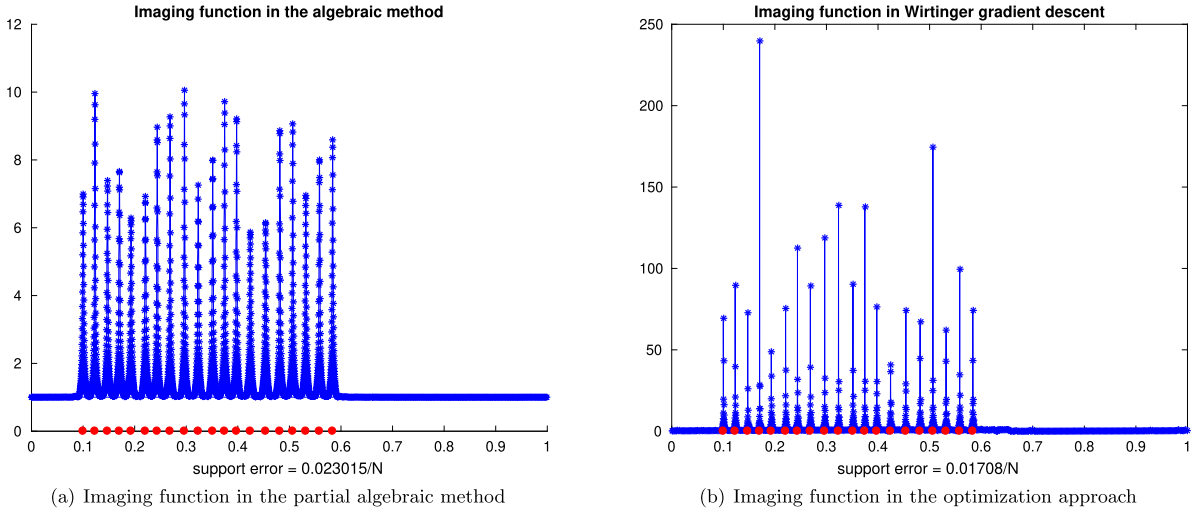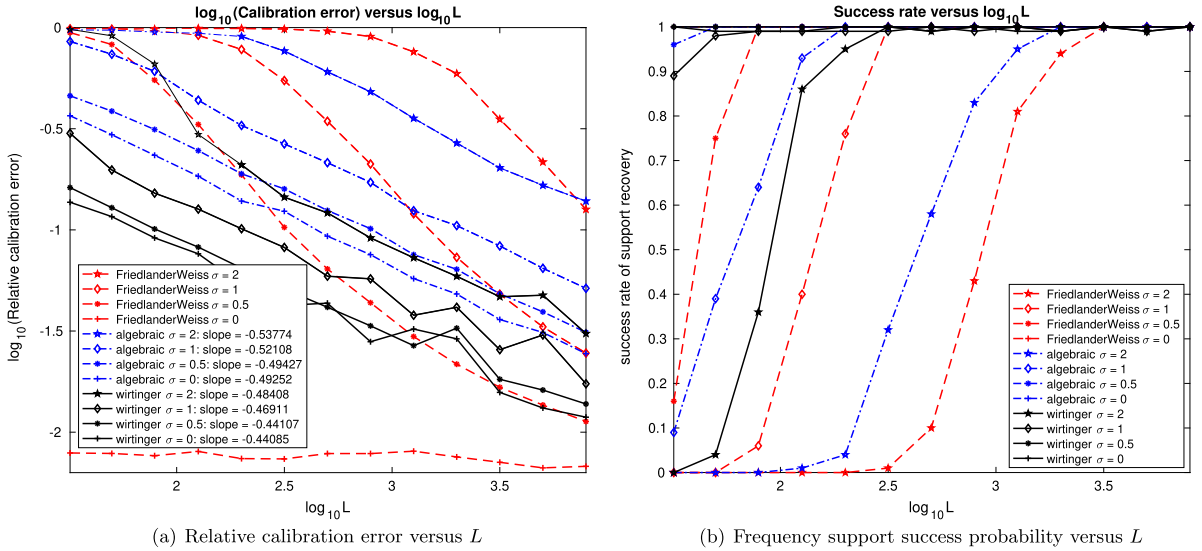
(a) Imaging function in the partial algebraic method

(b) Imaging function in the optimization approach

Fig. 1. Imaging functions (after a proper translation) in the MUSIC algorithm for the partial algebraic method (a) and the optimization approach (b). Red dots represent the locations of true frequencies. The two methods are applied on the same set of measurements generated by 20 frequencies separated by $2/N$, when $DR_\gamma = DR_g = 2$, $L = 500$ and $\sigma = 0.5$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

(a) Relative calibration error versus L

(b) Frequency support success probability versus L

Fig. 2. (a) and (b) show the average relative reconstruction error of calibration parameters and the success probability of support recovery in 100 independent experiments versus L in a logarithmic scale. We choose 20 frequencies separated by $2/N$, $DR_\gamma = DR_g = 2$, and noise level $\sigma = 0, 0.5, 1, 2$.

5.3. Sensitivity to noise

To test the sensitivity of the various approaches to noise, we perform reconstructions on 20 frequencies separated by $2/N$ when σ increases from 10^{-1} to 10. We set $DR_\gamma = DR_g = 2$, and let $L = 500, 1000$ respectively. The frequency support is successfully recovered if $d(\mathcal{S}, \widehat{\mathcal{S}}) \leq 0.2/N$. Fig. 3 displays the average reconstruction error of calibration parameters and the success probability of support recovery in 100 independent experiments versus σ in a logarithmic scale. We observe that, (1) the reconstruction errors of calibration parameters for the partial algebraic method and the optimization approach increase like $O(\sigma)$ when $\log_{10} \sigma$ varies from -0.5 to 0.6 since the slopes in Fig. 3 (a) are roughly 1; (2) in terms of stability to noise, the alternating algorithm in [15] yields the smallest calibration error when σ is small. As the noise
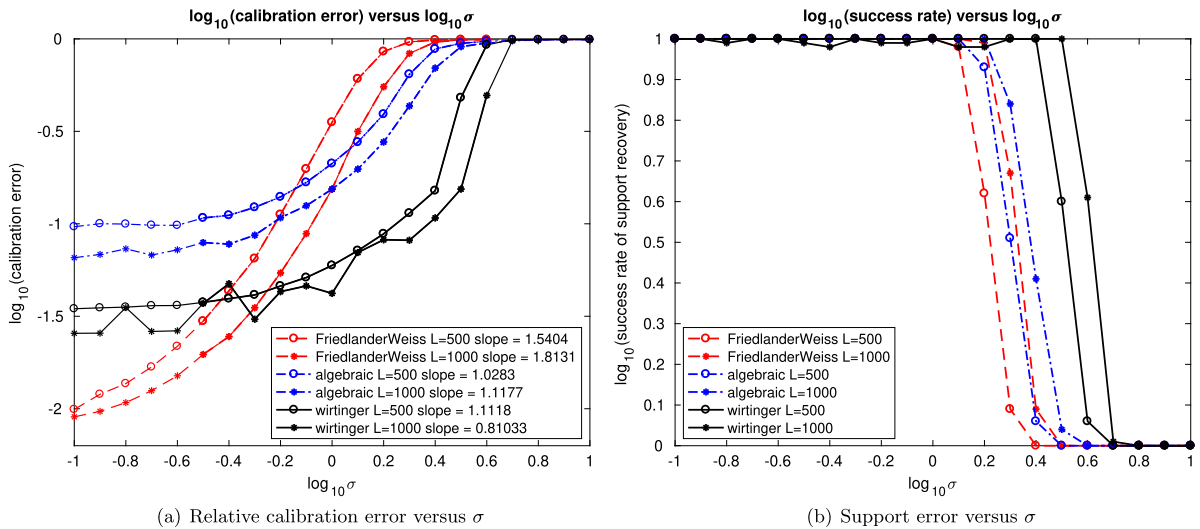
(a) Relative calibration error versus σ (b) Support error versus σ

Fig. 3. Reconstruction errors of the calibration parameters and the support versus σ in $\log_{10}$ scale. We choose 20 frequencies separated by $2/N$, $\mathrm{DR}_\gamma = \mathrm{DR}_g = 2$, $L = 2000, 5000$, and σ varies from 0.1 to 10.

level increases, our optimization approach becomes the best performer, while the partial algebraic method is the second best. Notice that the reconstruction errors do not necessarily approach 0 when σ decreases to 0 due to deviation of the empirical covariance matrix from the true covariance matrix caused by the finite number of snapshots.

6. Conclusion and future research

This paper studies sensor calibration in spectral estimation with multiple snapshots. We assume the true frequencies are located on a continuous domain and each sensor has an unknown calibration parameter. Uniqueness of the calibration parameters and frequencies, up to a trivial ambiguity, is proved with infinite snapshots of noiseless measurements, based on the algebraic methods in [27,42]. A sensitivity analysis of the partial algebraic method [42] with respect to the number of snapshots and noise is presented. While only partial measurements are exploited in the algebraic method, we propose an optimization approach to make full use of the measurements. Superior performance of our optimization approach is demonstrated through numerical comparisons with the partial algebraic method [42] and the alternating algorithm [15].

Several interesting questions are left for future investigations. First, uniqueness in the current paper holds with infinite snapshots of noiseless measurements. It is interesting to study uniqueness with a minimal number of snapshots. Second, global convergence of the Wirtinger gradient descent in our optimization approach is not proved in this paper, even though we have observed its superior numerical performance. The recent work in [20] guarantees global convergence of a non-convex optimization for the sensor calibration problem modeled by (3) where measurements are bilinear. In our problem, the covariance matrix is quadratic in g and linear in f, which makes the local regularity condition [6,20] harder to prove.

Acknowledgments

Appendix A. Sensitivity of the partial algebraic method (Proof of Theorem 2)

The proof of Theorem 2 relies on the following matrix Bernstein inequalities.

Proposition 3 *([37, Theorem 7.3.1]). Consider a finite sequence $\{X_k\}$ of random Hermitian matrices that satisfy*

$$\mathbb{E}X_k = 0 \quad and \quad \|X_k\| \leq R.$$

Define the random matrix $Y = \sum_k X_k$. Suppose $\mathbb{E}Y^2 \preceq V$ for some positive semidefinite matrix V and let the intrinsic dimension of V be $\mathrm{intdim}(V) := trace(V)/\|V\|$. Then for any $t \geq \|V\|^{1/2} + R/3$,

$$\mathbb{P}\{\|Y\| \geq t\} \leq 4 \cdot \mathrm{intdim}(V) \cdot \exp\left(\frac{-t^2/2}{\|V\| + Rt/3}\right)$$

$$\mathbb{E}\|Y\| \leq \sqrt{2\|V\|\log(4 \cdot \mathrm{intdim}(V))} + \frac{1}{3}R\log(4 \cdot \mathrm{intdim}(V)).$$

Proof of Theorem 2. In this proof, we assume $\|g\|$ is known, and $\beta_0 = \beta_{N-1} = 0$ to remove trivial ambiguities; otherwise, our estimate gives an upper bound on $\min_{c_0>0, c_1, c_2 \in \mathbb{R}} \|c_0\widehat{g} - e^{ic_1}\mathrm{diag}(\{e^{inc_2}\}_{n=0}^{N-1})g\|_\infty$.

In the case of finite snapshots, the sampled covariance matrix $\widetilde{R}_e^y$ deviates from R_e^y by:

$$\|R_e^y - \widetilde{R}_e^y\| \leq \|GA(R^x - \widetilde{R}^x)A^*G^* + GA(R^{xe} - \widetilde{R}^{xe}) + (R^{ex} - \widetilde{R}^{ex})A^*G^* + R^e - \widetilde{R}^e\|$$

$$\leq \sigma_{\max}^2(G)\sigma_{\max}^2(A)\|R^x - \widetilde{R}^x\| + \sigma_{\max}(G)\sigma_{\max}(A)\|R^{xe} - \widetilde{R}^{xe}\|$$

$$+ \sigma_{\max}(G)\sigma_{\max}(A)\|R^{ex} - \widetilde{R}^{ex}\| + \|R^e - \widetilde{R}^e\|,$$

where we use the same notations as (8). We will estimate $\|R^x - \widetilde{R}^x\|$ using the matrix Bernstein inequality in Proposition 3. Let $\widetilde{R}_{x,t} = \frac{1}{L}(x(t)x^*(t) - R^x)$ which satisfies $\|\widetilde{R}_{x,t}\| \leq \frac{1}{L}\left(\max_{t \in \Gamma}\|x(t)\|^2 + \gamma_{\max}^2\right)$ for $t \in \Gamma$. Then $\widetilde{R}^x - R^x = \sum_{t \in \Gamma}\widetilde{R}_{x,t}$. We observe that

$$\mathbb{E}(\widetilde{R}^x - R^x)^2 = \sum_{t \in \Gamma}\mathbb{E}\widetilde{R}_{x,t}^2 = \frac{1}{L^2}\sum_{t \in \Gamma}\mathbb{E}(x(t)x^*(t) - R^x)(x(t)x^*(t) - R^x)$$

$$\preceq \frac{1}{L}\left(\max_t\|x(t)\|^2 R^x - (R^x)^2\right) \preceq \frac{\max_t\|x(t)\|^2}{L}R^x,$$

where R^x has the intrinsic dimension $\mathrm{intdim}(R^x) \leq s$. Applying Proposition 3, we obtain that for any $\eta \geq \frac{\gamma_{\max}\max_t\|x(t)\|}{\sqrt{L}} + \frac{\max_{t \in \Gamma}\|x(t)\|^2 + \gamma_{\max}^2}{3L}$, we have

$$\mathbb{P}\{\|R^x - \widetilde{R}^x\| \geq \eta\} \leq 4s \cdot \exp\left(\frac{-\eta^2/2}{\frac{\gamma_{\max}^2\max_t\|x(t)\|^2}{L} + \frac{\gamma_{\max}^2 + \max_{t \in \Gamma}\|x(t)\|^2}{3L}\eta}\right), \tag{38}$$

$$\mathbb{E}\|R^x - \widetilde{R}^x\| \leq \frac{\gamma_{\max}\max_t\|x(t)\|\sqrt{2\log 4s}}{\sqrt{L}} + \frac{\gamma_{\max}^2 + \max_{t \in \Gamma}\|x(t)\|^2}{3L}\log 4s. \tag{39}$$

Similarly, for all $\eta > 0$, we have

$$\mathbb{P}\{\|R^{xe} - \widetilde{R}^{xe}\| \geq \eta\} \leq (N + s) \cdot \exp\left(\frac{-\eta^2/2}{\frac{N\sigma^2\gamma_{\max}^2}{L} + \frac{\max_{t \in \Gamma}\|x(t)\|\|e(t)\|}{3L}\eta}\right), \tag{40}$$

$$\mathbb{P}\{\|R^e - \widetilde{R}^e\| \geq \eta\} \leq 2N \cdot \exp\left(\frac{-\eta^2/2}{\frac{\sigma^2\max_t\|e(t)\|^2}{L} + \frac{\sigma^2 + \max_{t \in \Gamma}\|e(t)\|^2}{3L}\eta}\right), \tag{41}$$

and

$$\mathbb{E}\|R^{xe} - \widetilde{R}^{xe}\| \lesssim \frac{\sigma \gamma_{\max} \sqrt{2N \log(N+s)}}{\sqrt{L}} + \frac{\max_{t \in \Gamma} \|x(t)\| \|e(t)\|}{3L} \log(N+s), \qquad (42)$$

$$\mathbb{E}\|R^e - \widetilde{R}^e\| \lesssim \frac{\sigma \max_t \|e(t)\| \sqrt{2 \log 2N}}{\sqrt{L}} + \frac{\sigma^2 + \max_{t \in \Gamma} \|e(t)\|^2}{3L} \log 2N, \qquad (43)$$

where we apply the matrix Bernstein inequality for the non-Hermitian case [37, Theorem 1.6.2] to estimate $\|R^{xe} - \widetilde{R}^{xe}\|$. Our estimator of R^y is $\widehat{R}^y = \widetilde{R}^y_e - \widehat{\sigma}^2 I_N$, which has the following error

$$\|R^y - \widehat{R}^y\| = \|(R^y_e - \sigma^2 I_N) - (\widetilde{R}^y_e - \widehat{\sigma}^2 I_N)\| \leq \|R^y_e - \widetilde{R}^y_e\| + |\sigma^2 - \widehat{\sigma}^2| \leq 2\|R^y_e - \widetilde{R}^y_e\|, \qquad (44)$$

where the last inequality follows from the Weyl's inequality [41]. Combining (39), (42), (43) and (44) gives rise to $\mathbb{E}\|R^y - \widehat{R}^y\| \leq \Delta R^y$ with ΔR^y defined in (21).

Define the event

$$\mathcal{E} := \left\{ \max(\alpha^2_{\max} \sigma^2_{\max}(A) \|R^x - \widetilde{R}^x\|, \alpha_{\max} \sigma_{\max}(A) \|R^{xe} - \widetilde{R}^{xe}\|, \|R^e - \widetilde{R}^e\|) \leq \frac{\alpha^2_{\min} |f_1|}{16} \right\}$$

under which we have

$$\|R^y - \widehat{R}^y\| \leq \frac{1}{2} \alpha^2_{\min} |f_1| \leq \frac{1}{2} \alpha^2_{\min} f_0.$$

This implies

$$\text{trace}(\widehat{R}^y) \geq \text{trace} R^y - N\|R^y - \widehat{R}^y\| = \|g\|^2 f_0 - \frac{1}{2} N \alpha^2_{\min} f_0 \geq \frac{1}{2} \|g\|^2 f_0 = \frac{1}{2} \text{trace}(R^y),$$

and $\text{trace}(\widehat{R}^y) \leq 3/2 \text{trace}(R^y)$. We first perform all estimates under the event $\mathcal{E}$ and consider $\mathcal{E}^c$ later.

Condition on $\mathcal{E}$

In the partial algebraic method, if $\|g\|$ is known, then the calibration amplitudes α can be recovered without any scaling ambiguity. The exact and recovered calibration phases are:

$$\alpha^2_n = \frac{R^y_{n,n}}{\text{trace}(R^y)} \|g\|^2 \qquad \widehat{\alpha}^2_n = \frac{\widehat{R}^y_{n,n}}{\text{trace}(\widehat{R}^y)} \|g\|^2. \qquad (45)$$

Hence

$$|\alpha^2_n - \widehat{\alpha}^2_n| \leq \left| \frac{R^y_{n,n}}{\text{trace}(R^y)} - \frac{\widehat{R}^y_{n,n}}{\text{trace}(\widehat{R}^y)} \right| \|g\|^2 = \left| \frac{R^y_{n,n} \text{trace}(\widehat{R}^y) - \widehat{R}^y_{n,n} \text{trace}(R^y)}{\text{trace}(R^y) \text{trace}(\widehat{R}^y)} \right| \|g\|^2$$

$$\leq \frac{R^y_{n,n} |\text{trace}(\widehat{R}^y) - \text{trace}(R^y)| + \text{trace}(R^y) |\widehat{R}^y_{n,n} - R^y_{n,n}|}{\text{trace}(R^y) \text{trace}(\widehat{R}^y)} \|g\|^2$$

$$\leq \frac{|g_n|^2 f_0 N + \|g\|^2 f_0}{\|g\|^2 f_0 \|g\|^2 f_0 / 2} \|g\|^2 \|R^y - \widehat{R}^y\| = 2 \frac{N |g_n|^2 + \|g\|^2}{\|g\|^2 f_0} \|R^y - \widehat{R}^y\|.$$

On the other hand,

$$\alpha^2_n = \frac{|g_n|^2 f_0}{\text{trace} R^y} \|g\|^2 \qquad \widehat{\alpha}^2_n \geq \frac{R^y_{n,n} - \|R^y - \widehat{R}^y\|}{\frac{3}{2} \text{trace} R^y} \|g\|^2 \geq \frac{|g_n|^2 f_0 - \frac{1}{2} |g_n|^2 f_0}{\frac{3}{2} \text{trace} R^y} \|g\|^2 \geq \frac{\alpha^2_n}{3},$$

in the event $\mathcal{E}$ and then

$$|\alpha_n - \widehat{\alpha}_n| = \frac{|\alpha_n^2 - \widehat{\alpha}_n^2|}{\alpha_n + \widehat{\alpha}_n} \leq \frac{1}{4\alpha_n/3} \cdot 2\frac{\|g\|^2 + N|g_n|^2}{\|g\|^2 f_0} \cdot \|R^y - \widehat{R}^y\|,$$

$$\|\alpha - \widehat{\alpha}\|_\infty \leq \frac{3}{2\alpha_{\min}} \cdot \frac{\|g\|^2 + N\alpha_{\max}^2}{\|g\|^2 f_0} \cdot \|R^y - \widehat{R}^y\|.$$

Next we estimate $\|\beta - \widehat{\beta}\|_\infty$. To remove trivial ambiguities, we assume the exact calibration phases β satisfy

$$\Phi\beta = b \quad \text{where } b_0 = b_{N-1} = 0, \ b_n = \angle \frac{R_{n+1,n}^y}{R_{n,n-1}^y}, n = 1, \ldots, N-2.$$

Our recovered calibration phases $\widehat{\beta}$ are:

$$\Phi\widehat{\beta} = \widehat{b} \quad \text{where } \widehat{b}_0 = \widehat{b}_{N-1} = 0, \ \widehat{b}_n = \angle \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y}, n = 1, \ldots, N-2.$$

Recall that $R_{n,n-1}^y = \alpha_n \alpha_{n-1} e^{i(\beta_n - \beta_{n-1})} f_1$, so $\alpha_{\min}^2 |f_1| \leq |R_{n,n-1}^y| \leq \alpha_{\max}^2 |f_1|$. In the event $\mathcal{E}$, we have $\alpha_{\min}^2 |f_1|/2 \leq |\widehat{R}_{n,n-1}^y| \leq 3\alpha_{\max}^2 |f_1|/2$, and

$$\left| \frac{R_{n+1,n}^y}{R_{n,n-1}^y} - \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y} \right| = \frac{|R_{n+1,n}^y \widehat{R}_{n,n-1}^y - R_{n,n-1}^y \widehat{R}_{n+1,n}^y|}{|R_{n,n-1}^y \widehat{R}_{n,n-1}^y|} \leq 4\frac{\alpha_{\max}^2}{\alpha_{\min}^4 |f_1|} \|R^y - \widehat{R}^y\|.$$

For any $z, \widehat{z} \in \mathbb{C}$, by a simple geometric argument, we have

$$|(\angle z - \angle \widehat{z}) \mod 2\pi| \leq \frac{4|z - \widehat{z}|}{\min(|z|, |\widehat{z}|)} \tag{46}$$

whenever $|z - \widehat{z}| \leq \min(|z|, |\widehat{z}|)$. Whenever $\|R^y - \widehat{R}^y\| \leq \frac{\alpha_{\min}^6 |f_1|}{12\alpha_{\max}^4}$ (This is guaranteed for sufficiently large L),

$$\left| \frac{R_{n+1,n}^y}{R_{n,n-1}^y} - \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y} \right| \leq \frac{\alpha_{\min}^2}{3\alpha_{\max}^2} \leq \min\left(\left| \frac{R_{n+1,n}^y}{R_{n,n-1}^y} \right|, \left| \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y} \right| \right).$$

Hence

$$\|b - \widehat{b}\|_\infty = \max_n \left| \left(\angle \frac{R_{n+1,n}^y}{R_{n,n-1}^y} - \angle \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y} \right) \mod 2\pi \right|$$

$$\leq \max_n \frac{4\left| \frac{R_{n+1,n}^y}{R_{n,n-1}^y} - \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y} \right|}{\min\left(\left| \frac{R_{n+1,n}^y}{R_{n,n-1}^y} \right|, \left| \frac{\widehat{R}_{n+1,n}^y}{\widehat{R}_{n,n-1}^y} \right| \right)} \leq \frac{48\alpha_{\max}^4}{\alpha_{\min}^4} \cdot \frac{\|R^y - \widehat{R}^y\|}{\alpha_{\min}^2 |f_1|}.$$

The infinity norm of the matrix Φ^{-1} is upper bounded by (see [18, Chapter 2]):

$$\|\Phi^{-1}\|_\infty = \max_j \sum_{i=0}^{N-1} |\Phi_{i,j}^{-1}| \leq 3N^2.$$

Therefore

$$\|\beta - \widehat{\beta}\|_\infty \leq \|\Phi^{-1}\|_\infty \|b - \widehat{b}\|_\infty \leq 144N^2 \frac{\alpha_{\max}^4}{\alpha_{\min}^4} \cdot \frac{\|R^y - \widehat{R}^y\|}{\alpha_{\min}^2 |f_1|}.$$

Combining the estimates of $\|\alpha - \widehat{\alpha}\|_\infty$ and $\|\beta - \widehat{\beta}\|_\infty$ gives rise to

$$\|g - \widehat{g}\|_\infty = \max_n |g_n - \widehat{g}_n| \leq \max_n |\alpha_n||e^{i\beta_n} - e^{i\widehat{\beta}_n}| + |\alpha_n - \widehat{\alpha}_n| \leq \|\alpha - \widehat{\alpha}\|_\infty + \alpha_{\max}\|\beta - \widehat{\beta}\|_\infty.$$

As for the input matrix for the MUSIC algorithm, we have

$$F = \mathrm{diag}(g)^{-1} R^y \mathrm{diag}(\bar{g})^{-1} \qquad \widehat{F} = \mathrm{diag}(\widehat{g})^{-1} \widehat{R}^y \mathrm{diag}(\bar{\widehat{g}})^{-1}.$$

Then

$$\|F - \widehat{F}\| \leq \frac{1}{\widehat{\alpha}_{\min}^2} \|R^y - \widehat{R}^y\| + \frac{\|R^y\|}{\alpha_{\min}} \max_n \left| \frac{1}{g_n} - \frac{1}{\widehat{g}_n} \right| + \frac{\|R^y\|}{\widehat{\alpha}_{\min}} \max_n \left| \frac{1}{g_n} - \frac{1}{\widehat{g}_n} \right|$$

$$\leq 9 \frac{\|R^y - \widehat{R}^y\|}{\alpha_{\min}^2} + 12 \frac{\|R^y\|}{\alpha_{\min}^3} \|g - \widehat{g}\|_\infty \leq 9 \frac{\|R^y - \widehat{R}^y\|}{\alpha_{\min}^2} + 12 \frac{\alpha_{\max}^2 \gamma_{\max}^2 \sigma_{\max}^2(A)}{\alpha_{\min}^3} \|g - \widehat{g}\|_\infty.$$

When the input of MUSIC is $\widehat{F}$, Proposition 2 provides an estimate on the perturbation of the noise-space correlation function:

$$|\widehat{\mathcal{R}}(\omega) - \mathcal{R}(\omega)| \leq \frac{2}{\gamma_{\min}^2 \sigma_{\min}^2(A)} \cdot \|F - \widehat{F}\|$$

as long as $2\|F - \widehat{F}\| < \gamma_{\min}^2 \sigma_{\min}^2(A)$.

Conditioning on the event $\mathcal{E}$, we have

$$\mathbb{E}(\|g - \widehat{g}\|_\infty | \mathcal{E}) \leq \mathbb{E}(\|\alpha - \widehat{\alpha}\|_\infty | \mathcal{E}) + \alpha_{\max} \mathbb{E}(\|\beta - \widehat{\beta}\|_\infty | \mathcal{E})$$

$$\leq \frac{3(\|g\|^2 + N\alpha_{\max}^2)}{2\alpha_{\min}\|g\|^2 f_0} \Delta R^y + 144N^2 \frac{\alpha_{\max}^5}{\alpha_{\min}^6 |f_1|} \Delta R^y$$

and

$$\mathbb{E}(\|F - \widehat{F}\| | \mathcal{E}) \leq 9 \frac{\Delta R^y}{\alpha_{\min}^2} + \frac{12\alpha_{\max}^2 \gamma_{\max}^2 \sigma_{\max}^2(A)}{\alpha_{\min}^3} \left(\frac{3(\|g\|^2 + N\alpha_{\max}^2)}{2\alpha_{\min}\|g\|^2 f_0} + 144N^2 \frac{\alpha_{\max}^5}{\alpha_{\min}^6 |f_1|} \right) \Delta R^y.$$

Condition on $\mathcal{E}^c$

Finally we consider the event $\mathcal{E}^c$ which occurs with small probability when L is sufficiently large:

$$\mathbb{P}\{\mathcal{E}^c\} \leq \mathbb{P}\left\{ \|R^x - \widetilde{R}^x\| \geq \frac{\alpha_{\min}^2 |f_1|}{16\alpha_{\max}^2 \sigma_{\max}^2(A)} \right\} + \mathbb{P}\left\{ \|R^{xe} - \widetilde{R}^{xe}\| \geq \frac{\alpha_{\min}^2 |f_1|}{16\alpha_{\max} \sigma_{\max}(A)} \right\}$$

$$+ \mathbb{P}\left\{ \|R^e - \widetilde{R}^e\| \geq \frac{\alpha_{\min}^2 |f_1|}{16} \right\}$$

$$\leq 4N e^{-LC(\alpha_{\max}, \alpha_{\min}, \gamma_{\max}, \sigma_{\max}(A), |f_1|, \sigma, \max_t \|x(t)\|, \max_t \|e(t)\|)}$$

for some positive constant $C(\alpha_{\max}, \alpha_{\min}, \gamma_{\max}, \sigma_{\max}(A), |f_1|, \sigma, \max_t \|x(t)\|, \max_t \|e(t)\|)$. In any case, $\|g - \widehat{g}\|_\infty \leq \|g\|_\infty + \|\widehat{g}\|_\infty \leq \|g\|_\infty + \|g\| \leq 2\|g\|$, where $\|\widehat{g}\|_\infty \leq \|g\|$ due to (45). Therefore,

$$\mathbb{E}\|g - \widehat{g}\|_\infty \leq \mathbb{E}(\|g - \widehat{g}\|_\infty | \mathcal{E})\mathbb{P}\{\mathcal{E}\} + 2\|g\|\mathbb{P}\{\mathcal{E}^c\}$$

$$\leq \frac{3(\|g\|^2 + N\alpha_{\max}^2)}{2\alpha_{\min}\|g\|^2 f_0}\Delta R^y + 144N^2 \frac{\alpha_{\max}^5}{\alpha_{\min}^6 |f_1|^2}\Delta R^y + 8N\|g\|_2 e^{-LC}. \tag{47}$$

Since the first two terms in (47) is $O(1/L)$ and the last term is $O(e^{-CL})$, we can guarantee (22) when L is sufficiently large. A similar estimate holds for $\|F - \widehat{F}\|$. $\square$

Appendix B. Proof of Lemma 2

According to (12), we have

$$|g_n|^2 = \frac{R_{n,n}^y}{f_0}, \ n = 0, \dots, N - 1,$$

and for any $k = 1, \dots, N - 1$

$$|f_k|^2 = \frac{|R_{n+k,n}^y|^2}{|g_{n+k}|^2 |g_n|^2} = \frac{f_0^2 |R_{n+k,n}^y|^2}{R_{n+k,n+k}^y R_{n,n}^y} \quad \text{for any } 0 \leq n \leq N - k - 1.$$

Therefore

$$\|g\|^2 = \frac{\sum_{n=0}^{N-1} R_{n,n}^y}{f_0}$$

and

$$\|f\| = \sqrt{\sum_{k=0}^{N-1} |f_k|^2} = \sqrt{f_0^2 + \sum_{k=1}^{N-1} \frac{1}{N-k} \sum_{n=0}^{N-k-1} \frac{f_0^2 |R_{n+k,n}^y|^2}{R_{n+k,n+k}^y R_{n,n}^y}}$$

$$= f_0 \sqrt{1 + \frac{1}{N-k} \sum_{k=1}^{N-1} \sum_{n=0}^{N-k-1} \frac{|R_{n+k,n}^y|^2}{R_{n+k,n+k}^y R_{n,n}^y}}$$

which gives rise to Lemma 2.

Appendix C. Proof of Theorem 3

We first show that $\nabla \tilde{\mathcal{L}}$, restricted within $\mathcal{N}_{\widehat{n}_0}$, is a Lipchitz function. Notice that g are the exact calibration parameters and f is defined in (7).

Lemma 3. *For any $z := (\mathbf{g}; \mathbf{f})$ and $\Delta \mathbf{z} := (\Delta \mathbf{g}; \Delta \mathbf{f})$ such that $\mathbf{z}, \mathbf{z} + \Delta \mathbf{z} \in \mathcal{N}_{\widehat{n}_0}$, $\nabla \tilde{\mathcal{L}}$ is Lipchitz such that*

$$\|\nabla \tilde{\mathcal{L}}(\mathbf{z} + \Delta \mathbf{z}) - \nabla \tilde{\mathcal{L}}(\mathbf{z})\| \leq C_{\text{Lip}}\|\Delta \mathbf{z}\|$$

with

$$C_{\text{Lip}} \leq 146\widehat{n}_0 \max(\sqrt{\widehat{n}_0}, \sqrt[4]{\widehat{n}_0}) + 8\widehat{n}_0 + 16\max(\sqrt{\widehat{n}_0}, \sqrt[4]{\widehat{n}_0})\|R^y - \widehat{R}^y\|_F + \frac{8\rho}{\min(\widehat{n}_0, \sqrt{\widehat{n}_0})}$$

where $\rho \geq \frac{3\widehat{n}_0 + \|R^y - \widehat{R}^y\|_F}{(\sqrt{2}-1)^2}$.

Proof of Lemma 3. The Wirtinger gradient of $\tilde{\mathcal{L}}$ is

$$\nabla\tilde{\mathcal{L}} = (\nabla_{\mathbf{g}}\tilde{\mathcal{L}};\ \nabla_{\mathbf{f}}\tilde{\mathcal{L}}) = (\nabla_{\mathbf{g}}\mathcal{L} + \nabla_{\mathbf{g}}\mathcal{G};\ \nabla_{\mathbf{f}}\mathcal{L} + \nabla_{\mathbf{f}}\mathcal{G}). \tag{48}$$

Part 1: We estimate $\|\nabla_{\mathbf{g}}\mathcal{L}(\mathbf{z} + \mathbf{w}) - \nabla_{\mathbf{g}}\mathcal{L}(\mathbf{z})\|$. Recall that $R^y = \mathbb{E}y(t)y(t)^*$, and

$$\nabla_{\mathbf{g}}\mathcal{L}(z) = 2\mathrm{diag}\Big[\overline{\mathcal{T}(\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}})\Big(\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}}) - \mathrm{diag}(g)\mathcal{T}(f)\mathrm{diag}(\bar{g}) + R^y - \widehat{R}^y\Big)}\Big].$$

Notice that for any $\mathbf{f}, \mathbf{g}, \mathbf{g}_1, \mathbf{g}_2, \mathbf{h} \in \mathbb{C}^N$ and $X \in \mathbb{C}^{N \times N}$, we have

$$\|\mathrm{diag}[\mathcal{T}(\mathbf{f})^*\mathrm{diag}(\mathbf{g}_1)\mathcal{T}(\mathbf{h})\mathrm{diag}(\mathbf{g}_2)]\| \le \|\mathbf{f}\|\|\mathbf{g}_1\|\|\mathbf{h}\|\|\mathbf{g}_2\|$$
$$\|\mathrm{diag}\left[\mathcal{T}(\mathbf{f})^*\mathrm{diag}(\mathbf{g})X\right]\| \le \sqrt{2}\|\mathbf{f}\|\|\mathbf{g}\|\|X\|_F.$$

For any $\mathbf{z}, \mathbf{z} + \Delta\mathbf{z} \in \mathcal{N}_{\widehat{n}_0}$, we have

$$\|\nabla_{\mathbf{g}}\mathcal{L}(\mathbf{z} + \Delta\mathbf{z}) - \nabla_{\mathbf{g}}\mathcal{L}(\mathbf{z})\|$$
$$\le 2\Big\|\mathrm{diag}\Big[\mathcal{T}(\mathbf{f} + \Delta\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}} + \overline{\Delta\mathbf{g}})\mathrm{diag}(\mathbf{g} + \Delta\mathbf{g})\mathcal{T}(\mathbf{f} + \Delta\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}} + \overline{\Delta\mathbf{g}})$$
$$\quad - \mathcal{T}(\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}})\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}})\Big]\Big\|$$
$$\quad + 2\Big\|\mathrm{diag}\Big[\mathcal{T}(\mathbf{f} + \Delta\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}} + \overline{\Delta\mathbf{g}})\mathrm{diag}(g)\mathcal{T}(f)\mathrm{diag}(\bar{g}) - \mathcal{T}(\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}})\mathrm{diag}(g)\mathcal{T}(f)\mathrm{diag}(\bar{g})\Big]\Big\|$$
$$\quad + 2\Big\|\mathrm{diag}\Big[\mathcal{T}(\mathbf{f} + \Delta\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}} + \overline{\Delta\mathbf{g}})(R^y - \widehat{R}^y) - \mathcal{T}(\mathbf{f})^*\mathrm{diag}(\bar{\mathbf{g}})(R^y - \widehat{R}^y)\Big]\Big\|$$
$$\le 2\Big(\|\Delta\mathbf{f}\|\|\mathbf{g} + \Delta\mathbf{g}\|^3\|\mathbf{f} + \Delta\mathbf{f}\| + \|\mathbf{f}\|\|\Delta\mathbf{g}\|\|\mathbf{g} + \Delta\mathbf{g}\|^2\|\mathbf{f} + \Delta\mathbf{f}\| + \|\mathbf{f}\|\|\mathbf{g}\|\|\Delta\mathbf{g}\|\|\mathbf{g} + \Delta\mathbf{g}\|\|\mathbf{f} + \Delta\mathbf{f}\|$$
$$\quad + \|\mathbf{f}\|\|\mathbf{g}\|^2\|\Delta\mathbf{f}\|\|\mathbf{g} + \Delta\mathbf{g}\| + \|\mathbf{f}\|^2\|\mathbf{g}\|^2\|\Delta\mathbf{g}\| + \|\Delta\mathbf{f}\|\|\mathbf{g} + \Delta\mathbf{g}\|\|g\|^2\|f\| + \|\mathbf{f}\|\|\Delta\mathbf{g}\|\|g\|^2\|f\|\Big)$$
$$\quad + 2\sqrt{2}\Big(\|\Delta\mathbf{f}\|\mathbf{g} + \Delta\mathbf{g}\|\|R^y - \widehat{R}^y\|_F + \|\mathbf{f}\|\|\Delta\mathbf{g}\|\|R^y - \widehat{R}^y\|_F\Big)$$
$$\le 64\widehat{n}_0\sqrt{\widehat{n}_0}\|\Delta\mathbf{g}\| + 24\sqrt{2}\widehat{n}_0\widehat{n}_0^{\frac{1}{4}}\|\Delta\mathbf{f}\| + 4\widehat{n}_0^{\frac{1}{4}}\|R^y - \widehat{R}^y\|_F\|\Delta\mathbf{f}\| + 4\sqrt{2\widehat{n}_0}\|\Delta\mathbf{g}\|\|R^y - \widehat{R}^y\|_F. \tag{49}$$

Part 2: We estimate $\|\nabla_{\mathbf{f}}\mathcal{L}(\mathbf{z} + \Delta\mathbf{z}) - \nabla_{\mathbf{f}}\mathcal{L}(\mathbf{z})\|$. Notice that for any $\mathbf{g}_1, \mathbf{g}_2, \mathbf{f} \in \mathbb{C}^N$ and $X \in \mathbb{C}^{N \times N}$, we have

$$\|\mathcal{T}^a[\mathrm{diag}(\mathbf{g}_1)\mathcal{T}(\mathbf{f})\mathrm{diag}(\mathbf{g}_2)]\| \le 2\|\mathbf{g}_1\|\|\mathbf{f}\|\|\mathbf{g}_2\|,$$
$$\|\mathcal{T}^a[\mathrm{diag}(\mathbf{g}_1)X\mathrm{diag}(\mathbf{g}_2)]\| \le 2\|\mathbf{g}_1\|\|\mathbf{g}_2\|\|X\|_F.$$

By using triangle inequalities, we obtain

$$\|\nabla_{\mathbf{f}}\mathcal{L}(\mathbf{z} + \Delta\mathbf{z}) - \nabla_{\mathbf{f}}\mathcal{L}(\mathbf{z})\|$$
$$\le \Big\|\mathcal{T}^a\Big[\mathrm{diag}(\overline{\mathbf{g} + \Delta\mathbf{g}})\mathrm{diag}(\mathbf{g} + \Delta\mathbf{g})\mathcal{T}(\mathbf{f} + \Delta\mathbf{f})\mathrm{diag}(\overline{\mathbf{g} + \Delta\mathbf{g}})\mathrm{diag}(\mathbf{g} + \Delta\mathbf{g})$$
$$\quad - \mathrm{diag}(\bar{\mathbf{g}})\mathrm{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\mathrm{diag}(\bar{\mathbf{g}})\mathrm{diag}(\mathbf{g})\Big]\Big\|$$
$$\quad + \Big\|\mathcal{T}^a\Big[\mathrm{diag}(\overline{\mathbf{g} + \Delta\mathbf{g}})\mathrm{diag}(g)\mathcal{T}(f)\mathrm{diag}(\bar{g})\mathrm{diag}(\mathbf{g} + \Delta\mathbf{g}) - \mathrm{diag}(\bar{\mathbf{g}})\mathrm{diag}(g)\mathcal{T}(f)\mathrm{diag}(\bar{g})\mathrm{diag}(\mathbf{g})\Big]\Big\|$$
$$\quad + \Big\|\mathcal{T}^a\Big[\mathrm{diag}(\overline{\mathbf{g} + \Delta\mathbf{g}})(R^y - \widehat{R}^y)\mathrm{diag}(\mathbf{g} + \Delta\mathbf{g}) - \mathrm{diag}(\bar{\mathbf{g}})(R^y - \widehat{R}^y)\mathrm{diag}(\mathbf{g})\Big]\Big\|$$
$$\le 2\Big\|\|\mathbf{g} + \Delta\mathbf{g}\|^2 - \|\mathbf{g}\|^2\Big\| \cdot \|\mathbf{f} + \Delta\mathbf{f}\| \cdot \Big\|\|\mathbf{g} + \Delta\mathbf{g}\|^2\Big\| + 2\Big\|\|\mathbf{g}\|^2\Big\| \cdot \|\Delta\mathbf{f}\| \cdot \Big\|\|\mathbf{g} + \Delta\mathbf{g}\|^2\Big\|$$

$$+ 2\left|\|\mathbf{g}|^2\| \right| \cdot \|\mathbf{f}\| \cdot \left| \|\mathbf{g} + \Delta\mathbf{g}|^2 - |\mathbf{g}|^2 \right| + 2\|\Delta\mathbf{g}\| \cdot \|g\|^2 \cdot \|f\| \cdot \|\mathbf{g} + \Delta\mathbf{g}\| + 2\|\mathbf{g}\| \cdot \|g\|^2 \cdot \|f\| \cdot \|\Delta\mathbf{g}\|$$

$$+ 2\|\Delta\mathbf{g}\| \cdot \|R^y - \widehat{R}^y\|_F \cdot \|\mathbf{g} + \Delta\mathbf{g}\| + 2\|\mathbf{g}\| \cdot \|R^y - \widehat{R}^y\|_F \cdot \|\Delta\mathbf{g}\|.$$

Whenever $\mathbf{z}, \mathbf{z} + \Delta\mathbf{z} \in \mathcal{N}_{\widehat{n}_0}$, we have

$$\|\Delta\mathbf{g}\| \le 2\sqrt{2}\widehat{n}_0^{\frac{1}{4}}, \quad \left| \|\mathbf{g} + \Delta\mathbf{g}|^2 - |\mathbf{g}|^2 \right| \le 2\sqrt{2}\widehat{n}_0^{\frac{1}{4}}\|\Delta\mathbf{g}\|$$

and therefore

$$\|\nabla_{\mathbf{f}}\mathcal{L}(\mathbf{z} + \Delta\mathbf{z}) - \nabla_{\mathbf{f}}\mathcal{L}(\mathbf{z})\| \le 48\sqrt{2}\widehat{n}_0\widehat{n}_0^{\frac{1}{4}}\|\Delta\mathbf{g}\| + 8\widehat{n}_0\|\Delta\mathbf{f}\| + 4\sqrt{2}\widehat{n}_0^{\frac{1}{4}}\|R^y - \widehat{R}^y\|_F\|\Delta\mathbf{g}\|. \tag{50}$$

Part 3: We estimate $\|\nabla_{\mathbf{f}}\mathcal{G}(\mathbf{z}+\Delta\mathbf{z}) - \nabla_{\mathbf{f}}\mathcal{G}(\mathbf{z})\|$ and $\|\nabla_{\mathbf{g}}\mathcal{G}(\mathbf{z}+\Delta\mathbf{z}) - \nabla_{\mathbf{g}}\mathcal{G}(\mathbf{z})\|$. Notice that $\mathcal{G}_0'(z) = 2\max(z - 1, 0)$ and hence

$$|\mathcal{G}_0'(z_1) - \mathcal{G}_0'(z_2)| \le 2|z_1 - z_2|, \quad \mathcal{G}_0'(z) \le 2|z|, \quad \forall z_1, z_2, z \in \mathbb{R}.$$

For any $\mathbf{z}, \mathbf{z} + \Delta\mathbf{z} \in \mathcal{N}_{\widehat{n}_0}$, we have

$$\|\nabla_{\mathbf{f}}\mathcal{G}(\mathbf{z} + \Delta\mathbf{z}) - \nabla_{\mathbf{f}}\mathcal{G}(\mathbf{z})\| = \frac{\rho}{2\widehat{n}_0} \left\| \mathcal{G}_0'\left(\frac{\|\mathbf{f} + \Delta\mathbf{f}\|^2}{2\widehat{n}_0}\right)(\mathbf{f} + \Delta\mathbf{f}) - \mathcal{G}_0'\left(\frac{\|\mathbf{f}\|^2}{2\widehat{n}_0}\right)\mathbf{f} \right\|$$

$$\le \frac{\rho}{2\widehat{n}_0} \left| \mathcal{G}_0'\left(\frac{\|\mathbf{f} + \Delta\mathbf{f}\|^2}{2\widehat{n}_0}\right) - \mathcal{G}_0'\left(\frac{\|\mathbf{f}\|^2}{2\widehat{n}_0}\right) \right| \|\mathbf{f} + \Delta\mathbf{f}\| + \frac{\rho}{2\widehat{n}_0}\mathcal{G}_0'\left(\frac{\|\mathbf{f}\|^2}{2\widehat{n}_0}\right)\|\Delta\mathbf{f}\|$$

$$\le \frac{\rho}{2\widehat{n}_0} \frac{2(\|\mathbf{f} + \Delta\mathbf{f}\| + \|\mathbf{f}\|)(\|\mathbf{f} + \Delta\mathbf{f}\| - \|\mathbf{f}\|)}{2\widehat{n}_0}\|\mathbf{f} + \Delta\mathbf{f}\| + \frac{\rho}{2\widehat{n}_0} \cdot \frac{2\|\mathbf{f}\|^2}{2\widehat{n}_0}\|\Delta\mathbf{f}\| \le \frac{6\rho}{\widehat{n}_0}\|\Delta\mathbf{f}\|, \tag{51}$$

and

$$\|\nabla_{\mathbf{g}}\mathcal{G}(\mathbf{z} + \Delta\mathbf{z}) - \nabla_{\mathbf{g}}\mathcal{G}(\mathbf{z})\| \le \frac{6\rho}{\sqrt{\widehat{n}_0}}\|\Delta\mathbf{g}\|. \tag{52}$$

Combining (48), (49), (50), (51), (52) gives rise to

$$\|\nabla\tilde{\mathcal{L}}(\mathbf{z} + \Delta\mathbf{z}) - \nabla\tilde{\mathcal{L}}(\mathbf{z})\|$$

$$\le \left(64\widehat{n}_0\sqrt{\widehat{n}_0} + 48\sqrt{2}\widehat{n}_0\sqrt[4]{\widehat{n}_0} + 4\sqrt{2\widehat{n}_0}\|R^y - \widehat{R}^y\|_F + 4\sqrt{2}\sqrt[4]{\widehat{n}_0}\|R^y - \widehat{R}^y\|_F + \frac{6\rho}{\sqrt{\widehat{n}_0}} \right) \|\Delta\mathbf{g}\|$$

$$+ \left(24\sqrt{2}\widehat{n}_0\sqrt[4]{\widehat{n}_0} + 8\widehat{n}_0 + 4\sqrt[4]{\widehat{n}_0}\|R^y - \widehat{R}^y\|_F + \frac{6\rho}{\widehat{n}_0} \right) \|\Delta\mathbf{f}\|$$

$$\le \left(166\widehat{n}_0\max(\sqrt{\widehat{n}_0}, \sqrt[4]{\widehat{n}_0}) + 8\widehat{n}_0 + 16\max(\sqrt{\widehat{n}_0}, \sqrt[4]{\widehat{n}_0})\|R^y - \widehat{R}^y\|_F + \frac{12\rho}{\min(\widehat{n}_0, \sqrt{\widehat{n}_0})} \right) \|\Delta\mathbf{z}\|. \quad \square$$

The proof of Theorem 3 is given below.

Proof of Theorem 3. This proof consists of two parts. In Part 1, we will prove that $(\mathbf{g}^k, \mathbf{f}^k) \in \mathcal{N}_{\widehat{n}_0}$ for every k, so $\nabla\tilde{\mathcal{L}}$ always satisfies the Lipchitz property in Lemma 3. In Part 2, we prove the convergence of the gradient descent algorithm.

Part 1: In the optimization approach, we assume $\widehat{n}_0 = \|g\|^2\|f\|$ is known, and start with an initial point $(\mathbf{g}^0, \mathbf{f}^0)$ satisfying $\|\mathbf{g}^0\| \le \sqrt[4]{2\widehat{n}_0}, \|\mathbf{f}^0\| \le \sqrt{2\widehat{n}_0}$. Notice that for any $\mathbf{g}, \mathbf{f}, \mathbf{h} \in \mathbb{C}^N$, we have

$$\|\text{diag}(\mathbf{g})\mathcal{T}(\mathbf{f})\text{diag}(\mathbf{h})\|_F \le \|\mathbf{g}\|\|\mathbf{f}\|\|\mathbf{h}\|,$$

and then

$$\tilde{\mathcal{L}}(\mathbf{g}^0, \mathbf{f}^0) = \mathcal{L}(\mathbf{g}^0, \mathbf{f}^0) = \|\mathrm{diag}(g^0)\mathcal{T}(\mathbf{f}^0)\mathrm{diag}(\overline{\mathbf{g}^0}) - \mathrm{diag}(g)\mathcal{T}(f)\mathrm{diag}(\overline{g}) + R^y - \widehat{R}^y\|_F$$
$$\leq \|\mathbf{g}^0\|^2\|\mathbf{f}^0\| + \|g\|^2\|f\| + \|R^y - \widehat{R}^y\|_F \leq 3\widehat{n}_0 + \|R^y - \widehat{R}^y\|_F. \tag{53}$$

Our gradient descent algorithm guarantees $\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k) \leq \tilde{\mathcal{L}}(\mathbf{g}^0, \mathbf{f}^0), k = 1, 2, \ldots$ (see (54)). We will prove $(\mathbf{g}^k, \mathbf{f}^k) \in \mathcal{N}_{\widehat{n}_0}$ by contradiction. Assume that $(\mathbf{g}^k, \mathbf{f}^k) \notin \mathcal{N}_{\widehat{n}_0}$ for some k. Then

$$\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k) \geq \rho\left[\mathcal{G}_0\left(\frac{\|\mathbf{f}^k\|^2}{2\widehat{n}_0}\right) + \mathcal{G}_0\left(\frac{\|\mathbf{g}^k\|^2}{\sqrt{2\widehat{n}_0}}\right)\right] > \rho\mathcal{G}_0(\sqrt{2}) = \rho(\sqrt{2} - 1)^2.$$

By taking $\rho \geq \frac{3\widehat{n}_0 + \|R^y - \widehat{R}^y\|_F}{(\sqrt{2}-1)^2}$, we would have $\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k) > 3\widehat{n}_0 + \|R^y - \widehat{R}^y\|_F$ which contradicts (53). We conclude that $(\mathbf{g}^k, \mathbf{f}^k) \in \mathcal{N}_{\widehat{n}_0}$ at every iteration k.

Part 2: Let $\mathbf{z} = (\mathbf{g}, \mathbf{f})$ and $\Delta\mathbf{z} = (\Delta\mathbf{g}, \Delta\mathbf{f})$. Notice that $\tilde{\mathcal{L}}$ is continuously differentiable and real-valued. Suppose $\mathbf{z}, \mathbf{z} + \Delta\mathbf{z} \in \mathcal{N}_{\widehat{n}_0}$. Then $\mathbf{z} + t\Delta\mathbf{z} \in \mathcal{N}_{\widehat{n}_0}$ due to convexity of $\mathcal{N}_{\widehat{n}_0}$.

It follows from Lemma 6.1 in [20] that, if $h(t) := \tilde{\mathcal{L}}(\mathbf{z} + t\Delta\mathbf{z})$, then

$$\frac{dh(t)}{dt} = (\Delta\mathbf{z})^T\frac{\partial\tilde{\mathcal{L}}}{\partial\mathbf{z}}(\mathbf{z} + t\Delta\mathbf{z}) + (\Delta\bar{\mathbf{z}})^T\frac{\partial\tilde{\mathcal{L}}}{\partial\bar{\mathbf{z}}}(\mathbf{z} + t\Delta\mathbf{z}) = 2\mathrm{Re}\left((\Delta\mathbf{z})^T\overline{\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{z} + t\Delta\mathbf{z})}\right).$$

By the Fundamental Theorem of Calculus, we have

$$\tilde{\mathcal{L}}(\mathbf{z} + \Delta\mathbf{z}) - \tilde{\mathcal{L}}(\mathbf{z}) = \int_0^1 \frac{dh(t)}{dt}dt = 2\int_0^1 \mathrm{Re}\left((\Delta\mathbf{z})^T\overline{\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{z} + t\Delta\mathbf{z})}\right)dt$$

$$\leq 2\mathrm{Re}\left((\Delta\mathbf{z})^T\overline{\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{z})}\right) + 2\|\Delta\mathbf{z}\|\int_0^1 \|\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{z} + t\Delta\mathbf{z}) - \nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{z})\|dt$$

$$\leq 2\mathrm{Re}\left((\Delta\mathbf{z})^T\overline{\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{z})}\right) + C_{\mathrm{Lip}}\|\Delta\mathbf{z}\|^2.$$

At the kth iteration, we let $\mathbf{z} = (\mathbf{g}^k, \mathbf{f}^k)$, and $\Delta\mathbf{z} = -\eta^k\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)$, and then

$$\tilde{\mathcal{L}}(\mathbf{g}^{k+1}, \mathbf{f}^{k+1}) \leq \tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k) - (2 - C_{\mathrm{Lip}}\eta^k)\eta^k\|\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)\|^2. \tag{54}$$

As long as $\|\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)\| > 0$ and $\eta^k < 2/C_{\mathrm{Lip}}$, we have

$$\tilde{\mathcal{L}}(\mathbf{g}^{k+1}, \mathbf{f}^{k+1}) < \tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k),$$

which implies $\|\nabla_{\mathbf{z}}\tilde{\mathcal{L}}(\mathbf{g}^k, \mathbf{f}^k)\| \to 0$ as $k \to \infty$. This captures the proof that the Wirtinger gradient descent converges to a critical point. $\square$

References

[1] A. Ahmed, L. Demanet, Leveraging diversity and sparsity in blind deconvolution, IEEE Trans. Inform. Theory 64 (6) (2018) 3975–4000.
[2] A. Ahmed, B. Recht, Justin Romberg, Blind deconvolution using convex programming, IEEE Trans. Inform. Theory 60 (3) (2014) 1711–1732.
[3] T. Bendory, Y.C. Eldar, N. Boumal, Non-convex phase retrieval from STFT measurements, IEEE Trans. Inform. Theory 64 (1) (2018) 467–484.
[4] C. Bilen, G. Puy, R. Gribonval, L. Daudet, Convex optimization approaches for blind sensor calibration using sparsity, IEEE Trans. Signal Process. 62 (18) (2014) 4847–4856.

[5] E.J. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Comm. Pure Appl. Math. 67 (6) (2014) 906–956.
[6] E.J. Candes, X. Li, M. Soltanolkotabi, Phase retrieval via Wirtinger flow: theory and algorithms, IEEE Trans. Inform. Theory 61 (4) (2015) 1985–2007.
[7] Y. Chi, Guaranteed blind sparse spikes deconvolution via lifting and convex optimization, IEEE J. Sel. Top. Signal Process. 10 (4) (2016) 782–794.
[8] Y. Chi, L.L. Scharf, A. Pezeshki, A.R. Calderbank, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Process. 59 (5) (2010) 2182–2195.
[9] A. Cosse, From blind deconvolution to blind super-resolution through convex programming, preprint, arXiv:1709.09279, 2017.
[10] M. Duarte, R.G. Baraniuk, Spectral compressive sensing, Appl. Comput. Harmon. Anal. 35 (1) (2013) 111–129.
[11] Y.C. Eldar, Sampling Theory: Beyond Bandlimited Systems, Cambridge University Press, 2015.
[12] A. Fannjiang, W. Liao, Mismatch and Resolution in Compressive Imaging, Wavelets and Sparsity XIV, vol. 8138, International Society for Optics and Photonics, 2011.
[13] A. Fannjiang, W. Liao, Coherence pattern-guided compressive sensing with unresolved grids, SIAM J. Imaging Sci. 5 (1) (2012) 179–202.
[14] A.C. Fannjiang, T. Strohmer, P. Yan, Compressed remote sensing of sparse objects, SIAM J. Imaging Sci. 3 (3) (2010) 595–618.
[15] B. Friedlander, A.J. Weiss, Eigenstructure methods for direction finding with sensor gain and phase uncertainties, Circuits Systems Signal Process. 9 (3) (1990) 271–300.
[16] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94.
[17] K. Lee, N. Tian, J. Romberg, Fast and guaranteed blind multichannel deconvolution under a bilinear system model, IEEE Trans. Inform. Theory 64 (7) (2018) 4792–4818.
[18] R.J. LeVeque, Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems, Society for Industrial and Applied Mathematics, 2007.
[19] R.C. Li, Relative perturbation theory: II. Eigenspace and singular subspace variations, SIAM J. Matrix Anal. Appl. 20 (2) (1998) 471–492.
[20] X. Li, S. Ling, T. Strohmer, K. Wei, Rapid, robust, and reliable blind deconvolution via nonconvex optimization, Appl. Comput. Harmon. Anal. 47 (3) (2019) 893–934, https://doi.org/10.1016/j.acha.2018.01.001.
[21] Y. Li, M.H. Er, Theoretical analyses of gain and phase error calibration with optimal implementation for linear equispaced array, IEEE Trans. Signal Process. 54 (2) (2006) 712–723.
[22] Y. Li, K. Lee, Y. Bresler, Optimal sample complexity for blind gain and phase calibration, IEEE Trans. Signal Process. 64 (21) (2016) 5549–5556.
[23] Y. Li, K. Lee, Y. Bresler, Blind gain and phase calibration for low-dimensional or sparse signal sensing via power iteration, arXiv:1712.00111, 2017.
[24] W. Liao, A. Fannjiang, MUSIC for single-snapshot spectral estimation: stability and super-resolution, Appl. Comput. Harmon. Anal. 40 (1) (2016) 33–67.
[25] S. Ling, T. Strohmer, Self-calibration and biconvex compressive sensing, Inverse Probl. 31 (11) (2015) 115002.
[26] J. Nocedal, S. Wright, Numerical Optimization, Springer Science & Business Media, 2006.
[27] A. Paulraj, Thomas Kailath, Direction of Arrival Estimation by Eigenstructure Methods with Unknown Sensor Gain and Phase, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 10, IEEE, 1985.
[28] G.R.B. de Prony, Essai Experimentale et Analytique, J. Éc. Polytech. 2 (1795) 24–76.
[29] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[30] R.O. Schmidt, A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation, Ph.D. thesis, Stanford Univ., Stanford, CA, Nov. 1981.
[31] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas and Propagation 34 (3) (1986) 276–280.
[32] G.W. Stewart, J.G. Sun, Matrix Perturbation Theory, Computer Science and Scientific Computing, 1990.
[33] P. Stoica, R.L. Moses, Introduction to Spectral Analysis, Upper Saddle River, vol. 1, Prentice Hall, 1997.
[34] J. Sun, Q. Qu, J. Wright, A geometric analysis of phase retrieval, in: 2016 IEEE International Symposium on Information Theory (ISIT), 2016.
[35] J. Sun, Q. Qu, John Wright, Complete dictionary recovery over the sphere I: overview and the geometric picture, IEEE Trans. Inform. Theory 63 (2) (2017) 853–884.
[36] G. Tang, B.N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Trans. Inform. Theory 59 (11) (2013) 7465–7490.
[37] J.A. Tropp, An introduction to matrix concentration inequalities, Found. Trends Mach. Learn. 8.1-2 (2015) 1–230.
[38] R. Vershynin, High-Dimensional Probability an Introduction with Applications in Data Science, Cambridge University Press, 2017, in press.
[39] G. Wang, G.B. Giannakis, Y.C. Eldar, Solving systems of random quadratic equations via truncated amplitude flow, IEEE Trans. Inform. Theory 64 (2) (2018) 773–794.
[40] P.Å. Wedin, Perturbation bounds in connection with singular value decomposition, BIT Numer. Math. 12 (1) (1972) 99–111.
[41] H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung), Math. Ann. 71 (4) (1912) 441–479.
[42] M.P. Wylie, S. Roy, R.F. Schmitt, Self-calibration of linear equi-spaced (LES) arrays, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 1 (1993) 281–284.

[43] M.P. Wylie, S. Roy, H. Messer, Joint DOA estimation and phase calibration of linear equispaced (LES) arrays, IEEE Trans. Signal Process. 42 (12) (1994) 3449–3459.
[44] D. Yang, G. Tang, M. Wakin, Super-resolution of complex exponentials from modulations with unknown waveforms, IEEE Trans. Inform. Theory 62 (10) (2016) 5809–5830.
[45] Z. Tuo, Z. Wang, Han Liu, A non-convex optimization framework for low rank matrix estimation, Adv. Neural Inf. Process. Syst. (2015).