Signal Processing 198 (2022) 108600

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Subspace estimation along a frequency band through projection matrix approximation☆

J. Selva

Department of Physics, Systems Engineering and Signal Theory (DFISTS), University of Alicante, Alicante E-03080, Spain

ARTICLE INFO

Article history:
Received 16 December 2021
Revised 13 April 2022
Accepted 21 April 2022
Available online 27 April 2022

ABSTRACT

In this paper, we present a method to estimate the signal subspace at all the frequencies in a given band, which is computed from the usual set of frequency-bin sample covariance matrices in wideband subspace estimation. Fundamentally, the method exploits the similarity between the signal subspace at any two near-by frequencies to produce an improved subspace estimate along the band. Its key idea consists of modeling the signal subspace by means of a projection matrix function which is approximated by a polynomial. The method provides two improvements: a reduced-size representation of the signal subspace along the frequency band, and a quality improvement in wideband direction-of-arrival (DOA) estimators such as Incoherent Multiple Signal Classification (IC-MUSIC) and Modified Test of Orthogonality of Projected Subspaces (MTOPS). The paper includes the derivation of asymptotic bounds for the bias and root-mean-square (RMS) error of the projection matrix estimate, and a numerical assessment of the method and its combination with the previous two DOA estimators.

1. Introduction

In array processing, the estimation of the subspace spanned by several sources is a fundamental step in DOA estimation [1, Ch. 9]. This estimation relies on the so-called narrowband condition, i.e, the array response must be constant in the spectral band covered by the sources.

In practice, however, this condition is often unrealistic in applications involving high data rates or acoustic or seismic signals, in which the array response varies with frequency significantly. In the literature on DOA estimation, these cases are classified as "wideband" and addressed by dividing the signals' band into bins in which the array response is approximately constant. The problem is then the way the data from all bins should be combined in order to produce a single set of DOA estimates. In the literature, there are fundamentally two approaches for this combination. The first is the coherent approach in which the data from all bins are linearly combined [2–5], and the second is the incoherent approach in which the combination is performed by other means, such as averaging the DOA narrowband estimates from each bin or adding up the bin pseudo-spectrum functions (as is done in IC-MUSIC), [6]. Additionally, there exist other ways to process the bin data derived from general principles such as Maximum Likelihood (ML) [7–11], polynomial matrix decompositions [12–15], group sparsity [16–18], or variants of the Test of Orthogonality of Projected Subspaces (TOPS) method, [19–21].

In this estimation problem, there is a feature that seems to be overlooked in the literature: the signal subspace spanned by the impinging waves varies smoothly along the band covered by the frequency bins and, as a consequence, the subspaces at any two near-by frequencies are similar. In this paper, we present a method for exploiting this smoothness in order to improve the quality of the subspace estimation at the frequency bins and, by extension, at any frequency in the band covered by them. The method is based on modeling the signal subspace using a projection matrix function, which must be estimated from the data signal subspaces computed at the frequency bins.

The paper has been organized as follows. In the next section, we present a signal model for wideband subspace estimation in which the initial data processing is interpreted as a bank on narrowband filters. Then, in Section 3, we characterize the variation of the signal subspace along the frequency band using a projection matrix function and justify this characterization. After that, we in-

☆ This work was supported by the Spanish Ministry of Science and Innovation, the State Agency of Research (AEI) and the European Funds for Regional Development (EFRD) under Project PID2020-117303GB-C22.

E-mail address: jesus.selva@ua.es

https://doi.org/10.1016/j.sigpro.2022.108600

troduce an estimator for this last function in Section 4 and apply it to wideband DOA estimation in Section 5. Finally, we evaluate the proposed method in Section 7 numerically.

1.1. Notation and main symbols

We use the following notation and basic concepts:

- We write vectors in lower case ($\boldsymbol{a}$, $\boldsymbol{x}$) and matrices in upper case, ($\boldsymbol{A}$, $\boldsymbol{Y}$).
- $\boldsymbol{I}_M$ is an identity matrix of size $M \times M$ and $\boldsymbol{0}_M$ the $M \times M$ zero matrix.
- $[\boldsymbol{a}]_m$ and $[\boldsymbol{A}]_{m,k}$ respectively denote the mth and (m,k) components of $\boldsymbol{a}$ and $\boldsymbol{A}$. Also, $[\boldsymbol{A}]_{\cdot,k}$ denotes the kth column of $\boldsymbol{A}$ and $[\boldsymbol{A}]_{\cdot,m:k}$ denotes the matrix form by its columns m to k.
- $\boldsymbol{A}^H$ is the Hermitian of $\boldsymbol{A}$ and $\boldsymbol{A}^\dagger$ its pseudo-inverse.
- The operator '$\equiv$' introduces new symbols.
- '$*$' denotes convolution: $(a * b)(t)$ is the convolution of $a(t)$ and $b(t)$.
- '$\mathcal{E}\{\cdot\}$' and 'Var$\{\cdot\}$' denote the expectation and variance operators.
- $\delta(t)$ and δ_n respectively denote the Dirac and Kronecker delta.
- In the paper, a given $M \times M$ matrix $\boldsymbol{P}$ is said to be a projection matrix if $\boldsymbol{P} = \boldsymbol{P}^H$ and $\boldsymbol{P}^2 = \boldsymbol{P}$.
- $\|\boldsymbol{B}\|_F$ and $\|\boldsymbol{B}\|_2$ denote the Frobenius and 2-norm of a given matrix $\boldsymbol{B}$ respectively.
- $O(\cdot)$ is the big-O notation. A function $g(N)$ is $O(1/N)$ if there are positive numbers N_o and A such that $|g(N)| \leq A/N$ if $N > N_o$.
- $o(\cdot)$ is the little-o notation. A function $g(N)$ is $o(1/N)$ if for any $\epsilon > 0$ there is positive N_o such that $|g(N)| \leq \epsilon/N$ if $N > N_o$.
- In some contexts, '$\cong$' marks an equality that holds if an $o(1/N)$ term is added.

2. Signal model for wideband subspace estimation

Consider an array of M sensors and K impinging waves carrying signals $s_k(t)$, $k = 1, 2, \ldots, K$, with spectra lying in a pass band $[f_A, f_B]$. Let $a_{m,k}(f)$ denote the response of the mth sensor to the kth DOA as a function of the frequency variable f, and $\boldsymbol{A}(f)$ denote the matrix formed by these MK responses, i.e,

$$[\boldsymbol{A}(f)]_{m,k} \equiv a_{m,k}(f), \ m = 1, \ldots, M, \ k = 1, \ldots, K. \tag{1}$$

In a narrowband scenario, $\boldsymbol{A}(f)$ is approximately constant in $[f_A, f_B]$ and this allows us to regard the signals $\boldsymbol{s}(t)$ as concentrated at a single and fixed frequency f_1 lying in $[f_A, f_B]$. As a consequence, the $M \times 1$ vector $\boldsymbol{x}(t)$ formed by the received signals can be written as

$$\boldsymbol{x}(t) = \boldsymbol{A}(f_1)\boldsymbol{s}(t) + \boldsymbol{\eta}(t), \tag{2}$$

where $\boldsymbol{s}(t)$ is the $K \times 1$ vector formed by the impinging signals,

$$[\boldsymbol{s}(t)]_k \equiv s_k(t), \tag{3}$$

and $\boldsymbol{\eta}(t)$ is formed by independent, circularly-symmetric, complex-white components of zero mean and variance σ^2. As can be readily seen in (2), the signal component of the vectors $\boldsymbol{x}(t)$ lie in the column span of $\boldsymbol{A}(f_1)$ and this fact allows us to estimate this last span from a set of samples of $\boldsymbol{x}(t)$. Let us recall the procedure for performing this estimation.

If the components of $\boldsymbol{s}(t)$ are zero-mean stationary processes, then the covariance of $\boldsymbol{s}(t)$ has the form

$$\boldsymbol{R}(f_1) \equiv \mathcal{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}(t)^H\right\} = \boldsymbol{A}(f_1)\boldsymbol{R}_s(f_1)\boldsymbol{A}(f_1)^H + \sigma^2\boldsymbol{I}_M, \tag{4}$$

where $\boldsymbol{R}_s(f)$ denotes the covariance matrix of $\boldsymbol{s}(t)$,

$$\boldsymbol{R}_s(f_1) \equiv \mathcal{E}\left\{\boldsymbol{s}(t)\boldsymbol{s}(t)^H\right\}. \tag{5}$$

$\boldsymbol{R}(f_1)$ can be estimated from a sample covariance matrix given by

$$\widehat{\boldsymbol{R}}(f_1) \equiv \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}(nT)\boldsymbol{x}(nT)^H, \tag{6}$$

where T denotes the sampling period and N the number of samples taken. Additionally, the span of $\boldsymbol{A}(f_1)$ can be estimated from the K eigenvectors of $\widehat{\boldsymbol{R}}(f_1)$ corresponding to its K largest eigenvalues, provided $\boldsymbol{R}_s(f_1)$ is non-singular, i.e, provided the K signal components are not completely coherent. More precisely, let $\widehat{\boldsymbol{Q}}_{1,K}$ denote the $M \times K$ matrix formed by the eigenvectors corresponding to the largest K eigenvalues of $\widehat{\boldsymbol{R}}(f_1)$ and following $\widehat{\boldsymbol{Q}}_{1,K}^H\widehat{\boldsymbol{Q}}_{1,K} = \boldsymbol{I}_K$. Then, the span of $\widehat{\boldsymbol{Q}}_{1,K}$ is an estimate of the span of $\boldsymbol{A}(f_1)$.

Next, let us consider the wideband scenario. If $\boldsymbol{A}(f)$ shows a significant variation in $[f_A, f_B]$ then (2) is no longer valid. However, it is valid at the output of any filter selecting a narrow band centered at a frequency f inside $[f_A, f_B]$, provided $\boldsymbol{A}(f)$ is approximately constant in the filter's output band. So, if we apply a bank of R such filters centered at distinct frequencies f_r, $(f_r < f_{r+1})$, lying in $[f_A, f_B]$, we obtain R models akin to (2)

$$\boldsymbol{x}_r(t) = \boldsymbol{A}(f_r)\boldsymbol{s}_r(t) + \boldsymbol{\eta}_r(t), \ r = 1, \ldots, R, \tag{7}$$

where $\boldsymbol{x}_r(t)$, $\boldsymbol{s}_r(t)$, and $\boldsymbol{\eta}_r(t)$ denote the signals at the rth filter's output. Here, we assume non-overlapping filter bands of equal width, so that these last models are independent and all components of $\boldsymbol{\eta}_r(t)$ have the same variance σ^2.

The narrowband subspace estimation that we recalled before can now be performed at each filter output $\boldsymbol{x}_r(t)$, in order to obtain a set of sample covariance matrices $\widehat{\boldsymbol{R}}(f_r)$, which are estimates of the corresponding covariance matrices $\boldsymbol{R}(f_r)$, $r = 1, 2, \ldots, R$. Additionally, we obtain matrices $\widehat{\boldsymbol{Q}}_{r,K}$ whose spans are estimates of the corresponding spans of the signature matrices $\boldsymbol{A}(f_r)$, $r = 1, 2, \ldots, R$.

Up to this point, we have described the first processing steps in wideband subspace estimation in terms of a filter bank. In practice, such a bank is implemented using the Fast Fourier Transform (FFT) snapshot model, i.e, by computing the FFT of the data samples in consecutive time slots. Additionally, these steps are usually followed by further processing of either the matrices $\widehat{\boldsymbol{R}}(f_r)$ or $\widehat{\boldsymbol{Q}}_{r,K}$ in order to estimate the desired parameters, (which often consist of the waves' DOAs). In this paper, we are interested in exploiting the similarity between the signal subspaces at any two near-by frequencies lying in $[f_A, f_B]$ as shown in the next section.

3. Wideband subspace characterization in terms of a projection matrix function

We can make two basic observations about the processing steps described in the previous section. The first is that we may expect the span of two consecutive signature matrices, $\boldsymbol{A}(f_r)$ and $\boldsymbol{A}(f_{r+1})$, to be close to each other if the difference $f_{r+1} - f_r$ is not too large. More precisely, if $\boldsymbol{P}(f)$ denotes the orthogonal projection matrix associated with the span of $\boldsymbol{A}(f)$, i.e, the matrix,

$$\boldsymbol{P}(f) = \boldsymbol{A}(f)\left(\boldsymbol{A}(f)^H\boldsymbol{A}(f)\right)^{-1}\boldsymbol{A}(f)^H \tag{8}$$

then, given a fixed unit-norm $M \times 1$ vector $\boldsymbol{y}$, we may expect $\boldsymbol{P}(f_r)\boldsymbol{y}$ to be close to $\boldsymbol{P}(f_{r+1})\boldsymbol{y}$. Since $\boldsymbol{y}$ is arbitrary, we may re-state this condition in more precise terms by saying that the magnitude

$$\sup_{\|\boldsymbol{y}\|_2=1} \|(\boldsymbol{P}(f_{r+1}) - \boldsymbol{P}(f_r))\boldsymbol{y}\|_2 \tag{9}$$

is small. However, note two features of this argument. First, this last expression is the definition of the matrix 2-norm of $\boldsymbol{P}(f_{r+1}) - \boldsymbol{P}(f_r)$, i.e,

$$\|\boldsymbol{P}(f_{r+1}) - \boldsymbol{P}(f_r)\|_2 = \sup_{\|y\|_2=1} \|(\boldsymbol{P}(f_{r+1}) - \boldsymbol{P}(f_r))\boldsymbol{y}\|_2, \tag{10}$$

[22, Section 2.3]. And second, we can repeat the argument for f_r and any frequency f between f_r and f_{r+1} and, more generally, for any frequency f and the closest bin frequency f_r. So, in summary, this first observation leads us to view the projection matrix $\boldsymbol{P}(f)$ as a "signal" whose value at any f in $[f_A, f_B]$ can be approximated from its "samples" $\boldsymbol{P}(f_1), \boldsymbol{P}(f_2), \ldots, \boldsymbol{P}(f_R)$.

The second observation is that the estimates $\widehat{\boldsymbol{Q}}_{r,K}$ already provide a set of noisy estimates of $\boldsymbol{P}(f_1), \boldsymbol{P}(f_2), \ldots, \boldsymbol{P}(f_R)$, namely, the projection matrices

$$\widehat{\boldsymbol{P}}_0(f_r) \equiv \widehat{\boldsymbol{Q}}_{r,K}\widehat{\boldsymbol{Q}}_{r,K}^H, \quad r = 1, 2, \ldots, R. \tag{11}$$

Besides, we can repeat the argument that led to (10), but for $\widehat{\boldsymbol{P}}_0(f_r)$ and $\boldsymbol{P}(f_r)$, to show that $\widehat{\boldsymbol{P}}_0(f_r)$ approximates $\boldsymbol{P}(f_r)$ in the sense that $\|\widehat{\boldsymbol{P}}_0(f_r) - \boldsymbol{P}(f_r)\|_2$ is small.

So, we have that the estimation of the signal subspace at any frequency in $[f_A, f_B]$ can be viewed as a signal approximation problem, in which the true signal is $\boldsymbol{P}(f)$ and its noisy samples are $\widehat{\boldsymbol{P}}_0(f_r)$, $r = 1, 2, \ldots, R$. In order to estimate $\boldsymbol{P}(f)$ in this setting, we need to specify two aspects: a model for this matrix function and a quality estimate for the samples $\widehat{\boldsymbol{P}}_0(f_r)$.

Regarding the model, the fact is that $\boldsymbol{P}(f)$ is a very regular function, due to the regularity of $\boldsymbol{A}(f)$, that can be approximated by a Q-order polynomial of the form

$$\boldsymbol{P}(f) = \left(\sum_{q=0}^{Q} \boldsymbol{G}_q f^q\right) + \boldsymbol{\Phi}_Q(f) \tag{12}$$

in $[f_A, f_B]$, where $\boldsymbol{G}_q$ is a set of $M \times M$ coefficient matrices and $\boldsymbol{\Phi}_Q(f)$ is the error term. The norm of $\boldsymbol{\Phi}_Q(f)$ can be made arbitrarily small by increasing Q. The justification of (12) consists of a chain of results available in the literature:

1. The components of $\boldsymbol{A}(f)$ are typical band-limited functions due to the finite extent of the sensor array and, therefore, they are analytic functions for any complex f and, in particular, in range $[f_A, f_B]$.[1]
2. A consequence of (1) is that the components of the Hermitian matrix $\boldsymbol{A}(f)\boldsymbol{A}(f)^H$ are also analytic in $[f_A, f_B]$.
3. $\boldsymbol{P}(f)$ is the projection matrix associated with the non-zero eigenvalues of $\boldsymbol{A}(f)\boldsymbol{A}(f)^H$.
4. According to the Perturbation Theory result in [23, Ch.2, Th. 1.10], (2) and (3) imply that $\boldsymbol{P}(f)$ is also analytic in $[f_A, f_B]$.
5. A consequence of 4) is that $\boldsymbol{P}(f)$ is approximable by a polynomial of the form in (12) for fixed DOAs, (Weiertrass theorem [24, Th. 2.4.1]).
6. Since the set of DOAs can be parameterized using a set of variables whose domain is a compact set, (set of angles of arrival, for example), (5) implies that there is a finite order Q valid for all possible DOAs.

The determination of Q given a bound for $\|\boldsymbol{\Phi}_Q(f)\|_2$ seems to be a cumbersome problem from a theoretical point of view. However, Q can be determined numerically for any specific sensor array by computing the interpolation error of a standard interpolation method such as Chebyshev's. We do this in Section 7.2 for a ten-sensor Uniform Linear Array (ULA).

Regarding the quality of $\widehat{\boldsymbol{P}}_0(f_r)$, we have that it can be analyzed by extending the known results on the perturbation of the eigenvalues and eigenvectors of sample covariance matrices in [25,26]. In Appendix A we show that $\widehat{\boldsymbol{P}}_0(f_r)$ is an asymptotically unbiased estimate of $\boldsymbol{P}(f)$, i.e,

$$\mathcal{E}\{\widehat{\boldsymbol{P}}_0(f_r)\} = \boldsymbol{P}(f_r) + O(1/N). \tag{13}$$

[1] Any of the components of $\boldsymbol{A}(f)$ is the Fourier transform of a signal which is time-limited, given that the delay between any two sensors is bounded for any impinging wave, due to the finite spatial extent of the sensor array.

We also show in that appendix that $\widehat{\boldsymbol{P}}_0(f_r)$ is a consistent estimate of $\boldsymbol{P}(f_r)$. More precisely, let the eigenvalue decomposition of $\boldsymbol{R}(f_r)$ be given by

$$\boldsymbol{R}(f_r) = \sum_{m=1}^{M} \lambda_m(f_r)\boldsymbol{q}_m(f_r)\boldsymbol{q}_m(f_r)^H, \quad r = 1, 2, \ldots, R, \tag{14}$$

where $\lambda_m(f_r)$ denote the eigenvalues, sorted in decreasing order, and $\boldsymbol{q}_m(f_r)$ the corresponding eigenvectors. The asymptotic quadratical error for any component (m, m') of $\boldsymbol{P}(f_r)$ admits the bound

$$\mathcal{E}\left\{\left|[\widehat{\boldsymbol{P}}_0(f_r) - \boldsymbol{P}(f_r)]_{m,m'}\right|^2\right\} \leq \frac{8(M-K)\sigma^2}{N(\lambda_K(f_r)-\sigma^2)^2}\sum_{k=1}^{K}\lambda_k(f_r) + o\left(\frac{1}{N}\right). \tag{15}$$

A corollary of this bound is that the variance of any component of $\widehat{\boldsymbol{P}}_0(f_r)$ is $O(1/N)$, i.e,

$$\mathrm{Var}\{\widehat{\boldsymbol{P}}_0(f_r)\} = O(1/N). \tag{16}$$

(13) and this last formula imply that $\widehat{\boldsymbol{P}}_0(f_r)$ is an asymptotically unbiased and consistent estimator of $\boldsymbol{P}(f_r)$.

The bound in (15) has an interesting interpretation. Since $\lambda_K(f_r) - \sigma^2$ is the power gap between the signal and noise subspaces, the factor

$$\frac{1}{\lambda_K(f_r)-\sigma^2}\sum_{k=1}^{K}\lambda_k(f_r) \tag{17}$$

is the signal power measured in number of gaps and

$$\frac{(M-K)\sigma^2}{\lambda_K(f_r)-\sigma^2} \tag{18}$$

is the noise power again measured in number of gaps. Therefore, the bound in (A.14) is proportional to the product of these two relative powers and decreases as $1/N$.

In the next section, we derive an estimator of $\boldsymbol{P}(f)$ based on the results in this section.

4. Estimation of the projection matrix function

If the mismatch $\boldsymbol{\Phi}_Q(f)$ in (12) is small then we may estimate the coefficient matrices $\boldsymbol{G}_q$, by posing a linear regression model with data $\widehat{\boldsymbol{P}}_0(f_r)$ component-wise, i.e,

$$[\widehat{\boldsymbol{P}}_0(f_r)]_{m,m'} \approx \sum_{q=0}^{Q}[\boldsymbol{G}_q]_{m,m'}f_r^q, \quad (m, m' = 1, 2, \ldots, M; \ \ r = 1, 2 \ldots, R) \tag{19}$$

in which $[\boldsymbol{G}_q]_{m,m'}$, $q = 0, 1, \ldots, Q$, is the set of unknown parameters. The samples $[\widehat{\boldsymbol{P}}_0(f_r)]_{m,m'}$ are independent given that the bands of the filter banks used to produce $\boldsymbol{x}_r(t)$ in (7) do not overlap. Besides, from (13) and (15), they are asymptotically unbiased and consistent. All these features suggest to pose a weighted least-squares cost function for the estimation of the coefficients $[\boldsymbol{G}_q]_{m,m'}$ with weights w_r, i.e, the function

$$\sum_{r=1}^{R} w_r^2\left|[\widehat{\boldsymbol{P}}_0(f_r)]_{m,m'} - \sum_{q=0}^{Q}[\widehat{\boldsymbol{G}}_q]_{m,m'}f_r^q\right|^2. \tag{20}$$

(We present a possible set of coefficients w_r in the next Section.) The minimizer of this last cost function is the estimate

$$[\widehat{\boldsymbol{G}}_q]_{m,m'} \equiv \sum_{r=1}^{R}[\boldsymbol{B}]_{q,r}[\widehat{\boldsymbol{P}}_0(f_r)]_{m,m'}, \tag{21}$$

where $(r, r' = 1, 2, \ldots, R;\ q = 0, 1, \ldots, Q)$

$$\boldsymbol{B} \equiv (\boldsymbol{V}^H \boldsymbol{\Lambda}^2 \boldsymbol{V})^{-1} \boldsymbol{V}^H \boldsymbol{\Lambda}^2, \quad [\boldsymbol{V}]_{r,q+1} \equiv f_r^q, \tag{22}$$

$$[\boldsymbol{\Lambda}]_{r,r'} \equiv w_r \delta_{r-r'}. \tag{23}$$

Since $\boldsymbol{B}$ is independent of m and m', we may write (21) as

$$\widehat{\boldsymbol{G}}_q \equiv \sum_{r=1}^{R} [\boldsymbol{B}]_{q,r} \widehat{\boldsymbol{P}}_0(f_r) \tag{24}$$

and the estimate of $\boldsymbol{P}(f)$ at any f is

$$\widehat{\boldsymbol{P}}_1(f) \equiv \sum_{q=0}^{Q} \widehat{\boldsymbol{G}}_q f^q. \tag{25}$$

$\widehat{\boldsymbol{P}}_1(f)$ can be expressed in terms of $\widehat{\boldsymbol{P}}_0(f_r)$ by substituting (24) into (25). We have

$$\widehat{\boldsymbol{P}}_1(f) \equiv \sum_{r=1}^{R} d_r(f) \widehat{\boldsymbol{P}}_0(f_r) \tag{26}$$

where

$$d_r(f) = \sum_{q=0}^{Q} [\boldsymbol{B}]_{q,r} f^q. \tag{27}$$

Let us now analyze the quality of the estimate $\widehat{\boldsymbol{P}}_1(f)$ by assessing its expected value and variance. But first, let us derive an expression for the mismatch. (26) can be viewed as an interpolator that is exact for the sum in (12), i.e,

$$\sum_{q=0}^{Q} \boldsymbol{G}_q f^q = \sum_{r=1}^{R} d_r(f) \sum_{q=0}^{Q} \boldsymbol{G}_q f_r^q, \tag{28}$$

given that the minimizer of (20) involves an (over-)determined linear system. Using (12) to solve for the sums in this last equation, we have

$$\begin{aligned} \boldsymbol{P}(f) - \boldsymbol{\Phi}_Q(f) &= \sum_{r=1}^{R} d_r(f) \Big(\boldsymbol{P}(f_r) - \boldsymbol{\Phi}_Q(f_r) \Big) \\ \boldsymbol{P}(f) &= \sum_{r=1}^{R} d_r(f) \boldsymbol{P}(f_r) - \sum_{r=1}^{R} d_r(f) \boldsymbol{\Phi}_Q(f_r) + \boldsymbol{\Phi}_Q(f). \end{aligned} \tag{29}$$

The last two terms on the right-hand side are the mismatch that can be made arbitrarily small by increasing Q.

Next, let us derive a formula for the expected value of $\widehat{\boldsymbol{P}}_1(f)$. From (26) and (29), we have

$$\begin{aligned} \widehat{\boldsymbol{P}}_1(f) = \boldsymbol{P}(f) &+ \sum_{r=1}^{R} d_r(f) (\widehat{\boldsymbol{P}}_0(f_r) - \boldsymbol{P}(f_r)) \\ &+ \sum_{r=1}^{R} d_r(f) \boldsymbol{\Phi}_Q(f_r) - \boldsymbol{\Phi}_Q(f). \end{aligned} \tag{30}$$

So, taking expectations and using (13), we obtain

$$\mathcal{E}\left\{ \widehat{\boldsymbol{P}}_1(f) \right\} = \boldsymbol{P}(f) + \sum_{r=1}^{R} d_r(f) \boldsymbol{\Phi}_Q(f_r) - \boldsymbol{\Phi}_Q(f) + O(1/N). \tag{31}$$

And finally, from (16) and (26), the variance is

$$\mathrm{Var}\{\widehat{\boldsymbol{P}}_1(f)\} = O(1/N). \tag{32}$$

In summary, if the mismatch term in (31) in negligible, (31) and (32) imply that $\widehat{\boldsymbol{P}}_1(f)$ is a consistent estimate of $\boldsymbol{P}(f)$.

For any f in $[f_A, f_B]$, $\widehat{\boldsymbol{P}}_1(f)$ is a projection matrix only approximately, due to the mismatch in (31) and to the noise. This drawback can be mitigated by taking the rank-K projection matrix lying closest to $\widehat{\boldsymbol{P}}_1(f)$ as the final signal projection matrix estimate at any frequency; i.e, the final signal projection matrix estimate is

$$\widehat{\boldsymbol{P}}_2(f) \equiv \arg \min_{\substack{\text{rank } K \\ \text{proj. matrix } \boldsymbol{P}}} \|\widehat{\boldsymbol{P}}_1(f) - \boldsymbol{P}\|_F^2. \tag{33}$$

It can be easily checked that $\widehat{\boldsymbol{P}}_2(f)$ is just the signal projection matrix of $\widehat{\boldsymbol{P}}_1(f)$, [i.e, the projection matrix associated with the K largest eigenvalues of $\widehat{\boldsymbol{P}}_1(f)$]. Though the computation of $\widehat{\boldsymbol{P}}_2(f)$ from $\widehat{\boldsymbol{P}}_1(f)$ requires an additional eigenvalue decomposition, we may expect a small number of such decompositions in practice, given that $\widehat{\boldsymbol{P}}(f)$ can be well approximated in $[f_A, f_B]$ by a Q-order polynomial. We assess this point in Secs. 7.5 and 7.6 numerically.

4.1. A possible choice for the coefficients w_r

The bound in (15) provides a possible choice of coefficients w_r if we replace $\lambda(f_r)$ and σ^2 with their corresponding estimates obtained from $\widehat{\boldsymbol{R}}(f_r)$. Specifically, if $\widehat{\lambda}(f_r)$ denotes the rth eigenvalue of $\widehat{\boldsymbol{R}}(f_r)$ in the eigen-decomposition akin to (14), then a possible choice $\widehat{w}_r$ for the coefficients w_r is specified by the equation

$$\frac{1}{\widehat{w}_r^2} = \frac{8(M-K)\widehat{\sigma}^2(f_r)}{N(\widehat{\lambda}_K(f_r) - \widehat{\sigma}^2(f_r))^2} \sum_{k=1}^{K} \widehat{\lambda}_k(f_r), \tag{34}$$

where

$$\widehat{\sigma}^2(f_r) \equiv \frac{1}{N-K} \sum_{m=K+1}^{M} \widehat{\lambda}_m(f_r). \tag{35}$$

5. Application of the proposed method to wideband DOA estimation

In this section, we apply the method in Section 4 to wideband DOA estimation in a ULA. Let us first particularize the signal model in Section 2 to this type of array, then recall two wideband DOA estimators, IC-MUSIC and MTOPS, and finally adapt these estimators to the method proposed in this paper.

In a ULA, the time-domain response of the mth sensor to the kth DOA is

$$\tilde{a}_{k,m}(t) = e^{-j2\pi f_o \tau_m \gamma_k} \delta(t - \tau_m \gamma_k) \tag{36}$$

where

- f_o is the array's central frequency, so that the sensor spacing is $c/(2f_o)$ where c is the propagation speed,
- $\gamma_k \equiv \sin(\theta_k)$ and θ_k is the angle of arrival relative to the broadside,
- τ_m is the delay associated with the mth sensor along the array. If $\tau_{\max}$ denotes the array diameter, measured as a delay, then

$$\tau_m \equiv -\frac{\tau_{max}}{2} + (m-1)\frac{\tau_{max}}{M-1},\ m = 1, 2, \ldots, M. \tag{37}$$

The array response $\boldsymbol{A}(f)$ is just the Fourier transform of (36), $(m = 1, 2, \ldots, M,\ k = 1, 2, \ldots, K)$,

$$[\boldsymbol{A}(f, \gamma)]_{m,k} \equiv e^{-j2\pi (f_o + f)\tau_m \gamma_k}, \tag{38}$$

where

$$[\boldsymbol{\gamma}]_k \equiv \gamma_k,\ k = 1, 2, \ldots, K, \tag{39}$$

and where we have written $\boldsymbol{A}(f, \gamma)$ rather than $\boldsymbol{A}(f)$ to show the dependence on the parameters γ_k explicitly. Finally, we may write the model in (4) but for the rth frequency bin as

$$\boldsymbol{R}(f_r) = \boldsymbol{A}(f_r, \gamma) \boldsymbol{R}_s(f_r) \boldsymbol{A}(f_r, \gamma)^H + \sigma^2 \boldsymbol{I}_M. \tag{40}$$

The problem of estimating the angles of arrival θ_k can now be cast as the problem of estimating the parameters γ_k, given that there is a one-to-one relationship between θ_k and γ_k, $\gamma_k = \sin(\theta_k)$.

In this paper, we consider the estimation of the DOA parameters γ_k by means of the following two methods:

- **IC-MUSIC**, [27, **Section 4.4.3]**. In this estimator, the K DOA estimates are given by the abscissa of the main K local maxima of the pseudo-spectrum

$$\phi(\gamma) \equiv \sum_{r=1}^{R} \|\widehat{\boldsymbol{P}}_0(f_r)\boldsymbol{a}(f_r,\gamma)\|^2, \tag{41}$$

where

$$[\boldsymbol{a}(f,\gamma)]_m \equiv e^{-j2\pi(f_o+f)\tau_m\gamma},\ m=1,\,2,\ldots,\,M. \tag{42}$$

- **MTOPS,** [20]. This estimator starts by computing a set of $M\times K$ matrices $\widehat{\boldsymbol{U}}(f_r)$ and another set of $M\times(M-K)$ matrices $\widehat{\boldsymbol{V}}(f_r)$, $r=1,\,2,\ldots,\,R$, whose columns are ortho-normal bases of the signal and noise subspaces respectively; i.e, the columns of $\widehat{\boldsymbol{U}}(f_r)$ span the subspace associated with $\widehat{\boldsymbol{P}}_0(f_r)$ and

$$[\widehat{\boldsymbol{U}}(f_r),\widehat{\boldsymbol{V}}(f_r)]^H[\widehat{\boldsymbol{U}}(f_r),\widehat{\boldsymbol{V}}(f_r)] = \boldsymbol{I}_M. \tag{43}$$

Note that we may view the matrix pair $(\widehat{\boldsymbol{U}}(f_r),\widehat{\boldsymbol{V}}(f_r))$ as a function of $\widehat{\boldsymbol{P}}_0(f_r)$, given that one such pair can be easily computed from $\widehat{\boldsymbol{P}}_0(f_r)$. The MTOPS estimator uses the pseudo-spectrum

$$\mu(\gamma) \equiv \left\|\left[\boldsymbol{E}_2(\gamma),\boldsymbol{E}_3(\gamma),\ldots,\boldsymbol{E}_R(\gamma)\right]\right\|_F^2, \tag{44}$$

where $(r=2,\,3,\ldots,R)$

$$\boldsymbol{E}_r(\gamma) \equiv \widehat{\boldsymbol{U}}(f_1)^H\mathrm{diag}(\boldsymbol{a}(f_r-f_1,\gamma))\widehat{\boldsymbol{V}}(f_r). \tag{45}$$

The MTOPS DOA estimates are the smallest K local minima of $\mu(\gamma)$.

These estimators can be easily adapted to the method proposed in this paper, simply by applying them to a set of samples of the projection function estimate $\widehat{\boldsymbol{P}}_2(f)$; i.e, the matrices $\widehat{\boldsymbol{P}}_0(f_r)$, $(r=1,\,2,\ldots,\,R)$, in IC-MUSIC and MTOPS would be replaced by a set of matrices $\widehat{\boldsymbol{P}}_2(f_r')$, $(r=1,\,2,\ldots,\,R')$, where the frequencies f_r' are regularly spaced in $[f_A,f_B]$ and R' is close to $Q+1$. Note that the computation of the matrices $\widehat{\boldsymbol{P}}_2(f_r')$ involves R' eigenvalue decompositions. However, this additional complexity is small if $R\gg Q$, given that we may expect R' to be close to Q.

6. Computational burden

The proposed method consists of four computation steps whose complexities are the following:

1. **Sample projection matrices $\widehat{\boldsymbol{P}}_0(f_r)$.** The computation of the R projection matrices $\widehat{\boldsymbol{P}}_0(f_r)$, $r=1,\,2,\ldots,\,R$, requires R subspace decompositions and the total complexity is $O(M^3R)$.
2. **Polynomial coefficient matrices $\boldsymbol{G}_q$.** The coefficient matrices $\widehat{\boldsymbol{G}}_q$, $q=1,\,2,\ldots,\,Q$, are computed using (21) and involve $O(M^2R)$ operations.
3. **First estimate $\widehat{\boldsymbol{P}}_1(f)$.** The computation of $\widehat{\boldsymbol{P}}_1(f)$ is done through (25) with cost $O(M^2Q)$.
4. **Second estimate $\widehat{\boldsymbol{P}}_2(f)$.** This estimate is obtained from the subspace decomposition of $\widehat{\boldsymbol{P}}_1(f)$ and involves $O(M^3)$ operations.

Step (1) is the most expensive, though it must be performed in any method combining signal subspaces (such as IC-MUSIC and MTOPS, for example). Steps (2) and (3) have less complexity than step 1), given that they are just linear combinations of several matrices. Finally, step (4) has cost $O(M^3)$ and may be necessary at a set of R' frequencies in order to refine the projection matrix estimates. For this last point, see the numerical examples in Secs. 7.5 and 7.6.

7. Numerical examples

We have carried out several numerical examples for a 10-sensor ULA following the model in Section 5. The sub-sections that follow contain the main results:

- Section 7.1 contains a list of the main parameters in the numerical examples.
- In Section 7.2, we assess the selection of Q using Chebyshev interpolation in order to upper-bound the mismatch of the polynomial approximation in (12) for several values of Q.
- In Section 7.3, we evaluate the approximation error of one component of $\boldsymbol{P}(f)$ using the corresponding component of $\widehat{\boldsymbol{P}}_1(f)$.
- In Section 7.4, we evaluate the same error but for the whole matrix $\widehat{\boldsymbol{P}}_1(f)$ using the error measure in (52) and in the RMS sense.
- Finally, in Secs. 7.5 and 7.6, we combine the proposed method with IC-MUSIC and MTOPS for DOA estimation.

7.1. Main parameters in the numerical examples

The parameters in the numerical examples were the following:
Sensor array. Linear array of $M=10$ sensors with half wavelength spacing.

Central frequency. $f_o = 2.4$ GHz.

Maximum delay along the array.

$$\tau_{max} = \frac{M-1}{2f_o} = 1.875 \text{ nsec}.$$

DOA parameters. The DOA was parameterized in terms of γ rather than θ, where $\gamma=\sin(\theta)$, following the approach in Section 5.

Received signals. Linearly-modulated signals of the form

$$\sum_{n=-\infty}^{\infty} a_n g(t-nT_{ch}) \tag{46}$$

where

- a_n are zero-mean, independent complex Gaussian noise samples of variance equal to 1.
- $g(t)$ is a raised-cosine pulse with chip period $T_{ch}\equiv 2.6$ nsec and roll-off factor $\beta\equiv 0.25$.

Sampling period. $T\equiv T_{ch}/2$.

Signals' bandwidth relative to f_o. The signals' two-sided bandwidth B followed $B/f_o=0.2$, i.e, $B=0.48$ GHz. However, in the numerical examples, only the band in which $g(t)$ has flat spectrum was used, i.e, the band $[-B_1/2,B_1/2]$, where $B_1\equiv(1-\beta)/T_{ch}=0.288$ GHz. So, the relative bandwidth employed was $B_1/f_o=0.12$.

Number of slots. $N_{sl}=50$.

Number of samples per slot. $N=1024$.

Number of frequency bins. The number of frequency bins was fixed to $R=41$ and they were equally spaced in $[-B_1/2,B_1/2]$.

Directions of arrival (DOAs). Two cases have been assessed:

- Three DOAs given by

$$\boldsymbol{\gamma} = [0.1,0.27,0.82]^T. \tag{47}$$

- Two DOAs of the form

$$\boldsymbol{\gamma} = [0.1,0.1+\Delta\gamma]^T, \tag{48}$$

where the increment $\Delta\gamma$ is a simulation parameter.

Signal-to-noise ratio (SNR). The SNRs in the numerical examples are equal to the total signal energy at frequency f_o divided by the corresponding noise energy.

Least squares weights w_r. Equal to 1.

Table 1
Chebyshev interpolation error using measure in (50).

QK	1	2	3	4
0	0.1486	0.2624	0.3313	0.3766
1	0.06228	0.09993	0.1106	0.114
2	0.01739	0.02558	0.03154	0.03407
3	0.003732	0.007473	0.01038	0.01325
4	0.0006299	0.00211	0.003046	0.00411
5	0.00008891	0.0007461	0.001155	0.00163

Table 2
Chebyshev interpolation error using measure in (51).

QK	1	2	3	4
0	0.5051	0.6319	0.6578	0.6794
1	0.1552	0.1986	0.2247	0.2454
2	0.03203	0.0652	0.07654	0.08669
3	0.005908	0.02142	0.02551	0.03059
4	0.0008885	0.006993	0.00854	0.01067
5	0.0001158	0.00227	0.002881	0.003813

Generation method for numerical trials. The numerical trials were generated using the domain model in (7) for reducing the simulation time. However, the results were validated by generating these trials in the time domain and checking that the results are consistent if $N \to \infty$.

Number of Monte Carlo trials. 2200.

7.2. Polynomial order Q versus mismatch for a ten-sensor uniform linear array

We have computed an upper bound for the approximation error in (12) by means of Chebyshev interpolation applied to $\boldsymbol{P}(f)$ in range $f \in [f_A, f_B]$ with $Q+1$ nodes, where the array response matrix is given by (38), [28, Ch. 6]. This interpolation scheme delivers a polynomial approximation in the f variable of the form

$$\boldsymbol{P}(f) \approx \sum_{q=0}^{Q} \boldsymbol{G}_q(Q, \gamma_1, \ldots, \gamma_K) f^q \tag{49}$$

for f in $[f_A, f_B]$ and fixed Q, γ_1, ..., γ_K. Specifically, we have evaluated the error in (49) for any value of the variables involved. Table 1 shows the maximum error in (49) for several values of Q and K, i.e the error measure

$$\max_{m,m',f,\gamma_1,\ldots,\gamma_K} \left| [\boldsymbol{P}(f) - \sum_{q=0}^{Q} \boldsymbol{G}_q(Q, \gamma_1, \ldots, \gamma_K) f^q]_{m,m'} \right|. \tag{50}$$

Note that, since the components of $\boldsymbol{P}(f)$ are bounded by one, this interpolation scheme provides significant accuracy even for small values of Q. Table 2 shows the error measure

$$\sup_{f,\gamma_1,\ldots,\gamma_K} \|\boldsymbol{P}(f) - \sum_{q=0}^{Q} \boldsymbol{G}_q(Q, \gamma_1, \ldots, \gamma_K) f^q\|_F \tag{51}$$

where, again, the supremum is taken over all variables involved. This is the maximum error that would be incurred if $\boldsymbol{P}(f)$ were replaced by its polynomial approximation in the computation of $\boldsymbol{P}(f)\boldsymbol{x}$ for unit-norm vectors $\boldsymbol{x}$. The conclusion is the same.

7.3. Approximation of $\hat{P}_1(f_r)$ and $\hat{P}_2(f)$ to $P(f)$

In this section, we assess qualitatively the error in approximating the true projection matrix $\boldsymbol{P}(f)$ using either $\widehat{\boldsymbol{P}}_0(f_r)$, $\widehat{\boldsymbol{P}}_1(f)$ or $\widehat{\boldsymbol{P}}_2(f)$ for a single component of these matrices and the DOAs in (47). The objective of this assessment is to show the improvement that can be achieved graphically.

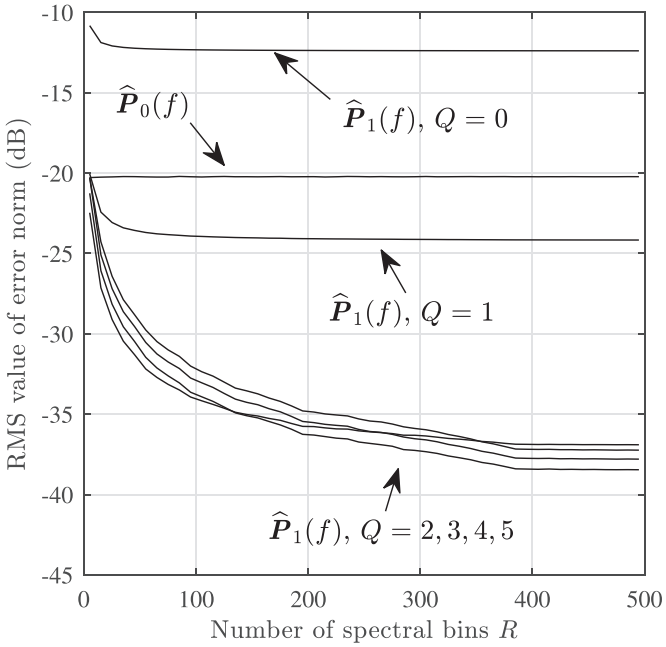


Fig. 1. Real and imaginary part of component $[\boldsymbol{P}(f)]_{1,10}$ (thick lines) and its associated noisy estimates from $[\widehat{\boldsymbol{P}}_0(f_r)]_{1,10}$ (thin lines).

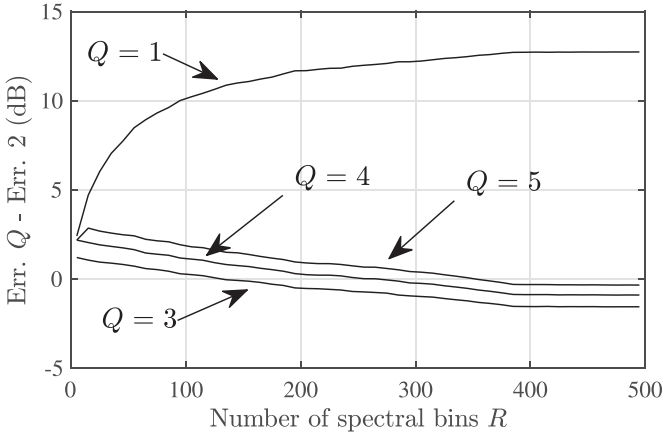


Fig. 2. Result of fitting a polynomial of order $Q=3$ to $R=41$ equally-spaced estimates $\mathrm{Re}\{[\widehat{\boldsymbol{P}}_0(f_r)]_{1,10}\}$.

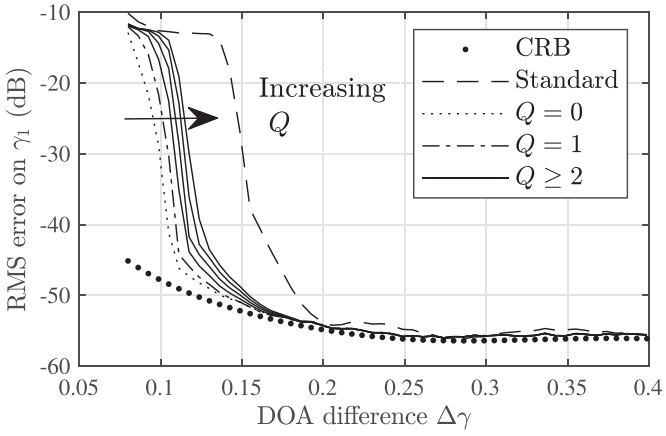


Fig. 3. Error in approximating $\mathrm{Re}\{[\boldsymbol{P}(f)]_{1,10}\}$ using either $\mathrm{Re}\{[\widehat{\boldsymbol{P}}_1(f)]_{1,10}\}$ or $\mathrm{Re}\{[\widehat{\boldsymbol{P}}_2(f)]_{1,10}\}$.

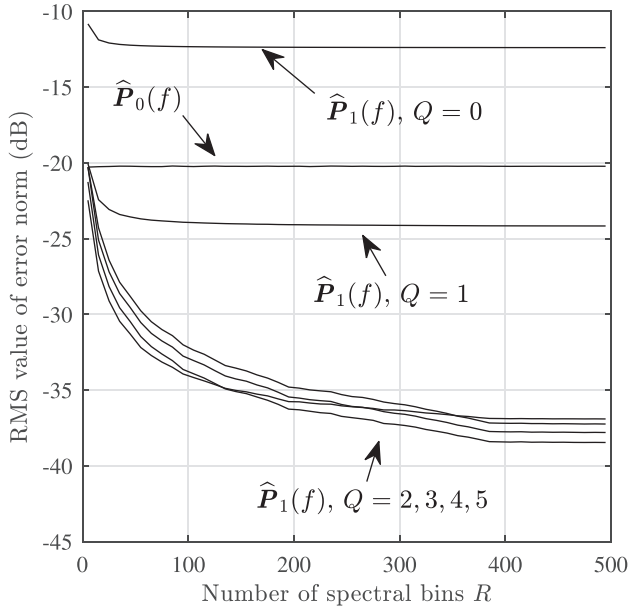


Fig. 4. RMS value of error norm in (52) versus the number of spectral bins R.

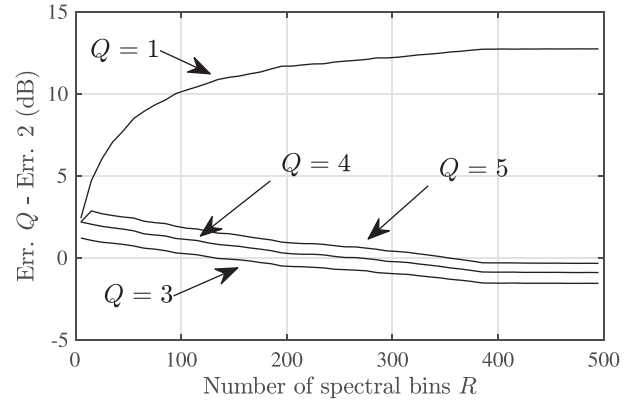
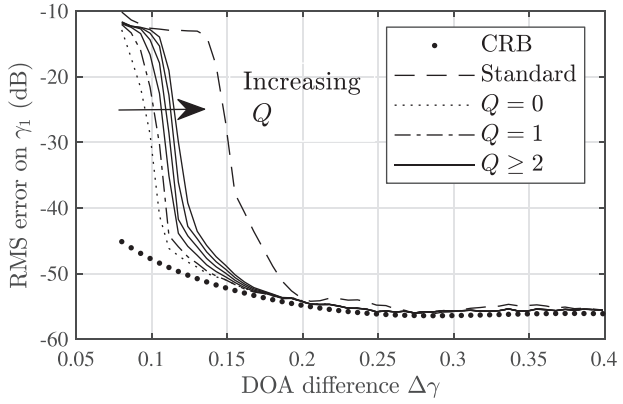


Fig. 5. Difference in dBs between the RMS error of $\widehat{\boldsymbol{P}}_1(f)$ for $Q=2$ and the same error for $Q\neq 2$ versus the number of covariance matrices R.

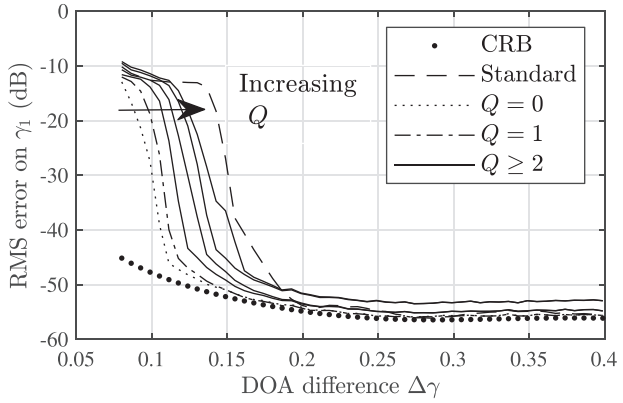
Fig. 1 shows the real and imaginary parts of component $[\boldsymbol{P}(f)]_{1,10}$ and the state-of-the-art estimate $[\widehat{\boldsymbol{P}}_0(f_r)]_{1,10}$ in one realization of the numerical example for SNR = 8 dB. The smooth thick curves are the components of $[\boldsymbol{P}(f)]_{1,10}$ and the noisy thin curves the corresponding components of $[\widehat{\boldsymbol{P}}_0(f_r)]_{1,10}$. Note that $[\widehat{\boldsymbol{P}}_0(f_r)]_{1,10}$ approximates $[\boldsymbol{P}(f)]_{1,10}$ but with some error, fundamentally due to the variation of the received signals' sample spectra. Though this figure only shows one component of $\boldsymbol{P}(f)$ and $\widehat{\boldsymbol{P}}_0(f_r)$, the trends also hold for the rest of components, i.e, the whole matrix $\boldsymbol{P}(f)$ is a smooth function of f and $\boldsymbol{P}(f_r)\approx\widehat{\boldsymbol{P}}_0(f_r)$.

Fig. 2 shows the result of fitting a polynomial of order $Q=3$ to the real part of $R=41$ equally-spaced samples $\mathrm{Re}\{[\widehat{\boldsymbol{P}}_0(f_r)]_{1,10}\}$. The continuous curve is the true value $[\boldsymbol{P}(f)]_{1,10}$ and the dashed curve the fitted value $[\widehat{\boldsymbol{P}}_1(f)]_{1,10}$. Note that the fitted value is a significantly better estimate of $[\boldsymbol{P}(f)]_{1,10}$ along the frequency band than the initial estimates (dots).

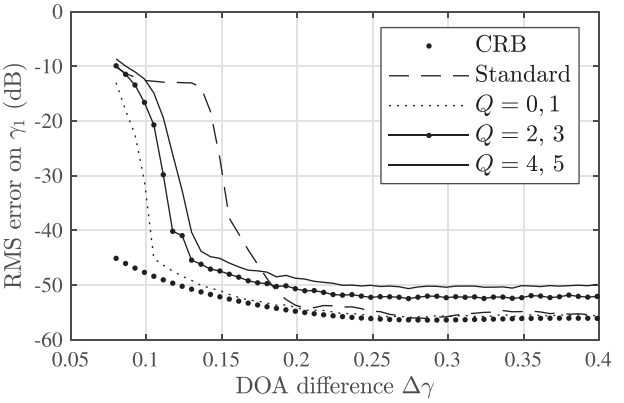
The final correction in (33) for obtaining $\widehat{\boldsymbol{P}}_2(f)$ from $\widehat{\boldsymbol{P}}_1(f)$ produces a slight variation, that can be readily seen in Fig. 3 for the real part of component (1,10). This figure shows the error in approximating $\boldsymbol{P}(f)$ using either $\widehat{\boldsymbol{P}}_1(f)$ or $\widehat{\boldsymbol{P}}_2(f)$ for component (1,10). Note that the curve is smooth for $\widehat{\boldsymbol{P}}_1(f)$ and $\widehat{\boldsymbol{P}}_2(f)$ and that the approximation error is small in both cases.



(a) $R'=41$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.



(b) $R'=5$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.



(c) $R'=1$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.

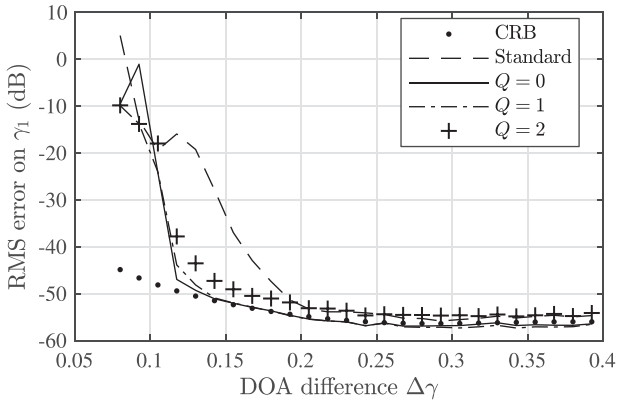
Fig. 6. RMS DOA error of IC-MUSIC in the estimation of γ_1 for the DOAs in (48) versus the DOA difference $\Delta\gamma$.

7.4. RMS approximation error of $\widehat{\boldsymbol{P}}_1(f)$ versus the number of sample covariance matrices R

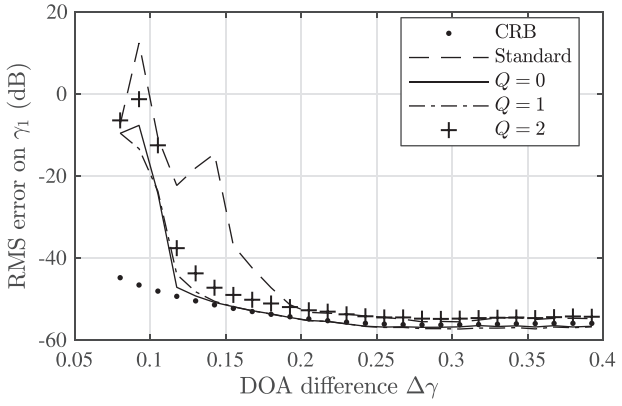
Fig. 4 shows the approximation error for the whole projection matrix in the example of the previous sub-section, where the error norm is

$$\left(\frac{1}{R}\sum_{r=1}^{R}\|\boldsymbol{P}(f_r)-\widehat{\boldsymbol{P}}_1(f_r)\|_2^2\right)^{1/2}. \tag{52}$$

Note that, except for $Q=0$, $\widehat{\boldsymbol{P}}_1(f)$ outperforms $\widehat{\boldsymbol{P}}_0(f)$ by a significant margin that can reach 17 dB for a high number of bins R.



(a) $R' = 41$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.



(b) $R' = 5$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.

Fig. 7. RMS DOA error of MTOPS in the estimation of γ_1 for the DOAs in (48) versus the DOA difference $\Delta\gamma$.

Fig. 5 shows the curves in Fig. 4 for $Q \neq 2$ minus the curve for $Q = 2$ in dBs. This figure allows us to see what value of Q performs best versus the number of bins R. As can be readily seen, $Q = 1$ is the best choice up to $R = 10$ (value below zero in "$Q = 1$" curve), while $Q = 2$ is the best choice between $R = 11$ and $R = 150$, and $Q = 3$ is the best choice for $R > 150$.

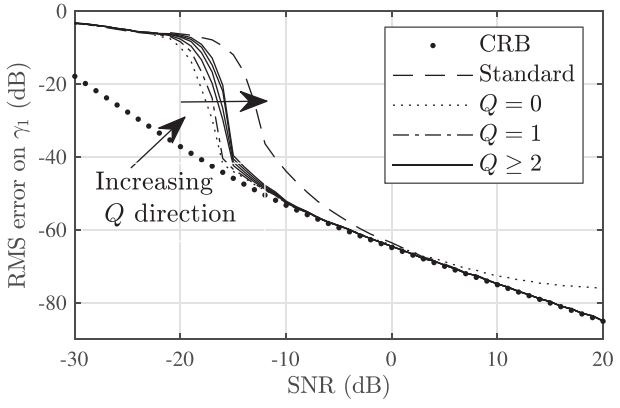
7.5. Improvement in DOA separation

In order to assess the effect of the proposed method on the separation of DOA estimates, we have evaluated the RMS error in the estimation of γ_1 for varying $\Delta\gamma$ in (48) using several variants of IC-MUSIC and MTOPS for SNR $= -10$ dB. The variants were the following:

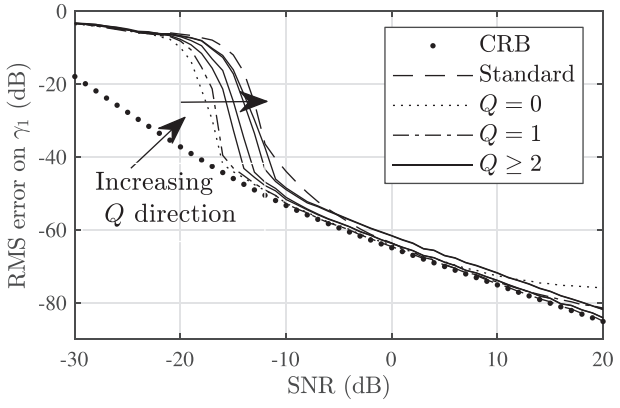
- **Standard IC-MUSIC**. Standard IC-MUSIC estimator using the pseudo-spectrum in (41).
- **41-bin IC-MUSIC**. Standard IC-MUSIC estimator but with projection matrices $\widehat{\boldsymbol{P}}_0(f_r)$ replaced with their estimates $\widehat{\boldsymbol{P}}_2(f_r)$.
- **5-bin IC-MUSIC**. IC-MUSIC estimator applied to $R' = 5$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$ as proposed at the end of Section 5. The frequencies f_r' formed a regular grid covering the band $[f_o - B_1/2, f_o + B_1/2]$ and the IC-MUSIC pseudo-spectrum was (41) but with frequencies f_r', $r = 1, 2, \ldots, R'$.
- **1-bin IC-MUSIC**. The same estimator but computed from the single projection matrix $\widehat{\boldsymbol{P}}_2(f_o)$, i.e, with $f_1' = f_o$ and $R' = 1$.

And the MTOPS estimators were:

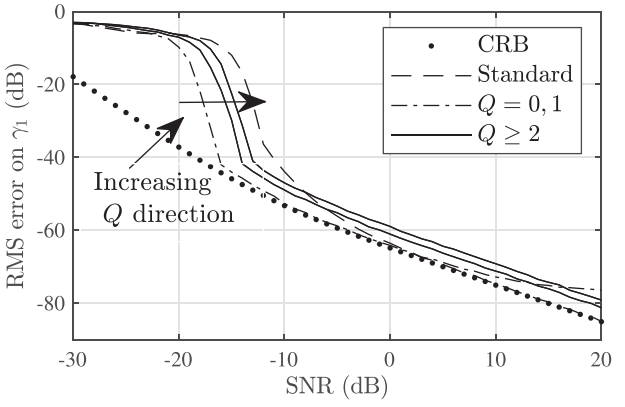
- **Standard MTOPS**. MTOPS estimator computed from the pseudo-spectrum in (44). Since this pseudo-spectrum has numerous local peaks, the peak lying closest to 1-bin IC-MUSIC was selected as estimate. (See [20] for comments on this drawback of MTOPS.)
- **41-bin MTOPS**. Standard MTOPS estimator but with projection matrices $\widehat{\boldsymbol{P}}_0(f_r)$ replaced with their estimates $\widehat{\boldsymbol{P}}_2(f_r)$.
- **5-bin MTOPS**. MTOPS estimator applied to $R' = 5$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$, (Section 5), where the frequencies f_r' formed a regular grid covering the band $[f_o - B_1/2, f_o + B_1/2]$.



(a) $R' = 41$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.



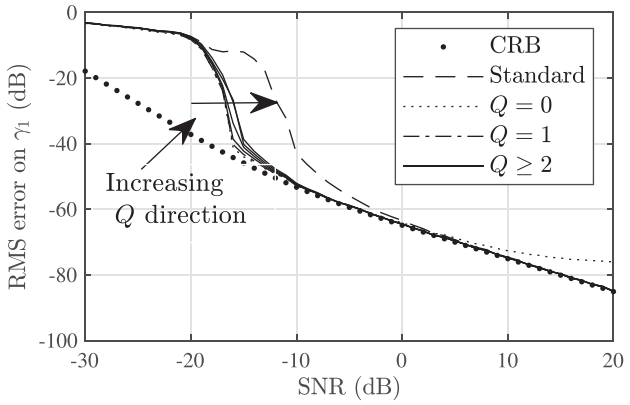
(b) $R' = 5$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.



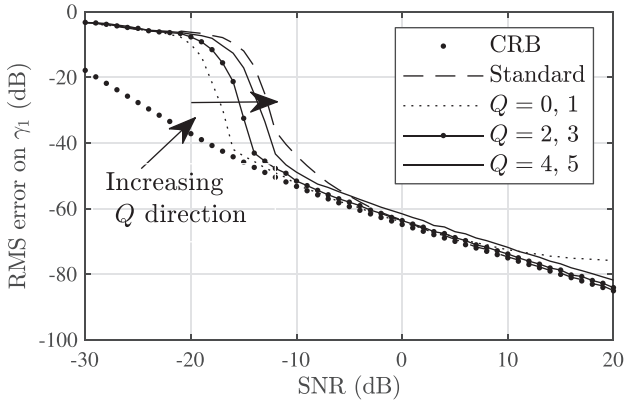
(c) $R' = 1$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.

Fig. 8. RMS DOA error of IC-MUSIC in the estimation of γ_1 for the DOAs in (47) versus the SNR.

Fig. 6a shows the performance of 41-bin IC-MUSIC versus that of standard IC-MUSIC. Note that the proposed method improves the smallest $\Delta\gamma$ for which the two DOAs are separable. With the



(a) $R' = 41$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.



(b) $R' = 5$ projection matrices $\widehat{\boldsymbol{P}}_2(f_r')$.

Fig. 9. RMS DOA error of MTOPS in the estimation of γ_1 for the DOAs in (47) versus the SNR.

standard estimator, we have $\Delta\gamma = 0.15$, while the adapted method with $Q = 1$ provides $\Delta\gamma = 0.11$ roughly. Larger values of Q imply a larger threshold for $\Delta\gamma$. We can explain this feature by the fact that a smaller Q implies that less noise gets into the computation of $\widehat{\boldsymbol{P}}_1(f)$. However, if Q is too small then we may expect a bias due to the model mismatch. As shown in the next sub-section, this bias becomes visible at high SNRs. Fig. 6b shows the same comparison for 5-bin IC-MUSIC and the behavior is the same, but with a more noticeable degradation with increasing Q. It is interesting to note that the best threshold is obtained for $Q = 0$ and that it degrades with increasing Q. This phenomenon can be explained by the fact that less noise gets into the estimate $\widehat{\boldsymbol{P}}_2(f)$ if Q is decreased. However, note that the SNR is fixed at -10 dB. At higher SNRs, a low value of Q introduces a bias due to the model mismatch; (see curves for $Q = 0$ at high SNRs in the next section).

Figs. 6c, 7a and b show the same comparison for 5-bin IC-MUSIC, 41-bin MTOPS and 5-bin MTOPS, respectively, and the conclusions that can be drawn are similar.

7.6. Performance improvement in SNR threshold

We have assessed the RMS error of γ_1 for varying SNR for the three DOAs in (47). Fig. 8a to c show the performance of 41-bin, 5-bin and 1-bin IC-MUSIC respectively, while Figs. 9a and b show the performance of 41-bin MTOPS and 5-bin MTOPS respectively. Note that the behavior is similar to the one in the previous sub-section, i.e, the method provides an improvement of roughly 7 dBs in the SNR threshold. Additionally, we can see that the mismatch for the polynomial approximation can be perceived in the high-SNR region and $Q = 0$ in all figures. Additionally, this mismatch can be seen for $Q = 1$ in Fig. 8c.

8. Conclusions

We have presented a method for estimating the signal subspace along a frequency band from a set of frequency-bin sample covariance matrices. Fundamentally, the method consists of fitting a polynomial to the signal projection matrices associated with these last matrices. The resulting polynomial provides an approximate signal projection matrix at any frequency in the band, which can then be used to improve the quality of wideband DOA estimators such as IC-MUSIC and MTOPS. We have presented asymptotic bounds for the bias and RMS error of the polynomial estimate and we have assessed its performance in several numerical examples.

Credit author statement

J. Selva: All credits correspond to this (the only) author

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Assessment of the first two moments of $\hat{\boldsymbol{P}}_0(f_r)$

In this appendix, we present asympotic expressions for the mean and for a bound on the variance of $\widehat{\boldsymbol{P}}_0(f_r)$ which prove (13) and (15). For simplicity, let us suppress the dependency on f_r in writing in the sequel, i.e, let us write $\boldsymbol{R}$ rather than $\boldsymbol{R}(f_r)$ and $\widehat{\boldsymbol{R}}$ rather than $\widehat{\boldsymbol{R}}(f_r)$, etc. Let us start by recalling the existing results on the perturbation of the eigenvalues of a sample covariance matrix [25,26]. For this, write the eigenvalue decomposition of $\boldsymbol{R}$ as

$$\boldsymbol{R} = \sum_{m=1}^{M} \lambda_m \boldsymbol{q}_m \boldsymbol{q}_m^H, \tag{A.1}$$

where λ_m is the mth eigenvalue in decreasing order and $\boldsymbol{q}_m$ is a corresponding eigenvector, $\boldsymbol{q}_m^H \boldsymbol{q}_{m'} = \delta_{m-m'}$, $(m,\ m' = 1, \ldots, M)$. Due to the model in (4), but applied to the rth frequency bin, we have $\lambda_m = \sigma^2$, $m = K+1, \ldots, M$. For the sample covariance matrix $\widehat{\boldsymbol{R}}$ in (6), but again applied to the rth bin, this same decomposition takes the form

$$\widehat{\boldsymbol{R}} = \sum_{m=1}^{M} \hat{\lambda}_m \hat{\boldsymbol{q}}_m \hat{\boldsymbol{q}}_m^H, \tag{A.2}$$

where "$\hat{\ }$" marks the estimates of the corresponding parameters. Let $\hat{\boldsymbol{\epsilon}}_m$ denote the estimation error for $\boldsymbol{q}_m$, i.e,

$$\hat{\boldsymbol{q}}_m = \boldsymbol{q}_m + \hat{\boldsymbol{\epsilon}}_m. \tag{A.3}$$

In order to recall the asymptotic expressions for the first two moments of $\hat{\epsilon}_m$, define the coefficients

$$b_{m,\ell} \equiv \begin{cases} \dfrac{\lambda_m \lambda_\ell}{N(\lambda_m - \lambda_\ell)^2} & \text{if } m \neq \ell \\ 0 & \text{otherwise} \end{cases} \tag{A.4}$$

From Eqs. (12) to (14) in [26], we have the following asymptotic expressions, where "$\cong$" denotes an $o(1/N)$ equality $(m,\ m' = 1,\ 2, \ldots, M)$:

$$\mathcal{E}\{\hat{\boldsymbol{\epsilon}}_m\} \cong \left(-\frac{1}{2} \sum_{\ell=1}^{M} b_{m,\ell} \right) \boldsymbol{q}_m, \tag{A.5}$$

$$\mathcal{E}\left\{\hat{\boldsymbol{\epsilon}}_m\hat{\boldsymbol{\epsilon}}_{m'}^H\right\} \cong \delta_{m-m'}\sum_{\ell=1}^{M} b_{m,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H, \tag{A.6}$$

$$\mathcal{E}\left\{\hat{\boldsymbol{\epsilon}}_m\hat{\boldsymbol{\epsilon}}_{m'}^T\right\} \cong -(1-\delta_{m-m'})b_{m,m'}\boldsymbol{q}_{m'}\boldsymbol{q}_m^T. \tag{A.7}$$

In App. Appendix B, we use these expressions to prove the formula

$$\mathcal{E}\{\widehat{\boldsymbol{P}}_0\} = \boldsymbol{P} + \boldsymbol{Q}\boldsymbol{\Lambda}_b\boldsymbol{Q}^H + o(1/N), \tag{A.8}$$

where $[\boldsymbol{Q}]_{\cdot,m} \equiv q_m$, $(m = 1, 2, \ldots, M)$ and $\boldsymbol{\Lambda}_b$ is the diagonal matrix

$$[\boldsymbol{\Lambda}_b]_{m,m'} \equiv \begin{cases} -\delta_{m-m'}\sum_{m=K+1}^{M} b_{m,\ell} & \text{if } m \le K \\ \delta_{m-m'}\sum_{k=1}^{K} b_{k,\ell} & \text{if } K < m \le M. \end{cases} \tag{A.9}$$

Since $b_{k,\ell}$ is $O(1/N)$, (A.8) implies that $\widehat{\boldsymbol{P}}_0$ is asymptotically unbiased provided there is some separation between the signal and noise subspaces, i.e, $\lambda_K - \lambda_{K+1} > 0$. To see this last point, note that all coefficients $b_{k,\ell}$ in (A.9) depend on one signal eigenvalue and one noise eigenvalue and, therefore, we have

$$b_{k,\ell} = \frac{\lambda_k\lambda_\ell}{N(\lambda_k-\lambda_\ell)^2} \le \frac{\lambda_1\lambda_{K+1}}{N(\lambda_K-\lambda_{K+1})^2}. \tag{A.10}$$

The variance of the components of $\widehat{\boldsymbol{P}}_0$ can be computed using the results in [25,26] through a laborious derivation and the result seems to depend on $\boldsymbol{P}$. However, we may proceed in an indirect way by bounding $\|\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}\|_F^2$, given that

$$\max_{m,m'}\left|[\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}]_{m,m'}\right|^2 \le \|\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}\|_F^2. \tag{A.11}$$

In App. Appendix C, we prove the following asymptotic bound

$$\mathcal{E}\{\|\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}\|_F^2\} \le \frac{8}{N(\lambda_K-\lambda_{K+1})^2}\sum_{k=1}^{K}\lambda_k\sum_{\ell=1}^{M-K}\lambda_{K+\ell} + o\left(\frac{1}{N}\right). \tag{A.12}$$

Recalling that $\lambda_m = \sigma^2$ if $K < m \le M$, we may combine the last two inequalities to obtain a bound on the quadratic error for any component (m, m'),

$$\mathcal{E}\left\{\left|[\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}]_{m,m'}\right|^2\right\} \le \frac{8(M-K)\sigma^2}{N(\lambda_K-\sigma^2)^2}\sum_{k=1}^{K}\lambda_k + o\left(\frac{1}{N}\right). \tag{A.13}$$

Since, from (A.8), the mean of $[\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}]_{m,m'}$ is $O(1/N)$, (A.13) is also valid for the variance, i.e,

$$\mathrm{Var}\{[\widehat{\boldsymbol{P}}_0 - \boldsymbol{P}]_{m,m'}\} \le \frac{8(M-K)\sigma^2}{N(\lambda_K-\sigma^2)^2}\sum_{k=1}^{K}\lambda_k + o\left(\frac{1}{N}\right). \tag{A.14}$$

Appendix B. First- and second-order moments of sample covariance matrix eigenvectors

Let us first compute the first- and second-order moments of $\hat{q}_m$. From (A.5), we have

$$\mathcal{E}\left\{\hat{\boldsymbol{q}}_m\right\} \cong \boldsymbol{q}_m + \left(-\frac{1}{2}\sum_{\ell=1}^{M} b_{m,\ell}\right)\boldsymbol{q}_m = c_m\boldsymbol{q}_m, \tag{B.1}$$

where we define the coefficients

$$c_m \equiv 1 - \frac{1}{2}\sum_{\ell=1}^{M} b_{m,\ell}. \tag{B.2}$$

From (A.6) and (B.1), we have

$$\begin{aligned}\mathcal{E}\left\{\hat{\boldsymbol{q}}_m\hat{\boldsymbol{q}}_{m'}^H\right\} &= \mathcal{E}\left\{\left(\boldsymbol{q}_m + \hat{\boldsymbol{\epsilon}}_m\right)\left(\boldsymbol{q}_{m'} + \hat{\boldsymbol{\epsilon}}_{m'}\right)^H\right\} \\ &= \boldsymbol{q}_m\boldsymbol{q}_{m'}^H + \boldsymbol{q}_m E\left\{\hat{\boldsymbol{\epsilon}}_{m'}^H\right\} + E\left\{\hat{\boldsymbol{\epsilon}}_m\right\}\boldsymbol{q}_{m'}^H + E\left\{\hat{\boldsymbol{\epsilon}}_m\hat{\boldsymbol{\epsilon}}_{m'}^H\right\} \\ &= (1 + c_m + c_{m'})\boldsymbol{q}_m\boldsymbol{q}_{m'}^H + \delta_{m-m'}\sum_{\ell=1}^{M} b_{m,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H. \end{aligned} \tag{B.3}$$

If $m = m'$ this expression reduces to

$$\mathcal{E}\left\{\hat{\boldsymbol{q}}_m\hat{\boldsymbol{q}}_m^H\right\} = (1 + 2c_m)\boldsymbol{q}_m\boldsymbol{q}_m^H + \sum_{\ell=1}^{M} b_{m,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H. \tag{B.4}$$

Adding up (B.4) for $m = 1, 2, \ldots, K$, we obtain a formula for $\mathcal{E}\{\widehat{\boldsymbol{P}}_0\}$, noting that $\boldsymbol{P} = \sum_{k=1}^{K}\boldsymbol{q}_k\boldsymbol{q}_k^H$:

$$\begin{aligned}\mathcal{E}\left\{\hat{\boldsymbol{P}}_0\right\} &\cong \sum_{k=1}^{K}\left((1 + 2c_k)\boldsymbol{q}_k\boldsymbol{q}_k^H + \sum_{\ell=1}^{M} b_{k,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H\right) \\ &= \sum_{k=1}^{K}\left(\left(1 - \sum_{\ell=1}^{M} b_{k,\ell}\right)\boldsymbol{q}_k\boldsymbol{q}_k^H + \sum_{\ell=1}^{M} b_{k,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H\right) \\ &= \boldsymbol{P} - \sum_{k=1}^{K}\sum_{\ell=1}^{M} b_{k,\ell}\boldsymbol{q}_k\boldsymbol{q}_k^H + \sum_{k=1}^{K}\sum_{\ell=1}^{M} b_{k,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H. \end{aligned} \tag{B.5}$$

And doing the same for $m = K+1, \ldots, M$, we obtain

$$\begin{aligned}\mathcal{E}\left\{\boldsymbol{I}_M - \hat{\boldsymbol{P}}_0\right\} &\cong \boldsymbol{I}_M - \boldsymbol{P} \\ &\quad - \sum_{m=K+1}^{M}\sum_{\ell=1}^{M} b_{m,\ell}\boldsymbol{q}_m\boldsymbol{q}_m^H + \sum_{m=K+1}^{M}\sum_{\ell=1}^{M} b_{m,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H \end{aligned} \tag{B.6}$$

The last two formulas involve coefficients $b_{k,\ell}$ in which both k and ℓ are either smaller than $K+1$ or larger than K, i.e, coefficients computed from pairs of eigenvalues associated to either the signal or noise subspace. Such coefficients $b_{k,\ell}$ can be arbitrarily large, given that the only condition on the eigenvalues λ_m is that there is a significant gap $\lambda_K - \lambda_{K+1}$ between the signal and noise eigenvalues. We solve this drawback by writing $\mathcal{E}\{\widehat{\boldsymbol{P}}_0\}$ in terms of $\mathcal{E}\{(\boldsymbol{I}_M - \boldsymbol{P})\widehat{\boldsymbol{P}}_0\}$ and $\mathcal{E}\{\boldsymbol{P}(\boldsymbol{I}_M - \widehat{\boldsymbol{P}}_0)\}$. We have from (B.5),

$$\mathcal{E}\left\{(\boldsymbol{I}_M - \boldsymbol{P})\hat{\boldsymbol{P}}_0\right\} \cong \sum_{k=1}^{K}\sum_{\ell=K+1}^{M} b_{k,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H \tag{B.7}$$

and from (B.6),

$$\mathcal{E}\left\{\boldsymbol{P}\left(\boldsymbol{I}_M - \hat{\boldsymbol{P}}_0\right)\right\} \cong \sum_{m=K+1}^{M}\sum_{\ell=1}^{K} b_{m,\ell}\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H. \tag{B.8}$$

Finally, combining the last two equations, we have

$$\begin{aligned}\mathcal{E}\left\{\hat{\boldsymbol{P}}_0\right\} &= \mathcal{E}\left\{\boldsymbol{P} + (\boldsymbol{I}_M - \boldsymbol{P})\hat{\boldsymbol{P}}_0 - \boldsymbol{P}\left(\boldsymbol{I}_M - \hat{\boldsymbol{P}}_0\right)\right\} \\ &= \boldsymbol{P} + \mathcal{E}\left\{(\boldsymbol{I}_M - \boldsymbol{P})\hat{\boldsymbol{P}}_0\right\} - \mathcal{E}\left\{\boldsymbol{P}\left(\boldsymbol{I}_M - \hat{\boldsymbol{P}}_0\right)\right\} \\ &\cong \boldsymbol{P} + \sum_{\ell=K+1}^{M}\left(\sum_{k=1}^{K} b_{k,\ell}\right)\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H \\ &\quad - \sum_{\ell=1}^{K}\left(\sum_{m=K+1}^{M} b_{m,\ell}\right)\boldsymbol{q}_\ell\boldsymbol{q}_\ell^H. \end{aligned} \tag{B.9}$$

The matrix form of this expression is (A.8).

Appendix C. Derivation of bound on expected quadratical error

The eigenvalue decomposition of $\boldsymbol{R}$ in (A.1) can be written as

$$\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H \tag{C.1}$$

where $\boldsymbol{Q}$ is unitary, $[\boldsymbol{Q}]_m \equiv \boldsymbol{q}_m$, and $\boldsymbol{\Lambda}$ is a diagonal matrix with components $[\boldsymbol{\Lambda}]_{m,m'} = \delta_{m-m'}\lambda_m$, $m, m' = 1, 2, \ldots, M$. The sample

covariance matrix $\widehat{\boldsymbol{R}}$ can be described as the average of N independent, complex normal $M\times 1$ vectors $\boldsymbol{x}_n$, of zero mean and covariance $\boldsymbol{R}$,

$$\widehat{\boldsymbol{R}}=\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^H. \tag{C.2}$$

Next, consider the vectors $\boldsymbol{s}_n\equiv\boldsymbol{Q}^H\boldsymbol{x}_n$ which are also independent, complex normal and of zero mean, but of covariance matrix $\boldsymbol{\Lambda}$. The sample covariance matrix of these vectors is

$$\widehat{\boldsymbol{C}}\equiv\frac{1}{N}\sum_{n=1}^{N}\widehat{\boldsymbol{s}}_n\widehat{\boldsymbol{s}}_n^H. \tag{C.3}$$

where $\widehat{\boldsymbol{s}}_n\equiv\boldsymbol{Q}^H\widehat{\boldsymbol{x}}_n$. Given a realization $\widehat{\boldsymbol{R}}$ and its corresponding $\widehat{\boldsymbol{C}}$, let us bound the error in approximating $\boldsymbol{P}$ using $\widehat{\boldsymbol{P}}_0$ by resorting to a perturbation theory result in [29]. First, define the following measure for the dissimilarity between the spans of $\boldsymbol{P}$ and $\widehat{\boldsymbol{P}}_0$:

$$\widehat{\eta}\equiv\frac{\|\widehat{\boldsymbol{C}}_{sn}\|_F^2}{(\lambda_K-\lambda_{K+1})^2}, \tag{C.4}$$

where $\widehat{\boldsymbol{C}}_{sn}$ is the block formed by the intersection of the first K rows and last $M-K$ columns of $\widehat{\boldsymbol{C}}$, i.e, $\widehat{\boldsymbol{C}}_{sn}\equiv[\widehat{\boldsymbol{C}}]_{1:K,K+1:M}$. Combining theorems 2.1 and 3.1 of [29], we have that if $\widehat{\eta}<1/4$ then

$$\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2\leq 8\,\widehat{\eta}. \tag{C.5}$$

(See also comments on page 232 of [29].)

In order to turn (C.5) into an asymptotic inequality, let us first compute the first two moments of the components of $\widehat{\boldsymbol{C}}_{sn}$ and then the mean and variance of $\widehat{\eta}$. For simplicity, let $c_{k,\ell}$ and $s_{n,k}$ denote $[\widehat{\boldsymbol{C}}]_{k,\ell}$ and $[\boldsymbol{s}_n]_k$ respectively. For any indices k, ℓ, p and q, lying between 1 and M and following $k\neq\ell$ and $p\neq q$, we have:

- $\mathcal{E}\{c_{k,\ell}\}=0$ given that $k\neq\ell$ and
$$\mathcal{E}\{c_{k,\ell}\}=[\mathcal{E}\{\widehat{\boldsymbol{C}}\}]_{k,\ell}=[\boldsymbol{\Lambda}]_{k,\ell}=0. \tag{C.6}$$

- The formula
$$\mathcal{E}\{c_{k,\ell}c_{p,q}\}=\frac{1}{N}\delta_{k-q}\delta_{\ell-p}\lambda_k\lambda_\ell. \tag{C.7}$$
To prove this result, recall the formula for the expectation of the product of four complex normal random variables a_k, [30]:
$$\begin{aligned}\mathcal{E}\{a_1a_2a_3a_4\}&=\mathcal{E}\{a_1a_2\}\mathcal{E}\{a_3a_4\}+\mathcal{E}\{a_1a_3\}\mathcal{E}\{a_2a_4\}\\&\quad+\mathcal{E}\{a_1a_4\}\mathcal{E}\{a_2a_3\}-\mathcal{E}\{a_1\}\mathcal{E}\{a_2\}\mathcal{E}\{a_3\}\mathcal{E}\{a_4\}.\end{aligned} \tag{C.8}$$
The proof is the following. We have
$$\begin{aligned}\mathcal{E}\{c_{k,\ell}c_{p,q}\}&=\mathcal{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{s}_n\boldsymbol{s}_n^H\right]_{k,\ell}\left[\frac{1}{N}\sum_{n'=1}^{N}\boldsymbol{s}_{n'}\boldsymbol{s}_{n'}^H\right]_{p,q}\right\}\\&=\frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\mathcal{E}\{s_{n,k}s_{n,\ell}^*s_{n',p}s_{n',q}^*\}\\&=\frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\Big(\mathcal{E}\{s_{n,k}s_{n,\ell}^*\}\mathcal{E}\{s_{n',p}s_{n',q}^*\}\\&\quad+\mathcal{E}\{s_{n,k}s_{n',p}\}\mathcal{E}\{s_{n,\ell}^*s_{n',q}^*\}+\mathcal{E}\{s_{n,k}s_{n',q}^*\}\mathcal{E}\{s_{n,\ell}^*s_{n',p}\}\\&\quad-\mathcal{E}\{s_{n,k}\}\mathcal{E}\{s_{n,\ell}\}\mathcal{E}\{s_{n',p}\}\mathcal{E}\{s_{n',q}\}\Big)\end{aligned} \tag{C.9}$$
In this parenthesis, we have:
  - The first term is zero because $\mathcal{E}\{s_{n,k}s_{n,\ell}^*\}=[\boldsymbol{\Lambda}]_{k,\ell}=0$.
  - If $n\neq n'$ the second term is zero because $\boldsymbol{s}_n$ and $\boldsymbol{s}_{n'}$ are independent and $\mathcal{E}\{\boldsymbol{s}_n\}=0$. If $n=n'$ this term is zero because $\mathcal{E}\{\boldsymbol{s}_n\boldsymbol{s}_n^T\}=\boldsymbol{0}_M$.
  - The third term is zero if $n\neq n'$ because $\boldsymbol{s}_n$ is independent of $\boldsymbol{s}_{n'}$ and $\mathcal{E}\{\boldsymbol{s}_n\}=0$. And, if $n=n'$, then it is also zero if $k\neq q$ or $\ell\neq p$, because $\mathcal{E}\{\boldsymbol{s}_n\boldsymbol{s}_n^H\}=\boldsymbol{\Lambda}$ is a diagonal matrix. Thus, we have that the third term is equal to
$$\begin{aligned}&\delta_{n-n'}\delta_{k-q}\delta_{\ell-p}\mathcal{E}\{s_{n,k}s_{n,k}^*\}\mathcal{E}\{s_{n,\ell}s_{n,\ell}^*\}\\&\quad=\delta_{n-n'}\delta_{k-q}\delta_{\ell-p}\lambda_k\lambda_\ell.\end{aligned} \tag{C.10}$$
  - The fourth term is zero because $\mathcal{E}\{\boldsymbol{s}_n\}=0$.

So, in summary, we have
$$\begin{aligned}\mathcal{E}\{c_{k,\ell}c_{p,q}\}&=\frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\delta_{n-n'}\delta_{k-q}\delta_{\ell-p}\lambda_k\lambda_\ell\\&=\frac{1}{N}\delta_{k-q}\delta_{\ell-p}\lambda_k\lambda_\ell.\end{aligned} \tag{C.11}$$

- We also have
$$\mathcal{E}\{c_{k,\ell}c_{p,q}^*\}=\frac{1}{N}\delta_{k-p}\delta_{\ell-q}\lambda_k\lambda_\ell. \tag{C.12}$$
The proof is the following. We have
$$\begin{aligned}\mathcal{E}\left\{c_{k,\ell}c_{p,q}^*\right\}&=\mathcal{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{s}_n\boldsymbol{s}_n^H\right]_{k,\ell}\left[\frac{1}{N}\sum_{n'=1}^{N}\boldsymbol{s}_{n'}^*\boldsymbol{s}_{n'}^T\right]_{p,q}\right\}\\&=\frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\mathcal{E}\left\{s_{n,k}s_{n,\ell}^*s_{n',p}^*s_{n',q}\right\}\\&=\frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\mathcal{E}\left\{s_{n,k}s_{n,\ell}^*s_{n',q}s_{n',p}^*\right\}\end{aligned} \tag{C.13}$$
Comparing this expression with the second line of (C.9), we can readily see that
$$\mathcal{E}\{c_{k,\ell}c_{p,q}^*\}=\mathcal{E}\{c_{k,\ell}c_{q,p}\}. \tag{C.14}$$
Therefore, from (C.11), we obtain (C.12).

Let us now compute the mean of $\widehat{\eta}$. From (C.4) and (C.12), we have
$$\begin{aligned}\mathcal{E}\{\widehat{\eta}\}&=\frac{1}{(\lambda_K-\lambda_{K+1})^2}\sum_{k=1}^{K}\sum_{\ell=K+1}^{M}\mathcal{E}\{c_{k,\ell}c_{k,\ell}^*\}\\&=\frac{1}{N(\lambda_K-\lambda_{K+1})^2}\sum_{k=1}^{K}\sum_{\ell=K+1}^{M}\lambda_k\lambda_\ell.\end{aligned} \tag{C.15}$$
Regarding the second-order moment, it follows the formula:
$$\begin{aligned}\mathcal{E}\{\widehat{\eta}^2\}&=\frac{1}{(\lambda_K-\lambda_{K+1})^4}\\&\quad\cdot\sum_{k=1}^{K}\sum_{\ell=K+1}^{M}\sum_{k'=1}^{K}\sum_{\ell'=K+1}^{M}\mathcal{E}\{c_{k,\ell}c_{k,\ell}^*c_{k',\ell'}^*c_{k',\ell'}\}.\end{aligned} \tag{C.16}$$
Expanding the summand using (C.8), we have
$$\begin{aligned}\mathcal{E}\{c_{k,\ell}c_{k,\ell}^*c_{k',\ell'}^*c_{k',\ell'}\}&=\mathcal{E}\{c_{k,\ell}c_{k,\ell}^*\}\mathcal{E}\{c_{k',\ell'}^*c_{k',\ell'}\}\\&\quad+\mathcal{E}\{c_{k,\ell}c_{k',\ell'}^*\}\mathcal{E}\{c_{k,\ell}^*c_{k',\ell'}\}+\mathcal{E}\{c_{k,\ell}c_{k',\ell'}\}\mathcal{E}\{c_{k,\ell}^*c_{k',\ell'}^*\}.\end{aligned} \tag{C.17}$$
Note that this is a sum of products consisting of factors of the form in either (C.7) or (C.12). As a consequence, all these products are $O(1/N^2)$ and we have
$$\mathcal{E}\left\{\widehat{\eta}^2\right\}=O\left(1/N^2\right). \tag{C.18}$$
Finally, let us derive the asymptotic inequality. Start by decomposing the expectation of $\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2$ by conditioning on $\widehat{\eta}^2<1/4$:

$$\begin{aligned}\mathcal{E}\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2\} &= \mathcal{E}\Big\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2 \mid \widehat{\eta}^2\le 1/4\Big\}\Big(1-\mathcal{P}(\widehat{\eta}^2>1/4)\Big)\\ &\quad+\mathcal{E}\Big\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2 \mid \widehat{\eta}^2>1/4\Big\}\mathcal{P}(\widehat{\eta}^2>1/4)\\ &= \mathcal{E}\Big\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2 \mid \widehat{\eta}^2\le 1/4\Big\}\\ &\quad+\Big(\mathcal{E}\Big\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2 \mid \widehat{\eta}^2>1/4\Big\}\\ &\quad-\mathcal{E}\Big\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2 \mid \widehat{\eta}^2\le 1/4\Big\}\Big)\cdot\mathcal{P}(\widehat{\eta}^2>1/4).\end{aligned} \tag{C.19}$$

From (C.5), we have that the first term follows

$$\mathcal{E}\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2 \mid \widehat{\eta}^2\le 1/4\}\le 8\,\mathcal{E}\{\widehat{\eta}\}. \tag{C.20}$$

Regarding the second term, the expectations inside the parenthesis are bounded, because projection matrices have components bounded by one. Besides, we may apply Markov's inequality and use (C.17) to obtain

$$\mathcal{P}(\widehat{\eta}>1/4)=\mathcal{P}(\widehat{\eta}^2>1/16)<16\mathcal{E}\{\widehat{\eta}^2\}=O(1/N^2). \tag{C.21}$$

So, we have that the whole second term in (C.18) is $O(1/N^2)$ and, therefore, is $o(1/N)$. In summary, recalling (C.14), we obtain the asymptotic inequality

$$\mathcal{E}\{\|\widehat{\boldsymbol{P}}_0-\boldsymbol{P}\|_F^2\}\le\frac{8}{N(\lambda_K-\lambda_{K+1})^2}\sum_{k=1}^{K}\sum_{\ell=K+1}^{M}\lambda_k\lambda_\ell+o\Big(\frac{1}{N}\Big). \tag{C.22}$$

References

[1] H.L. van Trees, Detection, Estimation, and Modulation Theory. Part IV Optimum Array Processing, 1st, John Wiley & Sons, Inc,, 2002.
[2] H. Wang, M. Kaveh, Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources, IEEE Trans. Acoust. 33 (4) (1985) 823–831.
[3] S. Valaee, P. Kabal, Wideband array processing using a two-sided correlation transformation, IEEE Trans. Signal Process. 43 (1) (1995) 160–172.
[4] T.K. Yasar, T.E. Tuncer, Wideband DOA estimation for nonuniform linear arrays with Wiener array interpolation, in: in: 2008 5th IEEE Sensor Array and Multichannel Signal Processing Workshop, 2008, pp. 207–211.
[5] W.J. Zeng, X.L. Li, High-resolution multiple wideband and nonstationary source localization with unknown number of sources, IEEE Trans. Signal Process. 58 (6) (2010) 3125–3136.
[6] M. Wax, T.J. Shan, T. Kailath, Spatio-temporal spectral analysis by eigenstructure methods, IEEE Trans. Acoust. 32 (4) (1984) 817–827.
[7] M.A. Doron, A.J. Weiss, H. Messer, Maximum-likelihood direction finding of wide-band sources, IEEE Trans. Signal Process. 41 (1) (1993) 411–414.
[8] L. Yip, J.C. Chen, R.E. Hudson, K. Yao, Cramer-Rao Bound Analysis of Wideband Source Localization and DOA estimation, in: in: International Symposium on Optical Science and Technology, International Society for Optics and Photonics, 2002, pp. 304–316.
[9] J.C. Chen, R.E. Hudson, K. Yao, Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field, IEEE Trans. Signal Process. 50 (8) (2002) 1843–1854.
[10] L. Yip, C.E. Chen, R.E. Hudson, K. Yao, DOA estimation method for wideband color signals based on least-squares joint approximate diagonalization, in: Proceedings of Sensor Array and Multichannel Signal Processing, 2008, pp. 104–107.
[11] J. Selva, Efficient wideband DOA estimation through function evaluation techniques, IEEE Trans. Signal Process. 66 (12) (2018) 3112–3123.
[12] J.G. McWhirter, P.D. Baxter, T. Cooper, S. Redif, J. Foster, An EVD Algorithm for Para-Hermitian Polynomial Matrices, IEEE Trans. Signal Process. 55 (5) (2007) 2158–2169, doi:10.1109/TSP.2007.893222.
[13] M. A. Alrmah, S. Weiss, S. Lambotharan, An extension of the MUSIC algorithm to broadband scenarios using a polynomial Eigenvalue decomposition, in: in: 2011 19th European Signal Processing Conference, 2011, pp. 629–633.
[14] S. Weiss, M. Alrmah, S. Lambotharan, J.G. McWhirter, M. Kaveh, Broadband angle of arrival estimation methods in a polynomial matrix decomposition framework, in: in: 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013, pp. 109–112. Doi:10.1109/CAMSAP.2013.6714019
[15] S. Redif, S. Weiss, J.G. McWhirter, Relevance of polynomial matrix decompositions to broadband blind signal separation, Signal Process. 134 (2017) 76–86, doi:10.1016/j.sigpro.2016.11.019.
[16] P.T. Boufounos, P. Smaragdis, B. Raj, Joint sparsity models for wideband array processing, in: Proceedings SPIE 8138, 2011, pp. 1–10.
[17] Q. Shen, W. Liu, W. Cui, S. Wu, Y.D. Zhang, M.G. Amin, Group sparsity based wideband DOA estimation for co-prime arrays, in: in: 2014 IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP), 2014, pp. 252–256.
[18] Q. Shen, W. Liu, W. Cui, S. Wu, Y.D. Zhang, M.G. Amin, Low-complexity direction-of-arrival estimation based on wideband co-prime arrays, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (9) (2015) 1445–1456.
[19] Y.-S. Yoon, L.M. Kaplan, J.H. McClellan, TOPS: new DOA estimator for wideband signals, IEEE Trans. Signal Process. 54 (6) (2006) 1977–1989.
[20] A.K. Shaw, Improved wideband DOA estimation using modified TOPS (mTOPS) algorithm, IEEE Signal Process. Lett. 23 (12) (2016) 1697–1701.
[21] H. Hayashi, T. Ohtsuki, DOA Estimation for wideband signals based on weighted squared TOPS, EURASIP J. Wirel. Commun. Netw. 2016 (1) (2016).
[22] G.H. Golub, C.F.V. Loan, Matrix Computations, The Johns Hopkins University Press, 1996.
[23] T. Kato, Perturbation Theory for Linear Operators, Springer, 1995.
[24] G.M. Phillips, Interpolation and Approximation by Polynomials, Springer, 2003.
[25] D.R. Brillinger, Classics in Applied Mathematics, Time Series: Data Analysis and Theory, SIAM, 2001.
[26] M. Kaveh, A.J. Barabell, The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise, IEEE Trans. Acoust. Speech Signal Process. ASSP 34 (2) (1986) 331–341.
[27] E. Tuncer, Classical and Modern Direction-of-Arrival Estimation, Elsevier, 2009.
[28] J.C. Mason, D.C. Handscomb, Chebyshev Polynomials, CRC Press, 2002.
[29] G.W. Stewart, J.G. Sun, Matrix Perturbation Theory, Academic Press, Inc., 1990.
[30] P. Janssen, P. Stoica, On the expectation of the product of four matrix-valued Gaussian random variables, IEEE Trans. Automat. Contr. 33 (9) (1988) 867–870.