Low Complexity Subspace Approach for Unbiased Frequency Estimation of a Complex Single-tone

Alireza Pourafzal [a], Pavel Skrabanek [b,*], Michael Cheffena [a], Sule Yildirim [c], Thomas Roi-Taravella [d]

[a] *Faculty of Engineering, Norwegian University of Science and Technology, Gjøvik, Norway*

[b] *Institute of Automation and Computer Science, Brno University of Technology, Brno, Czech Republic*

[c] *The Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway*

[d] *National Graduate School of Engineering & Research Center, Caen, France*

Abstract

We propose a single-tone frequency estimator of a one-dimensional complex signal in complex white Gaussian noise. The estimator is based on the subspace approach and the unitary transformation. Due to its low space and time-complexity, we name the estimator as Low complexity Unitary Principal-singular-vector Utilization for Model Analysis (LUPUMA). Regardless of the observation length, LUPUMA provides a uniform estimation variance over the whole frequency range, while achieving the lowest time-complexity among subspace methods. The proposed estimator asymptotically reaches the Cramér-Rao Lower Bound. For short observations, the signal-to-noise ratio threshold of LUPUMA corresponds to the threshold of the maximum likelihood estimator. The low space and time-complexity along with the stable and state-of-the-art estimation performance for short observations make LUPUMA an ideal candidate for applications with a limited number of signal samples, limited computational power, limited memory, and for applications that require rapid processing time (low latency).

Keywords: Frequency Estimation, Complex Single-Tone, Subspace Method, Short Observation Interval

1. Introduction

Estimation of a single-tone frequency from a finite number of noisy discrete-time observations of a complex sinusoid signal is of great importance among others in telecommunications [1], microwave sensors [2], and power systems [3]. In some applications, the estimation must be based on a limited number of samples. The short observation length can naturally result from physical limitations of measurement systems (channel estimation in fifth-generation communications for high-speed train systems [1], abrupt changes of voltage in a three-phase power system [4], etc.) or it can be subject to hardware limitations (e.g. the low processing power of hardware in Internet-of-Things (IoT) applications, where the devices can only process a limited number of samples [3]). Thus, developing an unbiased frequency estimator over the whole frequency range which complies with the short observation time constraint is a highly relevant challenge.

Let the k-th sample of a received one-dimensional (1D) continuous signal $r(t)$ is given as

$$r[k] = b_0 e^{j\omega k} + q[k], \qquad k = 0,1,\dots,K-1, \tag{1}$$

where t stands for time, b_0 is a constant unknown amplitude, $\omega \in (-\pi,\pi)$ is an unknown frequency, $q[k]$ is the k-th sample of the zero-mean complex white Gaussian noise $q(t)$ with an unknown variance σ^2, $r[k], q[k] \in \mathbb{C}$, and K is the total number of samples. By the single-tone frequency estimation, we are interested in real-time and unbiased estimation of the frequency ω over the whole frequency range regardless of the number of available samples (whether the observation is short ($8 \le K < 256$) or long ($K \ge 256$)). Noting that the variance of an unbiased estimator must be independent of the actual value of the frequency over the whole frequency range.

The information modulated in the frequency of the transmitted signal $r(t)$ can be estimated using a maximum likelihood estimator. The frequency estimation problem is reformed into a non-linear nonconvex multidimensional optimization problem [5]. It has theoretically the optimal performance in terms of the Signal-to-Noise Ratio (SNR) threshold and the estimation accuracy (it attains the Cramér-Rao Lower Bound (CRLB) for a wide SNR range [6]).

However, obtaining the exact solution demands numerical methods with high time-complexities. To reduce the time-complexity of the maximum likelihood estimator, a two-stage approach of coarse search/fine search is employed. In the coarse search, a frequency bin associated with the highest magnitude of the Discrete Fourier Transform (DFT) of the signal is selected. Then, the residual fractional frequency is estimated using dichotomous search or interpolation refinement schemes.

In recent years, interpolation schemes are preferred due to their lower time-complexity and easy implementation [7-11]. An interpolation scheme can be done using direct methods and iterative methods. Iterative DFT-based frequency estimators shift the peak of the DFT coefficient at each iteration until the algorithm converges. Within this class, the A&M algorithm [5] shifts the periodogram around half of the DFT bin resolution, providing the analytical performance of variance $\pi^4/96 \approx 1.015$ of CRLB [5]. By introducing smaller fractions than half, estimation performance can be improved [6]. Nevertheless, the iterative methods suffer from higher time-complexity compared to direct methods [7]. Moreover, each step of iterative methods must be done sequentially and cannot be implemented in a parallel fashion [7]. Furthermore, their refinement scheme is only accurate when there is a large number of samples available [8]. Thus, the iterative methods are not suitable for real-time applications with a limited number of samples.

Direct methods reuse the calculated DFT coefficients in the coarse search to estimate the fine resolution frequency. Within this class of estimators, the Candan estimator [9] has the lowest time-complexity whereas Weighted Least Squares (WLS) estimator [7] has the best estimation performance. The CRLB of frequency estimation based on available DFT coefficients is a function of residual fractional frequency and of the number of reused coefficients [7]. This results in unbiased estimates over the whole frequency range even for a limited number of samples.

Subspace-based estimators such as Principal-singular-vector Utilization for Modal Analysis (PUMA) [10] and Unitary-PUMA [11] use the linear prediction property of sinusoidal signals achieving better frequency resolution than the DFT-based estimators [12]. Even for short observations, PUMA shows uniform estimation performance over the whole frequency range with SNR thresholds comparable with thresholds of the DFT-based estimators [10].

Both PUMA and Unitary-PUMA reduce the effect of the additive noise on the received signal by separating signal and noise subspaces. For this purpose, they reshape the K samples of the received signal $r(t)$ into a received signal matrix

$$\mathbf{R} = \begin{bmatrix} r[0] & r[M] & \dots & r[M(N-1)] \\ r[1] & r[M+1] & \cdots & r[M(N-1)+1] \\ \vdots & \vdots & \ddots & \vdots \\ r[M-1] & r[2M-1] & \dots & r[MN-1] \end{bmatrix}, \tag{2}$$

where the factorization parameters M and N are arbitrary natural numbers satisfying the condition $K = MN$ [10], and $\mathbf{R} \in \mathbb{C}^{M\times N}$.

The first left and right-singular vectors of the rank-one matrix $\mathbf{R}$ obtained using the Singular Values Decomposition (SVD) have a linear prediction property corresponding to the frequency ω [13]. By taking advantage of this property, the PUMA estimator uses WLS to estimate the unknown frequency ω, where the optimal setting of the weights in a weighting matrix is the result of an iterative procedure. To reach the CRLB, PUMA estimates the frequency from the matrix $\mathbf{R}$. PUMA indicates an unbiased estimation with a variance approximately equal to the CRLB for the whole frequency range; however, it suffers from a high time-complexity due to the SVD of the complex matrix $\mathbf{R}$, and the iterative procedure of WLS (PUMA calculates the inverse of the weighting matrix in each iteration to obtain the best linear unbiased estimate [14]).

The PUMA, unlike the DFT-based methods, allows sufficiently accurate estimation of the frequency for short observations ($K < 256$). However, the high time-complexities of the PUMA limit its utilization in applications with low processing power requirements (such as IoT devices) or in applications with real-time data processing requirements. To reach a time-complexity lower than PUMA, the Unitary-PUMA [11] maps the matrix $\mathbf{R}$ and its Hermitian transpose $\mathbf{R}^H$ onto their codomain real-valued matrices using the unitary transformation $\varphi(\cdot)$ and calculates the SVDs of the resulting real value matrices. This is due the fact that applying a proper unitary transformation $\varphi(\cdot)$ on the complex matrix $\mathbf{R}$, one can reduce the time-complexity of SVD calculations [15], even though the size of the resultant matrix $\varphi(\mathbf{R}) \in \mathbb{R}^{M\times 2N}$ is doubled. Unitary-PUMA calculates two SVDs and two matrix inversion operations within each iteration. For sufficiently high SNR values, Unitary-PUMA converges with only one iteration, providing a lower time-complexity than PUMA (two real-valued SVDs and two- real-valued matrix inversions). However, the simulation results presented in this article show that the variance of Unitary-PUMA's estimates is a function of the frequency. Meaning that the estimator experiences an abrupt increase in variance for specific frequencies which remains even in high SNR values. Moreover, the Unitary-PUMA suffers from high space-complexity which is not preferred for applications with limited memory.

Considering the above-stated facts, we conclude that there is not a general estimator for both short and long observations which can achieve accurate and unbiased frequency estimation over the whole frequency range, and yet suffice the time and space-complexity requirement. In this dilemma, the complexity and the estimation performance must be preferentially prioritized based on the application. Motivated by this, we develop a subspace method with lower space and time-complexity than other subspace methods, yet near-to-uniform estimation performance over the whole frequency range even for short observations.

The key contributions of this article are as follows:

- A new subspace-based frequency estimator is proposed. A substantial property of the estimator is the ability to provide uniform frequency estimation over the whole frequency range for short observation lengths ($8 \leq K < 256$). The SNR thresholds of the estimator are comparable with thresholds of state-of-the-art estimators. Its space-complexity is the lowest among time-domain and DFT-based methods. Its time-complexity is linear, and it is comparable to DFT-based methods (even for short observations).
- An analytical proof that the proposed estimator is unbiased and with a variance asymptotically equals to the CRLB is presented.
- A dependence of variance of Unitary-PUMA's estimates on the frequency is shown.

2. Materials and Methods

Notations

Throughout the text, we use boldface lowercase and uppercase letters for vectors and matrices, respectively. $[\mathbf{A}]_{i,j}$ is the (i,j)-th element of the matrix $\mathbf{A}$, $\mathbf{I}_{m\times m}$ is the $m\times m$ identity matrix, $\mathbf{\Pi}_{m\times m}$ is the $m\times m$ exchange matrix (matrix with ones on its antidiagonal and zeros elsewhere), and $\mathbf{0}_{m\times n}$ and $\mathbf{1}_{m\times n}$ are $m\times n$ matrices of all zeros and ones, respectively. We denote diagonal matrices as $\mathrm{diag}(\cdot)$. Superscripts $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^\dagger$ represent transpose, Hermitian transpose, and Moore–Penrose inverse, respectively. The symbols $\bullet$, $\otimes$, $\odot$, $\oplus$ and $\mathrm{vec}(\cdot)$ stand for transposed Khatri-Rao product [16], Kronecker product, Hadamard product, direct-sum, and matrix vectorization, respectively. We use $\overline{(\cdot)}$, $\angle$., $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ for the complex conjugate, the phase, the real and the imaginary part of a complex number, respectively. The symbol $\lfloor\cdot\rfloor$ denotes rounding to the nearest integer toward $-\infty$. If x is a random variable, then $\mathrm{E}(x)$ and $\mathrm{var}(x)$ represent expectation and variance, respectively, and $\hat{x}$ denotes the estimate of x.

2.1. Complex to real mapping

Any complex $p\times q$ matrix $\mathbf{G}\in\mathbb{C}^{p\times q}$ can be transformed into its real-valued counterpart according to

$$\varphi(\mathbf{G}) = \mathbf{T}_{p\times p}^H\begin{bmatrix}\mathbf{G} & \mathbf{\Pi}_{p\times p}\overline{\mathbf{G}}\mathbf{\Pi}_{q\times q}\end{bmatrix}\mathbf{T}_{2q\times 2q}, \tag{3-a}$$

where $\varphi(\mathbf{G})\in\mathbb{R}^{p\times 2q}$, and $\varphi(\cdot)$ denotes the unitary transformation [11, 15]. The $X\times X$ unitary matrices $\mathbf{T}_{X\times X}$ are given as

$$\mathbf{T}_{X\times X} = \begin{cases} \dfrac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{I}_{x\times x} & j\mathbf{I}_{x\times x}\\ \mathbf{\Pi}_{x\times x} & -j\mathbf{\Pi}_{x\times x}\end{bmatrix}, & \text{for } X = 2x,\\[2ex] \dfrac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{I}_{x\times x} & \mathbf{0}_{x\times 1} & j\mathbf{I}_{x\times x}\\ \mathbf{0}_{x\times 1}^T & \sqrt{2} & \mathbf{0}_{x\times 1}^T\\ \mathbf{\Pi}_{x\times x} & \mathbf{0}_{x\times 1} & -j\mathbf{\Pi}_{x\times x}\end{bmatrix}, & \text{for } X = 2x+1.\end{cases} \tag{3-b}$$

Let $\mathbf{G}$ be partitioned as

$$\mathbf{G} = \begin{bmatrix}\mathbf{G}_1\\ \mathbf{g}^T\\ \mathbf{G}_2\end{bmatrix}, \tag{4}$$

where the block matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ have the same size. Then, the real value matrix $\varphi(\mathbf{G})$ is given as [15]

$$\varphi(\mathbf{G}) = \begin{bmatrix} \mathrm{Re}\left(\mathbf{G}_1 + \mathbf{\Pi}_{\lfloor\frac{p}{2}\rfloor\times\lfloor\frac{p}{2}\rfloor}\overline{\mathbf{G}}_2\right) & -\mathrm{Im}\left(\mathbf{G}_1 - \mathbf{\Pi}_{\lfloor\frac{p}{2}\rfloor\times\lfloor\frac{p}{2}\rfloor}\overline{\mathbf{G}}_2\right) \\ \sqrt{2}\mathrm{Re}(\mathbf{g}^T) & -\sqrt{2}\mathrm{Im}(\mathbf{g}^T) \\ \mathrm{Im}\left(\mathbf{G}_1 + \mathbf{\Pi}_{\lfloor\frac{p}{2}\rfloor\times\lfloor\frac{p}{2}\rfloor}\overline{\mathbf{G}}_2\right) & \mathrm{Re}\left(\mathbf{G}_1 - \mathbf{\Pi}_{\lfloor\frac{p}{2}\rfloor\times\lfloor\frac{p}{2}\rfloor}\overline{\mathbf{G}}_2\right) \end{bmatrix}. \tag{5}$$

Note that the central row is dropped for even p.

2.2. The explicit form of real-valued noise-free signal

We expect the k-th sample of the received signal to be the linear combination (1) of the k-th sample of a noise-free signal $s[k] = b_0 e^{j\omega k}$, and of the k-th sample of the Gaussian noise $q(k)$. We reshape the samples of the noise-free signal for $k = 0, \ldots, K-1$ into a matrix [10]

$$\mathbf{S} = \begin{bmatrix} s[0] & s[M] & \ldots & s[M(N-1)] \\ s[1] & s[M+1] & \cdots & s[M(N-1)+1] \\ \vdots & \vdots & \ddots & \vdots \\ s[M-1] & s[2M-1] & \ldots & s[MN-1] \end{bmatrix}, \tag{6}$$

where $\mathbf{S} \in \mathbb{C}^{M\times N}$, $M + N + \tau, = K, M, N \in \mathbb{N}^+$, and $\tau \in \mathbb{N}$. Without loss of generality, let M be an even number. According to (5), the real-valued mapping of this matrix is given as

$$\varphi(\mathbf{S}) = \begin{bmatrix} \mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} \\ \mathbf{\Phi}_{21} & \mathbf{\Phi}_{22} \end{bmatrix}, \tag{7-a}$$

where

$$\begin{aligned} \mathbf{\Phi}_{11} &= \mathrm{Re}\left(\mathbf{S}_1 + \mathbf{\Pi}_{\frac{M}{2}\times\frac{M}{2}}\overline{\mathbf{S}}_2\right), \\ \mathbf{\Phi}_{12} &= \mathrm{Im}\left(\mathbf{\Pi}_{\frac{M}{2}\times\frac{M}{2}}\overline{\mathbf{S}}_2 - \mathbf{S}_1\right), \\ \mathbf{\Phi}_{21} &= \mathrm{Im}\left(\mathbf{S}_1 + \mathbf{\Pi}_{\frac{M}{2}\times\frac{M}{2}}\overline{\mathbf{S}}_2\right), \\ \mathbf{\Phi}_{22} &= \mathrm{Re}\left(\mathbf{S}_1 - \mathbf{\Pi}_{\frac{M}{2}\times\frac{M}{2}}\overline{\mathbf{S}}_2\right), \end{aligned} \tag{7-b}$$

and according to (5), $\mathbf{S} = [\mathbf{S}_1 \quad \mathbf{S}_2]^T$.

As the (m, n)-th element of the matrix $\mathbf{S}$ is

$$[\mathbf{S}]_{m,n} = b_0 \mathrm{e}^{j\omega(m+Mn)},$$

the compact forms of the matrices $\mathbf{S}_1$ and $\mathbf{\Pi}_{\frac{M}{2}\times\frac{M}{2}}\overline{\mathbf{S}}_2$ can be written as

$$\begin{aligned} [\mathbf{S}_1]_{m,n} &= b_0 e^{j\omega(m+Mn)}, \\ \left[\mathbf{\Pi}_{\frac{M}{2}\times\frac{M}{2}}\overline{\mathbf{S}}_2\right]_{m,n} &= b_0 e^{-j\omega(M(n+1)-(m+1))}, \end{aligned} \qquad \begin{aligned} m &= 0, \ldots, \frac{M}{2} - 1, \\ n &= 0, \ldots N - 1. \end{aligned} \tag{8}$$

With the help of the Euler's formula for complex numbers, we substitute (8) into (7-b) as

$$\begin{aligned} [\mathbf{\Phi}_{11}]_{m,n} &= 2b_0\left[\cos\left(\frac{\omega}{2}(M(2n+1)-1)\right)\cos\left(\frac{\omega}{2}(M-2m-1)\right)\right], \\ [\mathbf{\Phi}_{12}]_{m,n} &= -2b_0\left[\sin\left(\frac{\omega}{2}(M(2n+1)-1)\right)\cos\left(\frac{\omega}{2}(M-2m-1)\right)\right], \\ [\mathbf{\Phi}_{21}]_{m,n} &= -2b_0\left[\cos\left(\frac{\omega}{2}(M(2n+1)-1)\right)\sin\left(\frac{\omega}{2}(M-2m-1)\right)\right], \\ [\mathbf{\Phi}_{22}]_{m,n} &= 2b_0\left[\sin\left(\frac{\omega}{2}(M(2n+1)-1)\right)\sin\left(\frac{\omega}{2}(M-2m-1)\right)\right]. \end{aligned}$$

We can express each submatrix (7-b) as a rank one matrix of the form

$$\begin{aligned} \mathbf{\Phi}_{11} &= 2\tilde{\mathbf{u}}_{\mathrm{L}}\tilde{\mathbf{v}}_{\mathrm{L}}^T, \\ \mathbf{\Phi}_{12} &= 2\tilde{\mathbf{u}}_{\mathrm{L}}\tilde{\mathbf{v}}_{\mathrm{R}}^T, \\ \mathbf{\Phi}_{21} &= 2\tilde{\mathbf{u}}_{\mathrm{R}}\tilde{\mathbf{v}}_{\mathrm{L}}^T, \\ \mathbf{\Phi}_{22} &= 2\tilde{\mathbf{u}}_{\mathrm{R}}\tilde{\mathbf{v}}_{\mathrm{R}}^T, \end{aligned} \tag{9-a}$$

in which

$$
\begin{aligned}
[\tilde{\mathbf{u}}_{\mathrm{L}}]_k &= \cos\left(\frac{\omega}{2}(M-2k-1)\right), \text{for } k = 0, \dots \frac{M}{2}-1, \\
[\tilde{\mathbf{u}}_{\mathrm{R}}]_k &= -\sin\left(\frac{\omega}{2}(M-2k-1)\right), \text{for } k = 0, \dots \frac{M}{2}-1, \\
[\tilde{\mathbf{v}}_{\mathrm{L}}]_k &= \cos\left(\frac{\omega}{2}(M(2k+1)-1)\right), \text{for } k = 0, \dots N-1, \\
[\tilde{\mathbf{v}}_{\mathrm{R}}]_k &= -\sin\left(\frac{\omega}{2}(M(2k+1)-1)\right), \text{for } k = 0, \dots N-1.
\end{aligned} \tag{9-b}
$$

Let

$$
\begin{aligned}
\tilde{\mathbf{u}} &= [\tilde{\mathbf{u}}_{\mathrm{L}} \quad \tilde{\mathbf{u}}_{\mathrm{R}}]^T, \\
\tilde{\mathbf{v}} &= [\tilde{\mathbf{v}}_{\mathrm{L}} \quad \tilde{\mathbf{v}}_{\mathrm{R}}]^T,
\end{aligned} \tag{10}
$$

then $\varphi(\mathbf{S})$ can be written as

$$\varphi(\mathbf{S}) = b_0 \tilde{\mathbf{u}} \tilde{\mathbf{v}}^T. \tag{11}$$

2.3. Approximation of the factorized form of the real-valued noise-free signal

The same way as (2) and (6), we can write the samples of the noise $q[k]$ in a noise matrix $\mathbf{Q} \in \mathbb{C}^{M \times N}$. According to (1), it holds that

$$\mathbf{R} = \mathbf{S} + \mathbf{Q}. \tag{12}$$

Using the complex-to-real mapping (5), we obtain a real-valued matrix form of the received signal

$$\varphi(\mathbf{R}) = \varphi(\mathbf{S}) + \varphi(\mathbf{Q}).$$

The real-valued noise-free signal (11) can be factorized using the SVD of $\varphi(\mathbf{R})$ [11] which is given as

$$\varphi(\mathbf{R}) = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T, \tag{13}$$

where $\mathbf{U} = [\mathbf{u_0} \quad \mathbf{u_1} \quad \dots \quad \mathbf{u}_{M-1}]$ and $\mathbf{V} = [\mathbf{v_0} \quad \mathbf{v_1} \quad \dots \quad \mathbf{v}_{2N-1}]$ are $M \times M$ and $2N \times 2N$ real orthogonal matrices, respectively, the column vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th left and right-singular vectors of $\varphi(\mathbf{R})$, respectively, and $\boldsymbol{\Lambda}$ is an $M \times 2N$ rectangular diagonal matrix with non-negative real numbers λ (singular values) arranged descending on the diagonal. According to (11), the rank of $\varphi(\mathbf{S})$ is one. Assuming $\mathbf{R} \approx \mathbf{S}$ (i.e., $||\mathbf{Q}||_2^2 \to 0$), we can approximate $\varphi(\mathbf{R})$ as a perturbed rank-one matrix, with the first-order approximation written as [17]

$$\varphi(\mathbf{R}) = \lambda_0 \mathbf{u}_0 \mathbf{v}_0^T + \mathbf{U}_{\mathbf{Q}} \boldsymbol{\Lambda}_{\mathbf{Q}} \mathbf{V}_{\mathbf{Q}}^T \approx \lambda_0 (\mathbf{u}_0 + \Delta\mathbf{u})(\mathbf{v}_0 + \Delta\mathbf{v})^T, \tag{14-a}$$

where λ_0 is the first singular value of $\varphi(\mathbf{R})$, $\mathbf{u}_0$ and $\mathbf{v}_0$ are the first vectors of the matrices $\mathbf{U}$ and $\mathbf{V}$, respectively, and $\mathbf{U_Q}$, $\mathbf{V_Q}$, and $\boldsymbol{\Lambda}_{\mathbf{Q}}$ are matrixes obtained by removing the first columns of the matrixes $\mathbf{U}, \mathbf{V}$, and $\boldsymbol{\Lambda}$, respectively. The estimation error vectors $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$ are given as

$$
\begin{aligned}
\Delta\mathbf{u} &= -\frac{1}{\lambda_0} \mathbf{U_Q} \mathbf{U_Q}^T \varphi(\mathbf{Q}) \mathbf{v}, \\
\Delta\mathbf{v} &= -\frac{1}{\lambda_0} \mathbf{V_Q}^T \mathbf{V_Q} \varphi(\mathbf{Q})^{\boldsymbol{T}} \mathbf{u}.
\end{aligned} \tag{14-b}
$$

We define approximations of the left and of the right vectors of the factorized real-valued noise-free signal $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$, $\mathbf{u}$ and $\mathbf{v}$, respectively, such that

$$\mathbf{u} \triangleq k_{\mathbf{u}} \mathbf{u}_0, \; \mathbf{v} \triangleq k_{\mathbf{v}} \mathbf{v}_0. \tag{15}$$

As $\mathbf{u}_0^T \mathbf{u}_0 = \mathbf{v}_0^T \mathbf{v}_0 = 1$, $\tilde{\mathbf{u}}^T \tilde{\mathbf{u}} = k_{\mathbf{u}}^2$ and $\tilde{\mathbf{v}}^T \tilde{\mathbf{v}} = k_{\mathbf{v}}^2$ [13], the unknown coefficients $k_{\mathbf{u}}$ and $k_{\mathbf{v}}$ are given as

$$k_{\mathbf{u}} = \sqrt{\frac{M}{2}}, \; k_{\mathbf{v}} = \sqrt{N}.$$

2.4. Phasal transformation

2.4.1. Definition

Let us define selection matrices for a vector $\mathbf{x} \in \mathbb{R}^{2p \times 1}, p \in \mathbb{N}$ such that

$$\mathbf{J}_{\mathbf{x}}^{r} = \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{L}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{L} + \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{R}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{R}, \\ \mathbf{J}_{\mathbf{x}}^{i} = \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{L}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{R} - \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{R}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{L}, \tag{16}$$

where

$$\mathbf{J}_{\mathbf{x}}^{L} \triangleq [\mathbf{I}_{p\times p} \quad \mathbf{0}_{p\times p}], \\ \mathbf{J}_{\mathbf{x}}^{R} \triangleq [\mathbf{0}_{p\times p} \quad \mathbf{I}_{p\times p}], \\ \mathbf{J}_{\mathbf{x}}^{0} \triangleq [\mathbf{I}_{(p-1)\times(p-1)} \quad \mathbf{0}_{(p-1)\times 1}], \\ \mathbf{J}_{\mathbf{x}}^{1} \triangleq [\mathbf{0}_{(p-1)\times 1} \quad \mathbf{I}_{(p-1)\times(p-1)}]. \tag{17}$$

We define the phasal transformation as

$$\Phi(\mathbf{x}) = (\mathbf{J}_{\mathbf{x}}^{r} + j\mathbf{J}_{\mathbf{x}}^{i})(\mathbf{x} \otimes \mathbf{x}), \tag{18}$$

where $\Phi(\mathbf{x}) \in \mathbb{C}^{(p-1)\times 1}$.

2.4.2. Low time-complexity version of the transformation

The calculation of the phasal transformation according to (18) is due to the matrix multiplications inappropriate for practical application. The calculation can be simplified by utilization of the Khatri-Rao transposed product property:

Lemma1: For any arbitrary matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$, and vectors $\mathbf{c} \in \mathbb{R}^{p\times 1}$ and $\mathbf{d} \in \mathbb{R}^{p\times 1}$, we have

$$(\mathbf{A}\bullet\mathbf{B})(\mathbf{c} \otimes \mathbf{d}) = (\mathbf{A}\mathbf{c})\odot(\mathbf{B}\mathbf{d}). \tag{19}$$

Proof: By expanding $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_q]^T, \\ \mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_q]^T,$$

in which $\mathbf{a}_i, \mathbf{b}_j \in \mathbb{R}^{p\times 1}$, we can express the transposed Khatri-Rao product of the matrices $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A}\bullet\mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \ldots \quad \mathbf{a}_p \otimes \mathbf{b}_p]^T.$$

The i-the element of (19) left side can be then expressed as

$$[(\mathbf{A}\bullet\mathbf{B})(\mathbf{c} \otimes \mathbf{d})]_i = (\mathbf{a}_i^T \otimes \mathbf{b}_i^T)(\mathbf{c} \otimes \mathbf{d}) = (\mathbf{a}_i^T\mathbf{c}) \otimes (\mathbf{b}_i^T\mathbf{d}) = (\mathbf{a}_i^T\mathbf{c})(\mathbf{b}_i^T\mathbf{d}).$$

Thus, we can say that

$$\mathbf{A}\bullet\mathbf{B} = (\mathbf{A}\mathbf{c})\odot(\mathbf{B}\mathbf{d}). \blacksquare$$

Considering lemma 1, we can write the real part of the phasal transformation (18) as

$$\mathrm{Re}\big(\Phi(\mathbf{x})\big) = (\mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{L}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{L} + \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{R}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{R})(\mathbf{x} \otimes \mathbf{x}) = \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{L}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{L}(\mathbf{x} \otimes \mathbf{x}) + \mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{R}\bullet\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{R}(\mathbf{x} \otimes \mathbf{x}) \\ = (\mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{L}\mathbf{x})\odot(\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{L}\mathbf{x}) + (\mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{R}\mathbf{x})\odot(\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{R}\mathbf{x}).$$

In the same way, the imaginary part is given as

$$\mathrm{Im}\big(\Phi(\mathbf{x})\big) = (\mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{L}\mathbf{x})\odot(\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{R}\mathbf{x}) + (\mathbf{J}_{\mathbf{x}}^{0}\mathbf{J}_{\mathbf{x}}^{R}\mathbf{x})\odot(\mathbf{J}_{\mathbf{x}}^{1}\mathbf{J}_{\mathbf{x}}^{L}\mathbf{x}).$$

So, the i-the element of the vector $\Phi(\mathbf{x})$ is

$$[\Phi(\mathbf{x})]_i = [\mathbf{x}]_i[\mathbf{x}]_{i+1} + [\mathbf{x}]_{p+i}[\mathbf{x}]_{p+i+1} \\ + j[\mathbf{x}]_i[\mathbf{x}]_{p+i+1} - j[\mathbf{x}]_{p+i}[\mathbf{x}]_{i+1}, \tag{20}$$

where $i \in \{0, 1, \ldots, p-2\}$.

2.5. Proposed estimation of the frequency

The vectors $\mathbf{u}$ and $\mathbf{v}$ carry information that allows estimation of the desired frequency ω. We can formulate its estimation as [18]

$$\widehat{\omega} = \beta\widehat{\omega}_{\mathbf{u}} + (1-\beta)\widehat{\omega}_{\mathbf{v}}, \tag{21}$$

where $\widehat{\omega}$ is the final estimate of the desired frequency ω, and $\widehat{\omega}_{\mathbf{u}}$ and $\widehat{\omega}_{\mathbf{v}}$ are estimates of ω based on the vectors $\mathbf{u}$ and $\mathbf{v}$, respectively. The weighting coefficient β is given as

$$\beta = \frac{\mathrm{var}(\widehat{\omega}_{\mathbf{v}})}{\mathrm{var}(\widehat{\omega}_{\mathbf{u}}) + \mathrm{var}(\widehat{\omega}_{\mathbf{v}})}. \tag{22}$$

The variance of $\widehat{\omega}$ is given as [18]

$$\mathrm{var}(\hat{\omega}) = \beta^2 \mathrm{var}(\hat{\omega}_{\mathbf{u}}) + (1-\beta)^2 \mathrm{var}(\hat{\omega}_{\mathbf{v}}). \tag{23}$$

The estimation of the frequency using the vector **u** or using the vector **v** can be handled as a search for the frequency resulting in the smallest sum of squares of residual errors. Let us consider the vector **u** at first. A vector of residual errors for **u** is given as

$$\mathbf{e}_{\mathbf{u}} = a_{\mathbf{u}} \mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} - \mathbf{y}_{\mathbf{u}},$$

where $\mathbf{y}_{\mathbf{u}} = \Phi(\mathbf{u})$, and $a_{\mathbf{u}} \triangleq e^{j\omega}$. Note that the phasal transformation (18) of the noise-free signal results in a vector of constant values. As shown in *Appendix A,* $\Phi(\tilde{\mathbf{u}}) = e^{j\omega} \mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}$ hence $a_{\mathbf{u}} \triangleq e^{j\omega}$.

Considering the Gauss-Markov Theorem [14], we formulize the estimation of $a_{\mathbf{u}}$ as a WLS problem to ensure that the residual errors are uncorrelated. The estimate of $a_{\mathbf{u}}$ is given as

$$\hat{a}_{\mathbf{u}} = \underset{a_{\mathbf{u}}}{\mathrm{argmin}}\, \mathbf{e}_{\mathbf{u}}^H \mathbf{W}_{\mathbf{u}} \mathbf{e}_{\mathbf{u}}, \tag{24}$$

where $\mathbf{W}_{\mathbf{u}} \triangleq \mathbf{C}_{\mathbf{e}}^{-1}$ is the weighting matrix, and $\mathbf{C}_{\mathbf{e}} = \mathrm{E}(\mathbf{e}_{\mathbf{u}}\mathbf{e}_{\mathbf{u}}^H)$ is the covariance matrix of the residual errors. Note that $\mathbf{C}_{\mathbf{e}}$ is a positive semidefinite matrix, thus its Cholesky decomposition exists as $\mathbf{C}_{\mathbf{e}} = \mathbf{L}\mathbf{L}^H$. By transforming the error vector $\mathbf{e}_{\mathbf{u}}$ with the matrix $\mathbf{L}^{-1}$, we can update the covariance matrix as

$$\mathrm{E}\big((\mathbf{L}^{-1}\mathbf{e}_{\mathbf{u}})(\mathbf{L}^{-1}\mathbf{e}_{\mathbf{u}})^H\big) = \mathbf{L}^{-1}\mathrm{E}(\mathbf{e}_{\mathbf{u}}\mathbf{e}_{\mathbf{u}}^H)(\mathbf{L}^{-1})^H = \mathbf{L}^{-1}\mathbf{L}\mathbf{L}^H(\mathbf{L}^{-1})^H = \mathbf{I}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}.$$

Thus, $\mathbf{W}_{\mathbf{u}}$ is the whitening filter of the residual error. The variance of $\hat{a}_{\mathbf{u}}$ is [19]

$$\mathrm{var}(\hat{a}_{\mathbf{u}}) = \frac{1}{\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times 1} \mathbf{W}_{\mathbf{u}} \mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}}. \tag{25}$$

The matrix $\mathbf{W}_{\mathbf{u}}$ is not a priori known. Considering this fact, we propose a second-order approximation of $\mathbf{W}_{\mathbf{u}}$. Based on the explicit forms of the real-valued noise-free signal (8-b) and (9), the approximation is (see *Appendix B*)

$$\widehat{\mathbf{W}}_{\mathbf{u}} = \frac{b_0^2 N}{\sigma^2}\left(\mathbf{I}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)} + \mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times 1}\right). \tag{26}$$

Applying the approximation to the optimization problem (24), we get the analytical solution

$$\hat{a}_{\mathbf{u}} = \frac{2}{M-2}\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{y}_{\mathbf{u}}. \tag{27}$$

Considering (25), the variance of the estimator (27) is

$$\mathrm{var}(\hat{a}_{\mathbf{u}}) \approx \frac{\sigma^2}{b_0^2}\frac{4}{NM(M-2)}. \tag{28}$$

Similarly, the vector of residual errors for the vector **v** is given as

$$\mathbf{e}_{\mathbf{v}} = a_{\mathbf{v}}\mathbf{1}_{(N-1)\times 1} - \mathbf{y}_{\mathbf{v}}, \tag{29}$$

where $a_{\mathbf{v}} \triangleq e^{jM\omega}$, and $\mathbf{y}_{\mathbf{v}} = \overline{\Phi(\mathbf{v})}$. The estimate of $a_{\mathbf{v}}$ is given as

$$\hat{a}_{\mathbf{v}} = \underset{a_{\mathbf{v}}}{\mathrm{argmin}}\, \mathbf{e}_{\mathbf{v}}^H \mathbf{W}_{\mathbf{v}} \mathbf{e}_{\mathbf{v}}, \tag{30}$$

with the variance

$$\mathrm{var}(\hat{a}_{\mathbf{v}}) = \frac{1}{\mathbf{1}^T_{(N-1)\times 1}\mathbf{W}_{\mathbf{v}}\mathbf{1}_{(N-1)\times 1}}, \tag{31}$$

where $\mathbf{W}_{\mathbf{v}} \triangleq \mathrm{E}(\mathbf{e}_{\mathbf{v}}\mathbf{e}_{\mathbf{v}}^H)^{-1}$.

We propose a second-order approximation of the whitening filter $\mathbf{W}_{\mathbf{v}}$

$$\widehat{\mathbf{W}}_{\mathbf{v}} = \frac{b_0^2 M}{2\sigma^2}\left(\mathbf{I}_{(N-1)\times(N-1)} + \mathbf{1}_{(N-1)\times 1}\mathbf{1}^T_{(N-1)\times 1}\right), \tag{32}$$

which leads to the analytical solution of the optimization problem (30)

$$\hat{a}_{\mathbf{v}} = \frac{1}{N-1}\mathbf{1}^T_{(N-1)\times 1}\mathbf{y}_{\mathbf{v}}. \tag{33}$$

Based on (31), the variance of the estimator (33) is

$$\mathrm{var}(\hat{a}_{\mathbf{v}}) \approx \frac{\sigma^2}{b_0^2}\frac{2}{MN(N-1)}. \tag{34}$$

The estimates $\hat{a}_{\mathbf{u}}$ and $\hat{a}_{\mathbf{v}}$ allow us to calculate the estimates of the desired frequency. We first calculate the estimate of the desired frequency from the vector **u** which is given as

$$\widehat{\omega}_{\mathbf{u}} = \angle\hat{a}_{\mathbf{u}}.$$

The estimate $\hat{a}_{\mathbf{v}}$ corresponds to $2\left\lfloor\frac{M}{2}\right\rfloor + 1$ possible frequencies [10]

$$\breve{\omega}_{\mathbf{v}} \in \breve{\Omega}_{\mathbf{v}} = \left\{\frac{\angle\hat{a}_{\mathbf{v}} + 2\pi i}{M}\middle| i = -\left\lfloor\frac{M}{2}\right\rfloor, -\left\lfloor\frac{M}{2}\right\rfloor + 1, \dots, \left\lfloor\frac{M}{2}\right\rfloor\right\}.$$

We consider the frequency $\breve{\omega}_{\mathbf{v}}$ with the lowest Euclidean distance to $\widehat{\omega}_{\mathbf{u}}$ to be the estimate of ω based on the vector **v**, i.e.

$$\widehat{\omega}_{\mathbf{v}} = \underset{\breve{\omega}_{\mathbf{v}} \in \breve{\Omega}_{\mathbf{v}}}{\operatorname{argmin}} \|\widehat{\omega}_{\mathbf{u}} - \breve{\omega}_{\mathbf{v}}\|_2,$$

where $\breve{\Omega}_{\mathbf{v}}$ is the set of all possible frequencies $\breve{\omega}_{\mathbf{v}}$.

According to (21) and (22), the final estimate $\widehat{\omega}$ depends on the variance of each estimator. If $M, N \gg 1$, variances of $\widehat{\omega}_{\mathbf{u}}$ and $\widehat{\omega}_{\mathbf{v}}$ can be approximated as functions of var($\hat{a}_{\mathbf{u}}$) and var($\hat{a}_{\mathbf{v}}$), respectively [10, 20]:

$$\operatorname{var}(\widehat{\omega}_{\mathbf{u}}) \approx \frac{\operatorname{var}(\hat{a}_{\mathbf{u}})}{2} \approx \frac{\sigma^2}{b_0^2}\frac{2}{NM(M-2)}, \tag{35-a}$$

$$\operatorname{var}(\widehat{\omega}_{\mathbf{v}}) \approx \frac{\operatorname{var}(\hat{a}_{\mathbf{v}})}{2M^2} \approx \frac{\sigma^2}{b_0^2}\frac{1}{M^3N(N-1)}. \tag{35-b}$$

Then we can approximate the calculation of the weighting coefficient β (22) as

$$\beta \approx \frac{(M-2)}{2M^2(N-1) + (M-2)}.$$

Consequently, the approximation of the final estimate of the desired frequency (23) can be expressed as

$$\widehat{\omega} \approx \frac{(M-2)\widehat{\omega}_{\mathbf{u}} + 2M^2(N-1)\widehat{\omega}_{\mathbf{v}}}{2M^2(N-1) + (M-2)}. \tag{36}$$

As shown in *Appendix C*, the estimator (36) is unbiased in small noise scenarios. It follows from (23) that the asymptotic variance of $\widehat{\omega}$ (for large values of M, N and K) is

$$\operatorname{var}(\widehat{\omega}) \approx \frac{\sigma^2}{b_0^2}\frac{2}{MN\left(2M^2(N-1) + (M-2)\right)}. \tag{37}$$

2.6. LUPUMA implementation

The key components of the LUPUMA estimator are the complex to real mapping of the received signal $\varphi(\mathbf{R})$, SVD of the resulting real-valued matrix, and the low time-complexity version of the phasal transformation (20). We implement the method as a function LUPUMA, and we summarize the implementation in Table I. Inputs of the function are factorization parameters M and N, and a vector $\mathbf{r}$ of K samples of the received signal, where $\mathbf{r} = [r[0] \quad \cdots \quad r[K-1]]$. The function returns the final estimate of the desired frequency $\widehat{\omega}$.

2.7. The setting of the factorization parameters

The choice of the factorization parameters M and N influences the variance of LUPUMA (37), where the CRLB of the variance is [21]

$$\operatorname{var}(\widehat{\omega}) = \frac{6\sigma^2}{Kb_0^2(K^2-1)}. \tag{38}$$

As mentioned in subsection 2.2, the parameters M and N must be positive natural numbers respecting the number of the received signal samples K. The phasal transformation (18) introduces an additional and more stringent restriction on the parameters. Specifically, $2 < M < K$ and $1 < N < K$. This restriction implies that LUPUMA requires at least 8 samples of the received signal for the frequency estimation.

To express the dependence of the variance on the factorization of the received signal samples, we define an auxiliary factorization parameter $\alpha \triangleq M/K$. Then the variance (37) can be written as

$$\operatorname{var}(\widehat{\omega}, \alpha) \approx \frac{\sigma^2}{b_0^2}\frac{6}{K(6(\alpha - \alpha^2)K^2 + (3\alpha)K - 6)}. \tag{39}$$

TABLE I
PSEUDOCODE OF LUPUMA

	function LUPUMA($\mathbf{r}\ M\ N$)
Require: vector $\mathbf{r}$ of K samples of the received signal $r(t)$, factorization parameters M and N, where $M, N \in \mathbb{N}^{+}$ and $MN \leq K$	
Ensure: the final estimate of the desired frequency $\widehat{\omega}$	
1:	$\mathbf{R} \leftarrow \begin{bmatrix} [\mathbf{r}]_0 & \cdots & [\mathbf{r}]_{M(N-1)} \\ \vdots & \ddots & \vdots \\ [\mathbf{r}]_{M-1} & \cdots & [\mathbf{r}]_{MN-1} \end{bmatrix}$
2:	$\varphi(\mathbf{R}) \leftarrow \begin{bmatrix} \mathrm{Re}\left(\mathbf{R}_1 + \mathbf{\Pi}_{\frac{m}{2}\times\frac{m}{2}}\bar{\mathbf{R}}_2\right) & -\mathrm{Im}\left(\mathbf{R}_1 - \mathbf{\Pi}_{\frac{m}{2}\times\frac{m}{2}}\bar{\mathbf{R}}_2\right) \\ \mathrm{Im}\left(\mathbf{R}_1 + \mathbf{\Pi}_{\frac{m}{2}\times\frac{m}{2}}\bar{\mathbf{R}}_2\right) & \mathrm{Re}\left(\mathbf{R}_1 - \mathbf{\Pi}_{\frac{m}{2}\times\frac{m}{2}}\bar{\mathbf{R}}_2\right) \end{bmatrix}$, where $\mathbf{R} \leftarrow [\mathbf{R}_1 \quad \mathbf{R}_2]^T$
3:	$\mathbf{u}_0, \mathbf{v}_0 \leftarrow \mathrm{SVD}\big(\varphi(\mathbf{R})\big)$
4:	$\widehat{\omega}_{\mathbf{u}} \leftarrow \angle \sum_{i=0}^{\frac{M}{2}-1} [\mathbf{y}_{\mathbf{u}}]_i$, where $[\mathbf{y}_{\mathbf{u}}]_i \leftarrow [\mathbf{u}]_i[\mathbf{u}]_{i+1} + [\mathbf{u}]_{\frac{M}{2}+i}[\mathbf{u}]_{\frac{M}{2}+i+1} + j[\mathbf{u}]_i[\mathbf{u}]_{\frac{M}{2}+i+1} - j[\mathbf{u}]_{\frac{M}{2}+i}[\mathbf{u}]_{i+1}$, and $\mathbf{u} \leftarrow \sqrt{\frac{M}{2}}\mathbf{u}_0$
5:	$\breve{\Omega}_{\mathbf{v}} \leftarrow \left\{\frac{\angle \sum_{i=0}^{N-1}[\mathbf{y}_{\mathbf{v}}]_i + 2\pi i}{M} \mid i = -\left\lfloor\frac{M}{2}\right\rfloor, -\left\lfloor\frac{M}{2}\right\rfloor + 1, \dots, \left\lfloor\frac{M}{2}\right\rfloor\right\}$, where $[\mathbf{y}_{\mathbf{v}}]_i \leftarrow [\mathbf{v}]_i[\mathbf{v}]_{i+1} + [\mathbf{v}]_{N+i}[\mathbf{v}]_{N+i+1} - j[\mathbf{v}]_i[\mathbf{v}]_{N+i+1} + j[\mathbf{v}]_{N+i}[\mathbf{v}]_{i+1}$, and $\mathbf{v} \leftarrow \sqrt{N}\mathbf{v}_0$
6:	$\widehat{\omega}_{\mathbf{v}} \leftarrow \underset{\breve{\omega}_{\mathbf{v}} \in \breve{\Omega}_{\mathbf{v}}}{\mathrm{argmin}} \lVert \widehat{\omega}_{\mathbf{u}} - \breve{\omega}_{\mathbf{v}} \rVert_2$
7:	$\widehat{\omega} \leftarrow \frac{(M-2)\widehat{\omega}_{\mathbf{u}} + 2M^2(N-1)\widehat{\omega}_{\mathbf{v}}}{2M^2(N-1) + (M-2)}$

To reach CRLB at the lowest SNR value, we must find such α that the estimator variance (39) will be equal to (38). We formulate the search for α as an optimization problem

$$\hat{\alpha} = \underset{\alpha}{\mathrm{argmin}} \left(\frac{6\sigma^2}{Kb_0^2}\left(\frac{1}{6(\alpha - \alpha^2)K^2 + (3\alpha)K - 6} - \frac{1}{K^2 - 1}\right)\right)^2$$

subject to

$$\alpha \in (0,1).$$

The analytical solution to this problem is

$$\hat{\alpha} = \frac{(6K+3) \pm \sqrt{3(4K^2 + 12K - 53)}}{12K}. \tag{40}$$

As the factorization parameters M and N must be positive natural numbers, we estimate their optimal values according to

$$M^* = [\hat{\alpha}K], \quad N^* = \left\lfloor\frac{K}{M^*}\right\rfloor,$$

where $[\cdot]$ stands for rounding to the nearest integer. Note that we remove the last $(K - M^*N^*)$ samples for $M^*N^* < K$ (see Table I, operation number 1).

The optimization problem (39) has two feasible solutions. Concerning the robustness of SVD toward noise [22], $\hat{\alpha}$ resulting in smaller differences between M^* and N^* are preferred. For example, for $K \to \infty$, $\hat{\alpha} \in \{0.21, 0.79\}$. $\hat{\alpha} \approx 0.21$ is preferred as the constructed shape with this adjustment is closer to the squared matrix. Note that the variance for $\hat{\alpha} \approx 0.21$ is

$$\mathrm{var}(\widehat{\omega}) \approx \frac{\sigma^2}{b_0^2}\frac{6}{K(K-2.18)(K+2.8)}, \tag{41}$$

i.e., the variance is asymptotically equal to CRLB (38) for this factorization and $K \to \infty$.

2.8. LUPUMA time-complexity

We use the number of floating-point operations (FLOPs) to express the time complexity of LUPUMA. We summarize time-complexities of the LUPUMA operations (given in Table I) in Table II.

LUPUMA relies on one SVD and simple matrix operations. As the optimal setting of the factorization parameters (M, N) results in tall matrices, we use an SVD algorithm based on QR iteration [23]. The total operation counts of this algorithm depend on (M, N) (see Table III, operation 3). We show in Table III that for the optimal setting ($M = K/5$), the time-complexity of LUPUMA is linear.

2.9. Simulation experiments

We conduct simulation experiments aimed at the evaluation of LUPUMA and its comparison with PUMA [10], Unitary-PUMA [11], unbiased A&M estimator [5, 24], parabolic estimator [12, 25], and DFT-based weighted least squares (DFT-WLS) estimator [7]. In DFT-WLS, we use window lengths $L \in \{3, 5\}$ with their coefficients calculated and stored beforehand [7]. In A&M and PUMA, we employ up to five and three iterations, respectively, before the stopping criterion is met, while it is one iteration for Unitary-PUMA. In the parabolic estimator, we consider the distance of 1/10 between adjacent samples, identical to the value selected in [25]. For PUMA and Unitary-PUMA, we factorize the received signal by the factorization parameters set up $M \approx N$ (the optimal settings for PUMA and Unitary-PUMA).

We evaluate the estimation performances, time, and space-complexities of the estimators. For each experiment and for each estimator, we carry out 10000 and 100 simulations aimed at evaluations of the estimation performance and the time-complexity of the estimators, respectively. If not indicated otherwise, for each simulation run, we generate a new vector $\mathbf{r}$ of K signal samples with $\boldsymbol{\omega}$ drawn from uniform distribution $U(-\pi, \pi)$**.** Unless stated otherwise, we consider the signal affected by the Additive White Gaussian Noise (AWGN) (1), with amplitude $b_0 = e^{5j}$, and variance $\sigma^2 = b_0^2 10^{-0.1\mathrm{SNR}}$, where SNR is in dB.

To investigate the validity of the rank one approximation (14) and its influence on the estimation performance of LUPUMA, we carry out matrix error analysis for observation lengths $K \in \{8, 32, 128, 512\}$, and $\mathrm{SNR} \in \{2x | x \in \mathbb{Z}, -10 \le x \le 30\}$. We calculate the normalized error

$$\Psi(\mathbf{A}, \breve{\mathbf{A}}) = \frac{\left|\left|\mathbf{A} - \breve{\mathbf{A}}\right|\right|_2}{\left|\left|\mathbf{A}\right|\right|_2} \tag{42}$$

where $\mathbf{A}$ and $\breve{\mathbf{A}}$ are a matrix and its approximation, respectively. Here we take $\mathbf{A} = \varphi(\mathbf{R})$ (13) and $\breve{\mathbf{A}} = \lambda_0(\mathbf{u}_0 + \Delta\mathbf{u})(\mathbf{v}_0 + \Delta\mathbf{v})^T$.

Also, we observe the influence of rank one approximation on the estimation performance of LUPUMA by obtaining the Euclidean distance between estimates of the frequency based on observed singular vector $\mathbf{u}$ and $\breve{\mathbf{u}}$, $\widehat{\omega}_{\mathbf{u}}$ and $\widehat{\omega}_{\breve{\mathbf{u}}}$, respectively, where

$$\breve{\mathbf{u}} \triangleq \tilde{\mathbf{u}} - \boldsymbol{\Delta}\mathbf{u}, \tag{43}$$

is the approximated singular vector and $\tilde{\mathbf{u}}$ is defined based on (11). Note that the error analysis of $\mathbf{v}$ follows similar steps. Thus, for clarity purposes, we focus only on the analysis of $\mathbf{u}$.

TABLE II
TIME-COMPLEXITY OF LUPUMA OPERATIONS

Operation No.	Description of the operation	FLOPs count
1	reshaping	0
2	Complex-to-Real Transform	$2MN$
3	QR-SVD [22]	$12MN^2 + 48N^3$ for $M \ge 2N$ and $6NM^2 + 6M^3$ for $M < 2N$
4	estimation of $\omega_{\mathbf{u}}$	$8\left(\left\lfloor\frac{M}{2}\right\rfloor - 1\right) + 42$
5 and 6	estimation of $\omega_{\mathbf{v}}$	$5 \times 2\left\lfloor\frac{M}{2}\right\rfloor + 8(N-1) + 42$
7	estimation of ω	4

TABLE III

TIME-COMPLEXITY OF LUPUMA FOR VARIOUS SETTINGS OF THE FACTORIZATION PARAMETERS

Factorization	FLOPs count
$M \geq 2N$	$12MN^2 + 48N^3 + +2MN + 8N + \left(18\left\lfloor\frac{M}{2}\right\rfloor\right) + 72$
$M = K/2$	$30.5K + 472$
$M = K/5$	$63.8K + 6112$

To validate the legitimacy of ignoring the third-order variation in (B8), we calculate the normalized error (42) between the inverse of the covariance matrix ($\mathbf{W}_{\mathbf{u}}^{-1}$ defined in (B5)) as $\mathbf{A}$ in (42), and its second-order approximation (B8) as $\breve{\mathbf{A}}$ in (42) for $K = 512$.

To observe correlations of residual errors for various least squares-based estimators, we introduce an ordinary least squares (LS) frequency estimator [11] and a WLS frequency estimator [11]. We estimate the covariance matrices of the error $\mathbf{e}_{\mathbf{u}}$ (B1) for LS estimator, WLS estimator, and LUPUMA (24) with $\mathbf{W}_{\mathbf{u}}$ given by (26). In each case, we estimate the covariance matrix by taking an average over 2000 observations for observation length $K = 128$ and frequency $\omega = 0.2\pi$. Moreover, we obtain the estimation accuracy of the frequency estimate associated with the vector $\mathbf{u}$. For LS estimator [11], we estimate the frequency without considering the correlation between residual errors according to

$$\widehat{\omega}_{\mathbf{u},\mathrm{LS}} = 2\arctan\left((\mathrm{Re}(\mathbf{\Upsilon})\mathbf{u})^{\dagger}(\mathrm{Im}(\mathbf{\Upsilon})\mathbf{u})\right), \tag{44}$$

where $\mathbf{\Upsilon} = \mathbf{T}_{M-1}^{H}\mathbf{J}_{M-1}^{1}\mathbf{T}_{M-1}$, and $\mathbf{T}$ and $\mathbf{J}$ given by (3-b) and (17), respectively.

For the WLS estimator [11], we estimate the frequency using the covariance matrix approximation

$$\mathbf{W} \approx \left(\mathrm{Re}(\mathbf{\Upsilon}) - \tan\left(\frac{\omega}{2}\right)\mathrm{Im}(\mathbf{\Upsilon})\right)\left(\mathrm{Re}(\mathbf{\Upsilon}) - \tan\left(\frac{\omega}{2}\right)\mathrm{Im}(\mathbf{\Upsilon})\right)^{T},$$

according to

$$\widehat{\omega}_{\mathbf{u},\mathrm{WLS}} = 2\arctan\left((\mathbf{u}^{T}\mathrm{Re}(\mathbf{\Upsilon}^{T})\mathbf{W}^{-1}\mathrm{Re}(\mathbf{\Upsilon})\mathbf{u})^{-1}(\mathbf{u}^{T}\mathrm{Re}(\mathbf{\Upsilon}^{T})\mathbf{W}^{-1}\mathrm{Im}(\mathbf{\Upsilon})\mathbf{u})\right). \tag{45}$$

To verify the theoretical assumptions on the estimation performance of LUPUMA for different settings of the factorization parameters (M, N), we observe the dependence of the Mean Squared Error (MSE) on the SNR for $\mathrm{SNR} \in \{2x | x \in \mathbb{Z}, -10 \leq x \leq 20\}$,

$$(M, N) \in \left\{([K/5], 5), \left(\sqrt{K}, \sqrt{K}\right), \left(\sqrt{2K}, \sqrt{K}/2\right), (4, [K/4])\right\}, \tag{46}$$

and $K = 256$. MSE of Euclidean distances is known as one of the natural optimality criteria [26], extensively used in frequency estimation problems. Thus, selecting this criterion enables fair comparisons with the state-of-the-art methods proposed in the literature. We calculate the MSE as

$$\mathrm{MSE} = 10\log_{10}\left(\frac{1}{T}\sum_{t=0}^{T-1}([\boldsymbol{\omega}]_t - [\widehat{\boldsymbol{\omega}}]_t)^2\right)$$

where $T = 10000$.

To evaluate the convergence of LUPUMA for different observation lengths K, we observe its MSE for $\mathrm{SNR} \in \{2x | x \in \mathbb{Z}, -10 \leq x \leq 30\}$, and $K \in \{8,\ 16,\ 64,\ 256,\ 512\}$. To utilize the maximum number of available samples, we use $M \approx K/2$ for $K \in \{8, 16, 64\}$, and $M \approx K/5$ for $K \in \{256, 512\}$.

To allow a fair comparison of LUPUMA with the state-of-the-art estimators, we observe the MSEs of the estimators for $\mathrm{SNR} \in \{2x | x \in \mathbb{Z}, -10 \leq x \leq 15\}$, and $K \in \{10, 32, 256\}$. We consider $M \approx K/2$ and $M \approx K/5$ (optimal shapes according to (40)), for $K = 32$ and $K = \{10,\ 256\}$, respectively. We also consider $M \approx N$ which allows a fair comparison with PUMA.

Especially in high SNR values, the estimation variance of LUPUMA approaches very small values, which makes the comparison of the estimation performance of the evaluated estimators difficult. Hence, to obtain a more detailed comparison, we investigate for each estimator the dependency of simulated variances on CRLB for $\{2x | x \in \mathbb{Z}, 5 \leq x \leq 30\}$, and $\omega = 0.2\pi$. We also calculate for each estimator an average ratio of variance to CRLB

$$\mathrm{ratio} = \frac{1}{n}\sum_{i=0}^{n-1}\frac{\mathrm{var}(\widehat{\omega}, [\boldsymbol{\rho}]_i)}{\mathrm{CRLB}([\boldsymbol{\rho}]_i)}, \tag{47}$$

where $\boldsymbol{\rho}$ is a vector of investigated SNR values (i.e. $\boldsymbol{\rho} = [10, 12, \ldots, 60]$), n is the number of the SNR values and $\mathrm{CRLB}([\boldsymbol{\rho}]_i)$ and $\mathrm{var}(\widehat{\omega}, [\boldsymbol{\rho}]_i)$ are CRLB (38) and the variance of the estimate $\widehat{\omega}$, respectively, for the i-th SNR value.

To study the robustness of selected estimators toward changes in the frequency ω in the AWGN scenario, we calculate the MSE of LUPUMA, Unitary-PUMA, DFT-WLS, A&M, and the parabolic estimator for

$$\omega \in \left\{ \frac{2\pi\left(-\frac{l}{2}\right)}{l}, \frac{2\pi\left(-\frac{l}{2} + 0.25\right)}{l}, \frac{2\pi\left(-\frac{l}{2} + 0.5\right)}{l}, \ldots, \frac{2\pi\left(\frac{l}{2}\right)}{l} \right\},$$

$l = 32$, and $\mathrm{SNR} \in \{-20, -14, \ldots, 34, 40\}$. We consider $K \in \{32, 256\}$, $M \approx K/2$ for $K = 32$, and $M \approx K/5$ for $K = 256$ (optimal setting of the factorization parameters according to (40)).

LUPUMA is derived assuming the signal is disrupted by AWGN; however, this assumption might be violated in real-world applications. Considering this fact, we propose experiments to study the robustness of selected estimators toward changes in the frequency ω in a colored-noise scenario. We use the setting described for the AWGN scenario except for the observation length $K = 32$. The colored noise is described by an auto-regressive moving average model

$$q[k] = \sum_{i=1}^{3} [\mathbf{a}]_i q[k-i] + \sum_{i=1}^{3} [\mathbf{b}]_i \epsilon[k-i] + \epsilon[k],$$

where $\mathbf{a} = [1, -0.683, 0.82]$, $\mathbf{b} = [0.34, -0.11, 0.34]$, $\epsilon[k]$ is the k-th sample of a zero-mean excitation noise $\epsilon(t)$ with variance $\sigma_\epsilon^2 = b_0^2 / S_q(\omega) 10^{-0.1\mathrm{SNR}}$, and $S_q(\omega)$ is the power spectral density of the process.

To compare the time and space-complexities of the estimators, we measure the total numbers of FLOPs and allocated memories for observation lengths $K \in \{(2x)^2 | x \in 2, 3, \ldots, 11\}$, and SNR = 5 dB.

We implement the experiments in Python and C languages. To obtain the variance of algorithms, we use Python with Linear Algebra PACKage (LAPACK) [27] library. Moreover, double-precision FLOPs and allocated memory results were computed using double-precision operations with BLAS (v3.9.0), LAPACK (v3.9.0), and FFTW (v3.3.10) libraries in C language. We run the simulations on a computer with a 1.9 GHz quad-core Intel i7 processor with 16 GB of RAM.

3. Results

We show the obtained one-dimensional results (vectors) as sets of graphs, and two-dimensional results (matrices) as heat maps. In all graphs, the simulation results are depicted as sets of markers connected by solid line segments. Dashed lines are theoretical variances obtained according to (38) and a dash-dotted line indicates CRLB.

Figs. 1-7 show results obtained solely for LUPUMA. Within these figures, Figs. 1-5 illustrate the impact of introduced approximations on the estimation performance of LUPUMA. Fig. 1 indicates the dependence of the normalized error Ψ (42) of rank one approximation (14-a) on SNR values. Fig. 2 presents the d ependence of

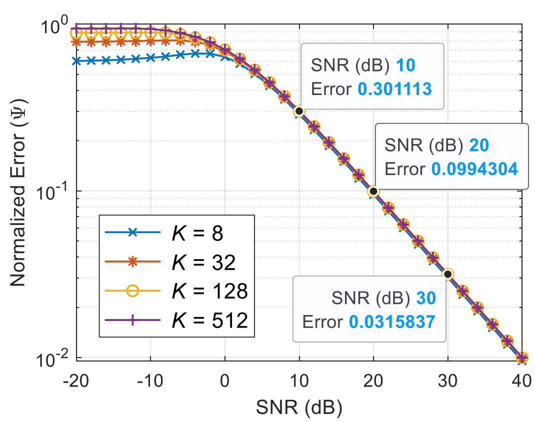


Fig1. Dependencies of normalized error Ψ of LUPUMA's rank one approximation on signal-to-noise ratio (SNR) for various observation lengths K. Marked data points are normalized errors at SNR $\in \{10, 20, 30\}$ dB.

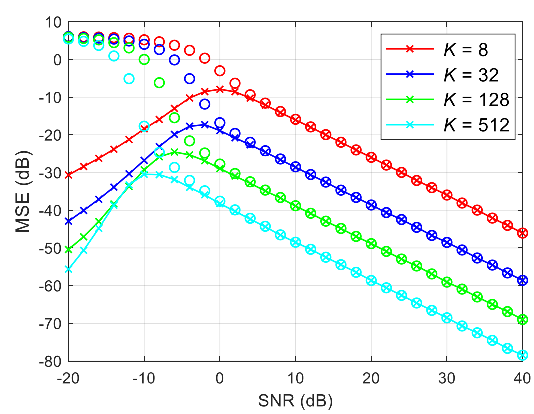


Fig2. Dependencies of LUPUMA's mean squared error (MSE) on signal-to-noise ratio (SNR) for various observation lengths K. The MSE is calculated between the estimated frequency based on $\tilde{\mathbf{u}}$ and $\mathbf{u}$ (marker ('x')) and between $\tilde{\mathbf{u}}$ and $\breve{\mathbf{u}}$ (marker ('o')), respectively.

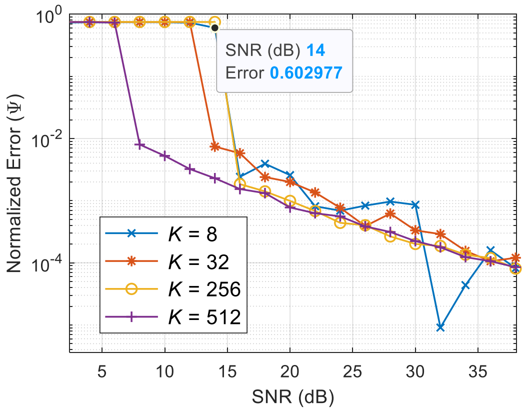


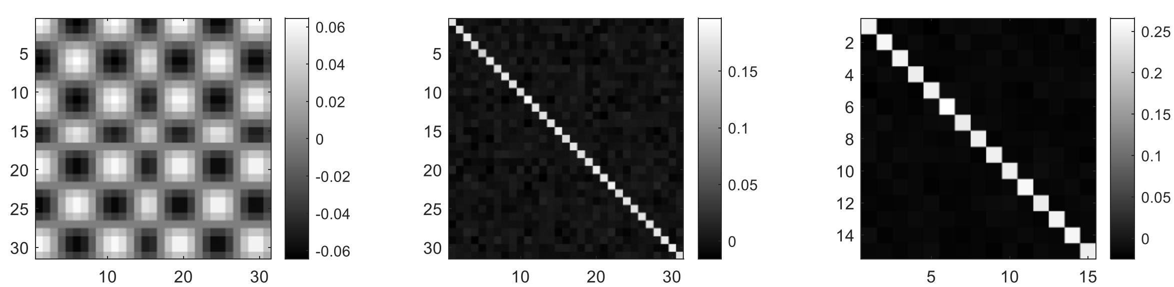
Fig3. Dependencies of normalized error Ψ of LUPUMA's weighting matrix approximation on signal-to-noise ratio (SNR) for various observation lengths K. The marked data point is related to the observation length $K = 8$.

estimation MSE on SNR for frequency estimates $\widehat{\omega}_{\mathbf{u}}$ and $\widehat{\omega}_{\breve{\mathbf{u}}}$, obtained from $\mathbf{u}$ (15) and $\breve{\mathbf{u}}$ (43), respectively. The data points with the marker ('o') and lines with the marker ('x') are associated with $\widehat{\omega}_{\mathbf{u}}$ and $\widehat{\omega}_{\breve{\mathbf{u}}}$, respectively. Fig. 3 presents the dependency of normalized error Ψ (42) between (B5) and (B8) on SNR and for various observation lengths. Fig. 4 (a), (b), and (c) represent three heat maps to explore the diagonality of error covariance matrices derived from LS, WLS, and LUPUMA based on Table I, line 4, respectively. Fig. 5 shows for LS, WLS, and LUPUMA estimators the dependency of MSE of $\widehat{\omega}_{\mathbf{u}}$ on SNR for $K = 128$, $M = 2N$, and $M = K/2$. Fig. 6 illustrates the dependence of MSE on SNR for $K = 256$ in which each behavior is associated with one setting of the factorization parameters (M, N). For each setting (46), we use one unique color. Fig. 7 displays the dependence of MSE on SNR for different observation lengths K. For each setting, we use one unique color. The graphs in Figs. 8-11 allow comparison of subspace methods and DFT-based frequency estimators with LUPUMA for two settings of the factorization parameters. For each estimator and setting, we use a unique color. Figs. 8 and 9 show the dependencies of MSE on SNR for subspace methods and DFT-based methods, respectively, for various observation lengths. Fig. 10 shows the dependency of the variance of the simulated estimates on CRLB for different SNR values. In this figure, for each estimator, ratios of variances to CRLB are marked with arrows. Fig. 11 displays average numbers of FLOPs for various observation lengths K. Fig. 12 presents the dependence of allocated memory on the observation length K.

Figs. 13-15 illustrate the dependence of MSE on the normalized frequency ω/π and SNR under AWGN (Fig. 13 for $K = 32$, and Fig. 14 for $K = 256$) and colored-noise assumptions (Fig. 15), respectively. The subplots (a), (b), (c), (d), (e), and (f) in the figures show contour plots for LUPUMA, Unitary-PUMA DFT-WLS ($L = 3$), DFT-WLS ($L = 5$), the parabolic estimator and A&M, respectively. In each contour plot, lines with higher color contrast have lower MSE (dB), and the MSE values are written with the same color.

4. Discussion

Within the development of LUPUMA, we used rank one approximation (14-a), weighting matrix approximation (B8), and approximated values of variances (35-a) and (35-b) to combine individual estimates $\widehat{\omega}_{\mathbf{u}}$ and $\widehat{\omega}_{\mathbf{v}}$ in (36). The results shown in Figs. 1-3 validate the rank one and weighting matrix approximations. The matrix $\varphi(\mathbf{R})$ (13) seems to be well-explained by the approximation (14-a) for $\mathrm{SNR} \geq 20\,\mathrm{dB}$ (see Fig. 1). Nevertheless, regardless of the observation length, the impact of rank one approximation (14-a) is negligible for $\mathrm{SNR} \geq 5\,\mathrm{dB}$ (see Fig. 2, the



(a) (b) (c)

Fig4. Covariance matrices of residual errors for (a) least squares, (b) weighted least squares, and (c) LUPUMA (24) with $\mathbf{W}_{\mathbf{u}}$ defined in (26). The x- and y-axis are associated with the column and row of the represented matrix, respectively. The values are presented using heat maps. The color bars map the values to grayscales.

convergence of data points associated with MSE of $\mathbf{u}$ and $\breve{\mathbf{u}}$ for $\mathrm{SNR} \geq 5$ dB). In addition, in the worst-case scenario ($K = 8$), the normalized errors in weighting matrix approximation (B8) are insignificant (i.e., $\leq 10^{-2}$) for $\mathrm{SNR} \geq 15$ dB (Fig. 3). The correlation of LUPUMA residual errors (Fig. 4 (c)) indicates low correlation property of the estimator [14], close to WLS (Fig. 4 (b)), and significantly lower than LS (Fig. 4 (a)). LUPUMA achieves lower MSE in the estimation of the frequency associated with the singular vector $\mathbf{u}$, $\omega_{\mathbf{u}}$, than both the WLS estimator (45) and LS estimator (44), regardless of the factorization of the matrix (Fig. 5). Remarking that the impact of correlation of residual errors is more severe in low-SNR regimes and tall matrix factorizations. Hence LS estimator exhibits lower performance in comparison with the WLS estimator when $\mathrm{SNR} \in (-10, 20]$ dB, and $M = K/2$ (Fig. 5, marker 'x' and marker '*').

Figs. 6 and 7 evaluate the theoretical convergence of LUPUMA for a wide range of the observation lengths ($K \in \{8, 16, \ldots, 512\}$). For the optimal setting of the factorization parameters (M, N), MSE of LUPUMA reaches to CRLB. For short observation lengths ($K < 256$), SNR thresholds of LUPUMA, regardless of the setting (M, N), are similar to the thresholds of the state-of-the-art estimators (Fig. 8 (a-b) and Fig. 9 (a-b)). For long observations ($K \geq 256$), the setting of the factorization parameters (M, N) becomes important for the performance of LUPUMA. For the optimal setting of the parameters ($M^* \approx K/5$), the SNR threshold of LUPUMA is higher than the threshold of PUMA (Fig. 8 (c)), DFT-WLS, A&M, and the parabolic estimator (Fig. 9 (c)). Nevertheless, the estimation variance of LUPUMA for $M = K/5$ is 1.29 times CRLB (Fig. 10), which is the best ratio achieved among the subspace methods and third-best among all the evaluated estimators (first and second are the parabolic estimator and A&M with the ratio

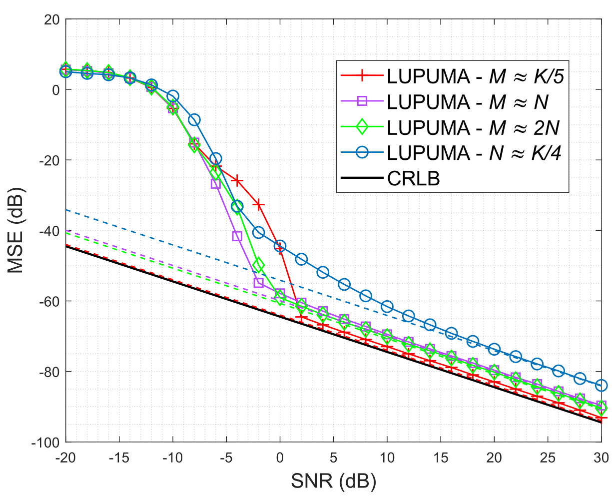


Fig6. Dependencies of LUPUMA's mean squared error (MSE) on signal-to-noise ratio (SNR) for various settings of the factorization parameters M and N, the observation length $K = 256$, and AWGN constraint. Dashed lines and the black dash-and-dot line indicate LUPUMA's theoretical variances and CRLB, respectively.

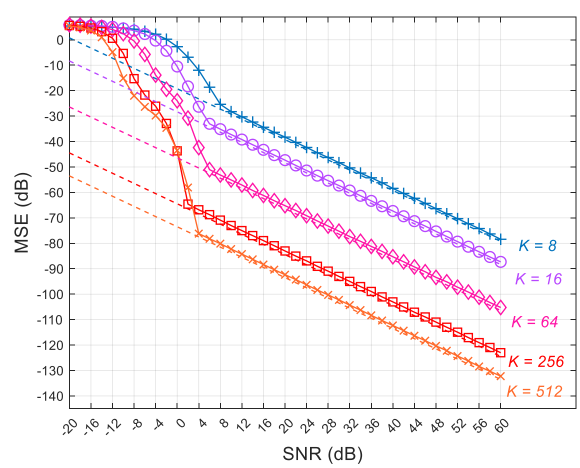


Fig7. Dependencies of LUPUMA's mean squared error (MSE) on signal-to-noise ratio (SNR) for different observation lengths K, the desired frequency $\omega = 0.2\pi$, and AWGN constraint. For $K \in \{8, 16, 64\}$ and $K \in \{256, 512\}$, we use $M \approx K/2$ and $M \approx K/5$, respectively. Dash-dotted lines indicate CRLB for each observation length.

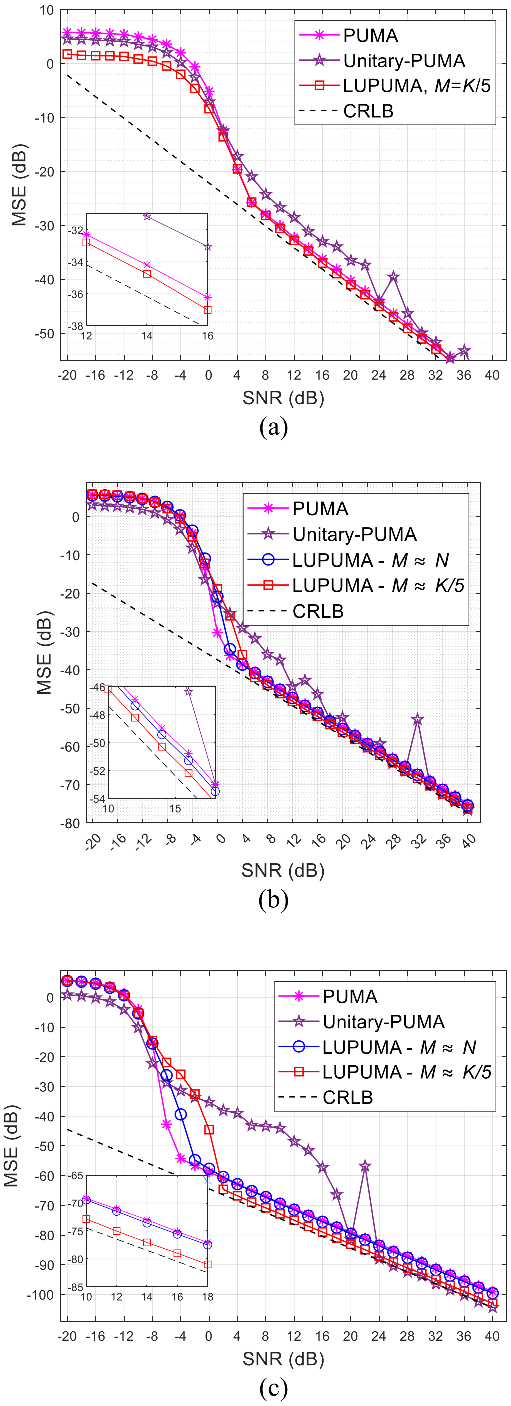


Fig8. Dependencies of mean squared error (MSE) on signal-to-noise ratio (SNR) of LUPUMA and subspace estimators for AWGN constraint and for observation lengths (a) $K = 10$, (b) $K = 32$, and (c) $K = 256$, respectively.

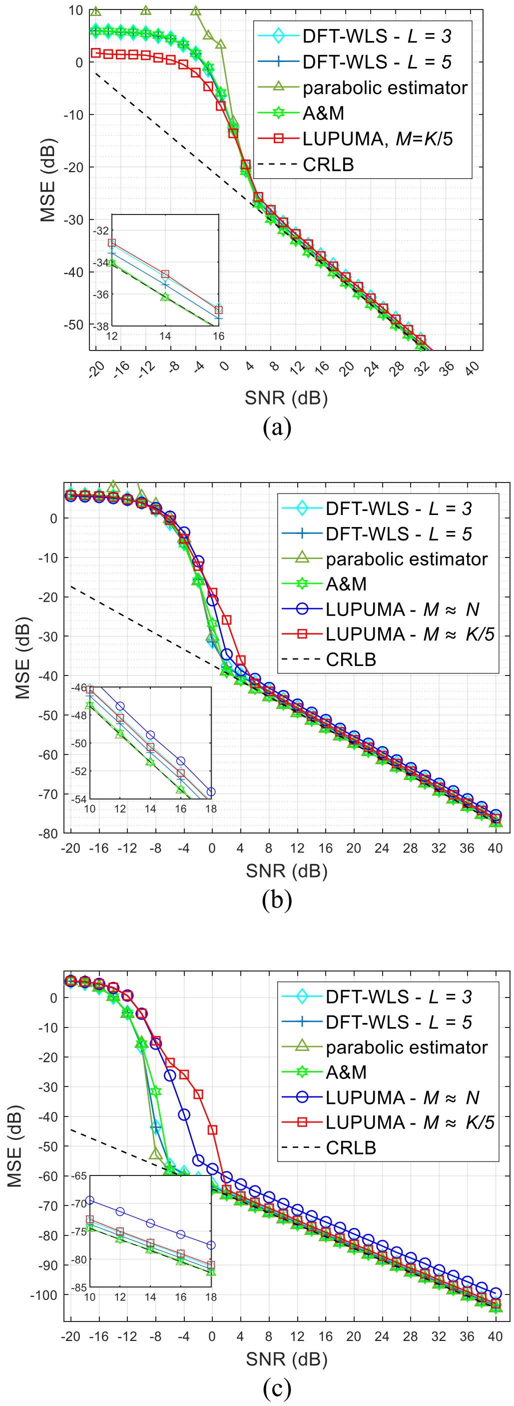


Fig9. Dependencies of mean squared error (MSE) on signal-to-noise ratio (SNR) of LUPUMA and DFT-based estimators for AWGN constraint and observation lengths (a) $K = 10$, (b) $K = 32$, and (c) $K = 256$, respectively.

1.0052 and 1.012, respectively (see Fig. 10)). In fact, choosing the optimal setting of (M, N) according to (40) results in a tall matrix which degrades the performance of SVD in noise [22]. For $M \approx N$, LUPUMA yields SNR threshold

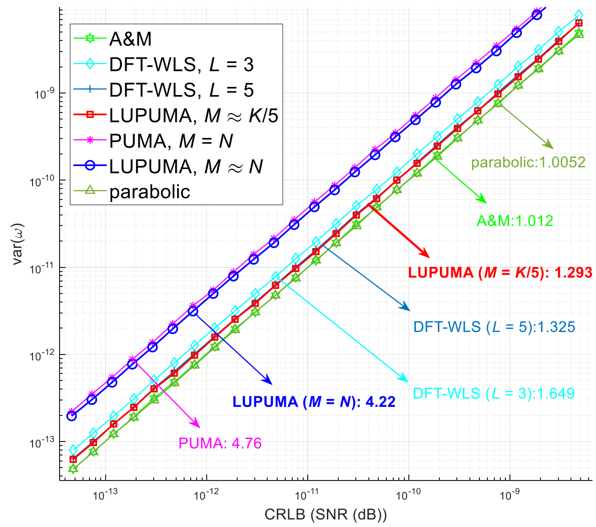


Fig10. Dependencies of variance of the estimators $\text{var}(\omega)$ on the values of Cramer-Rao Lower Bound (CRLB) for various signal-to-noise ratios (SNRs). Indicated value of each arrow is the slope of line for mentioned estimator.

similar to the thresholds of the state-of-the-art estimators (Fig. 8 (c) and Fig. 9 (c)), but its variance is 4.22 times CRLB (Fig. 10). The trade-off between the SNR-threshold and the divergence of MSE from the CRLB hinder the application of LUPUMA in cases where the frequency is estimated from long observations with low SNR-values. However, LUPUMA is fully competitive with the state-of-the-art estimators in terms of frequency estimation from short observations.

We recognize LUPUMA to be robust with respect to the desired frequency (Fig. 13 and Fig. 14). The dependence of LUPUMA's MSE on the frequency is negligible for $\text{SNR} \geq 10$ dB (Fig. 13 (a)). For lower SNR values, LUPUMA shows near-to-uniform estimation performance over a wide range of frequencies. In the case of long observations ($K = 256$), LUPUMA shows near to uniform estimation performance over the whole frequency range and a wide range of frequencies for $\text{SNR} \geq -2.5$ dB and $\text{SNR} < -2.5$ dB, respectively (Fig. 14 (a)). In low SNR regimes ($\text{SNR} < 5$), the parabolic estimator experiences the highest fluctuations among DFT-based methods and LUPUMA (compare Fig. 13 (e) with Fig. (a) and (c-d, f)). MSE of DFT-WLS varies for $\text{SNR} \geq -5$ dB with the magnitude of about 3 dB (Fig. 13 (c)) and 1 dB (Fig. 13 (d)) for $L = 3$ and $L = 5$, respectively. A&M exhibits the best performance in between DFT-based methods (Fig. 13 (f)), yet for both the short and the long observations, LUPUMA performance is the least fluctuating among the evaluated estimators.

Moreover, our results point out a previously unknown fact that the MSE of the Unitary-PUMA estimator heavily depends on the frequency. The variance of the Unitary-PUMA abruptly increases at certain frequencies (the blind spots of the estimator) (see Fig. 13 (b) and 14 (b)). In this context, we would like to point out the fact that in the

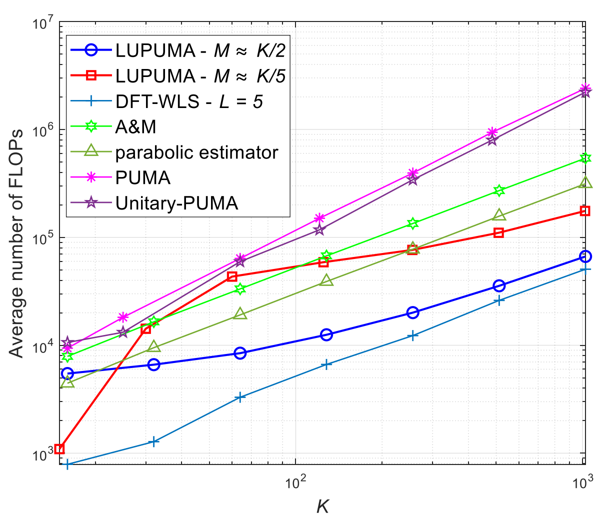


Fig11. Dependencies of the average number of FLOPs in simulations on the observation length K for various estimators, and signal-to-noise ratio SNR = 5 dB.

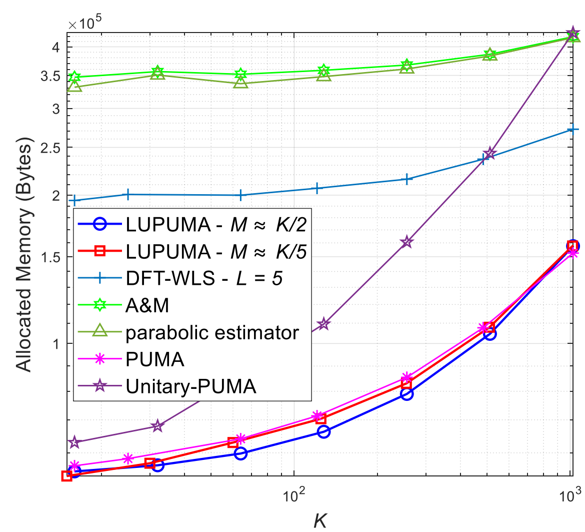
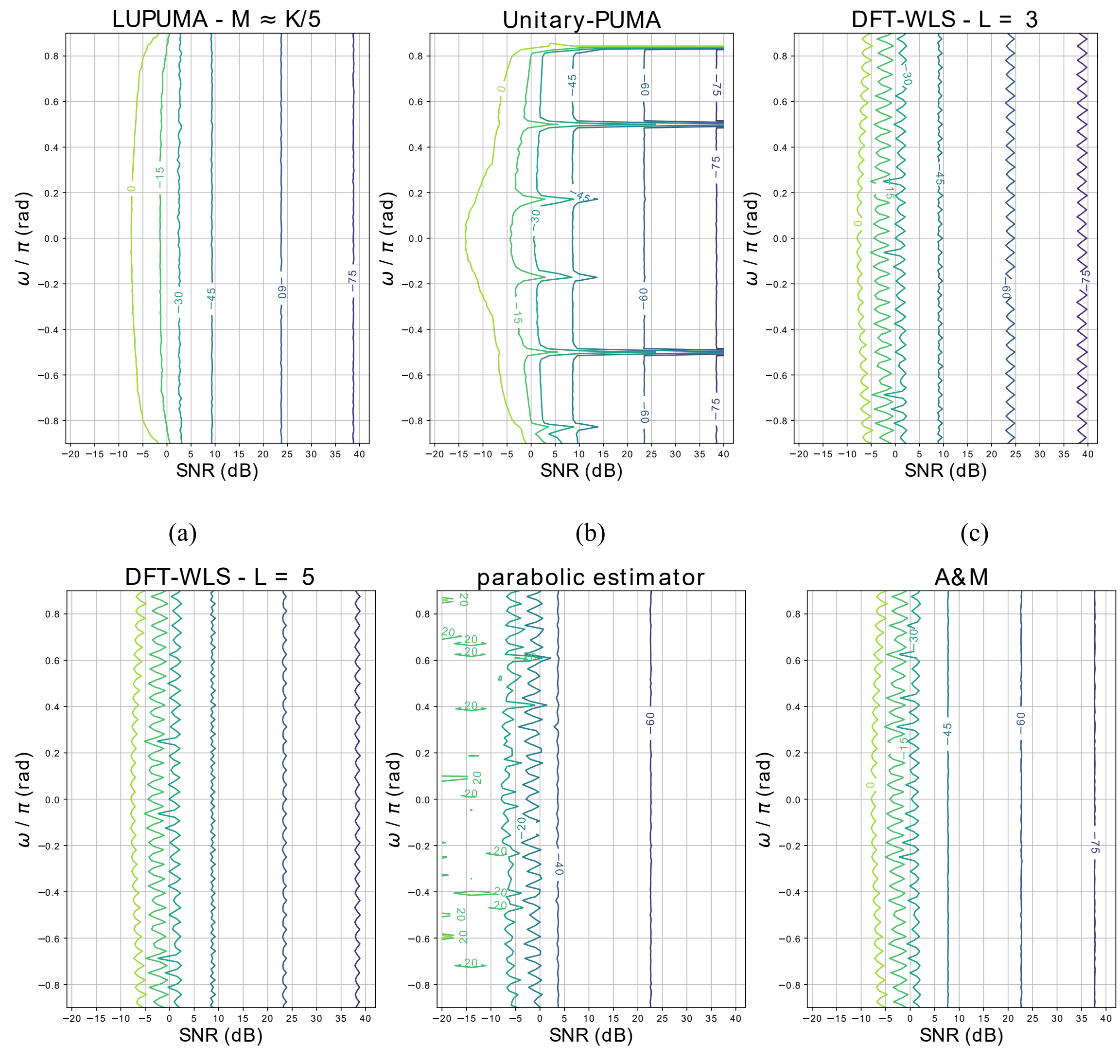


Fig12. Dependencies of the allocated memory in simulations on the observation length K for various estimators, and signal-to-noise ratio SNR = 5 dB.



(a) (b) (c)

(d) (e) (f)

Fig13. Dependencies of mean squared error (MSE) on signal-to-noise ratio (SNR) and on the normalized frequency ω/π of (a) LUPUMA, (b) Unitary-PUMA, (c) DFT-WLS, $L = 3$, (d) DFT-WLS, $L = 5$ (e) the parabolic estimator, (f) A&M for the observation length $K = 32$, and AWGN constraint.

original paper, the dependency of Unitary-PUMA on SNR is plotted for one specific frequency (see Fig.5 in [11]). We interpret the blind spots in the Unitary-PUMA as the violation of linear prediction property assumption in vectors $\mathbf{u}$ and $\mathbf{v}$ in (15). We can observe in (9) and (10) that the sub-vectors of vectors $\tilde{\mathbf{u}}$ ($\tilde{\mathbf{u}}_{\mathrm{L}}$ and $\tilde{\mathbf{u}}_{\mathrm{R}}$) and $\tilde{\mathbf{v}}$ ($\tilde{\mathbf{v}}_{\mathrm{L}}$ and $\tilde{\mathbf{v}}_{\mathrm{R}}$) individually have linear prediction property, but the block vectors of $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ do not share the property. Thus, the resultant matrix
first left and right-singular vectors of $\varphi(\mathbf{S})$ are not linearly predictable.

In the colored-noise scenario, LUPUMA achieves a smooth estimation variance following the power spectral density of the noise (compare Fig. 13 (a) and Fig. 15 (a)). The colored-noise influence the performance of Unitary-PUMA and DFT-WLS in the same way. The fluctuations that occur in the AWGN scenario are complemented by fluctuations caused by the colored noise (compare Fig. 13 (b-f) and Fig. 15 (b-f)). LUPUMA thus shows lower overall fluctuations than Unitary-PUMA, DFT-WLS, the parabolic estimator, and A&M in the colored-noise scenario (compare Fig. 15 (a-f)).

For the optimal setting of the factorization parameters (M, N), the theoretical time-complexity of LUPUMA is

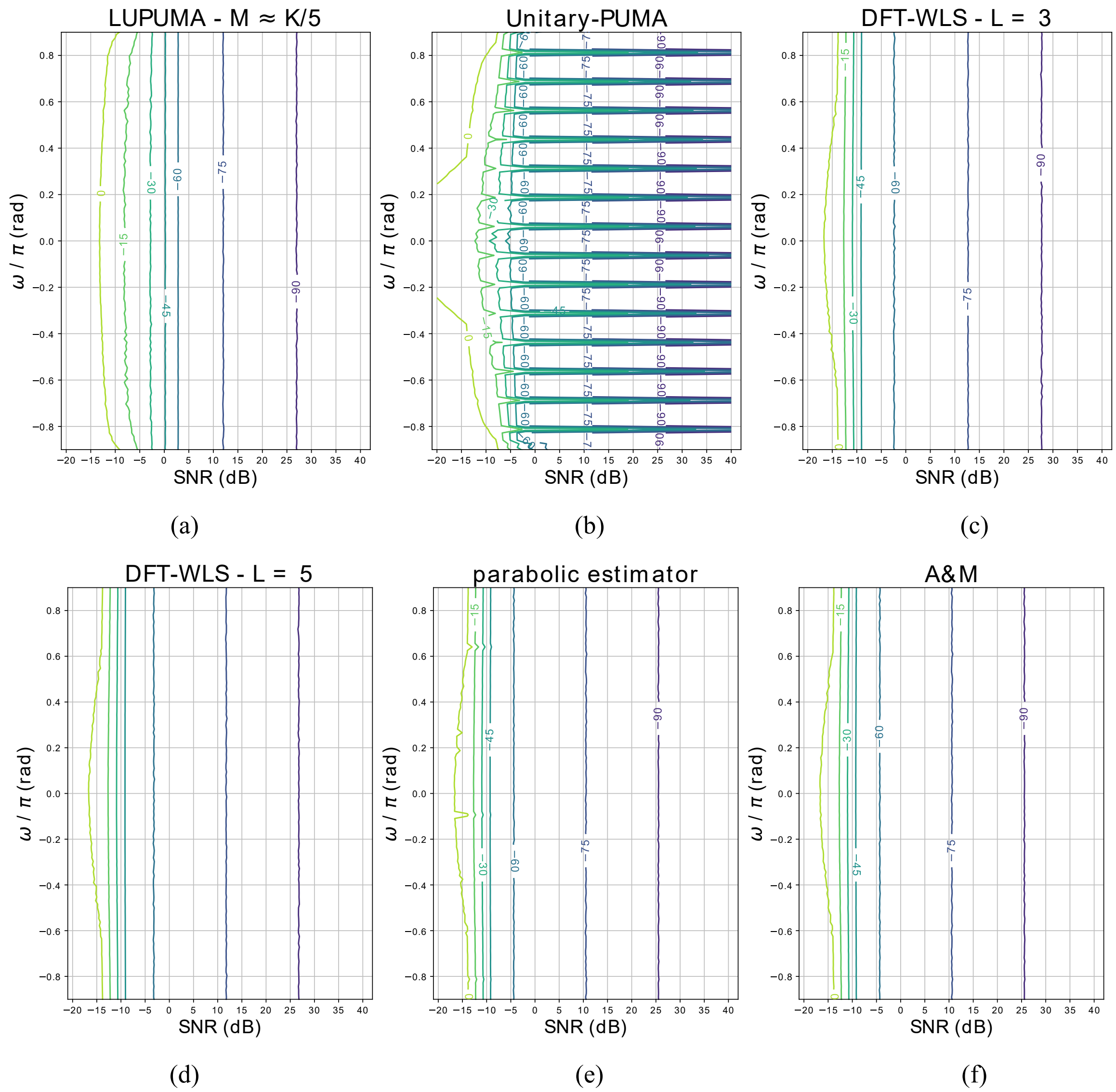


Fig14. Dependencies of mean squared error (MSE) on signal-to-noise ratio (SNR) and on the normalized frequency ω/π of (a) LUPUMA, (b) Unitary-PUMA, (c) DFT-WLS, $L = 3$, (d) DFT-WLS, $L = 5$ (e) parabolic estimator, (f) A&M for the observation length $K = 256$, and AWGN constraint.

$O(K)$ (see Table III, $M = K/2$ and $M = K/5$). The simulation results confirm this assumption for $M = K/2$ (Fig. 11). Time-complexity of LUPUMA is lower than time-complexities of PUMA, Unitary-PUMA, and A&M for both short and long observations. When considering short observations, LUPUMA has comparable time-complexity with the parabolic estimator. However, the time-complexity of LUPUMA is significantly lower than the time-complexity of the parabolic estimator in long observations. This is due to the linear time-complexity of LUPUMA versus $O(K \log K)$ of the parabolic estimator [25]. For long observations, LUPUMA's time-complexity is comparable with DFT-WLS time-complexity which is $O(K \log K)$ [7].

The space-complexity of LUPUMA corresponds to the space-complexity of PUMA. For all observation lengths, LUPUMA requires significantly less allocated memory than DFT-WLS, the parabolic estimator, A&M, and Unitary-PUMA (Fig. 12).

Our goal was to develop a time-domain frequency estimator of low time and space-complexity with minimum variance and unbiased frequency estimates over the whole frequency range $\omega \in (-\pi, \pi)$. Considering the SNR threshold, the estimation variance, the linear time-complexity, and the low space-complexity of LUPUMA, we conclude that LUPUMA fully meets the requirements on the accurate and yet time and space efficient estimator for

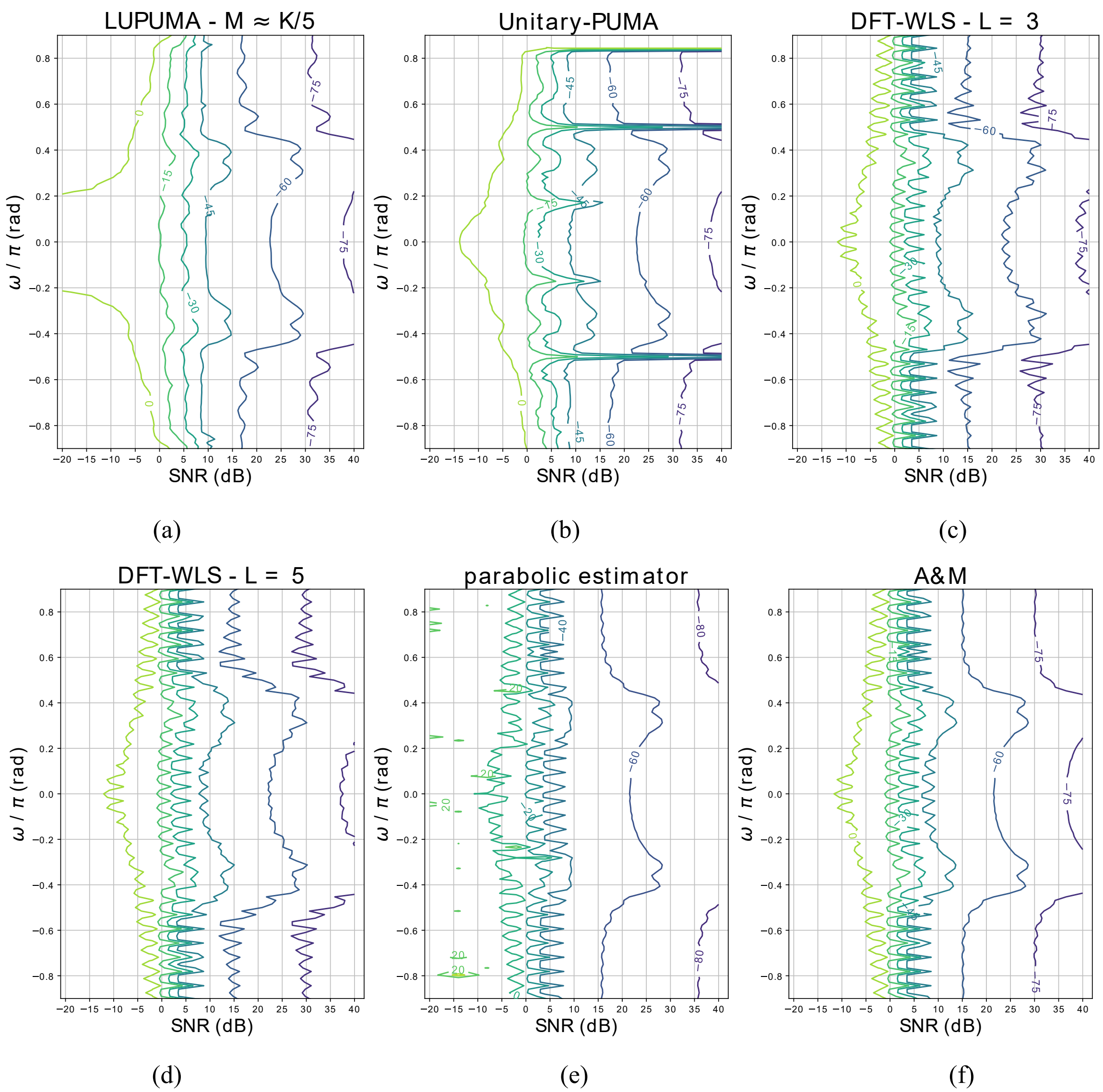


Fig15. Dependencies of mean squared error (MSE) on signal-to-noise ratio (SNR) and on the normalized frequency ω/π of (a) LUPUMA, (b) Unitary-PUMA, and (c) DFT-WLS, $L = 3$, (d) DFT-WLS, $L = 5$ (e) the parabolic estimator, (f) A&M for the observation length $K = 32$, and colored-noise constraint.

the short observations of the 1D complex signal in complex white Gaussian noise. The estimator also proved to be robust even if the white noise assumption is not met (as shown for a colored-noise case).

For short observations, PUMA, LUPUMA, DFT-WLS estimators, A&M, and the parabolic estimator demonstrate favorable SNR thresholds (Fig. 8 (a-b) and Fig. 9 (a-b)). The estimation performance of DFT-WLS, however, depends on the frequency (Fig. 13 (c-d)) which lowers the application potential of the DFT-WLS estimator. Note that the theoretical lower bound of direct DFT-based methods is a function of the frequency [7]. PUMA and A&M are robust in this regard; nevertheless, they have high time and space-complexity (Figs. 11-12). This makes PUMA and A&M inappropriate for applications or devices with limited computational power and memory. The parabolic estimator has comparable time-complexity with LUPUMA (Fig. 11). Nevertheless, it suffers from high space-complexity (Fig. 12), and high dependency on the frequency in low SNR regimes (Fig. 13 (e)). LUPUMA has none of these shortcomings and is thus convenient for these applications. Due to the low time-complexity and feed-forward process, LUPUMA is also suitable for real-time applications where the frequency estimation must be performed on a limited number of samples.

5. Conclusion

LUPUMA is the first single-tone frequency estimator with linear time-complexity which can reach the CRLB with a close to uniform performance over the whole frequency range. For a limited number of samples, LUPUMA is capable of fast and yet accurate frequency estimation, which is suitable for real-time applications such as frequency estimation in fast-varying propagation channels. The low space-complexity of LUPUMA makes the estimator to be optimal for applications with devices having limited computational power and memory, such as in wireless sensor nodes and IoT devices. Although A&M and parabolic frequency estimators outperform LUPUMA in statistical performance, the low time- and space-complexity, predictable performance across frequencies and potential for extension to multitone scenarios make LUPUMA interesting for practical applications.

6. Appendices

6.1. Appendix A

Let us consider the noiseless scenario in which $\varphi(\mathbf{Q}) = \mathbf{0}_{m\times n}$. It holds that $\mathbf{u} = \tilde{\mathbf{u}}$ and consequently $\Phi(\mathbf{u}) = \Phi(\tilde{\mathbf{u}})$. According to (18), the transformation $\Phi(\tilde{\mathbf{u}})$ is given as

$$\Phi(\tilde{\mathbf{u}}) = \left(\mathbf{J}^r_\mathbf{u} + j\mathbf{J}^i_\mathbf{u}\right)(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}}),$$

which can be expressed as

$$\mathbf{y}_{\tilde{\mathbf{u}}} = \mathbf{y}^r_{\tilde{\mathbf{u}}} + j\mathbf{y}^i_{\tilde{\mathbf{u}}},$$

where the real and the imaginary parts are given as $\mathbf{y}^r_{\tilde{\mathbf{u}}} = \mathbf{J}^r_\mathbf{u}(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})$ and $\mathbf{y}^r_{\tilde{\mathbf{u}}} = \mathbf{J}^i_\mathbf{u}(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})$, respectively. The selection matrices $\mathbf{J}^r_\mathbf{u}$ and $\mathbf{J}^i_\mathbf{u}$ are given by (16).

With respect to (17) and using *Lemma 1*, the k-th element of the vector real part is given as

$$\begin{aligned}
[\mathbf{y}^r_{\tilde{\mathbf{u}}}]_k &= [(\mathbf{J}^0_\mathbf{u}\mathbf{J}^L_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^L_\mathbf{u} + \mathbf{J}^0_\mathbf{u}\mathbf{J}^R_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^R_\mathbf{u})(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})]_k = [\mathbf{J}^0_\mathbf{u}\mathbf{J}^L_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^L_\mathbf{u}(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})]_k + [\mathbf{J}^0_\mathbf{u}\mathbf{J}^R_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^R_\mathbf{u}(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})]_k \\
&= [(\mathbf{J}^0_\mathbf{u}\mathbf{J}^L_\mathbf{u}\tilde{\mathbf{u}})\odot(\mathbf{J}^1_\mathbf{u}\mathbf{J}^L_\mathbf{u}\tilde{\mathbf{u}})]_k + [(\mathbf{J}^0_\mathbf{u}\mathbf{J}^R_\mathbf{u}\tilde{\mathbf{u}})\odot(\mathbf{J}^1_\mathbf{u}\mathbf{J}^R_\mathbf{u}\tilde{\mathbf{u}})]_k \\
&= \cos\left(\frac{\omega}{2}(M-2k-1)\right)\cos\left(\frac{\omega}{2}(M-2(k+1)-1)\right) \\
&\quad + \sin\left(\frac{\omega}{2}(M-2k-1)\right)\sin\left(\frac{\omega}{2}(M-2(k+1)-1)\right) \\
&= \cos\omega.
\end{aligned}$$

Similarly, the k-th element of the vector imaginary part is given as

$$\begin{aligned}
\left[\mathbf{y}^i_{\tilde{\mathbf{u}}}\right]_k &= [(\mathbf{J}^0_\mathbf{u}\mathbf{J}^L_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^R_\mathbf{u} - \mathbf{J}^0_\mathbf{u}\mathbf{J}^R_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^L_\mathbf{u})(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})]_k = [\mathbf{J}^0_\mathbf{u}\mathbf{J}^L_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^R_\mathbf{u}(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})]_k - [\mathbf{J}^0_\mathbf{u}\mathbf{J}^R_\mathbf{u}\bullet\mathbf{J}^1_\mathbf{u}\mathbf{J}^L_\mathbf{u}(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}})]_k \\
&= [(\mathbf{J}^0_\mathbf{u}\mathbf{J}^L_\mathbf{u}\tilde{\mathbf{u}})\odot(\mathbf{J}^1_\mathbf{u}\mathbf{J}^R_\mathbf{u}\tilde{\mathbf{u}})]_k - [(\mathbf{J}^0_\mathbf{u}\mathbf{J}^R_\mathbf{u}\tilde{\mathbf{u}})\odot(\mathbf{J}^1_\mathbf{u}\mathbf{J}^L_\mathbf{u}\tilde{\mathbf{u}})]_k = \\
&\quad -\cos\left(\frac{\omega}{2}(M-2k-1)\right)\sin\left(\frac{\omega}{2}(M-2(k+1)-1)\right) \\
&\quad + \sin\left(\frac{\omega}{2}(M-2k-1)\right)\cos\left(\frac{\omega}{2}(M-2(k+1)-1)\right) = \sin\omega
\end{aligned}$$

Thus, the transformation (18) for $\tilde{\mathbf{u}}$ is a column vector of $(M/2-1)$ complex numbers

$$\mathbf{y}_{\tilde{\mathbf{u}}} = \begin{bmatrix}\cos\omega \\ \vdots \\ \cos\omega\end{bmatrix} + j\begin{bmatrix}\sin\omega \\ \vdots \\ \sin\omega\end{bmatrix} = e^{j\omega}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}. \qquad \text{(A1)}$$

6.2. Appendix B

To obtain the second-order approximation (26) of the weighting matrix $\mathbf{W}_\mathbf{u}$, we expand $\mathbf{W}^{-1}_\mathbf{u} = \mathrm{E}(\mathbf{e}_\mathbf{u}\mathbf{e}^H_\mathbf{u})$. For the vector $\mathbf{u}$, the residual errors $\mathbf{e}_\mathbf{u}$ are defined as differences between expected values of the phasal transformation for a frequency ω and $\mathbf{y}_\mathbf{u} = \Phi(\mathbf{u})$. In the weighting matrix, we are interested in the difference between the phasal transformation of the noise free signal $\mathbf{y}_{\tilde{\mathbf{u}}}$ (A1) and $\mathbf{y}_\mathbf{u}$

$$\mathbf{e}_\mathbf{u} = a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} - \mathbf{y}_\mathbf{u}, \qquad \text{(B1)}$$

where $a_{\tilde{\mathbf{u}}} = e^{j\omega}$. Thus, $\mathbf{W}^{-1}_\mathbf{u}$ can be expressed as

$$\mathbf{W}_{\mathbf{u}}^{-1} = \mathrm{E}\left(\left(a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} - \mathbf{y}_{\mathbf{u}}\right)\left(a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} - \mathbf{y}_{\mathbf{u}}\right)^{H}\right) = |a_{\tilde{\mathbf{u}}}|^{2}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} - a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathrm{E}(\mathbf{y}_{\mathbf{u}}^{H}) - \bar{a}_{\tilde{\mathbf{u}}}\mathrm{E}(\mathbf{y}_{\mathbf{u}})\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} + \mathrm{E}(\mathbf{y}_{\mathbf{u}}\mathbf{y}_{\mathbf{u}}^{H}). \tag{B2}$$

Defining $\Delta\mathbf{u}$ as the projection of the complex to real mapping (4) of the noise $\mathbf{Q}$ on the desired signal basis vector, the left vector of the factorized real-valued signal $\mathbf{u}$ is given as

$$\mathbf{u} = \tilde{\mathbf{u}} + \Delta\mathbf{u}, \tag{B3}$$

where $\tilde{\mathbf{u}}$ is the left vector of the factorized real-valued noise-free signal. Accordingly, the phasal transformation $\mathbf{y}_{\mathbf{u}} = \Phi(\mathbf{u})$ is

$$\mathbf{y}_{\mathbf{u}} = \mathbf{y}_{\tilde{\mathbf{u}}} + \Delta\mathbf{y}_{\mathbf{u}}, \tag{B4}$$

where $\mathbf{y}_{\tilde{\mathbf{u}}} = a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}$. Thus, we can write (B2) as

$$\begin{aligned}\mathbf{W}_{\mathbf{u}}^{-1} &= |a_{\tilde{\mathbf{u}}}|^{2}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} - a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathrm{E}\left(\left(a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} + \Delta\mathbf{y}_{\mathbf{u}}\right)^{H}\right) - \\ &\bar{a}_{\tilde{\mathbf{u}}}\mathrm{E}\left(a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} + \Delta\mathbf{y}_{\mathbf{u}}\right)\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} + \mathrm{E}\left(\left(a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} + \Delta\mathbf{y}_{\mathbf{u}}\right)\left(a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1} + \Delta\mathbf{y}_{\mathbf{u}}\right)^{H}\right) = \\ &= |a_{\tilde{\mathbf{u}}}|^{2}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} - |a_{\tilde{\mathbf{u}}}|^{2}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} - a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}^{H}) \\ &\quad - |a_{\tilde{\mathbf{u}}}|^{2}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} - \bar{a}_{\tilde{\mathbf{u}}}\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}})\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} + |a_{\tilde{\mathbf{u}}}|^{2}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} \\ &\quad + a_{\tilde{\mathbf{u}}}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}^{H}) + \bar{a}_{\tilde{\mathbf{u}}}\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}})\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}^{T} + \mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}\Delta\mathbf{y}_{\mathbf{u}}^{H}),\end{aligned}$$

which results in

$$\mathbf{W}_{\mathbf{u}}^{-1} = \mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}\Delta\mathbf{y}_{\mathbf{u}}^{H}). \tag{B5}$$

The phasal transformation (18) of $\Delta\mathbf{u} = \mathbf{u} - \tilde{\mathbf{u}}$ can be expressed as

$$\Delta\mathbf{y}_{\mathbf{u}} = (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\mathbf{u}\otimes\mathbf{u}) - (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\tilde{\mathbf{u}}\otimes\tilde{\mathbf{u}}),$$

Considering (B3), it holds that

$$\Delta\mathbf{y}_{\mathbf{u}} = (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\tilde{\mathbf{u}}\otimes\Delta\mathbf{u}) + (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\Delta\mathbf{u}\otimes\tilde{\mathbf{u}}) + (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\Delta\mathbf{u}\otimes\Delta\mathbf{u}). \tag{B6}$$

This allows us to express the explicit form of $\mathbf{W}_{\mathbf{u}}^{-1}$ (B5) using $\Delta\mathbf{u}$ as

$$\begin{aligned}\mathbf{W}_{\mathbf{u}}^{-1} = &(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\mathrm{E}\big((\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})(\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})^{T}\big)(\mathbf{J}_{\mathbf{u}}^{r} - j\mathbf{J}_{\mathbf{u}}^{i})^{T} \\ &+(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\,\mathrm{E}\big((\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})(\Delta\mathbf{u}\otimes\tilde{\mathbf{u}})^{T}\big)(\mathbf{J}_{\mathbf{u}}^{r} - j\mathbf{J}_{\mathbf{u}}^{i})^{T} \\ &+(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\mathrm{E}\big((\Delta\mathbf{u}\otimes\tilde{\mathbf{u}})(\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})^{T}\big)(\mathbf{J}_{\mathbf{u}}^{r} - j\mathbf{J}_{\mathbf{u}}^{i})^{T} \\ &+(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\mathrm{E}\big((\Delta\mathbf{u}\otimes\tilde{\mathbf{u}})(\Delta\mathbf{u}\otimes\tilde{\mathbf{u}})^{T}\big)(\mathbf{J}_{\mathbf{u}}^{r} - j\mathbf{J}_{\mathbf{u}}^{i})^{T} \\ &+ o(\Delta\mathbf{u}^{3}).\end{aligned} \tag{B7}$$

Neglecting the terms associated with $o(\Delta\mathbf{u}^{3})$, we get $\mathbf{W}_{\mathbf{u}}^{-1}$ second-order approximation

$$\mathbf{W}_{\mathbf{u}}^{-1} \approx (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\mathbf{I}_{M^{2}\times M^{2}} + \mathbf{P})\mathrm{E}\big((\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})(\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})^{T}\big)(\mathbf{I}_{M^{2}\times M^{2}} + \mathbf{P})^{T}(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})^{T}, \tag{B8}$$

where $\mathbf{P}$ is the proper permutation $M^{2}\times M^{2}$ matrix defined as

$$\mathbf{P} = \sum_{i=0}^{M}(\mathbf{e}_{M\times 1}(i)\otimes\mathbf{I}_{M\times M})\otimes\mathbf{e}_{M\times 1}^{T}(i)$$

and $\mathbf{e}_{M\times 1}(i)$ is the unit vector with one on the i-th element and zero elsewhere. It holds that

$$\mathrm{E}\big((\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})(\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})^{T}\big) = (\tilde{\mathbf{u}}\tilde{\mathbf{u}}^{T})\otimes\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^{T}),$$

which allows to write the approximation (B8) as

$$\mathbf{W}_{\mathbf{u}}^{-1} \approx (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})(\mathbf{I}_{M^{2}\times M^{2}} + \mathbf{P})(\tilde{\mathbf{u}}\tilde{\mathbf{u}}^{T})\otimes\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^{\boldsymbol{T}})(\mathbf{I}_{M^{2}\times M^{2}} + \mathbf{P})^{T}(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})^{T}. \tag{B9}$$

Using the SVD of the real valued noise free signal $\varphi(\mathbf{S}) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{V}}^{T}$, we can approximate the projection $\Delta\mathbf{u}$ as [13, 28]

$$\Delta\mathbf{u} \approx \tilde{\lambda}_{0}^{-1}\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^{T}\varphi(\mathbf{Q})\tilde{\mathbf{v}}_{0} = \tilde{\lambda}_{0}^{-1}\tilde{\mathbf{v}}_{0}^{T}\otimes\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^{T}\,\mathrm{vec}\big(\varphi(\mathbf{Q})\big), \tag{B10}$$

where $\tilde{\lambda}_0$ is the first singular value of $\tilde{\boldsymbol{\Lambda}}$ given as $\tilde{\lambda}_0 = b_0\sqrt{2MN}$, $\tilde{\mathbf{v}}_0$ is the first right-singular vector of $\tilde{\mathbf{V}}$, $\tilde{\mathbf{U}}_{\mathbf{Q}}$ is the matrix of noise subspaces, and $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_{\mathbf{0}} \quad \tilde{\mathbf{U}}_{\mathbf{Q}}]$.

Using the approximation of the projection $\Delta\mathbf{u}$ (B10), we can express $\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^T)$ in (B9) as

$$\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^{\boldsymbol{T}}) = \tilde{\lambda}_{\mathbf{0}}^{-2}\tilde{\mathbf{v}}_0^T \otimes \tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T\, \mathrm{E}\left(\mathrm{vec}(\varphi(\mathbf{Q}))\mathrm{vec}(\varphi(\mathbf{Q}))^T\right)\, \tilde{\mathbf{v}}_0 \otimes \tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T. \tag{B11}$$

According to (3), $\varphi(\mathbf{Q}) = \mathbf{T}_{M\times M}^H[\mathbf{Q} \quad \mathbf{Q}_F]\mathbf{T}_{2N\times 2N}$, where $\mathbf{Q}_F$ is the flipped version of $\mathbf{Q}$ given as $\mathbf{Q}_F = \boldsymbol{\Pi}_{M\times M}\bar{\mathbf{Q}}\boldsymbol{\Pi}_{N\times N}$. We can write

$$\begin{aligned}\mathrm{E}\left(\mathrm{vec}(\varphi(\mathbf{Q}))\mathrm{vec}(\varphi(\mathbf{Q}))^T\right) &= \mathbf{T}_{2N\times 2N}^T \otimes \mathbf{T}_{M\times M}^H \mathrm{E}(\mathrm{vec}([\mathbf{Q} \quad \mathbf{Q}_F])\mathrm{vec}([\mathbf{Q} \quad \mathbf{Q}_F])^H)\bar{\mathbf{T}}_{2N\times 2N} \otimes \mathbf{T}_{M\times M} \\ &= \mathbf{T}_{2N\times 2N}^T \otimes \mathbf{T}_{M\times M}^H \sigma^2 \mathbf{I}_{2MN\times 2MN}\bar{\mathbf{T}}_{2N\times 2N} \otimes \mathbf{T}_{M\times M} = \sigma^2(\mathbf{T}_{2N\times 2N}^T \otimes \mathbf{T}_{M\times M}^H)(\bar{\mathbf{T}}_{2N\times 2N} \otimes \mathbf{T}_{M\times M}) \\ &= \sigma^2(\mathbf{T}_{2N\times 2N}^T\bar{\mathbf{T}}_{2N\times 2N}) \otimes (\mathbf{T}_{M\times M}^H\mathbf{T}_{M\times M}) = \sigma^2\mathbf{I}_{2N\times 2N} \otimes \mathbf{I}_{M\times M} = \sigma^2\mathbf{I}_{2MN\times 2MN}\end{aligned}$$

This allows us to write (B11) as

$$\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^{\boldsymbol{T}}) = \tilde{\lambda}_0^{-2}\sigma^2\tilde{\mathbf{v}}_0^T \otimes \tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T\tilde{\mathbf{v}}_0 \otimes \tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T = \tilde{\lambda}_0^{-2}\sigma^2(\tilde{\mathbf{v}}_0^T\tilde{\mathbf{v}}_0) \otimes (\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T) = \tfrac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T.$$

Considering that $\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T = \mathbf{I}_{M\times M} - \tilde{\mathbf{u}}_0\tilde{\mathbf{u}}_0^T$ [10], we get

$$\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^T) = \tfrac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2(\mathbf{I}_{M\times M} - \tilde{\mathbf{u}}_0\tilde{\mathbf{u}}_0^T). \tag{B12}$$

Considering (B12) and the fact that $\tilde{\mathbf{u}} = \sqrt{M/2}\,\tilde{\mathbf{u}}_0$, we can write the approximation (B9) as

$$\mathbf{W}_{\mathbf{u}}^{-1} \approx \tfrac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2(\mathbf{J}_{\mathbf{u}}^r + j\mathbf{J}_{\mathbf{u}}^i)(\mathbf{I}_{M^2\times M^2} + \mathbf{P})(\tilde{\mathbf{u}}\tilde{\mathbf{u}}^T) \otimes \left(\mathbf{I}_{M^2\times M^2} - \tfrac{2}{M}\tilde{\mathbf{u}}\tilde{\mathbf{u}}^T\right)(\mathbf{I}_{M^2\times M^2} + \mathbf{P})^T(\mathbf{J}_{\mathbf{u}}^r + j\mathbf{J}_{\mathbf{u}}^i)^T.$$

Using the properties of Kronecker product, we can write

$$\begin{aligned}\mathbf{W}_{\mathbf{u}}^{-1} &\approx -4\sigma^2\tilde{\lambda}_0^{-2}|a_{\mathbf{u}}|^2\mathbf{I}_{M^2\times M^2} \\ &+\frac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2(\mathbf{J}_{\mathbf{u}}^r + j\mathbf{J}_{\mathbf{u}}^i)(\tilde{\mathbf{u}} \otimes \mathbf{I}_{M\times M} + \mathbf{I}_{M\times M} \otimes \tilde{\mathbf{u}})\left((\mathbf{J}_{\mathbf{u}}^r + j\mathbf{J}_{\mathbf{u}}^i)(\tilde{\mathbf{u}} \otimes \mathbf{I}_{M\times M} + \mathbf{I}_{M\times M} \otimes \tilde{\mathbf{u}})\right)^H\end{aligned} \tag{B13}$$

It is obvious that the first term of (B13) is the multiplication of a non-squared matrix to its Hermitian transform.

Lemma2: For any arbitrary matrix $\mathbf{A} \in \mathbb{R}^{p_1\times p_2}$, $\mathbf{B} \in \mathbb{R}^{p_3\times p_4}$, $\mathbf{C} \in \mathbb{R}^{p_2\times p_5}$ and $\mathbf{D} \in \mathbb{R}^{p_4\times p_5}$, we have [29]

$$(\mathbf{A}\bullet\mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC})\bullet(\mathbf{BD}) \tag{B14}$$

Lemma3: For the vector $\mathbf{x} \in \mathbb{R}^{2p\times 1}, p \in \mathbb{N}$, and $\mathbf{J}_{\mathbf{x}}^r$ and $\mathbf{J}_{\mathbf{x}}^i$ defined in (16) and (17), the matrix $\mathbf{X} \triangleq (\mathbf{J}_{\mathbf{x}}^r + j\mathbf{J}_{\mathbf{x}}^i)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p} + \mathbf{I}_{2p\times 2p} \otimes \mathbf{x})$, can be written as the block matrix $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$ of two direct sums of matrices

$$\begin{aligned}\mathbf{X}_1 &= \bigoplus_{i\in I}\left[[\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L\mathbf{x} + j\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^R\mathbf{x}]_i \quad [\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\mathbf{x} - j\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^R\mathbf{x}]_i\right], \\ \mathbf{X}_2 &= \bigoplus_{i\in I}\left[[\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L\mathbf{x} - j\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^R\mathbf{x}]_i \quad [\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\mathbf{x} + j\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^R\mathbf{x}]_i\right],\end{aligned} \tag{B15}$$

where $I = \{0, 1, \dots, p-1\}$.

Proof: We expand matrix $\mathbf{X}$ as a summation of

$$\begin{aligned}\mathbf{X} &= (\mathbf{J}_{\mathbf{x}}^r + j\mathbf{J}_{\mathbf{x}}^i)(\mathbf{x} \oplus \mathbf{x}) = (\mathbf{J}_{\mathbf{x}}^r + j\mathbf{J}_{\mathbf{x}}^i)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p}) + (\mathbf{J}_{\mathbf{x}}^r + j\mathbf{J}_{\mathbf{x}}^i)(\mathbf{I}_{2p\times 2p} \otimes \mathbf{x}) \\ &= (\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p}) + (\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^R\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^R)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p}) + j(\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^R)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p}) + \\ &\quad j(\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^R\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p}) + (\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L)(\mathbf{I}_{2p\times 2p} \otimes \mathbf{x}) + (\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^R\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^R)(\mathbf{I}_{2p\times 2p} \otimes \mathbf{x}) \\ &\quad +j(\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^R)(\mathbf{I}_{2p\times 2p} \otimes \mathbf{x}) + j(\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^R\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L)(\mathbf{I}_{2p\times 2p} \otimes \mathbf{x}).\end{aligned}$$

In here, each element is a matrix-product of transposed Khatri-Rao and Kronecker products. By using Lemma2, we have

$$(\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\bullet\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L)(\mathbf{x} \otimes \mathbf{I}_{2p\times 2p}) = (\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\mathbf{x})\bullet(\mathbf{J}_{\mathbf{x}}^1\mathbf{J}_{\mathbf{x}}^L) = \left[\mathbf{0}_{(p-1)\times 1} \quad \mathrm{diag}(\mathbf{J}_{\mathbf{x}}^0\mathbf{J}_{\mathbf{x}}^L\mathbf{x}) \quad \mathbf{0}_{(p-1)\times p}\right].$$

In the same way,

$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R)(\mathbf{x}\otimes \mathbf{I}_{2p\times 2p}) = [\mathbf{0}_{(p-1)\times p} \quad \mathbf{0}_{(p-1)\times 1} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\mathbf{x})],$$
$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R)(\mathbf{x}\otimes \mathbf{I}_{2p\times 2p}) = [\mathbf{0}_{(p-1)\times p} \quad \mathbf{0}_{(p-1)\times 1} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\mathbf{x})],$$
$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L)(\mathbf{x}\otimes \mathbf{I}_{2p\times 2p}) = [\mathbf{0}_{(p-1)\times 1} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\mathbf{x}) \quad \mathbf{0}_{(p-1)\times p}],$$
$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L)(\mathbf{I}_{2p\times 2p}\otimes \mathbf{x}) = [\mathrm{diag}(\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\mathbf{x}) \quad \mathbf{0}_{(p-1)\times 1} \quad \mathbf{0}_{(p-1)\times p}],$$
$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R)(\mathbf{I}_{2p\times 2p}\otimes \mathbf{x}) = [\mathbf{0}_{(p-1)\times p} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\mathbf{x}) \quad \mathbf{0}_{(p-1)\times 1}],$$
$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R)(\mathbf{I}_{2p\times 2p}\otimes \mathbf{x}) = [\mathrm{diag}(\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\mathbf{x}) \quad \mathbf{0}_{(p-1)\times 1} \quad \mathbf{0}_{(p-1)\times p}],$$
$$(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R \bullet \mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L)(\mathbf{I}_{2p\times 2p}\otimes \mathbf{x}) = [\mathbf{0}_{(p-1)\times p} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\mathbf{x}) \quad \mathbf{0}_{(p-1)\times 1}].$$

520 Now, we write $\mathbf{X}$ as a block matrix

$$\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2],$$

where

$$\mathbf{X}_1 = [\mathbf{0}_{(p-1)\times 1} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\mathbf{x} - j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\mathbf{x})] + [\mathrm{diag}(\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\mathbf{x} + j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\mathbf{x}) \quad \mathbf{0}_{(p-1)\times 1}],$$
$$\mathbf{X}_2 = [\mathbf{0}_{(p-1)\times 1} \quad \mathrm{diag}(\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\mathbf{x} + j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\mathbf{x})] + [\mathrm{diag}(\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\mathbf{x} - j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\mathbf{x}) \quad \mathbf{0}_{(p-1)\times 1}].$$

521 Hence

$$\mathbf{X}_1 = \bigoplus_{i\in I}\left[[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\mathbf{x} + j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\mathbf{x}]_i \quad [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\mathbf{x} - j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\mathbf{x}]_i\right],$$
$$\mathbf{X}_2 = \bigoplus_{i\in I}\left[[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\mathbf{x} - j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\mathbf{x}]_i \quad [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\mathbf{x} + j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\mathbf{x}]_i\right],$$

for $I = \{0, 1, \dots, p-1\}$. ■

522 Let us define the matrix $\mathbf{H}$ as
$$\mathbf{H} \triangleq (\mathbf{J}_\mathbf{u}^r + j\mathbf{J}_\mathbf{u}^i)(\tilde{\mathbf{u}} \otimes \mathbf{I}_{M\times M} + \mathbf{I}_{M\times M} \otimes \tilde{\mathbf{u}}).$$

523 So, we can express $\mathbf{W}_\mathbf{u}^{-1}$ in (B13) as

$$\mathbf{W}_\mathbf{u}^{-1} \approx -4\sigma^2\tilde{\lambda}_0^{-2}|a_\mathbf{u}|^2\mathbf{I}_{M^2\times M^2} + \frac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2\,\mathbf{H}\mathbf{H}^H. \tag{B16}$$

524 Considering Lemma 3, and with respect to (16), matrix $\mathbf{H}$ can be written as the block matrix

$$\mathbf{H} \triangleq [\mathbf{H}_1 \quad \mathbf{H}_2],$$

525 where $\mathbf{H}_1$ and $\mathbf{H}_2$ are

$$\mathbf{H}_1 = \bigoplus_{i\in I}\left[[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i \quad [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i\right],$$
$$\mathbf{H}_2 = \bigoplus_{i\in I}\left[[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i \quad [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i\right],$$

and $I = \{0, 1, \dots, p-1\}$.

526 Also, the conjugate transpose form of $\mathbf{H}$ can be expressed as

$$\mathbf{H}^H = \begin{bmatrix}\mathbf{H}_3\\ \mathbf{H}_4\end{bmatrix},$$

$$\mathbf{H}_3 = \bigoplus_{i\in I}\begin{bmatrix}[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i\\ [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i\end{bmatrix},$$

$$\mathbf{H}_4 = \bigoplus_{i\in I}\begin{bmatrix}[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i\\ [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i\end{bmatrix}.$$

527 Thus, the multiplication of these two block matrices can be written as a block matrix itself as

$$\mathbf{H}\mathbf{H}^H = \mathbf{H}_1\mathbf{H}_3 + \mathbf{H}_2\mathbf{H}_4$$

528 Now using the distribution property of direct sum over matrix multiplication, we can say

$$\mathbf{H}\mathbf{H}^H = \bigoplus_{i\in I}\;[[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i \quad [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i]\begin{bmatrix}[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i\\ [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}}]_i\end{bmatrix}$$
$$+\bigoplus_{i\in I}\;[[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i \quad [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i]\begin{bmatrix}[\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} + j\mathbf{J}_\mathbf{x}^1\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i\\ [\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^R\tilde{\mathbf{u}} - j\mathbf{J}_\mathbf{x}^0\mathbf{J}_\mathbf{x}^L\tilde{\mathbf{u}}]_i\end{bmatrix}$$
$$= \mathrm{diag}\big((\mathbf{J}_\mathbf{u}^1\mathbf{J}_\mathbf{u}^L\tilde{\mathbf{u}})\odot(\mathbf{J}_\mathbf{u}^1\mathbf{J}_\mathbf{u}^L\tilde{\mathbf{u}}) + (\mathbf{J}_\mathbf{u}^1\mathbf{J}_\mathbf{u}^R\tilde{\mathbf{u}})\odot(\mathbf{J}_\mathbf{u}^1\mathbf{J}_\mathbf{u}^R\tilde{\mathbf{u}})\big)$$

$$+\operatorname{diag}\left((\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{L}\tilde{\mathbf{u}})\odot(\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{L}\tilde{\mathbf{u}})+(\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{R}\tilde{\mathbf{u}})\odot(\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{R}\tilde{\mathbf{u}})\right)$$
$$+\operatorname{diag}\left((\mathbf{J}_{\mathbf{u}}^{1}\mathbf{J}_{\mathbf{u}}^{R}\tilde{\mathbf{u}})\odot(\mathbf{J}_{\mathbf{u}}^{1}\mathbf{J}_{\mathbf{u}}^{R}\tilde{\mathbf{u}})+(\mathbf{J}_{\mathbf{u}}^{1}\mathbf{J}_{\mathbf{u}}^{L}\tilde{\mathbf{u}})\odot(\mathbf{J}_{\mathbf{u}}^{1}\mathbf{J}_{\mathbf{u}}^{L}\tilde{\mathbf{u}})\right)$$
$$+\operatorname{diag}\left((\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{R}\tilde{\mathbf{u}})\odot(\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{R}\tilde{\mathbf{u}})+(\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{L}\tilde{\mathbf{u}})\odot(\mathbf{J}_{\mathbf{u}}^{0}\mathbf{J}_{\mathbf{u}}^{L}\tilde{\mathbf{u}})\right)$$

which is a diagonal matrix. For $k=\left\{0,\dots,\frac{M}{2}-2\right\}$, the k-th diagonal element of $\mathbf{HH}^H$ is given as

$$[\operatorname{diag}(\mathbf{HH}^H)]_k=\left(\cos^2\left(\frac{\omega}{2}(M-2(k+1)-1)\right)+\sin^2\left(\frac{\omega}{2}(M-2(k+1)-1)\right)\right)+\left(\cos^2\left(\frac{\omega}{2}(M-2k-1)\right)+\sin^2\left(\frac{\omega}{2}(M-2k-1)\right)\right)+\left(\sin^2\left(\frac{\omega}{2}(M-2(k+1)-1)\right)+\cos^2\left(\frac{\omega}{2}(M-2(k+1)-1)\right)\right)+\left(\sin^2\left(\frac{\omega}{2}(M-2k-1)\right)+\cos^2\left(\frac{\omega}{2}(M-2k-1)\right)\right)=4\,.$$

Thus,

$$[\mathbf{HH}^H]_{m,n}=\begin{cases}4 & m=n\\ 0 & m\neq n\end{cases}\tag{B17}$$

Considering this, we can simplify the approximation (B16). It holds that

$$\mathbf{W}_{\mathbf{u}}^{-1}\approx 2\tilde{\lambda}_0^{-2}\sigma^2M\left(\mathbf{I}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}-\frac{2}{M}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times 1}\right).$$

Now, considering that the determinant is non-zero

$$\mathbf{1}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}-\frac{2}{M}\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}\mathbf{I}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}=2/M\neq 0,$$

the approximation is invertible and can be obtained using Sherman-Morrison Formula [30]

$$\mathbf{W}_{\mathbf{u}}\approx\frac{\tilde{\lambda}_0^2}{2\sigma^2M}\left(\mathbf{I}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}+\frac{\frac{2}{M}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times 1}}{1-\frac{2}{M}\mathbf{1}^T_{\left(\frac{M}{2}-1\right)\times 1}\mathbf{1}_{\left(\frac{M}{2}-1\right)\times 1}}\right).$$

The equation indicates a perturbed diagonal matrix. In fact, the second term can be interpreted as the correlation of the estimation residual error.

Considering that $\tilde{\lambda}_0=b_0\sqrt{2MN}$, we get the final form of the approximation

$$\mathbf{W}_{\mathbf{u}}\approx\frac{b_0^2N}{\sigma^2}\left(\mathbf{I}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}+\mathbf{1}_{\left(\frac{M}{2}-1\right)\times\left(\frac{M}{2}-1\right)}\right)$$

6.3. Appendix C

To prove the convergence of LUPUMA, we utilize equations (27) and (B4) to find the expectation of estimated $a_{\mathbf{u}}$ as

$$\mathrm{E}(\hat{a}_{\mathbf{u}})=\frac{2}{M-2}\mathbf{1}^T\mathrm{E}(\mathbf{y}_{\mathbf{u}})=a_{\tilde{\mathbf{u}}}+\frac{2}{M-2}\mathbf{1}^T\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}})\,.\tag{C1}$$

Based on (B6), we express $\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}})$ as

$$\begin{aligned}\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}})&=\mathrm{E}\left((\mathbf{J}_{\mathbf{u}}^{r}+j\mathbf{J}_{\mathbf{u}}^{i})(\tilde{\mathbf{u}}\otimes\Delta\mathbf{u})\right)+\mathrm{E}\left((\mathbf{J}_{\mathbf{u}}^{r}+j\mathbf{J}_{\mathbf{u}}^{i})(\Delta\mathbf{u}\otimes\tilde{\mathbf{u}})\right)+\mathrm{E}\left((\mathbf{J}_{\mathbf{u}}^{r}+j\mathbf{J}_{\mathbf{u}}^{i})(\Delta\mathbf{u}\otimes\Delta\mathbf{u})\right)\\&=(\mathbf{J}_{\mathbf{u}}^{r}+j\mathbf{J}_{\mathbf{u}}^{i})\operatorname{vec}(\tilde{\mathbf{u}}\,\mathrm{E}(\Delta\mathbf{u})^T)+(\mathbf{J}_{\mathbf{u}}^{r}+j\mathbf{J}_{\mathbf{u}}^{i})\operatorname{vec}(\mathrm{E}(\Delta\mathbf{u})\,\tilde{\mathbf{u}}^T)+(\mathbf{J}_{\mathbf{u}}^{r}+j\mathbf{J}_{\mathbf{u}}^{i})\operatorname{vec}\left(\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^T)\right)\end{aligned}\tag{C2}$$

According to (B10), $\mathrm{E}(\Delta\mathbf{u})$ is

$$\mathrm{E}(\boldsymbol{\Delta}\mathbf{u})=\tilde{\lambda}_0^{-1}\tilde{\mathbf{v}}_0^T\otimes\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T\,\mathrm{E}\left(\operatorname{vec}\left(\varphi(\mathbf{Q})\right)\right)=\tilde{\lambda}_0^{-1}\tilde{\mathbf{v}}_0^T\otimes\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T\,\mathbf{T}_{2N\times 2N}^T\otimes\mathbf{T}_{M\times M}^H\mathrm{E}(\operatorname{vec}([\mathbf{Q}\quad\mathbf{Q}_F]))\,.$$

For $\mathrm{E}(\operatorname{vec}([\mathbf{Q}\quad\mathbf{Q}_F]))$ we have

$$\begin{aligned}\mathrm{E}(\operatorname{vec}([\mathbf{Q}\quad\mathbf{Q}_F]))&=\mathrm{E}\left(\operatorname{vec}([\mathbf{Q}\quad\boldsymbol{\Pi}_{M\times M}\overline{\mathbf{Q}}\boldsymbol{\Pi}_{N\times N}])\right)\\&=\left[\mathrm{E}\left(\operatorname{vec}(\mathbf{Q})\right)\quad\boldsymbol{\Pi}_{N\times N}^T\otimes\boldsymbol{\Pi}_{M\times M}\;\mathrm{E}\left(\operatorname{vec}(\overline{\mathbf{Q}})\right)\right]=\left[\boldsymbol{\mu}_{\boldsymbol{q}}\quad\boldsymbol{\Pi}_{N\times N}^T\otimes\boldsymbol{\Pi}_{M\times M}\;\overline{\boldsymbol{\mu}_{\boldsymbol{q}}}\right],\end{aligned}$$

where $\boldsymbol{\mu}_{\boldsymbol{q}}$ is the mean of received noise. Assuming $\boldsymbol{\mu}_{\boldsymbol{q}}=\overline{\boldsymbol{\mu}_{\boldsymbol{q}}}=0$, we can say

$$\mathrm{E}(\boldsymbol{\Delta u})=\tilde{\lambda}_0^{-1}\tilde{\mathbf{v}}_0^T\otimes\tilde{\mathbf{U}}_{\mathbf{Q}}\tilde{\mathbf{U}}_{\mathbf{Q}}^T\,\mathbf{T}_{2N\times 2N}^T\otimes\mathbf{T}_{M\times M}^H\left[\boldsymbol{\mu}_{\boldsymbol{q}}\quad\boldsymbol{\Pi}_{N\times N}^T\otimes\boldsymbol{\Pi}_{M\times M}\;\overline{\boldsymbol{\mu}_{\boldsymbol{q}}}\right],\tag{C3}$$

$$\mathrm{E}(\boldsymbol{\Delta u}) = 0$$

Now by substituting equations (C3) and (B16) into (C2), we simplify $\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}})$ as

$$\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}) = (\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\mathrm{vec}\left(\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}^{T})\right)$$

$$\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}) = \tfrac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\mathrm{vec}\left((\mathbf{I}_{M\times M} - \tilde{\mathbf{u}}_0\tilde{\mathbf{u}}_0^{T})\right)$$

$$\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}) = -\tfrac{M}{2}\tilde{\lambda}_0^{-2}\sigma^2(\mathbf{J}_{\mathbf{u}}^{r} + j\mathbf{J}_{\mathbf{u}}^{i})\mathrm{vec}(\tilde{\mathbf{u}}_0\tilde{\mathbf{u}}_0^{T})$$

So, we can write (C1) as

$$\mathrm{E}(\hat{a}_{\mathbf{u}}) = \left(1 - \tfrac{M}{M-2}\tilde{\lambda}_0^{-2}\sigma^2\right)a_{\tilde{\mathbf{u}}}\,.$$

To find the relation between convergence of $\widehat{\omega}_{\mathbf{u}}$ and $\hat{a}_{\mathbf{u}}$, we define the function $g(.)$ as

$$g(\hat{a}_{\mathbf{u}}) = \widehat{\omega}_{\mathbf{u}},$$

In this way, we can expand $g(\hat{a}_{\mathbf{u}})$ using Taylor series as

$$g(\hat{a}_{\mathbf{u}}) = \omega_{\tilde{\mathbf{u}}} + g'(a_{\tilde{\mathbf{u}}})\left(\tfrac{2}{M-2}\mathbf{1}^{T}\Delta\mathbf{y}_{\mathbf{u}} + \left(\tfrac{M}{M-2}\tilde{\lambda}_0^{-2}\sigma^2\right)a_{\tilde{\mathbf{u}}}\right) + O(\Delta\mathbf{y}_{\mathbf{u}}^{2})\,,$$

Expected value of this function is written as

$$\begin{aligned}\mathrm{E}\left(g(\hat{a}_{\mathbf{u}})\right) &= \omega_{\tilde{\mathbf{u}}} + g'(a_{\tilde{\mathbf{u}}})\left(\frac{2}{M-2}\mathbf{1}^{T}\mathrm{E}(\Delta\mathbf{y}_{\mathbf{u}}) + \left(\frac{M}{M-2}\tilde{\lambda}_0^{-2}\sigma^2\right)a_{\tilde{\mathbf{u}}}\right) + \mathrm{E}\left(O(\Delta\mathbf{y}_{\mathbf{u}}^{2})\right)\\ &= \widehat{\omega}_{\tilde{\mathbf{u}}} + \mathrm{E}\left(O(\Delta\mathbf{y}_{\mathbf{u}}^{2})\right),\end{aligned}$$

which can be approximated as

$$\mathrm{E}\left(g(\hat{a}_{\mathbf{u}})\right) \approx \omega_{\tilde{\mathbf{u}}}. \tag{C4}$$

This approximation is accurate for high SNR values. Similarly, for $\mathbf{v}$, we can write

$$\mathrm{E}\left(g(\hat{a}_{\mathbf{v}})\right) \approx \omega_{\tilde{\mathbf{v}}}. \tag{C5}$$

Now, we substitute the equations of (C4) and (C5) into (36). we can say that for high SNR values, $\widehat{\omega}$ is unbiased as

$$\mathrm{E}(\widehat{\omega}) = \frac{(M-2)}{2M^2(N-1)+(M-2)}\mathrm{E}(\widehat{\omega}_{\mathbf{u}}) + \frac{+\,2M^2(N-1)}{2M^2(N-1)+(M-2)}\mathrm{E}(\widehat{\omega}_{\mathbf{v}})$$

$$\mathrm{E}(\widehat{\omega}) \approx \omega$$

7. References

[1] T. Kim, K. Ko, I. Hwang, D. Hong, S. Choi, and H. Wang, "RSRP-Based Doppler Shift Estimator Using Machine Learning in High-Speed Train Systems," *IEEE Transactions on Vehicular Technology,* vol. 70, no. 1, pp. 371-380, 2020.

[2] A. Pourafzal, T. Roi-Taravella, M. Cheffena, and S. Yildirim, "A Low-cost and Accurate Microwave Sensor System for Permittivity Characterization," *IEEE Sensors Journal,* pp. 1-1, 2022, doi: 10.1109/JSEN.2022.3225662.

[3] G. Campobello, A. Segreto, and N. Donato, "A novel low-complexity frequency estimation algorithm for industrial internet-of-things applications," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-10, 2020.

[4] C. Wu, M. E. Magaña, and E. Cotilla-Sánchez, "Dynamic frequency and amplitude estimation for three-phase unbalanced power systems using the unscented Kalman filter," *IEEE Transactions on Instrumentation and Measurement,* vol. 68, no. 9, pp. 3387-3395, 2018.

[5] E. Aboutanios and B. Mulgrew, "Iterative frequency estimation by interpolation on Fourier coefficients," *IEEE Transactions on signal processing,* vol. 53, no. 4, pp. 1237-1242, 2005.

[6] A. Serbes, "Fast and efficient sinusoidal frequency estimation by using the DFT coefficients," *IEEE Transactions on Communications,* vol. 67, no. 3, pp. 2333-2342, 2018.

[7] M. Morelli, M. Moretti, and A. A. D'Amico, "Single-Tone Frequency Estimation by Weighted Least-Squares Interpolation of Fourier Coefficients," *IEEE Transactions on Communications,* vol. 70, no. 1, pp. 526-537, 2021.

[8] K. Wu, J. A. Zhang, X. Huang, and Y. J. Guo, "Accurate frequency estimation with fewer DFT interpolations based on Padé approximation," *IEEE Transactions on Vehicular Technology,* vol. 70, no. 7, pp. 7267-7271, 2021.

[9] Ç. Candan, "A method for fine resolution frequency estimation from three DFT samples," *IEEE Signal processing letters,* vol. 18, no. 6, pp. 351-354, 2011.

[10] H. C. So, F. K. W. Chan, and W. Sun, "Subspace approach for fast and accurate single-tone frequency estimation," *IEEE Transactions on Signal Processing,* vol. 59, no. 2, pp. 827-831, 2010.

[11] C. Qian, L. Huang, H. C. So, N. D. Sidiropoulos, and J. Xie, "Unitary PUMA algorithm for estimating the frequency of a complex sinusoid," *IEEE Transactions on Signal Processing,* vol. 63, no. 20, pp. 5358-5368, 2015.

[12] S. Djukanović and V. Popović-Bugarin, "Efficient and accurate detection and frequency estimation of multiple sinusoids," *IEEE Access,* vol. 7, pp. 1118-1125, 2018.

[13] H.-C. So, F. K. Chan, W. H. Lau, and C.-F. Chan, "An efficient approach for two-dimensional parameter estimation of a single-tone," *IEEE Transactions on Signal Processing,* vol. 58, no. 4, pp. 1999-2009, 2009.

[14] C. E. Heckler, "Applied multivariate statistical analysis," ed: Taylor & Francis, 2005.

[15] M. Haardt and J. A. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE transactions on signal processing,* vol. 43, no. 5, pp. 1232-1242, 1995.

[16] C. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhyā: the Indian journal of statistics, series A,* pp. 167-180, 1968.

[17] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: unification, simplification, and observations," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 29, no. 4, pp. 1170-1184, 1993.

[18] J. M. Bates and C. W. Granger, "The combination of forecasts," *Journal of the Operational Research Society,* vol. 20, no. 4, pp. 451-468, 1969.

[19] S. Kay, "A fast and accurate single frequency estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 37, no. 12, pp. 1987-1990, 1989.

[20] Y.-X. Yao and S. M. Pandit, "Variance of least squares estimators for a damped sinusoidal process," *IEEE transactions on signal processing,* vol. 42, no. 11, pp. 3016-3025, 1994.

[21] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Transactions on information theory,* vol. 20, no. 5, pp. 591-598, 1974.

[22] B. De Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE transactions on signal processing,* vol. 41, no. 9, pp. 2826-2838, 1993.

[23] T. F. Chan, "An improved algorithm for computing the singular value decomposition," *ACM Transactions on Mathematical Software (TOMS),* vol. 8, no. 1, pp. 72-83, 1982.

[24] E. Aboutanios, "A modified dichotomous search frequency estimator," *IEEE Signal Processing Letters,* vol. 11, no. 2, pp. 186-188, 2004.

[25] S. Djukanović, T. Popović, and A. Mitrović, "Precise sinusoid frequency estimation based on parabolic interpolation," in *2016 24th Telecommunications Forum (TELFOR)*, 2016: IEEE, pp. 1-4.

[26] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[27] E. Anderson *et al.*, *LAPACK Users' guide*. SIAM, 1999.

[28] J. Liu, X. Liu, and X. Ma, "First-order perturbation analysis of singular vectors in singular value decomposition," *IEEE Transactions on Signal Processing,* vol. 56, no. 7, pp. 3044-3049, 2008.

[29] V. Slyusar, "A family of face products of matrices and its properties," *Cybernetics and Systems Analysis,* vol. 35, no. 3, pp. 379-384, 1999.

[30] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *The Annals of Mathematical Statistics,* vol. 21, no. 1, pp. 124-127, 1950.