Fast MUSIC Algorithm
for mm-Wave Massive-MIMO Radar

Bin Li[1,3], Shuseng Wang[2], Jun Zhang[3], Xainbin Cao[4], Chenglin Zhao[1]

Abstract—Millimeter-wave radar, in marriage with virtual massive multiple-input multiple-output (MIMO) system, provides great promises to super-resolution environment sensing and target detection in unmanned ground/aerial vehicle (UGA/UAV). However, the existing target detection algorithms suffer from either low resolution/accuracy or high time complexity. In this work, we propose two efficient approximate algorithms to significantly accelerate the high-resolution subspace algorithm. In particular, our fast MUST algorithms speed up the expensive signal subspace identification step using the randomized low-rank approximations. We also establish theoretical bounds to guarantee the quality of approximation. Specifically, we show that, in the case of high signal-to-noise ratio, the pseudo-spectrum produced by our algorithms are very close to the exact MUSIC. Our theoretical analysis is original, which does not simply follow from any existing work. Our comprehensive empirical study shows that our algorithms are tremendously faster than MUSIC, while their target detection accuracies in massive MIMO radar are almost as good as MUSIC. This work enables super-resolution automotive sensing in real time and thus has great practical value in emerging automotive applications.

Index Terms—Millimeter-wave radar, massive MIMO, automotive, fast MUSIC, randomized numerical linear algebra, the Nyström method.

I. INTRODUCTION

Unmanned aircrafts and vehicles are receiving much attention from both the industry and academia [1], [2] due to their great potential and usefulness in real-world applications [3]. The success of unmanned aircrafts and vehicles critically depends on the fusing of GPS, automotive radar, and other environmental sensors (e.g., lidar, ultrasound, and camera) [4], [5]. Millimeter-wave automotive radar is attractive to unmanned aircrafts/vehicles; in comparison to the other sensors, automotive radar has two advantages: first, it is immune to adverse environmental conditions such as dust, fog, and smoke, and second, it is robust to dazzling and no light [6]. The recent advancement in millimeter-wave (mm-wave) semiconductor circuit (e.g. 24 and 77GHz) [7]–[10] makes it possible to deploy large-scale arrays economically to unmanned aircrafts and vehicles. By cascading virtual Multiple-Input Multiple-Output (MIMO) or co-located MIMO systems [11], hundreds of reception channels are made available for achieving super-resolution environment sensing and target detection [12].

1 School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

2 Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ 07030, USA.

3 School of Information and Electronics, Beijing Institute of Technology, Beijing, 100081, China.

4 School of Electronic and Information Engineering, Beihang University, Beijing, 100191, China.

Unmanned vehicles typically use three types of mm-wave radars [7]: (1) long-range radar (LRR) for automotive cruise control (i.e. 10-250 m), (2) medium-range radar (LRR) for cross-traffic alert (i.e. 1-100 m), and (3) short-range radar (SRR) for park assist (i.e. 0.15-30 m); see the illustration in Figure 1(a). To meet such diverse requirements, various modulation waveforms have been developed for automotive radars [13], [14]. Successful examples include frequency-modulated continuous-waveform (FMCW) [15]–[17], pulsed continuous wave (CW) radar, and orthogonal frequency-division multiplexing (OFDM) [18], etc. In particular, FMCW can sweep a wide RF bandwidth (GHz) while keeping simultaneously a small IF bandwidth (MHz), and thereby permits high-resolution processing with low-cost circuits [12].

Suppose the radar system has M antennas and each antenna receives N discrete signals. The task is to use the $M \times N$ received signals to detect unknown targets. More specifically, if there are K targets, then we are interested to estimate the angle-of-arrival (AoA) and times-of-arrival (ToA) of all these K targets. Multiple Signal Classification (MUSIC) [19] and Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) [20] are the most widely used algorithms for such tasks. MUSIC and ESPRIT identify noise and signal subspaces via the singular value decomposition (SVD) of a co-variance matrix, and they can attain near-optimal performance [21]. However, the time complexity of SVD has a high-order dependence on M (or N). As for AoA estimation, the time complexities of the SVDs are $\mathcal{O}(M^3)$ for 1D and $\mathcal{O}(M^3 N^3)$ for 2D joint case, making MUSIC and ESPRIT prohibitive in high-resolution estimation tasks (e.g. $M > 200$)[1].

Despite the great advancements in the circuits and signal processing for automotive massive-MIMO radars [14], [22], high-resolution, yet real-time and low-complexity, algorithm is still lacking. In automotive scenarios, high-resolution environment sensing requires a large number of channels and massive samples [12] which incurs high computation cost. When deployed to unmanned vehicles, the high computation cost causes intolerable delays in decision making which would potentially leads to disastrous consequences. Thus, the deployment of high-resolution massive-MIMO radar to unmanned aircrafts/vehicles calls for efficient and real-time algorithms.

Alternatives to MUSIC and ESPRIT have been developed by prior work. By introducing a so-called propagator operator, the Propagator algorithm [23] reduces the time complexity to $\mathcal{O}(NMK)$ [24]; however, it unfortunately results in low-

resolution and uncertain estimations (especially in the range of $[70, 90]$ degree) [25]. Fast Fourier transform (FFT) reduces the time complexity even to $\mathcal{O}(M \log M)$ for 1-D estimation [26]–[28], and it has been adopted by some automotive radar systems [28], [29]. Despite its computational efficiency, the FFT algorithm fails to provide high accuracy and thus may be unsuitable for super-resolution automotive sensing [14].

To enable real-time sensing, in this work we propose to use the randomized low-rank approximation to overcome the computational bottleneck in high-resolution massive-MIMO radar. In particular, we propose to use the Nyström method [30]–[32] for approximating the $M \times M$ spatial covariance matrix $\mathbf{S}$ which is positive semi-definite. In this way, the SVD of $\mathbf{S}$, or the identification of signal subspace, can be performed very efficiently via our proposed algorithm.

Note that, using the Nyström method to speed up the SVD of positive semi-definite matrix is not a new idea; in fact, it has been widely studied in scientific computing, numerical linear algebra, and machine learning [33], [34]. However, exploiting the Nyström method to solve this classical radar signal-processing problem is not a trivial application at all, especially given the following reasons.

- From practical perspective, it was unclear whether low-rank approximation is applicable for MUSIC. The Nyström method, as well as other low-rank approximations, approximates the "top" K singular values and singular vector of $\mathbf{S}$. However, MUSIC relies commonly on the "bottom" singular vectors–noise subspace–for target detection. There is an obvious mismatch between Nyström and MUSIC. Before this work, low-rank approximation has not been used to speed up MUSIC.
- From theoretical perspective, the existing theories for the Nyström method do not apply. In this work, the objective of Nyström approximation is to preserve pseudo-spectrum $P_{\text{music}}(\theta)$. However, the existing theories of the Nyström method [30], [34]–[37] bound the matrix norm errors such as $\|\mathbf{S} - \tilde{\mathbf{S}}\|$, where $\tilde{\mathbf{S}}$ is a low-rank approximation to $\mathbf{S}$. Unfortunately, such matrix norm bounds do not lend any support to MUSIC+Nyström at all. We need show that with $\mathbf{S}$ replaced by $\tilde{\mathbf{S}}$, the change in $P_{\text{music}}(\theta)$ is bounded; this is much harder than showing the matrix norm bounds.

In sum, this work offers the following contributions.

- We propose to use the Nyström method for speeding up the SVD step in the MUSIC algorithm. Our fast MUSIC algorithm (namely MUSIC+Nyström) has $\mathcal{O}(Mp^2)$ time complexity, as opposed to MUSIC's $\mathcal{O}(M^3)$ complexity; here, p ($K \leq p \ll M$) is a user-specified parameter trading off accuracy and computation. Hence, our fast algorithm enables high-resolution target detection.
- We establish a theoretical bound for our fast MUSIC algorithm. We show that using the Nyström approximation with $p = \Omega(K \log K)$, $P_{\text{music}}(\theta)$ is at most

$$\mathcal{O}\left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})} \sqrt{\frac{M}{p}}\right)$$

times worse than the original. Here, $\sigma_k(\mathbf{S})$ is the k-th largest singular value of $\mathbf{S}$. If the signal-to-noise ratio
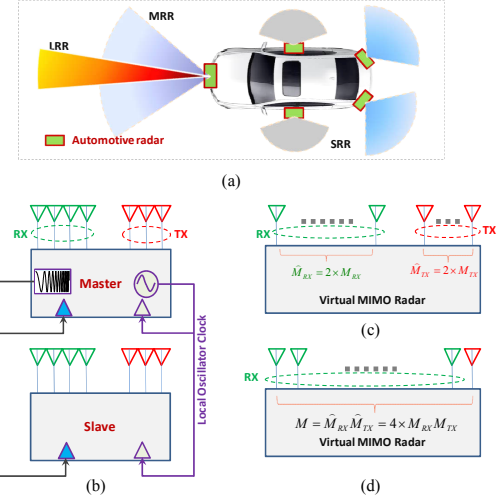

Fig. 1. **Millimeter-wave massive MIMO radar with cascaded sub-arrays**. (a) Typical automotive radars for autonomous driving systems, including LRR, MRR and SRR. (b) A schematic framework on cascaded MIMO radar, whereby the master generates the local oscillator signal and controls the slave via two synchronized clocks, i.e. FMCW clock and LO clock. Here, each co-located MIMO radar is equipped with $M_{TX} = 3$ transmitting antennas and $M_{RX} = 4$ receiving antennas. (c) The equivalent MIMO radar system of two cascaded sub-systems. (d) The cascaded-based massive MIMO radar system with the equivalent channels of $M = 4 \times M_{RX} M_{TX}$.

(SNR) is high, the spectral gap between the K-th and $(K + 1)$-th singular values is big, making our bound strong.

- We further extend our algorithm and theory to the other power method. We show that by performing t power iterations, $P_{\text{music}}(\theta)$ is at most

$$\mathcal{O}\left(\left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1} \sqrt{MK}\right)$$

times worse than the original. Obviously, the ratio converges to zero exponentially. If the SNR is high, several power iterations suffice for an almost optimal estimation.
- We conduct comprehensive experiments for evaluating the proposed algorithms and compare them with the counterparts such as MUSIC, ESPRIT, etc. The empirical studies demonstrate the efficiency and accuracy of our fast MUSIC algorithms.

The remaining of the work is structured as follows. In Section II, we briefly introduce the system model and existing algorithms for automotive sensing. In Section III, we propose the randomised low-rank approximation based fast MUSIC scheme. In Section IV, we derive some theoretical bounds of interests. In Section V, numerical simulation are provided. Finally, we conclude this work in Section VI.

II. NOTATION

This section defines the notation used throughout. The most commonly used notation is summarized in Table I. Let $\mathbf{A}$ be any $d_1 \times d_2$ matrix and $r = \text{rank}(\mathbf{A})$. The singular value decomposition (SVD) of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H = \sum_{i=1}^{r} \sigma_i(\mathbf{A}) \mathbf{u}_i \mathbf{v}_i^H,$$

TABLE I
COMMONLY USED NOTATION.

Notation	Definition
M	number of antennas
N	number of discrete samples
K	number of objects
$\mathbf{Y}$	$M \times N$ signal matrix
$\mathbf{S} = \mathbf{Y}\mathbf{Y}^T$	$M \times M$ spatial matrix

where $\sigma_i(\mathbf{A})$ (> 0) is the i-th diagonal entry of $\boldsymbol{\Sigma}$ $(r \times r)$, $\mathbf{u}_i$ is the i-th column of $\mathbf{U}$ $(d_1 \times r)$, and $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$ $(d_2 \times r)$. Suppose the singular values $\sigma_1(\mathbf{A}), \cdots, \sigma_r(\mathbf{A})$ are in the descending order. The spectral norm of $\mathbf{A}$ satisfies

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \sigma_1(\mathbf{A}).$$

The Moore-Penrose pseudo-inverse of $\mathbf{A}$ is

$$\mathbf{A}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H = \sum_{i=1}^{r} \sigma_i^{-1}(\mathbf{A})\, \mathbf{v}_i \mathbf{u}_i^H.$$

The best rank-k $(k < r)$ approximation to $\mathbf{A}$ is

$$\mathbf{A}_k \triangleq \sum_{i=1}^{k} \sigma_i(\mathbf{A})\, \mathbf{u}_i \mathbf{v}_i^H.$$

If $\mathbf{A}$ is a positive semi-definite matrix, then $\mathbf{U} = \mathbf{V}$ which means the SVD is also the eigenvalue decomposition of $\mathbf{A}$.

Let $\mathbf{U}_K$ be an $M \times K$ matrix with orthonormal columns. The row coherence of $\mathbf{U}_K$ is then defined as

$$\mu(\mathbf{U}_K) = \frac{M}{K} \max_i \left\|(\mathbf{U}_K)_{i:}\right\|_2^2 \in \left[1, \frac{M}{K}\right], \qquad (1)$$

where $(\mathbf{U}_K)_{i:}$ is the i-th row of $\mathbf{U}_K$. This coherence notation has been used in compressed sensing [38], matrix completion [39], and randomized linear algebra [30], [40], [41].

III. SYSTEM MODEL

In this section, we briefly introduce the FMCW radar system for autonomous vehicles and four widely used algorithms for objection detection and parameter estimation.

A. FMCW Radar System

Consider a massive-MIMO radar system (either virtual MIMO or real MIMO) consisting of M antennas, e.g. using the uniform linear array. For the m-th antenna, the emitted FMCW waveform within a symbol duration T_{sym} is [15], [28]:

$$s(t) = \exp\left[j\left(w_s t + \frac{\mu}{2}t^2\right)\right], \quad 0 \le t < T_{\text{sym}}, \qquad (2)$$

where w_s is the initial frequency and μ is the changing rate of the instantaneous frequency in chirp signal. Given the bandwidth of FMCW signal, w_B, and the emission period duration, T_{sym}, the changing rate is determined by $\mu = w_B/T_{\text{sym}}$.

Suppose there are K targets. We then seek to estimate the ToAs and AoAs, denoted as τ_k and θ_k, respectively, for $k = 0$
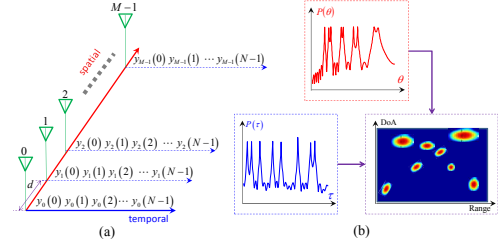

Fig. 2. **Schematic paradigm of unknown targets detection and estimation**. (a) Taking ULA for example, the data-flow of multiple channels (i.e. M) are reorganized to form a data matrix, $\mathbf{Y} \in \mathbb{C}^{M \times N}$. (b) The AoAs and time delays of K sources are computed using $\mathbf{S} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ and $\mathbf{T} = \frac{1}{M}\mathbf{Y}^H\mathbf{Y}$. Further with either a 2-D or the low-complexity 1-D paring mechanisms [42], [43], the range-angle (R-A) plane is attained.

to $K - 1$. After the signal calibration, the received signal at the m-th antenna reads:

$$y_m(t) = \sum_{k=0}^{K-1} \alpha_{k,m} s(t - \tau_k) \times \exp\left(j\frac{2\pi}{\lambda_s}md\sin\theta_k\right) + n_m(t).$$

Here, $r_k(t) \triangleq \alpha_{k,m} s(t - \tau_k)$ is the signal reflected from the k-th target; $\alpha_{k,m}$ is the complex channel gain between the k-th target and the m-th element; (for the far-filed case, $\alpha_{k,m} = \alpha_k$;) λ_s denotes the wavelength of a carrier signal; d is the spacing distance of two adjacent elements (we assume $d = \lambda_s/2$); $n_m(t) \sim \mathcal{N}(0, \sigma_n^2)$ is additive Gaussian noise.

An advantage of FMCW radars is that it directly utilizes a single mixer to demodulate signals [15], which enables the low-cost radio-frequency implementation [29]. To be specific, at the reception-end, the mixer-based de-chirping process is firstly applied to attain the beat signal $\tilde{y}_m(t)$, i.e.

$$\tilde{y}_m(t) = s(t) \times \left[\sum_{k=0}^{K-1} r_k(t)\exp\left(j\frac{2\pi}{\lambda_s}md\sin\theta_k\right) + n_m(t)\right].$$

With $s(t)$ substituting by (2) and after a low-pass filter (with impulse response $h(t)$), the beat signal becomes the composition of a group of target-modulated sinusoidal signals:

$$\tilde{y}_m(t) = \sum_{k=0}^{K-1} \alpha_k e^{j\left(\mu\tau_k t + \omega_s \tau_k - \frac{\mu}{2}\tau_m^2\right)} \cdot e^{j(\pi k\sin\theta_k)} + \tilde{n}_m(t).$$

Here, $\tilde{n}_m(t) = h(t) \otimes [n_m(t)s(t)]$ is the transformed additive Gaussian noise after the de-chirping processing, and $\otimes$ denotes convolution process. Then, the analog to digital convertor with a sampling frequency $f_s = 1/T_s$ outputs the discrete signal:

$$\tilde{y}_m(n) = \tilde{y}_m(nT_s),$$

for $m = 0$ to $M - 1$ and $n = 0$ to $N - 1$, where $N = T_{\text{sym}}/T_s$. On this basis, the discrete beat signal $y_m(n)$ involves the angle-induced phase shift ϑ_k, delay-induced phase shift κ_k, and the delay induced phase ρ_k, which can be then recovered

from $\tilde{y}_m(t)$ via the following parameters:

$$\vartheta_k = \exp\left[j\, \tfrac{2\pi}{\lambda_s}\, d\sin(\theta_k)\right], \tag{3}$$

$$\kappa_k = \exp(j\,\mu\,\tau_k\,T_{sym}), \tag{4}$$

$$\rho_k = \exp\left[j\left(-\tfrac{1}{2}\mu\,\tau_k^2 + w_s\,\tau_k\right)\right]. \tag{5}$$

With such estimated parameters, one can directly recover AoA, velocity, and range of unknown targets.

In the following analysis, we will focus on the high-resolution estimation of the unknown AoA, θ_k, and the temporal delay, τ_k. This is of great importance and remains the major challenge in massive-MIMO automotive radars.

B. Targets Detection & Estimation

In the automotive environment sensing, we need to accurately estimate the AoA, θ_k, and the temporal delay, τ_k. The most popular algorithms such as MUSIC [19], [44] and ESPRIT [20] require forming the temporal covariance matrix and the spatial covariance matrix, i.e.,

$$\mathbf{T} = \frac{1}{M}\mathbf{Y}^H\mathbf{Y} \quad\text{and}\quad \mathbf{S} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H, \tag{6}$$

where $\mathbf{Y} \in \mathbb{C}^{M\times N}$ is a spatial signal matrix arranged as:

$$\mathbf{Y} \triangleq \begin{bmatrix} \tilde{y}_0(0) & \tilde{y}_0(1) & \cdots & \tilde{y}_0(N-1) \\ \tilde{y}_1(0) & \tilde{y}_1(1) & \cdots & \tilde{y}_1(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_{M-1}(0) & \tilde{y}_{M-1}(1) & \cdots & \tilde{y}_{M-1}(N-1) \end{bmatrix}. \tag{7}$$

The $\mathbf{S}$ matrix is useful when extracting multiple AoAs, while the $\mathbf{T}$ matrix is critical in estimating ToAs (or ranges). Note that, although the involved matrix multiplications also have high time complexity, they can be efficiently implemented using parallel computing and hardware acceleration, which may not pose a challenge to massive-MIMO radars.

After forming the $\mathbf{S}$ and $\mathbf{T}$ matrices, subspace identification are performed for estimating AoAs and ToAs, which has the formidable time-complexity in massive-MIMO radars. In the rest of this paper, we discuss only the subspace-based estimation of AoA using a covariance matrix $\mathbf{S}$, and how to accelerate the computation using randomized approximate algorithms. The estimation of the ToA can be analogously performed using $\mathbf{T}$, and our developed fast algorithms naturally apply.

1) MUSIC: Here, we use the $\mathbf{S}$ matrix to estimate AoAs of K targets; we can use the $\mathbf{T}$ matrix to estimate ToA in the same way. First, we compute the SVD of $\mathbf{S}$:

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H = \mathbf{U}_K\mathbf{\Sigma}_K\mathbf{U}_K^H + \epsilon_n^2\mathbf{U}_{-K}\mathbf{U}_{-K}^H, \tag{8}$$

where $\mathbf{U}_K = [\mathbf{u}_1,\cdots,\mathbf{u}_K]$ corresponds to signals and $\mathbf{U}_{-K} = [\mathbf{u}_{K+1},\cdots,\mathbf{u}_M]$ corresponds to noise. As known, in this procedure the SVD of $\mathbf{S}$ has $\mathcal{O}(M^3)$ time complexity. Then, we uniformly partition the range $[0,\pi]$ to a grid of L values, $\theta_0,\theta_1,\cdots,\theta_{L-1}$. For every θ, we compute the M-dimensional steering vector:

$$\mathbf{a}(\theta) = \begin{bmatrix} a_0(\theta),\, a_1(\theta),\, \cdots,\, a_{M-1}(\theta) \end{bmatrix}^H$$

where

$$a_m(\theta) = \exp\left(\tfrac{j\,2\pi\,d\,m\cdot\sin(\theta)}{\lambda}\right), \quad\text{for}\quad m = 0,1,\cdots,M-1.$$

Finally, we evaluate the pseudo-spectrum:

$$P_{\text{music}}(\theta) = \frac{1}{\mathbf{a}(\theta)^H\mathbf{U}_{-K}\mathbf{U}_{-K}^H\mathbf{a}(\theta)} = \frac{1}{\mathbf{a}(\theta)^H(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}(\theta)}. \tag{9}$$

Note that $P_{\text{music}}(\theta)$ must be evaluated for every $\theta \in \{\theta_0,\cdots,\theta_{L-1}\}$. Given $\mathbf{U}$, the evaluation of $P_{\text{music}}(\theta)$ costs $\mathcal{O}(MKL)$ time. However, the evaluation of $P_{\text{music}}(\theta)$ for every θ is embarrassingly parallel, which means the actual time cost can be much lower than $\mathcal{O}(MKL)$.[2]

Figure 2(b) plots $P(\theta)$ against θ. The peaks of the pseudo-spectrum indicate detected targets, while the θ value at a peak is the AoA of the target. Only the peaks are useful to the application. If an algorithm is designed to approximate $P_{\text{music}}(\theta)$, the algorithm must preserve these peaks.

In the above, we discuss the two-stage 1-D MUSIC which computes ToAs and AoAs successively [42], [43]. There is another type of MUSIC–the 2-D MUSIC algorithm–which estimates ToAs and AoAs jointly [45], [46]. In this work, we focus on the 1-D MUSIC algorithm and develop approximate algorithms for speeding up the subspace identification step in the 1-D MUSIC. Our algorithms can be directly extended to the 2-D MUSIC, i.e. performing fast SVD on an extended $MN \times MN$ covariance matrix.

2) ESPRIT: Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) is an alternative to MUSIC. To be specific, ESPRIT adopts two sub-arrays, e.g., the first and the last $(M-1)$ elements, and uses them to formulate the following invariant equation [20]:

$$\mathbf{J}_1\mathbf{U}_K\mathbf{\Psi} = \mathbf{J}_2\mathbf{U}_K, \tag{10}$$

where $\mathbf{J}_1 \triangleq [\mathbf{I}_{M-1}\ \mathbf{0}_{(M-1)\times 1}] \in \{0,1\}^{(M-1)\times M}$ and $\mathbf{J}_2 \triangleq [\mathbf{0}_{(M-1)\times 1}\ \mathbf{I}_{M-1}] \in \{0,1\}^{(M-1)\times M}$ are two selection matrices on the signal sub-space. The matrix $\mathbf{\Psi} \in \mathbb{C}^{K\times K}$ can be computed by solving the least-square regression:

$$\begin{aligned}\mathbf{\Psi} &= \operatorname*{argmin}_{\mathbf{\Psi}\in\mathbb{C}^{K\times K}} \left\| \mathbf{J}_1\mathbf{U}_K\mathbf{\Psi} - \mathbf{J}_2\mathbf{U}_K \right\|_2^2 \\ &= (\mathbf{J}_1\mathbf{U}_K)^\dagger\mathbf{J}_2\mathbf{U}_K.\end{aligned} \tag{11}$$

In this manner, the exhaustive pseudo-spectrum search can be avoided. Given total K singular values of $\mathbf{\Psi}$, i.e. $\{\phi_k,\ k = 1,2,\cdots,K\}$, the AoA the k-th source is estimated via:

$$\hat{\theta}_k = -\arcsin\left(\operatorname{angle}(\phi_k)/\pi\right), \tag{12}$$

where $\operatorname{angle}(\phi_k)$ denotes the angle of complex-valued ϕ_k.

The ESPRIT algorithm relies also on the signal subspace [20], [47], $\mathbf{U}_K$, as well as MUSIC. ESPRIT thus needs the SVD of $\mathbf{S}$ and cost $\mathcal{O}(M^3)$ time. ESPRIT has a major shortcoming: ESPRIT requires exactly a prior knowledge of the number of targets, K, which is oftentimes unrealistic. When K is unknown beforehand, ESPRIT tends to be unreliable [43], and the AIC criterion [48] is needed to determine K. Note that MUSIC does also need K as input when determining the

---

[2]We mainly focus on the computation of $\mathbf{U}$. The efficient calculation of $P_{\text{music}}(\theta)$ (given $\mathbf{U}$) can be realized via the similar idea as the subsequent ESPRIT method, which, however, is out of our scope.

Method	MUSIC	Fast MUSIC	
		Nyström	Power Iteration
Compute $\mathbf{U}_K$	$\mathcal{O}(M^3)$	$\mathcal{O}(Mp^2)$	$\mathcal{O}(tM^2K)$

[1] Here, p ($K \leq p \ll M$) is a user-specified over-sampling parameter, and t ($t \geq 1$) is the number of power iterations.

dimension of the signal subspace; yet both MUSIC and our fast MUSIC are relatively robust to the the setting of K, as seen later.

3) Matrix-inversion algorithm: The work [43] proposed to approximate MUSIC using matrix inversion, which has more effective algorithms. It follows from (8) that

$$\mathbf{S}^{-1} = \mathbf{U}_K \mathbf{\Sigma}_K^{-1} \mathbf{U}_K^H + \frac{1}{\epsilon_n^2} \mathbf{U}_{-K} \mathbf{U}_{-K}^H \approx \frac{1}{\epsilon_n^2} \mathbf{U}_{-K} \mathbf{U}_{-K}^H. \quad (13)$$

Here, the approximation is high quality only if the SNR is high (in which case ϵ_n is small). It follows that

$$\mathbf{a}(\theta)^H \mathbf{U}_{-K} \mathbf{U}_{-K}^H \mathbf{a}(\theta) = \mathbf{a}(\theta)^H \left(\mathbf{U}_{-K} \mathbf{U}_{-K}^H \right)^2 \mathbf{a}(\theta)$$
$$\approx \epsilon_n^4 \, \mathbf{a}(\theta)^H \mathbf{S}^{-2} \mathbf{a}(\theta) = \epsilon_n^4 \left\| \mathbf{S}^{-1} \mathbf{a}(\theta) \right\|_2^2.$$

Based on the above observations, [43] used

$$P_{\text{inv}}(\theta) = \frac{1}{\epsilon_n^4 \|\mathbf{S}^{-1}\mathbf{a}(\theta)\|_2^2}$$

to approximate $P_{\text{music}}(\theta)$. The matrix inversion typically requires $\mathcal{O}(M^3)$ time complexity–the same as MUSIC and ESPRIT. Meanwhile, the matrix inversion algorithm requires a high SNR to make the approximation in (13) tight. In the case of low SNR (e.g. less sample length N), the approximation in (13) tends to be less accurate unfortunately, making $P_{\text{inv}}(\theta)$ a poor approximation to MUSIC.

4) Propagator Algorithm: Different from the SVD-based subspace partition, the Propagator algorithm directly resorts to a special-form partition of receiving signal matrix $\mathbf{Y}$, i.e.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \in \mathbb{C}^{K \times N} \\ \mathbf{Y}_2 \in \mathbb{C}^{(M-K) \times N} \end{bmatrix}.$$

Similar to the transform matrix in ESPRIT, a propagator is defined as a special transform matrix [23], i.e. $\mathbf{Y}_2 = \mathbf{P}_{\text{pro}}^H \mathbf{Y}_1$. It is then determined by:

$$\mathbf{P}_{\text{pro}} = \underset{\mathbf{P}}{\arg\min} \|\mathbf{Y}_2 - \mathbf{P}^H \mathbf{Y}_1\|_2^2 = (\mathbf{Y}_1 \mathbf{Y}_1^H)^{-1} \mathbf{Y}_1 \mathbf{Y}_2^H. \quad (14)$$

Relying on the above linear operator $\mathbf{P}_{\text{pro}}$, the signal sub-space and noise sub-space are alternatively identified via:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P}_{\text{pro}} \\ -\mathbf{I}_{M-K} \end{bmatrix}, \quad \mathbf{Q}^\perp = \begin{bmatrix} \mathbf{I}_K \\ \mathbf{P}_{\text{pro}}^H \end{bmatrix},$$

and one can easily verify $\mathbf{Q}^H \mathbf{Y} = \mathbf{0}$ and $\mathbf{Q}^H \mathbf{Q}^\perp = \mathbf{0}$. The propagator algorithm replaces the matrix $\mathbf{U}_{-K}$ in $P_{\text{music}}(\theta)$ by $\mathbf{Q}^\perp$. In this way, the spatial pseudo-spectrum is estimated via:

$$P_{\text{propa}}(\theta) = \frac{1}{\mathbf{a}(\theta)^H \mathbf{Q}^\perp \mathbf{Q}^{\perp H} \mathbf{a}(\theta)}. \quad (15)$$

Algorithm 1 Fast MUSIC via the Nyström Method.

1: **Input**: spatial covariance matrix $\mathbf{S}$ ($M \times M$), target rank K, and over-sampling parameter p ($\geq K$).
2: **// Step 1: The Nyström Method**
3: $\mathcal{I} \longleftarrow$ randomly sampling s indices from $\{0, 1, \cdots, M\}$;
4: $\mathbf{C}(M \times p) \longleftarrow$ the column of $\mathbf{S}$ indexed by $\mathcal{I}$;
5: $\mathbf{W}(p \times p) \longleftarrow$ the rows of $\mathbf{C}$ index by $\mathcal{I}$;
6: **// Step 2: Rank Restriction**
7: Compute the SVD of $\mathbf{W}$ and compute $(\mathbf{W}^\dagger)^{\frac{1}{2}}$;
8: Compute the SVD: $\mathbf{C}(\mathbf{W}^\dagger)^{\frac{1}{2}} = \tilde{\mathbf{U}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{V}}^H$;
9: $\tilde{\mathbf{U}}_K(M \times K) \longleftarrow$ the first K columns of $\tilde{\mathbf{U}}$ ($M \times p$);
10: **// Step 3: Approximate MUSIC**
11: Compute $\tilde{P}_{\text{music}}(\theta) = \frac{1}{\mathbf{a}(\theta)^H \mathbf{a}(\theta) - \mathbf{a}(\theta)^H \tilde{\mathbf{U}}_K \tilde{\mathbf{U}}_K^H \mathbf{a}(\theta)}$ for all θ.

Of course, the above spatial spectrum can be estimated via the orthonomal basis of $\mathbf{Q}^\perp$ [24]. This propagator algorithm has a low complexity $\mathcal{O}(MNK)$. Unfortunately, it provides low-resolution and uncertain results in specific spatial ranges.

IV. LOW-RANK APPROXIMATIONS

In view of high time complexity of the SVD of $\mathbf{S}$, we propose to use randomized low-rank approximations to speed up computation and enable real-time etection. In this section, we investigate two different low-rank approximation methods. Such methods has been widely used in linear algebra and scientific commutating, yet there is few effort in applying them to solve signal processing problems. Here, we use them to develop a fast MUSIC scheme, and furthermore, we establish the improved theoretical bounds for this new problem which would have independent interest. In Table II, we compare the time complexities for various subspace identification methods.

A. Computational Challenges

As mentioned, MUSIC and ESPRIT require the top K or the bottom $M - K$ singular vectors of the $M \times M$ matrix $\mathbf{S}$. One can use QR algorithm to compute all the M singular vectors using $\mathcal{O}(M^3)$ time. In order to avoid the high time complexity, we propose to use the Nyström method or the power method for efficiently and approximately decompose $\mathbf{S}$, and thereby obtain the subspace for high-resolution real-time target detection in automotive massive-MIMO radars.

B. The Nyström Method

In our fast MUSIC, we resort to the Nyström method [31], [32] for efficiently and approximately computing the top-K singular vectors of $\mathbf{S}$. To be specific, the Nyström method we use is the rank-restricted version of [36], [37]. We describe the rank-restricted Nyström method in the following.

We choose a parameter p ($K \leq p \ll M$) and sample p items from $\{0, 1, \cdots, M - 1\}$ to form the indexing set $\mathcal{I}$. Let $\mathbf{C}$ ($M \times p$) contain the columns of $\mathbf{S}$ indexed by $\mathcal{I} \subset \{0, 1, \cdots, M-1\}$. Then let $\mathbf{W}$ ($p \times p$) contain the rows of $\mathbf{C}$ indexed by $\mathcal{I}$. The Nyström approximation to $\mathbf{S}$ is thus given by

$$\tilde{\mathbf{S}} \triangleq \mathbf{C} \mathbf{W}^\dagger \mathbf{C}^H,$$

where $\mathbf{W}^\dagger$ is the Moore-Penrose pseudo-inverse of $\mathbf{W}$. Since $\tilde{\mathbf{S}}$ is a low-rank matrix, its SVD can be efficiently performed

Algorithm 2 Fast MUSIC via the Power Method.

1: **Input**: spatial covariance matrix $\mathbf{S}$ ($M \times M$) and target rank K.

2: **// Step 1: Power Iteration**
3: $\mathbf{\Pi} \longleftarrow M \times K$ matrix with entries i.i.d. drawn from $\mathcal{N}(0,1)$;
4: Repeat $\mathbf{A} \longleftarrow \mathbf{SV}$ and $\mathbf{V} \longleftarrow \text{orth}(\mathbf{A})$ for t times;
5: **// Step 2: The Nyström Method**
6: $\mathbf{C}(M \times K) \longleftarrow \mathbf{SV}$ and $\mathbf{W}(K \times K) \longleftarrow \mathbf{V}^T \mathbf{C}$;
7: Compute the SVD of $\mathbf{W}$ and compute $(\mathbf{W}^\dagger)^{\frac{1}{2}}$;
8: Compute the SVD: $\mathbf{C}(\mathbf{W}^\dagger)^{\frac{1}{2}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^H$;
9: **// Step 3: Approximate MUSIC**
10: Compute $\tilde{P}_{\text{music}}(\theta) = \frac{1}{\mathbf{a}(\theta)^H \mathbf{a}(\theta) - \mathbf{a}(\theta)^H \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H \mathbf{a}(\theta)}$ for all θ.

---

in $\mathcal{O}(Mp^2)$ time. Let the matrices $\tilde{\mathbf{U}}$ ($M \times p$) and $\tilde{\mathbf{\Lambda}}$ ($p \times p$) be computed in Line 8 of Algorithm 1. Then

$$\tilde{\mathbf{S}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^2\tilde{\mathbf{U}}^H$$

is the SVD of $\tilde{\mathbf{S}}$.

Algorithm 1 is the rank-restricted Nyström method described in the above. When it comes to the time complexity, we easily see that: Step 1 (Lines 2 to 5) costs $\mathcal{O}(Mp)$ time. Step 2 (Lines 6 to 9) costs $\mathcal{O}(Mp^2)$ time. Thus, the overall time complexity for subspace identification is $\mathcal{O}(Mp^2)$.

C. The Nyström Method with Power Iteration

The Nyström method is highly efficient and has good quality of approximation in practice. However, its worst-case error bound is weak; see Section V-C for the theories. If a high-precision approximation is required, one can alternatively implement the Nyström method with the power method [30].

In the following, we describe the Nyström method with power iteration. The main concept of the power iteration is to compute the more accurate projection matrix via an iterative process. To do so, we first construct an $M \times K$ matrix $\mathbf{\Pi}$ whose entries are i.i.d. sampled from the Gaussian distribution $\mathcal{N}(0,1)$. Then, we compute the $M \times K$ matrix $\mathbf{V} = \text{orth}(\mathbf{S\Pi})$ where orth means orthonormalizing the columns. We can repeat the following update for t iterations:

$$\mathbf{A} \longleftarrow \mathbf{SV} \qquad \text{and} \qquad \mathbf{V} \longleftarrow \text{orth}(\mathbf{A}). \qquad (16)$$

Finally, we obtain the $M \times K$ matrix $\mathbf{C} = \mathbf{SV}$ and $K \times K$ matrix $\mathbf{W} = \mathbf{V}^H \mathbf{C}$ (i.e. $p = K$). With $\mathbf{C}$ and $\mathbf{W}$ at hand, we obtain the singular vectors of $\mathbf{CW}^\dagger \mathbf{C}^H$ in the same way as the Nyström method. We show the details in Algorithm 2.

The overall time complexity of this power method is $\mathcal{O}(tM^2K)$, which is faster than the SVD of $\mathbf{S}$ and only slightly slower than the popular Propagator algorithm (e.g. $N = M$). Our theories in Section V-C show that t trades off the computation and accuracy. If the SNR is high, then the K-th singular value of $\mathbf{S}$ is much bigger than the $(K+1)$-th, making the algorithm very rapidly converge to the optimum.

V. PERFORMANCE BOUNDS FOR FAST-MUSIC

In this section, we analyze the theoretical performance of our low-rank approximation methods and their impact on the pseudo-spectrum, $P_{\text{music}}(\theta)$, which is originally defined by:

$$P_{\text{music}}(\theta) = \frac{1}{\mathbf{a}(\theta)^H (\mathbf{I}_M - \mathbf{U}_K \mathbf{U}_K^H)\mathbf{a}(\theta)}. \qquad (17)$$

Replacing $\mathbf{U}_K$ by some matrix $\tilde{\mathbf{U}}$ as in our fast MUSIC, we can approximate $P_{\text{music}}(\theta)$ by

$$\tilde{P}_{\text{music}}(\theta) = \frac{1}{\mathbf{a}(\theta)^H (\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}(\theta)}. \qquad (18)$$

Theorems 1 and 2 establish lower bounds for $\tilde{P}_{\text{music}}(\theta)$. Theorem 3 establishes an upper bound for $\tilde{P}_{\text{music}}(\theta)$.

A. Overview of the Theories

The pseudo-spectrum $\tilde{P}_{\text{music}}(\theta)$ (for $\theta \in [0, \pi]$) is what we obtained by running Algorithm 1 or its power iteration variant. As illustrate in Figure 2-b, the peaks in the spectrum indicates detected targets. The goal of theoretical analysis is to show that $\tilde{P}_{\text{music}}(\theta)$ preserves such peaks in $P_{\text{music}}(\theta)$. As seen later, in the context of massive-MIMO radar, these peaks are dramatically larger than noise baseline, our following bounds sufficiently characterizes the estimation accuracy of AoA.

On the one hand, we hope to establish an *lower bound* in the form: there is a bounded positive number α_l such that

$$\tilde{P}_{\text{music}}(\theta) \geq \frac{1}{1+\alpha_l} P_{\text{music}}(\theta)$$

for all θ. Such a lower bound is what we are most interested in because it guarantees that the matrix approximation will not make the target peaks disappear. Theorems 1 and 2 are such lower bounds.

On the other hand, we hope to establish an *upper bound* in the form: there is a bounded positive number α_u such that

$$\tilde{P}_{\text{music}}(\theta) \leq \frac{1}{1-\alpha_u} P_{\text{music}}(\theta)$$

for all θ. Such an upper bound guarantees that the approximation will not let a non-peak become a peak. Theorem 3 is such an upper bound for the power method. As for the Nyström method, we do not have such an upper bound at this stage; it will be left to our future work.

In the automotive application, the lower bound seems to be more interesting than the upper bounds. If a peak in the pseudo-spectrum disappears, a target will not be detected, and an unmanned vehicle accident may happen; the lower bound prevents such a false negative error. In contrast, if a non-peak becomes a peak, the system will "see" a target which does not exist; such a false positive (false alarm) is annoying but less harmful. The upper bound further prevents such a false alarm.

B. Lower Bounds for Nyström

Let $\mathbf{S}_K = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{U}_K^H$ be the best rank-K approximation to $\mathbf{S}$. Let $\tilde{\mathbf{S}} = \mathbf{CW}^\dagger \mathbf{C}^H$ be the Nyström approximation to $\mathbf{S}$ and $\tilde{\mathbf{U}}$ be the orthonormal bases of $\tilde{\mathbf{S}}$.

Theorem 1 establishes a lower bound on $P_{\text{music}}(\theta)$. It indicates that if $P_{\text{music}}(\theta)$ have multiple peaks, then $\tilde{P}_{\text{music}}(\theta)$ should also contain such peaks (which is what we care about). The relative approximation error is basically

$$\mathcal{O}\left(\sqrt{\frac{M}{p}} \cdot \frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right).$$

If the SNR is high, then $\sigma_K(\mathbf{S}) \gg \sigma_{K+1}(\mathbf{S}) = \epsilon_n^2$, and thereby the approximation error is small.

Theorem 1. *Let the notation be defined in the above. Let* $\delta \in (0,1)$ *be any user-specified constant and* $\mu(\mathbf{U}_K)$ *the coherence of* $\mathbf{U}_K$ *(defined in* (1)*). For*

$$p \geq 4.5\,\mu(\mathbf{U}_K)\,K \cdot \log \frac{K}{\delta},$$

the following inequality holds with probability at least $1 - \delta$:

$$\sqrt{\frac{P_{music}(\theta)}{\tilde{P}_{music}(\theta)}} \leq 1 + 2\sqrt{\frac{M}{p}}\,\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\,\frac{\left\|\mathbf{U}_K^H \mathbf{a}(\theta)\right\|_2}{\left\|(\mathbf{I}-\mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}(\theta)\right\|_2}.$$

C. Lower Bound for Power Method

Let $\mathbf{\Pi}$ be an $M \times K$ matrix with entries i.i.d. drawn from $\mathcal{N}(0,1)$. Let the $M \times K$ matrix $\mathbf{V}$ be the output of the power iteration in (16); that is, $\mathbf{V}$ is the orthonormal basis of $\mathbf{S}^t\mathbf{\Pi}$. By definition, $\mathbf{C} = \mathbf{S}\mathbf{V}$ and $\mathbf{W} = \mathbf{V}^T\mathbf{S}\mathbf{V}$, and $\tilde{\mathbf{S}} = \mathbf{C}\mathbf{W}^\dagger\mathbf{C}^H$ is an approximation to $\mathbf{S}$. Let $\tilde{\mathbf{U}}$ be the orthonormal bases of $\tilde{\mathbf{S}}$.

Theorem 2 establishes lower bounds for $\tilde{P}_{music}(\theta)$ which is obtained by running the power iteration (16) for t times. The relative approximation error is basically

$$\mathcal{O}\left(\sqrt{MK} \cdot \left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\right).$$

The approximation errors converge to zero exponentially. If the SNR is big, the power iteration converges to high precision in several iterations.

Theorem 2. *Let the notation be defined in the above. Repeat the power iteration* (16) *for* t ($t \geq 0$) *times. Let* $\delta \in (0,1)$ *be any constant. The following inequality holds with probability at least* $1 - \delta$:

$$\sqrt{\frac{P_{music}(\theta)}{\tilde{P}_{music}(\theta)}} \leq 1 + \frac{\sqrt{2MK}}{\delta}\left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\frac{\left\|\mathbf{U}_K^H \mathbf{a}(\theta)\right\|_2}{\left\|(\mathbf{I}-\mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}(\theta)\right\|_2}.$$

D. Upper Bound for Power Method

Theorem 3 establishes an upper bound for $\tilde{P}_{music}(\theta)$ which is computed by running the power iteration (16) for t times. The upper bound is useful because it ensures that if the optimal pseudo-spectrum $P_{music}(\theta)$ is not a peak, then the approximated pseudo-spectrum $\tilde{P}_{music}(\theta)$ is not a peak either. The relative approximation error is basically

$$\mathcal{O}\left(\sqrt{MK} \cdot \left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\right).$$

It converges to zero exponentially. In particular, if the SNR is high, then a few iterations will suffice for a high precision.

Theorem 3. *Let the notation be defined in Section V-C. Repeat the power iteration* (16) *for* t ($t \geq 0$) *times. Then, with probability at least* $1 - \delta$,

$$\sqrt{\frac{P_{music}(\theta)}{\tilde{P}_{music}(\theta)}} \geq 1 - \frac{\sqrt{2MK}}{\delta}\left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\frac{\|\mathbf{a}(\theta)\|_2}{\|(\mathbf{I}-\mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}(\theta)\|_2}.$$

VI. NUMERICAL SIMULATION & ANALYSIS

In the section, we conduct empirical evaluations of our proposed algorithms and compare them with the existing popular algorithms for massive-MIMO automotive radars.
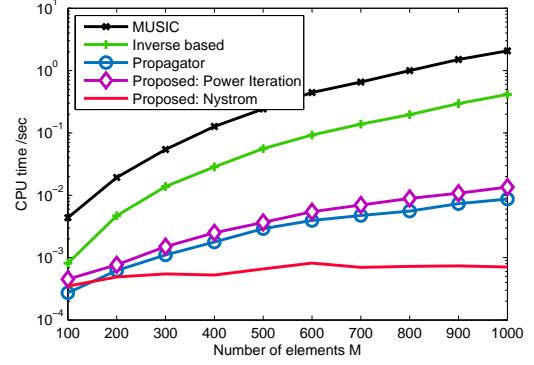

Fig. 3. Wall-clock runtime (second) against the number of antennas, M. We calculate only the runtime for computing the signal subspace which is the major bottleneck in automotive radar. Here we set $N = M$ and SNR = 0 dB.

A. Experiment Settings

Since we are interested in the subspace computation and the AoA estimation process, we directly simulate the post-processed signal matrix $\mathbf{Y}$ from M receiving elements. When it comes to time complexity, we implement our algorithm in MATLAB. Since there is no loop or recursion in our developed scheme, its efficiency can be fairly measured by the CPU runtime (basic frequency 3.5GHz, 4GB memory), which is proportional to the required number of multiplication flops. The averaged run time is then reported based on 100 independent repeats.

B. Complexity vs Accuracy

We compare the runtime of several target detection algorithms in computing the signal subspace. Figure 3 plots the runtime (seconds) against the number of antennas, M. We set $N = M$ and $K = 10$ in this analysis. For the Nyström method, we fix $p = 10$. For the power method, we fix $t = 2$.

Despite its high accuracy in high-resolution environment sensing, the exact MUSIC algorithm has at least $\mathcal{O}(M^3)$ time complexity, due to the SVD of the $M \times M$ matrix $\mathbf{S}$. As M increase, it will easily become impractical for real-time sensing application. For $M = 1000$, the runtime of the SVD alone exceeds 1 second, which is too much for real-time automotive sensing like unmanned aircrafts/vehicles.

Note that, we do not compare ESPRIT because it has exactly the same runtime as MUSIC in the subspace identification step. The Inversion-based algorithm and Propagator algorithm are approximations to MUSIC, as discussed in Section III-B. The Inversion-based algorithm is faster than MUSIC, but it may be still unsuitable for real-time target detection. The Propagator algorithm is highly efficient; however, later on, we will show it has the less attractive accuracy.

Our fast MUSIC based on Nyström method is the most efficient because its time complexity is mere $\mathcal{O}(Mp^2)$; for $M = 1000$, the Nyström MUSIC is 3000 times faster than the exact MUSIC! The power method has a complexity of $\mathcal{O}(tM^2K)$ which is higher than the Nyström method. Also, this power method is sufficiently efficient for real-world application: for $M = 1000$, its runtime is only 10^{-2} second.
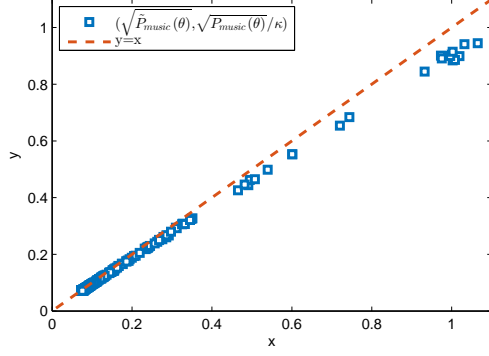
(a) Study of the Nyström method.
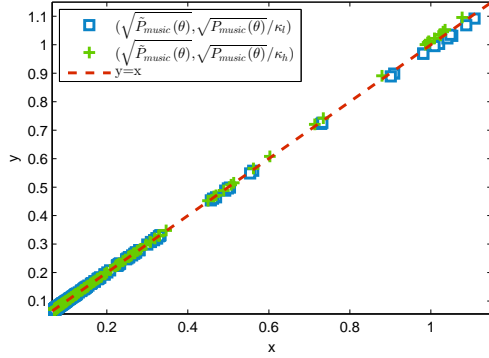
(b) Study of the power method.

Fig. 4. Comparing the theoretical bounds (Theorems 1, 2, and 3) with actual empirical results. The constants κ, κ_l, and κ_u are defined in Section VI-C.

C. Error Bound vs Actual Error

We investigate the quality of the error bounds (Theorems 1, 2, and 3), by comparing the pseudo-spectrum $\tilde{P}_{\text{music}}(\theta)$ approximated by our fast algorithm with the original one $P_{\text{music}}(\theta)$. We are interested in how much the bounds under-estimate or over-estimate the actual $P_{\text{music}}(\theta)$.

a) The Nyström Method: We first study Theorem 1, a lower bound for the Nyström method. We denote the righthand side of the bound in Theorem 1 by $\kappa = 1 + \alpha_l$. Theorem 1 shows that

$$\sqrt{\tilde{P}_{\text{music}}(\theta)} > \sqrt{P_{\text{music}}(\theta)}/\kappa \qquad (19)$$

We repeat our experiment to calculate a set of $\tilde{P}_{\text{music}}(\theta)$ and $P_{\text{music}}(\theta)$ (using different θs). We set $K = 11$, $M = N = 200$, and $p = K \log K$ in this analysis.

In Figure 4(a), we plot $\sqrt{P_{\text{music}}(\theta)}/\kappa$ (the y-axis) against $\sqrt{\tilde{P}_{\text{music}}(\theta)}$ (the x-axis) as blue squares. Ideally, if the bound is tight, the blue squares should fall on the line $y = x$. Since the lower bound (righthand side of (19)) under-estimate $\sqrt{\tilde{P}_{\text{music}}(\theta)}$, the blue squares can fall below the line with high probability. The empirical results match our theory: no blue square is above the line $y = x$.

If θ is in the non-target region, $P_{\text{music}}(\theta)$ is small. Figure 4(a) shows that our bound is very tight in the non-target region: the blue squares are almost on the line $y = x$.
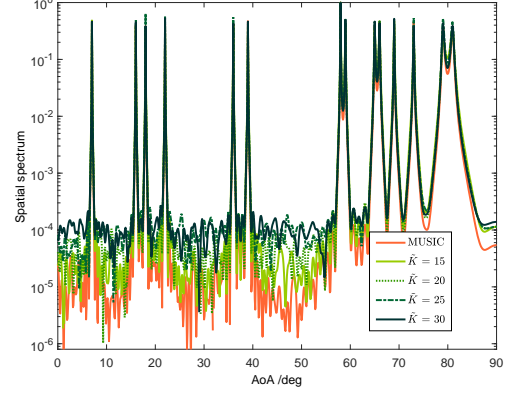
If θ is in the target region, $P_{\text{music}}(\theta)$ is large. Figure 4(a) shows that in the target region, the empirical results (blue squares) deviate slightly from the lower bound, indicating our lower bound is less tight in this region. Nevertheless, this is totally fine: big $\tilde{P}_{\text{music}}(\theta)$ in the target region is sufficiently good for target detection.

b) The Power Method: We then study the power method's bounds in Theorems 2 and 3. We denote the righthand side of the lower bound in Theorem 2 by $\kappa_l = 1 + \alpha_l$; denote the righthand side of the upper bound in Theorem 3 by $\kappa_u = 1 + \alpha_u$. Accordingly, we expect

$$\sqrt{\tilde{P}_{\text{music}}(\theta)} \geq \sqrt{P_{\text{music}}(\theta)}/\kappa_l, \quad \sqrt{\tilde{P}_{\text{music}}(\theta)} \leq \sqrt{P_{\text{music}}(\theta)}/\kappa_u.$$

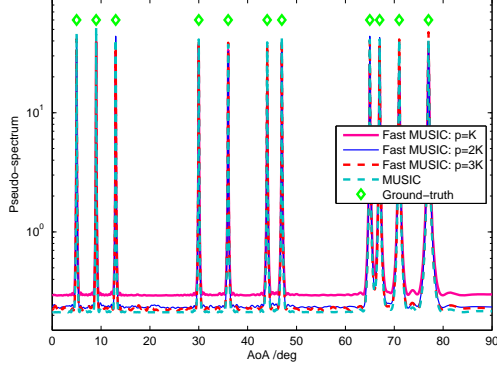Here, we fix $\delta = 0.05$ and $t = 1$.

In Figure 4(b), we plot $\sqrt{P_{\text{music}}(\theta)}/\kappa_l$ against $\sqrt{\tilde{P}_{\text{music}}(\theta)}$ as blue squares and $\sqrt{P_{\text{music}}(\theta)}/\kappa_u$ against $\sqrt{\tilde{P}_{\text{music}}(\theta)}$ as green crosses. The blue squares and green crosses lie on the line $y = x$ almost perfectly, indicating that our lower and upper bounds are very tight. To this end, theoretically, there is no false/missed peaks in our approximated pseudo-spectrum.

D. Robustness to Inaccurate K

As well as the exact MUSIC, our fast MUSIC algorithm (MUSIC+Nyström) needs prior knowledge of the number of targets, K, as the dimension of the signal subspace. Unfortunately, K is unknown beforehand; thus we have to use an estimate of K as the input of Algorithm 1.

In Figure 5, we study the effect of using inexact guess of K. We set $M = 200$, $N = 400$, $K = 14$, SNR = 0 dB, and $p = \text{round}(1.2 \times K)$. Every curve in the figure corresponds to a pseudo-spectrum (i.e., the plot of $P_{\text{music}}(\theta)$ against the AoA, θ). The red curve is obtained using the exact MUSIC with the exact knowledge of K (i.e., $K = 14$). The other curves are the plot of $\tilde{P}_{\text{music}}(\theta)$ using different guesses of K, i.e. $\tilde{K}$.

Figure 5 shows that our fast MUSIC algorithm is robust to an inexact guess of K. To be specific, if the guess (denote $\tilde{K}$) is reasonable larger than K, then our fast MUSIC algorithm

Fig. 5. Robustness of our fast MUSIC algorithm to inexact guess of K. Here, the actual number of targets is $K = 14$. The red curve plots MUSIC with the exact K. The other four curves plot our fast MUSIC algorithm with different guesses of K.

(a) Pseudo-spectrum of the Nyström method with different p.
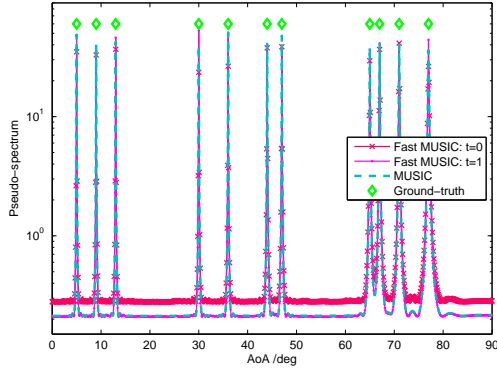
(b) Pseudo-spectrum of the power method with different t.

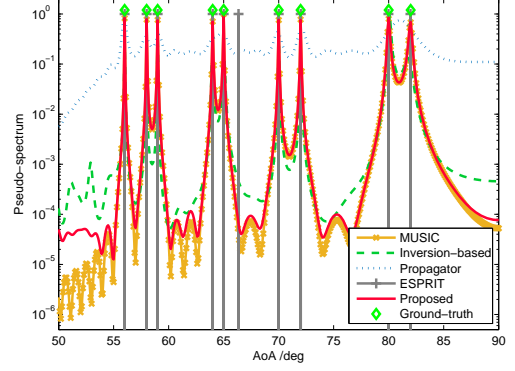Fig. 6. We plot the pseudo-spectra of MUSIC and the proposed algorithms with different tuning parameters.

(a) $M = 200$, $N = 800$, and SNR $= 0$ dB.

(b) $M = N = 200$ and SNR $= 0$ dB.

Fig. 7. Estimated pseudo-spectrum of various algorithms. Here, the proposed algorithm refers to MUSIC+Nyström with $p = 12$.

can produce almost the same pseudo-spectrum (especially at the peaks.)
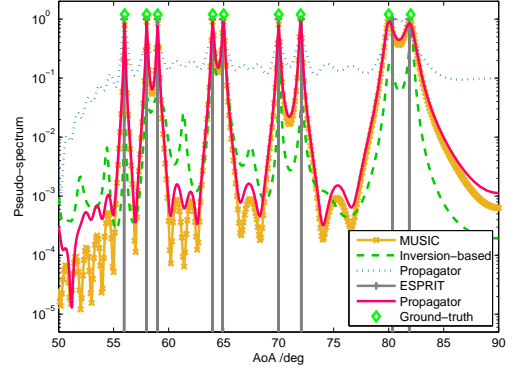
E. Tuning Parameters

The Nyström method (Algorithm 1) has a tuning paramter p ($K \leq p \ll M$) for trading off the accuracy and runtime; likewise, the power method (Algorithm 2) has a tuning parameter t (≥ 1). We now discuss the practical settings of p or t. In this analysis, we fix $M = N = 200$, $K = 11$, and SNR $= 0$ dB.

In Figure 6(a), we plot the pseudo-spectrum of MU-SIC+Nyström under different settings of p. We note that under all the three settings ($p = K$, $2K$, and $3K$), our algorithm successfully detected the targets. Define the distance between the two pseudo-spectrum as $\int_0^\pi \left[\tilde{P}_{\mathrm{music}}(\theta) - P_{\mathrm{music}}(\theta) \right]^2 d\theta$. When p is increased from K to $2K$, the distance drops from 0.59 to 0.11; when p is further increased to $3K$, the decrease in the distance becomes marginal.

In Figure 6(b), we plot the pseudo-spectrum of MU-SIC+Power method under different settings of t. We find that with $t = 1$, the MSE is mere 0.03 which is good enough. Thus, we can use $t = 1$ in practical automotive sensing.

F. Accuracy in AoA Estimation

We then evaluate the target detection accuracy and compare our algorithm with the existing ones. In Figure 7, the normalized pseudo-spectrum is defined as:

$$\check{P}(\theta) = \frac{P(\theta) - \min_\theta \{P(\theta)\}}{\max_\theta \{P(\theta)\} - \min_\theta \{P(\theta)\}}$$

where $P(\theta)$ can be the spectrum of MUSIC, ESPRIT, etc. We fix $M = 200$ and $K = 9$ and vary N. We use our MUSIC+Nyström algorithm with $p = 12$.

Figure 7 shows that among all the compared algorithms, the exact MUSIC's and our proposed algorithm's pseudo-spectrum are the closest to the ground truths. They are in the sense that its output results will not change when the temporal sample length N drops from $N = 800$ to $N = 200$. The pseudo-spectrum of our algorithm closely tracks the exact MUSIC.

ESPRIT is less accurate. The unexpected error (i.e. false targets) seems to be inevitable, especially when the exact number of sources remains unknown as in practical scenarios, see Figure 7(a) for one false source nearby 65 degrees.

Despite its low time complexity, the Propagator algorithm produces less attractive pseudo-spectrum. Figure 7 clearly shows that, in its pseudo-spectrum, a target's value is not much bigger than a non-target's value. When the SNR is not big enough, one can hardly distinguish a target from the non-
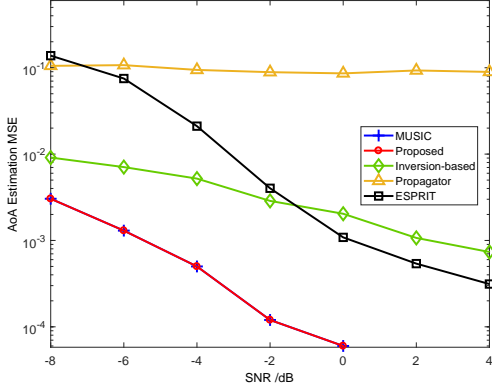
Fig. 8. Plot of AoA estimation MSE against SNRs.

targets based on its pseudo-spectrum. Its another disadvantage is the low resolution when unknown AoAs surpass 70 degrees. In Figure 7, it recognizes the two targets between 80 and 85 degrees as a single target. Note that modified Propagator algorithm has been proposed to improve its accuracy [25], which, however, relies on the special array structures (e.g. L-shape array) and can be less attractive for compact automotive radar devices.

The matrix-inversion based algorithm's target detection performance looks vary fine, although slightly worse than MUSIC and our proposed algorithm. It requires a large N (or high SNR) in order to make the estimation accurate. In Figure 7, when N drops from 800 to 200, the inversion algorithm's pseudo-spectrum around 60 degrees and 80 degrees looks problematic. For this matrix-inversion method, although its complexity in acquiring signal sub-space can be reduced to some extents, it still has one drawback. That is, its performance can be only guaranteed in the condition of high SNRs, so that the inverse approximated noise subspace is sufficiently accurate (as previously discussed in Section III-B). I.e., in order to accurately estimate AoAs, the required sample length N is relatively large. In the case of low SNR or less sample length $N \leqslant M$, its performance would be unstable, e.g. producing false/missed objects or AoA estimation error.

G. Performance under Varying SNRs

We finally evaluate the performance of various candidates for massive MIMO radar under varying SNRs. Our evaluation metric is the mean square error (MSE) of estimated AoAs:

$$\mathrm{MSE} \triangleq \mathbb{E}\left\{\sum_{k=1}^{K} \|\theta^{(k)} - \hat{\theta}^{(k)}\|_2^2\right\}. \tag{20}$$

Here, $\theta^{(k)}$ is the ground truth AoA of the kth target; $\hat{\theta}^{(k)}$ is the derived estimation. In Figure 8, we plot the MSE against SNR. All the compared algorithms improve as the SNR increases. Our proposed MUSIC+Nyström algorithm has the same MSEs as the exact MUSIC. That is, the fast MUSIC is much better than all the other compared algorithms.

VII. CONCLUSION

We studied the super-resolution target detection problem emerging from massive-MIMO radars. The classic algorithms such as MUSIC and ESPRIT are computationally expensive due to the subspace identification, whereas the Inverse and Propagator algorithms can have unsatisfactory accuracy; before this work, high resolution and real-time computation were mutually exclusive. To enable real-time yet super-resolution target detection and parameter estimation, we leveraged randomized low-rank matrix approximation and proposed two fast MUSIC algorithms. We theoretically proved that in the case of high SNR, our fast approximate MUSIC algorithms are almost as good as the optimal. Our empirical study demonstrated that while our proposed algorithms are faster than exact MUSIC by order-of-magnitude, yet they are nearly as accurate as MUSIC in the target detection tasks. As such, our proposed algorithms enjoy both low time complexity and high accuracy, as verified by our main theorems and experiments. Our fast MUSIC hence provides the great potential to high-resolution and real-time massive-MIMO radars in emerging automotive applications.

APPENDIX

Here, we show the detailed proofs of Theorems 1, 2, and 3. For simplicity, we leave out the θ in $P_{\mathrm{music}}(\theta)$, $\tilde{P}_{\mathrm{music}}(\theta)$, and $\mathbf{a}(\theta)$.

A. Analysis of Uniform Sampling

An $M \times p$ ($p < M$) matrix $\mathbf{\Pi}$ is called a uniform sampling matrix if its columns are sampled from the columns of $\frac{\sqrt{M}}{\sqrt{p}}\mathbf{I}_M$ uniformly at random. For any $N \times M$ matrix $\mathbf{X}$, the multiplication $\mathbf{X}\mathbf{\Pi}$ ($N \times p$) contains p uniformly sampled columns of $\mathbf{X}$.

Lemma 1. *Let $\mathbf{\Pi}$ be an $M \times p$ uniform sampling matrix. Then*

$$\|\mathbf{\Pi}\|_2^2 = M/p.$$

Lemma 2 ([33], [40]). *Let $\mathbf{U}_K$ be an $M \times K$ matrix with orthonormal columns. Let the row coherence of $\mathbf{U}_K$ be defined in (1). Let $\mathbf{\Pi}$ be an $M \times p$ uniform sampling matrix. For*

$$p \geq \frac{(6+2\eta)\mu(\mathbf{U}_K)K}{3\eta^2} \log \frac{K}{\delta}$$

the K singular values of $\mathbf{U}_K^H \mathbf{\Pi}\mathbf{\Pi}^H \mathbf{U}_K$ are at least $1-\eta$ with probability at least $1 - \delta$.

B. Analysis of Gaussian Projection

Lemma 3 ([49]). *Let $\mathbf{\Pi}$ be an $M \times K$ matrix whose entries are i.i.d. drawn from $\mathcal{N}(0,1)$. If M is substantially larger than K, the spectral norm of $\mathbf{\Pi}$ is bounded by*

$$\|\mathbf{\Pi}\|_2 \leq \sqrt{M} + \sqrt{K} + \mathcal{O}(1), \qquad almost\ surely.$$

Lemma 4 ([50], [51]). *Let $\mathbf{G}$ be a $K \times K$ matrix whose entries are i.i.d. drawn from $\mathcal{N}(0,1)$. Then the smallest singualr value of $\mathbf{G}$ satisfies*

$$\sigma_K(\mathbf{G}) \geq \frac{\delta}{\sqrt{K}}$$

with probability at least $1 - \delta - o(1)$.

C. Proof of Theorem 1

Lemma 5. *Let the notation be defined in Section V-B. Let $\mathbf{\Pi}$ be any matrix with $\mathrm{rank}(\mathbf{U}_K^H\mathbf{\Pi}) \geq K$. Then, for all $\mathbf{a}$,*

$$\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq \left\|(\mathbf{S} - \mathbf{S}_K)\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}\right\|_2,$$

Proof. Let $\tilde{\mathbf{U}}$ be the orthonormal bases of $\tilde{\mathbf{S}}$. Lemma 12 of [40] shows that

$$\left(\mathbf{C}\mathbf{W}^\dagger\mathbf{C}^H\right)\left(\mathbf{C}\mathbf{W}^\dagger\mathbf{C}^H\right)^\dagger\mathbf{C} = \mathbf{C}.$$

Thus $\mathbf{C} = \mathbf{S}\mathbf{\Pi}$ is in the subspace spanned by the columns of $\tilde{\mathbf{U}}$, and therefore,

$$\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{S}\mathbf{\Pi} = \mathbf{S}\mathbf{\Pi}. \tag{21}$$

It is not hard to prove that

$$\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} = \operatorname*{argmin}_{\mathbf{x}} \left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\mathbf{x}\right\|_2.$$

Thus, the inequality holds for all $\tilde{\mathbf{x}}$:

$$\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$= \min_{\mathbf{x}} \left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\mathbf{x}\right\|_2 \leq \left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{x}}\right\|_2. \tag{22}$$

We artificially construct

$$\tilde{\mathbf{x}} = \tilde{\mathbf{U}}^H\mathbf{S}\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}.$$

It follows from (21) that

$$\tilde{\mathbf{U}}\tilde{\mathbf{x}} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{S}\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}$$
$$= \mathbf{S}\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}$$
$$= \left(\mathbf{S}_K + \mathbf{S} - \mathbf{S}_K\right)\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}. \tag{23}$$

Since $\mathrm{rank}(\mathbf{U}_K^H\mathbf{\Pi}) \geq K$, we have $(\mathbf{U}_K^H\mathbf{\Pi})(\mathbf{U}_K^H\mathbf{\Pi})^\dagger = \mathbf{I}_K$. Thus

$$\mathbf{S}_K\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K\mathbf{a}$$
$$= \mathbf{U}_K\mathbf{\Sigma}_K\mathbf{U}_K^H\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}$$
$$= \mathbf{U}_K\mathbf{\Sigma}_K\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a} = \mathbf{U}_K\mathbf{U}_K^H\mathbf{a}. \tag{24}$$

It follows from (23) and (24) that

$$\tilde{\mathbf{U}}\tilde{\mathbf{x}} = \left[\mathbf{S}_K + (\mathbf{S} - \mathbf{S}_K)\right]\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}$$
$$= \mathbf{U}_K\mathbf{U}_K^H\mathbf{a} + (\mathbf{S} - \mathbf{S}_K)\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}. \tag{25}$$

It follows from (22) and (25) that

$$\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2 \leq \left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{x}}\right\|_2$$
$$= \left\|(\mathbf{S} - \mathbf{S}_K)\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}\right\|_2,$$

by which the lemma follows. $\square$

Complete the proof of Theorem 1:

Proof. The $M \times p$ matrix $\mathbf{\Pi}$ is a uniform sampling matrix. Then $\mathbf{C} = \mathbf{S}\mathbf{\Pi}$ and $\mathbf{W} = \mathbf{\Pi}^T\mathbf{S}\mathbf{\Pi}$. Lemma 1 shows the spectral norm of $\mathbf{\Pi}$ is bounded by

$$\|\mathbf{\Pi}\|_2 \leq \sqrt{\tfrac{M}{p}}.$$

Because $\mathbf{U}_K$ has orthonormal columns, Lemma 2 shows that for

$$p \geq \frac{(6 + 2\eta)\mu(\mathbf{U}_K)K}{3\eta^2}\log\frac{K}{\delta},$$

the K-th singular value of $\mathbf{U}_K^H\mathbf{\Pi}$ is bounded by

$$\sigma_K(\mathbf{U}_K^H\mathbf{\Pi}) \geq \sqrt{1-\eta}$$

with probability at least $1-\delta$. Let follows from Lemma 5 that

$$\left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq \left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq \left\|(\mathbf{S} - \mathbf{S}_K)\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-1}\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq \left\|(\mathbf{S} - \mathbf{S}_K)\right\|_2\left\|\mathbf{\Pi}\right\|_2\left\|(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\right\|_2\left\|\mathbf{\Sigma}_K^{-1}\right\|_2\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq \frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\frac{\sqrt{M/p}}{\sqrt{1-\eta}}\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2.$$

We set $\eta = 0.75$. Then, for

$$p \geq 4.5\mu(\mathbf{U}_K)K\cdot\log\frac{K}{\delta},$$

it holds with probability at least $1-\delta$ that

$$\left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2 \leq 2\sqrt{\tfrac{M}{p}}\,\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2.$$

It follows that

$$\left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2$$
$$= \left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)(\mathbf{U}_K\mathbf{U}_K^H + \mathbf{U}_{-K}\mathbf{U}_{-K}^H)\mathbf{a}\right\|_2$$
$$\leq \left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2 + \left\|\mathbf{U}_{-K}\mathbf{U}_{-K}^H\mathbf{a}\right\|_2.$$

Hence,

$$\left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2 - \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2$$
$$\leq \left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq 2\sqrt{\tfrac{M}{p}}\,\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2,$$

where the last inequality holds with probability at least $1-\delta$.

We leave out θ in $P_{\mathrm{music}}(\theta)$, $\tilde{P}_{\mathrm{music}}(\theta)$, and $\mathbf{a}(\theta)$. It follows from (17) and (18) that

$$\sqrt{\frac{P_{\mathrm{music}}}{\tilde{P}_{\mathrm{music}}}} = \frac{\left\|(\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2}{\left\|(\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2}$$
$$\leq 1 + 2\sqrt{\tfrac{M}{p}}\,\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\frac{\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2}{\left\|(\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2},$$

by which the theorem follows. $\square$

D. Proof of Theorem 2

Lemma 6. *Let the notation be defined in Section V-C. Let $\mathbf{\Pi}$ be any matrix satisfying that $\mathrm{rank}(\mathbf{U}_K^H\mathbf{\Pi}) \geq K$. Then, for all $\mathbf{a}$,*

$$\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2$$
$$\leq \left\|(\mathbf{S} - \mathbf{S}_K)^{t+1}\mathbf{\Pi}(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a}\right\|_2.$$

Proof. Since $\tilde{\mathbf{S}} = \mathbf{C}\mathbf{W}^\dagger\mathbf{C}^H$, it follows from [40, Lemma 12] that

$$\tilde{\mathbf{S}}\tilde{\mathbf{S}}^\dagger\mathbf{C} = \mathbf{C}.$$

Since $\tilde{\mathbf{U}}$ is an orthonormal bases of $\tilde{\mathbf{S}}$,

$$\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{C} = \tilde{\mathbf{S}}\tilde{\mathbf{S}}^\dagger\mathbf{C} = \mathbf{C}.$$

By the definition $\mathbf{C} = \mathbf{S}\mathbf{V}$ and that $\mathbf{V}$ is the orthonormal basis of $\mathbf{S}^t\mathbf{\Pi}$, we have that $\mathbf{C}$ and $\mathbf{S}^{t+1}\mathbf{\Pi}$ have the same column space.

$$\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{S}^{t+1}\mathbf{\Pi} = \mathbf{S}^{t+1}\mathbf{\Pi}. \tag{26}$$

In the same way as the proof of Lemma 5, we can prove that for all $\tilde{\mathbf{x}}$:

$$\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2 \leq \left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{x}}\right\|_2. \tag{27}$$

We artificially construct

$$\tilde{\mathbf{x}} = \tilde{\mathbf{U}}^H\mathbf{S}^{t+1}\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a}.$$

It follows from (26) that

$$\begin{aligned}\tilde{\mathbf{U}}\tilde{\mathbf{x}} &= \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{S}^{t+1}\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a} \\ &= \left(\mathbf{S}_K^{t+1} + \mathbf{S}^{t+1} - \mathbf{S}_K^{t+1}\right)\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a}.\end{aligned} \tag{28}$$

In the same way as the proof of Lemma 5, we can use $\mathrm{rank}(\mathbf{U}_K^H\mathbf{\Pi}) \geq K$ to show that

$$\begin{aligned}&\mathbf{S}_K^{t+1}\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a} \\ &= \mathbf{U}_K\mathbf{\Sigma}_K^{t+1}\mathbf{U}_K^H\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a} \\ &= \mathbf{U}_K\mathbf{\Sigma}_K^{t+1}\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a} = \mathbf{U}_K\mathbf{U}_K^H\mathbf{a}.\end{aligned}$$

It follows from (28) that

$$\tilde{\mathbf{U}}\tilde{\mathbf{x}} = \mathbf{U}_K\mathbf{U}_K^H\mathbf{a} + \left(\mathbf{S}^{t+1} - \mathbf{S}_K^{t+1}\right)\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a}.$$

It follows from (27) that

$$\begin{aligned}&\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2 \\ &\leq \left\|\left(\mathbf{S}^{t+1} - \mathbf{S}_K^{t+1}\right)\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a}\right\|_2,\end{aligned}$$

by which the lemma follows. $\qquad\square$

Complete the proof of Theorem 2:

Proof. Lemma 3 shows that the spectral norm of the $M \times K$ standard Gaussian matrix $\mathbf{\Pi}$ satisfies

$$\left\|\mathbf{\Pi}\right\|_2 \leq \sqrt{M} + \sqrt{K} + \mathcal{O}(1), \qquad \text{almost surely.}$$

Since $\mathbf{U}_K$ has orthonormal columns, the $K \times K$ matrix $\mathbf{U}_K^H\mathbf{\Pi}$ is a standard Gaussian matrix. It follows from Lemma 4 that

$$\sigma_K^{-1}(\mathbf{U}_K^H\mathbf{\Pi}) \leq \frac{\sqrt{K}}{\delta}$$

with probability at least $1 - \delta - o(1)$. It follows from Lemma 6 that

$$\begin{aligned}&\left\|\mathbf{U}_K\mathbf{U}_K^H\mathbf{a} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2 \\ &\leq \left\|\left(\mathbf{S}^{t+1} - \mathbf{S}_K^{t+1}\right)\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\mathbf{a}\right\|_2 \\ &\leq \left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\left\|\mathbf{\Pi}\right\|_2\left\|(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\right\|_2\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2 \\ &\leq \left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\frac{\sqrt{MK} + K + \mathcal{O}(\sqrt{K})}{\delta}\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2.\end{aligned}$$

The proof follows from the proof of Theorem 1 that

$$\begin{aligned}\sqrt{\frac{P_{\mathrm{music}}}{\tilde{P}_{\mathrm{music}}}} &= \frac{\left\|(\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2}{\left\|(\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2} \\ &\leq 1 + \frac{\left\|\left(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\right)\mathbf{U}_K\mathbf{U}_K^H\mathbf{a}\right\|_2}{\left\|(\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2}.\end{aligned}$$

Thus, it holds with probability at least $1 - \delta - o(1)$ that

$$\sqrt{\frac{P_{\mathrm{music}}}{\tilde{P}_{\mathrm{music}}}} \leq 1 + \left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\frac{\sqrt{MK}\left(1 + o(1)\right)}{\delta}\frac{\left\|\mathbf{U}_K^H\mathbf{a}\right\|_2}{\left\|(\mathbf{I} - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2},$$

by which the theorem follows. $\qquad\square$

Proof of Theorem 3:

Proof. It can be shown that for all $\tilde{\mathbf{X}}$:

$$\left\|\mathbf{U}_K\mathbf{U}_K^H - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\right\|_2 \leq \left\|\mathbf{U}_K\mathbf{U}_K^H - \tilde{\mathbf{U}}\tilde{\mathbf{X}}\right\|_2.$$

We artificially construct

$$\tilde{\mathbf{X}} = \tilde{\mathbf{U}}^H\mathbf{S}^{t+1}\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H.$$

In the same way as the proof of Lemma 6, we can show that

$$\tilde{\mathbf{U}}\tilde{\mathbf{X}} = \mathbf{U}_K\mathbf{U}_K^H + \left(\mathbf{S}^{t+1} - \mathbf{S}_K^{t+1}\right)\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H.$$

Thus

$$\begin{aligned}&\left\|\mathbf{U}_K\mathbf{U}_K^H - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\right\|_2 \\ &\leq \left\|\left(\mathbf{S}^{t+1} - \mathbf{S}_K^{t+1}\right)\mathbf{\Pi}\left(\mathbf{U}_K^H\mathbf{\Pi}\right)^\dagger\mathbf{\Sigma}_K^{-t-1}\mathbf{U}_K^H\right\|_2 \\ &\leq \left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\left\|\mathbf{\Pi}\right\|_2\left\|(\mathbf{U}_K^H\mathbf{\Pi})^\dagger\right\|_2.\end{aligned}$$

It follows from the proof of Theorem 2 that

$$\begin{aligned}&\left\|\mathbf{U}_K\mathbf{U}_K^H - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{U}_K\mathbf{U}_K^H\right\|_2 \\ &\leq \left(1 + o(1)\right)\frac{\sqrt{MK}}{\delta}\left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\end{aligned}$$

holds with probability at least $1 - \delta$. Since $\mathbf{U}_K$ and $\tilde{\mathbf{U}}$ are both $M \times K$, [52, Eqn 2.54] shows that

$$\left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{U}_K\right\|_2 = \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\tilde{\mathbf{U}}\right\|_2.$$

It follows that with probability at least $1 - \delta$,

$$\begin{aligned}\left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\tilde{\mathbf{U}}\right\|_2 &= \left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{U}_K\right\|_2 \\ &\leq \left(1 + o(1)\right)\frac{\sqrt{MK}}{\delta}\left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}.\end{aligned} \tag{29}$$

We have that

$$\begin{aligned}&\left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2 \\ &\leq \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{a}\right\|_2 \\ &\quad + \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2 \\ &\leq \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H\mathbf{a}\right\|_2 + \left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2.\end{aligned} \tag{30}$$

It follows from (29) and (30) that

$$\begin{aligned}&\left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2 - \left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2 \\ &\leq \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\tilde{\mathbf{U}}\right\|_2\left\|\mathbf{a}\right\|_2 \\ &\leq \left(1 + o(1)\right)\frac{\sqrt{MK}}{\delta}\left(\tfrac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\left\|\mathbf{a}\right\|_2,\end{aligned}$$

where the latter inequality holds with probability at least $1-\delta$. Equivalently,

$$\left\|(\mathbf{I}_M - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2 - \left\|(\mathbf{I}_M - \mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2$$
$$\geq -\left(1+o(1)\right)\frac{\sqrt{MK}}{\delta}\left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\left\|\mathbf{a}\right\|_2,$$

Thus, with probability at least $1-\delta$,

$$\sqrt{\frac{P_{\text{music}}}{\tilde{P}_{\text{music}}}} = \frac{\left\|(\mathbf{I}-\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)\mathbf{a}\right\|_2}{\left\|(\mathbf{I}-\mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2}$$
$$\geq 1 - \left(1+o(1)\right)\frac{\sqrt{MK}}{\delta}\left(\frac{\sigma_{K+1}(\mathbf{S})}{\sigma_K(\mathbf{S})}\right)^{t+1}\frac{\left\|\mathbf{a}\right\|_2}{\left\|(\mathbf{I}-\mathbf{U}_K\mathbf{U}_K^H)\mathbf{a}\right\|_2},$$

by which the theorem follows. $\qquad\square$

REFERENCES

[1] W. D. Jones, "Keeping cars from crashing," *IEEE Spectrum*, vol. 38, no. 9, pp. 40–45, 2001.

[2] D. Guizzo, "How google's self-driving car works," *IEEE Spectrum Online*, vol. 18, 2011.

[3] E. P. Blasch, A. Lakhotia, and G. Seetharaman, "Unmanned vehicles come of age: The darpa grand challenge," *Computer*, vol. 39, no. 12, pp. 26–29, 2006.

[4] K. Alonzo, A. Stentz, O. Amidi, M. Bode, D. Bradley, A. Diazcalderon, M. Happold, H. Herman, R. Mandelbaum, and T. Pilarski, "Toward reliable off road autonomous vehicles operating in challenging environments," in *Proc of International Symposium on Exprimental Robotics*, 2004.

[5] M. Murad, I. Bilik, M. Friesen, J. Nickolaou, J. Salinger, K. Geary, and J. S. Colburn, "Requirements for next generation automotive radars," pp. 1–6, 2013.

[6] C. Waldschmidt and H. H. Meinel, "Future trends and directions in radar concerning the application for autonomous driving," pp. 1719–1722, 2014.

[7] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-wave technology for automotive radar sensors in the 77 ghz frequency band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 845–860, 2012.

[8] A. Natarajan, A. Komijani, X. Guan, A. Babakhani, and A. Hajimiri, "A 77-ghz phased-array transceiver with on-chip antennas in silicon: Transmitter and local lo-path phase shifting," *IEEE Journal of Solid-state Circuits*, vol. 41, no. 12, pp. 2807–2819, 2006.

[9] I. Gresham, N. Jain, T. Budka, A. Alexanian, N. Kinayman, B. Ziegner, S. Brown, and P. Staecker, "A compact manufacturable 76-77-ghz radar module for commercial acc applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 1, pp. 44–58, 2001.

[10] W. Mayer, A. Gronau, W. Menzel, and H. Leier, "A compact 24 ghz sensor for beam-forming and imaging," pp. 1–6, 2006.

[11] J. Li and P. Stoica, "Mimo radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[12] I. Bilik, S. Villeval, D. Brodeski, H. Ringel, O. Longman, P. Goswami, C. Y. B. Kumar, S. Rao, P. Swami, A. Jain, *et al.*, "Automotive multimode cascaded radar data processing embedded system," in *Proceedings of 2018 IEEE Radar Conference*, pp. 1–6, 2018.

[13] F. Gini, A. D. Maio, and L. Patton, "Waveform design and diversity for advanced radar systems," p. 552, 2012.

[14] S. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017.

[15] A. G. Stove, "Linear fmcw radar techniques," *Proceedings of the IEEE*, 1992.

[16] C. H. Lin, Y. S. Wu, Y. L. Yeh, S. H. Weng, G. Y. Chen, C. H. Shen, and H. Y. Chang, "A 24-ghz highly integrated transceiver in 0.5-μm e/d-phemt process for fmcw automotive radar applications," in *Microwave Conference Proceedings*, pp. 512–515, 2011.

[17] G. Babur, O. A. Krasnov, A. Yarovoy, and P. Aubry, "Nearly orthogonal waveforms for mimo fmcw radar," *IEEE Transactions on Aerospace & Electronic Systems*, vol. 49, no. 3, pp. 1426–1437, 2013.

[18] P. Kumari, J. Choi, N. Gonzalezprelcic, and R. W. Heath, "Ieee 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3012–3027, 2018.

[19] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[20] R. H. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[21] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[22] D. Kok and J. S. Fu, "Signal processing for automotive radar," in *Radar Conference, 2005 IEEE International*, pp. 842–846, 2005.

[23] S. Marcos, A. Marsal, and M. Benidir, "The propagator method for source bearing estimation," *Signal Processing*, vol. 42, no. 2, pp. 121–138, 1995.

[24] S. Marcos, A. Marsal, and M. Benidir, "Performances analysis of the propagator method for source bearing estimation," in *Acoustics, Speech, & Signal Processing, on IEEE International Conference*, 1994.

[25] N. Tayem and H. M. Kwon, "L-shape 2-dimensional arrival angle estimation with propagator method," *IEEE Trans Antennas & Propagat*, vol. 1, no. 5, pp. 1622–1630, 2005.

[26] J. Wenger, "Automotive radar - status and perspectives," in *Compound Semiconductor Integrated Circuit Symposium, 2005 (CSIC '05)*, p. 4 pp., 2005.

[27] F. Engels, P. Heidenreich, A. M. Zoubir, F. K. Jondral, and M. Wintermantel, "Advances in automotive radar: A framework on computationally efficient high-resolution frequency estimation," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 36–46, 2017.

[28] M. Y. Keegan Garcia and A. Purkovic, "Robust traffic and intersection monitoring using millimeter wave sensors," tech. rep., Texas Instruments Incorporated (TI), May 2018.

[29] K. Ramasubramanian, "Using a complex-baseband architecture in FMCW radar systems," tech. rep., Texas Instruments Incorporated (TI), May 2017.

[30] A. Gittens and M. W. Mahoney, "Revisiting the Nystrom method for improved large-scale machine learning," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 3977–4041, 2016.

[31] E. J. Nyström, "Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben," *Acta Mathematica*, vol. 54, no. 1, pp. 185–204, 1930.

[32] C. K. I. Williams and M. W. Seeger, "Using the Nystrom method to speed up kernel machines," in *Proc. of Advances in Neural Information Processing Systems (NIPS)*, pp. 682–688, 2001.

[33] D. P. Woodruff, "Sketching as a tool for numerical linear algebra," *Foundations and Trends® in Theoretical Computer Science*, vol. 10, no. 1–2, pp. 1–157, 2014.

[34] P. Drineas and M. W. Mahoney, "On the Nyström method for approximating a gram matrix for improved kernel-based learning," *Journal of Machine Learning Research*, vol. 6, pp. 2153–2175, 2005.

[35] S. Kumar, M. Mohri, and A. Talwalkar, "Sampling methods for the Nystrom method," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 981–1006, 2012.

[36] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, "Fixed-rank approximation of a positive-semidefinite matrix from streaming data," in *Advances in Neural Information Processing Systems (NIPS)*, 2017.

[37] S. Wang, A. Gittens, and M. W. Mahoney, "Scalable kernel k-means clustering with Nystrom approximation: Relative-error bounds," *Journal of Machine Learning Research*, vol. 20, no. 12, pp. 1–49, 2019.

[38] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[39] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[40] S. Wang, L. Luo, and Z. Zhang, "SPSD matrix approximation vis column selection: Theories, algorithms, and extensions," *Journal of Machine Learning Research*, vol. 17, no. 49, pp. 1–49, 2016.

[41] S. Wang, A. Gittens, and M. W. Mahoney, "Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging," *Journal of Machine Learning Research*, vol. 18, no. 218, pp. 1–50, 2018.

[42] H. L. V. Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory*. New York: John Wiley & Sons, Inc., 2001.

[43] D. Oh and J. Lee, "Low-complexity range-azimuth fmcw radar sensor using joint angle and delay estimation without svd and evd," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 4799–4811, 2015.

[44] R. O. Schmidt, "A signal subspace approach to multiple emitter location spectral estimation," *Ph.d.thesis Stanford University*, 1981.

[45] J. W. Odendaal, E. Barnard, and C. W. I. Pistorius, "Two-dimensional superresolution radar imaging using the music algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 42, no. 10, pp. 1386–1391, 1994.

[46] F. Belfiori, W. V. Rossum, and P. Hoogeboom, "Application of 2d music algorithm to range-azimuth fmcw radar data," in *Proc. of Radar Conference*, pp. 242–245, 2013.

[47] S. Kim, D. Oh, and J. Lee, "Joint dft-esprit estimation for toa and doa in vehicle fmcw radars," *IEEE Antennas & Wireless Propagation Letters*, vol. 14, pp. 1710–1713, 2015.

[48] A. Hirotugu, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[49] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[50] M. Rudelson and R. Vershynin, "The Littlewood–Offord problem and invertibility of random matrices," *Advances in Mathematics*, vol. 218, no. 2, pp. 600–633, 2008.

[51] T. Tao and V. Vu, "Random matrices: The distribution of the smallest singular values," *Geometric And Functional Analysis*, vol. 20, no. 1, pp. 260–297, 2010.

[52] P. Arbenz, "Lecture notes on solving large scale eigenvalue problems," *D-MATH, EHT Zurich*, vol. 2, 2012.