RESEARCH ARTICLE

Convex Estimation of Sparse-Smooth Power Spectral Densities From Mixtures of Realizations With Application to Weather Radar

HIROKI KURODA[1], (Member, IEEE), **DAICHI KITAHARA**[2], (Member, IEEE),
EIICHI YOSHIKAWA[3,4], (Member, IEEE), **HIROSHI KIKUCHI**[5], (Member, IEEE),
AND TOMOO USHIO[2], (Senior Member, IEEE)

[1]Department of Information and Management Systems Engineering, Nagaoka University of Technology, Niigata 940-2188, Japan
[2]Division of Electrical, Electronic and Infocommunications Engineering, Osaka University, Osaka 565-0871, Japan
[3]Aeronautical Technology Directorate, Japan Aerospace Exploration Agency, Tokyo 181-0015, Japan
[4]Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80521, USA
[5]Center for Space Science and Radio Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan

Corresponding author: Hiroki Kuroda (kuroda@vos.nagaokaut.ac.jp)

ABSTRACT In this paper, we propose a convex optimization-based estimation of sparse and smooth power spectral densities (PSDs) of complex-valued random processes from mixtures of realizations. While the PSDs are related to the magnitude of the frequency components of the realizations, it has been a major challenge to exploit the smoothness of the PSDs, because penalizing the difference of the magnitude of the frequency components results in a nonconvex optimization problem that is difficult to solve. To address this challenge, we design the proposed model that jointly estimates the complex-valued frequency components and the nonnegative PSDs, which are respectively regularized to be sparse and sparse-smooth. By penalizing the difference of the nonnegative variable that estimates the PSDs, the proposed model can enhance the smoothness of the PSDs via convex optimization. Numerical experiments on the phased array weather radar, an advanced weather radar system, demonstrate that the proposed model achieves superior estimation accuracy compared to existing sparse estimation models, regardless of whether they are combined with a smoothing technique as a post-processing step or not.

INDEX TERMS Power spectral density estimation, random process, sparsity, smoothness, regularization, convex optimization, weather radar.

I. INTRODUCTION

Power spectral density (PSD) of a random process describes how power of the random process is distributed over frequency. Estimation of the PSD from realizations of a random process is a fundamental problem in science and engineering [1], [2], [3]. For weather radar applications, the PSD estimation is essential for the analysis of weather phenomena, because the PSD contains information pertaining to the precipitation intensity and the Doppler velocity distribution [4], [5], [6], [7], [8]. For example, the parabolic Doppler weather radar [4] transmits a pencil beam and

subsequently observes backscattered signals in a narrow range of elevation angles, which can be regarded as realizations of a single random process whose PSD reflects the weather condition in the narrow range.

We consider the estimation of PSDs from mixtures of realizations of random processes, which is much more challenging than the classical case of a single random process. Our primary interest is on the phased array weather radar (PAWR) [9], [10], [11], [12], [13], [14], [15], [16], which is developed to detect hazardous weather phenomena. The Doppler weather radar is not capable of detecting hazardous weather because of its mechanical vertical scan for observing backscattered signals in multiple elevation angles, which requires a long observation time. To shorten the observation

The associate editor coordinating the review of this manuscript and approving it for publication was Cheng Hu.

time, the PAWR transmits a fan beam and subsequently observes backscattered signals in a wide range of elevation angles. The backscattered signals observed by the PAWR can be modeled as mixtures of realizations of random processes whose PSDs reflect the weather conditions in finely-divided ranges. Thus, to obtain the weather condition in a fine resolution, the PAWR needs digital signal processing, recovering the PSDs in the finely-divided ranges.

Since the estimation of PSDs from mixtures of realizations is a challenging problem, the major existing methods employ a two-step approach that first estimates the frequency components of the realizations and then estimates the PSDs. For the frequency component estimation, sparsity-aware methods have achieved significant improvements on the estimation accuracy over the classical linear methods in many fields [16], [17], [18], [19], [20], [21], [22], [23]. Sparsity in the spatial domain is exploited in [17], [18], and [19] under the assumption that signals arrive from only a few angles. Unfortunately, this assumption is far from suitable for the PAWR because targets such as clouds and raindrops exist at many angles [16]. In [20], [21], [22], and [23], isolated sparse frequency components, called *line spectra*, are estimated based on the ℓ_1 regularization. It is demonstrated in [16] that a block-sparse regularization model using the mixed ℓ_2/ℓ_1 norm [24], [25], [26], [27], [28], [29], [30] is more effective for weather radar applications because the frequency components are clustered in a few blocks due to the narrow-bandwidth of the PSDs. After the frequency component estimation, the *periodogram*, i.e., the squared magnitude of the frequency components, is usually employed to estimate the PSD because of its asymptotic unbiasedness. However, the periodogram has the drawbacks of large variance and erratic oscillation [1], [2], [3], [4]. For the classical case of a single random process, smoothing techniques, e.g., those shown in [1] and [2], are often used to reduce the variance and the erratic oscillation. While the existing smoothing techniques can be used as a post-processing step, such a two-stage approach would be sub-optimal because the smoothness is not exploited when estimating the frequency components. Since the PSD is estimated by the periodogram, i.e., the squared magnitude of the frequency components, one may add a penalty for the difference between the magnitude of the frequency components in the frequency component estimation. However, due to the nonconvexity of this type of penalty (see, e.g., [31]), it is hard for this approach to obtain an optimal solution, and the performance dependency on the initial estimate and the optimization algorithm is difficult to elude.

Another line of studies derive approximated observation models between the realizations and the PSD, e.g., for a single random process [32], [33], [34], [35] and spatially independent random processes [36], [37]. Since the approximated observation model is written in terms of the (nonnegative) PSD, the smoothness of the PSD could be exploited via convex optimization. However, this approach takes the magnitude of the observation model to derive

the approximated observation models, implying that half of the information in the original observation model is lost as the phase information is discarded. In particular, this approach is not applicable to the PAWR because signals from different angles cannot be distinguished by the magnitude information (see (6) in Example 1).

In this paper, we propose a convex optimization-based method that simultaneously estimates block-sparse frequency components and block-sparse and smooth PSDs from mixtures of realizations.[1] To design the proposed method, we first apply the optimally structured block-sparse model of [38] for the frequency component estimation. Then, we newly leverage the latent variable of the designed model, which is originally introduced to optimize the block structure, for the PSD estimation. More precisely, we demonstrate that the latent variable is in fact related to the square root of the PSDs, enabling us to exploit the smoothness of the PSDs via convex optimization. The main contributions of this paper are summarized as follows.

- We present, for the first time in the literature, a convex optimization-based method that can exploit the smoothness of the PSDs for their estimation from mixtures of realizations.
- We show that many smoothness priors designed for real-valued signals, including the high-order total variation [39], [40] and the total generalized variation [41], can be directly incorporated into the proposed framework to enhance the smoothness of the PSDs of complex-valued random processes thanks to the nonnegative latent variable.
- We conduct thorough numerical simulations on the PAWR, which demonstrate that the proposed method achieves superior estimation accuracy to the existing sparse estimation models combined with or without post-smoothing, i.e., a smoothing technique applied as a post-processing step after the frequency component estimation.

The rest of this paper is organized as follows. In Section II, we formulate the estimation of PSDs from mixtures of realizations of random processes, and clarify its relation to weather radar applications. In Section III, we design the proposed convex optimization model that simultaneously estimates block-sparse frequency components and block-sparse and smooth PSDs. Section IV presents numerical experiments on the PAWR, followed by conclusion in Section V.

A preliminary short version of this paper was presented at a conference [42].

Notations: $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$, and $\mathbb{C}$ respectively denote the sets of nonnegative integers, real numbers, nonnegative real numbers, positive real numbers, and complex numbers. We use $\iota \in \mathbb{C}$ to denote the imaginary unit, i.e., $\iota = \sqrt{-1}$. For every $x \in \mathbb{C}$, x^* denotes the complex conjugate of x, and

---

[1]If a target is both sparse and smooth, it is also block-sparse since nonzero components are clustered in several blocks due to the smoothness.

$|x| := \sqrt{x^* x}$ denotes the absolute value of x. For matrices or vectors, we denote the transpose and the Hermitian transpose respectively by $(\cdot)^\top$ and $(\cdot)^H$. The identity matrix of order N is denoted by $\mathbf{I}_N \in \mathbb{R}^{N \times N}$. We denote the diagonal matrix with components of $\mathbf{w} \in \mathbb{C}^N$ on the main diagonal by $\mathrm{diag}(\mathbf{w}) \in \mathbb{C}^{N \times N}$. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The ℓ_2 (Euclidean) norm, the ℓ_1 norm, and the ℓ_0 pseudo-norm of $\mathbf{x} = (x_1, \ldots, x_N)^\top \in \mathbb{C}^N$ are respectively denoted by $\|\mathbf{x}\| := \sqrt{\sum_{n=1}^N |x_n|^2}$, $\|\mathbf{x}\|_1 := \sum_{n=1}^N |x_n|$, and $\|\mathbf{x}\|_0 := |\{n \in \{1, \ldots, N\} \mid x_n \neq 0\}|$. The expectation operator is denoted by $E[\cdot]$.

II. PROBLEM FORMULATION

We consider the estimation problem of power spectral densities (PSDs) of N random processes from noisy mixtures of their realizations. We denote the n-th complex-valued discrete-time random process ($n = 1, \ldots, N$) by

$$X_n^\star[\ell] \in \mathbb{C} \qquad (\ell = 0, \pm 1, \pm 2, \ldots). \tag{1}$$

To define the PSD of $X_n^\star[\ell]$, we assume that $X_n^\star[\ell]$ is zero-mean and wide-sense stationary, which imply that $E[X_n^\star[\ell]] = 0$ for any ℓ, and the auto-correlation $E[X_n^\star[m + \ell](X_n^\star[m])^*]$ does not depend on m for any ℓ. Under these assumptions, define the auto-correlation function by

$$R_n[\ell] := E[X_n^\star[m + \ell](X_n^\star[m])^*],$$

and suppose $\sum_{\ell=-\infty}^\infty |R_n[\ell]| < \infty$. Then, the PSD of $X_n^\star[\ell]$ is given by

$$S_n^\star(f) := \sum_{\ell=-\infty}^\infty R_n[\ell] e^{-\imath 2\pi f \ell} \qquad \left(f \in \left[-\frac{1}{2}, \frac{1}{2}\right)\right). \tag{2}$$

We denote L consecutive realizations of $X_n^\star[\ell]$ by

$$\bar{x}_{j,n}[\ell] \in \mathbb{C} \qquad (\ell = 1, \ldots, L), \tag{3}$$

where $j \in \{1, \ldots, J\}$ is the index of trials. Note that $\bar{x}_{j,n}[\ell]$ for $j = 1, \ldots, J$ are assumed to be realizations of the common random process $X_n^\star[\ell]$ (see Remark 1 for validity of this assumption). We define the observation model by

$$\mathbf{y}_j := \sum_{n=1}^N \mathbf{A}_n \bar{\mathbf{x}}_{j,n} + \boldsymbol{\varepsilon}_j \in \mathbb{C}^d \quad (j = 1, \ldots, J), \tag{4}$$

where the realizations in (3) are collectively denoted by

$$\bar{\mathbf{x}}_{j,n} := (\bar{x}_{j,n}[1], \bar{x}_{j,n}[2], \ldots, \bar{x}_{j,n}[L])^\top \in \mathbb{C}^L, \tag{5}$$

$\mathbf{A}_n \in \mathbb{C}^{d \times L}$ is the known matrix that models the observation process for the n-th source, and $\boldsymbol{\varepsilon}_j \in \mathbb{C}^d$ is the (unknown) observation noise. Our goal is to estimate the PSDs $S_n^\star(f)$ ($n = 1, \ldots, N$) from $\mathbf{y}_j$ ($j = 1, \ldots, J$) and $\mathbf{A}_n$ ($n = 1, \ldots, N$) in (4). Note that the classical PSD estimation problem for a single random process [1], [2], [3], e.g., for the Doppler weather radar [4], is a special instance of (4) for $N = 1$, $d = L$, and $\mathbf{A}_n = \mathbf{I}_L$. The generalized observation model (4) is introduced to cover the

PSD estimation for the PAWR [9], [10], [11], [12], [13], [14], [15], [16], which is our primary interest.

Example 1 (PAWR): For the PAWR, $X_n^\star[\ell]$ corresponds to the sum of backscattered signals in the angular interval $\left[\theta_n - \frac{\Delta\theta}{2}, \theta_n + \frac{\Delta\theta}{2}\right]$, where θ_n ($n = 1, \ldots, N$) are the equally spaced angles with a spacing of $\Delta\theta$. By using an M-element uniform linear array, the PAWR observes noisy mixtures of realizations by

$$\mathbf{y}_j[\ell] := \sum_{n=1}^N \mathbf{a}(\theta_n) \bar{x}_{j,n}[\ell] + \boldsymbol{\varepsilon}_j[\ell] \in \mathbb{C}^M \quad (\ell = 1, \ldots, L) \tag{6}$$

for each $j \in \{1, \ldots, J\}$, where $\mathbf{a}(\theta_n) \in \mathbb{C}^M$ is the known steering vector for the angle θ_n, and $\boldsymbol{\varepsilon}_j[\ell] \in \mathbb{C}^M$ is the white Gaussian noise. More precisely, $\mathbf{a}(\theta)$ is defined by

$$\mathbf{a}(\theta) := \left(1, e^{-\imath \frac{2\pi \Delta \sin\theta}{\lambda_{\mathrm{cw}}}}, \ldots, e^{-\imath \frac{2(M-1)\pi \Delta \sin\theta}{\lambda_{\mathrm{cw}}}}\right)^\top \in \mathbb{C}^M,$$

where Δ is the inter-element spacing of the uniform linear array, and λ_{cw} is the carrier wavelength. The observation model (6) for the PAWR can be written in the form of (4), i.e.,

$$\mathbf{y}_j^{(\mathrm{pawr})} = \sum_{n=1}^N \mathbf{A}_n^{(\mathrm{pawr})} \bar{\mathbf{x}}_{j,n} + \boldsymbol{\varepsilon}_j^{(\mathrm{pawr})} \quad (j = 1, \ldots, J),$$

by setting

$$\mathbf{y}_j^{(\mathrm{pawr})} := (\mathbf{y}_j[1]^\top, \mathbf{y}_j[2]^\top, \ldots, \mathbf{y}_j[L]^\top)^\top \in \mathbb{C}^{ML},$$

$$\boldsymbol{\varepsilon}_j^{(\mathrm{pawr})} := (\boldsymbol{\varepsilon}_j[1]^\top, \boldsymbol{\varepsilon}_j[2]^\top, \ldots, \boldsymbol{\varepsilon}_j[L]^\top)^\top \in \mathbb{C}^{ML},$$

and $\mathbf{A}_n^{(\mathrm{pawr})} \in \mathbb{C}^{ML \times L}$ to the block-diagonal matrix that contains L copies of $\mathbf{a}(\theta_n)$ on the diagonal blocks, i.e.,

$$\mathbf{A}_n^{(\mathrm{pawr})} := \begin{pmatrix} \mathbf{a}(\theta_n) & & & \\ & \mathbf{a}(\theta_n) & & \\ & & \ddots & \\ & & & \mathbf{a}(\theta_n) \end{pmatrix}.$$

The estimation of the PSDs $S_n^\star(f)$ of $X_n^\star[\ell]$ ($n = 1, \ldots, N$) is essential for the PAWR because $S_n^\star(f)$ contains information about the weather condition in the narrow angular interval $\left[\theta_n - \frac{\Delta\theta}{2}, \theta_n + \frac{\Delta\theta}{2}\right]$. More precisely, the weather condition can be obtained from $S_n^\star(f)$ as follows. First, the continuous-time PSD $S_n^{\star(\mathrm{ct})}(f)$ is recovered from $S_n^\star(f)$. When aliasing does not occur in $S_n^\star(f)$, $S_n^{\star(\mathrm{ct})}(f)$ can be simply obtained by $S_n^{\star(\mathrm{ct})}(f) = TS_n^\star(Tf)$ if $|f| \leq \frac{1}{2T}$ and $S_n^{\star(\mathrm{ct})}(f) = 0$ otherwise, where T is the pulse repetition time, i.e., the sampling interval of $X_n^\star[\ell]$. Even when aliasing occurs in $S_n^\star(f)$, $S_n^{\star(\mathrm{ct})}(f)$ can be recovered from $S_n^\star(f)$ unless the variation of wind velocity is extremely large, and thus the anti-aliasing filter is usually not employed in weather radar applications (see, e.g., [4, Chapter 5] and [16, Section II-A] for detail). Next, $S_n^{\star(\mathrm{ct})}(f)$ is decomposed as

$$S_n^{\star(\mathrm{ct})}(f) = P_n^\star q_n^\star(f),$$

where $P_n^\star := \int_{-\infty}^{\infty} S_n^{\star(\text{ct})}(f)df$ and $q_n^\star(f) := S_n^{\star(\text{ct})}(f)/P_n^\star$ respectively correspond to the precipitation intensity and the Doppler frequency distribution [4]. The Doppler frequency distribution can be converted to the distribution of Doppler velocity v, i.e., wind velocity parallel to the incident beam direction, by $v = \frac{\lambda_{\text{cw}}f}{2}$. For instance, the area of nonzero values of $q_n^\star(f)$ implies the existence of the corresponding wind velocity components. The precipitation intensity and the Doppler velocity distribution are useful for the analysis of weather phenomena, e.g., [5], [7], [8], and [43] for tornado, [6] for weather clutter, and [44] for raindrop size distribution. Since the estimation accuracy of the precipitation intensity and the Doppler velocity distribution heavily depends on that of the PSD, it is important for the analysis of weather phenomena based on the PAWR to realize a method that can accurately estimate the PSDs $S_n^\star(f)$ $(n = 1, \ldots, N)$ from the mixtures of realizations.

Remark 1 (Tradeoff between L and J): To derive the observation model in (4) where $\bar{x}_{j,n}[\ell]$ $(j = 1, \ldots, J)$ are realizations of the common random process $X_n^\star[\ell]$, similarly to the case of a single random process $(N = 1)$ [1], [2], [3], [4], we split all observations into J subsets. For weather radar applications, the total number K_{pls} of pulses is divided into J subsets, and thus we have $L = \frac{K_{\text{pls}}}{J}$. Since $\frac{1}{L} = \frac{J}{K_{\text{pls}}}$ is the frequency resolution, i.e., the sampling interval in the frequency domain (see (11)), increasing J sacrifices the frequency resolution [4, Chapter 5]. Note that we cannot increase K_{pls} unboundedly because K_{pls} corresponds to the observation time, and thus is set to be small enough to ensure that the statistics of targets such as clouds and raindrops are (approximately) unchanged. Typically, J is set to be very small for the sake of a fine frequency resolution, and $J = 1$ is of particular interest as the original frequency resolution $\frac{1}{K_{\text{pls}}}$ is preserved [4]. Note that, while $J = 1$ is a typical choice in practice, our method, which will be developed in Section III, is applicable to general J.

We rewrite (4) to an observation model in terms of frequency components of the time-domain realizations $\bar{x}_{j,n}$ in (5) because of their more direct relation to the PSD in (2) than $\bar{x}_{j,n}$. More precisely, we represent $\bar{x}_{j,n}$ as

$$\bar{x}_{j,n} = \mathbf{G}\bar{u}_{j,n} \tag{7}$$

for $j = 1, \ldots, J$ and $n = 1, \ldots, N$, where

$$\bar{u}_{j,n} := (\bar{u}_{j,n}[1], \bar{u}_{j,n}[2], \ldots, \bar{u}_{j,n}[L])^\top \in \mathbb{C}^L \tag{8}$$

is used as the frequency components, and $\mathbf{G} \in \mathbb{C}^{L \times L}$ is a suitable synthesis matrix. Substituting the representation (7) to the observation model (4), we have

$$\mathbf{y}_j = \sum_{n=1}^{N} \mathbf{A}_n \mathbf{G}\bar{u}_{j,n} + \boldsymbol{\varepsilon}_j \in \mathbb{C}^d \quad (j = 1, \ldots, J), \tag{9}$$

which is used as the observation model for the frequency components $\bar{u}_{j,n}$. The representation (7) covers popular frequency analysis methods used in weather radar applications,

e.g., the discrete Fourier transform (DFT) and the windowed DFT.

Example 2 (DFT): Let $\bar{u}_{j,n}^{(\text{DFT})}$ be the normalized DFT coefficients of $\bar{x}_{j,n}$. Define the normalized DFT matrix $\mathbf{F} \in \mathbb{C}^{L \times L}$ by

$$\mathbf{F} := \frac{1}{\sqrt{L}} \begin{pmatrix} 1 & e^{-\iota 2\pi f_1} & \cdots & e^{-\iota 2\pi f_1(L-1)} \\ 1 & e^{-\iota 2\pi f_2} & \cdots & e^{-\iota 2\pi f_2(L-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\iota 2\pi f_L} & \cdots & e^{-\iota 2\pi f_L(L-1)} \end{pmatrix}, \tag{10}$$

where

$$f_k := \frac{k - 1 - L/2}{L} \quad (k = 1, \ldots, L) \tag{11}$$

are uniform sampling points in $[-1/2, 1/2]$ used as a frequency grid, and L is assumed to be even for simplicity. Then, $\bar{u}_{j,n}^{(\text{DFT})}$ is given by

$$\bar{u}_{j,n}^{(\text{DFT})} = \mathbf{F}\bar{x}_{j,n}.$$

Due to the unitarity of $\mathbf{F}$, we have

$$\bar{x}_{j,n} = \mathbf{F}^{\text{H}}\bar{u}_{j,n}^{(\text{DFT})},$$

which corresponds to (7) with $\mathbf{G} = \mathbf{F}^{\text{H}}$.

Example 3 (Windowed DFT): To mitigate the frequency sidelobes, a window function $\mathbf{w} \in \mathbb{R}_{++}^L$ is applied to $\bar{x}_{j,n}$ before the DFT in some cases [3], [45]. The windowed DFT coefficients are given by

$$\bar{u}_{j,n}^{(\text{WDFT})} = \mathbf{FW}\bar{x}_{j,n},$$

where $\mathbf{F}$ is the DFT matrix in (10), and $\mathbf{W} := \text{diag}(\mathbf{w}) \in \mathbb{R}_{++}^{L \times L}$. Since $(\mathbf{FW})^{-1} = \mathbf{W}^{-1}\mathbf{F}^{\text{H}}$, we have

$$\bar{x}_{j,n} = \mathbf{W}^{-1}\mathbf{F}^{\text{H}}\bar{u}_{j,n}^{(\text{WDFT})},$$

which corresponds to (7) with $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\text{H}}$.

A. MAJOR CHALLENGE IN PSD ESTIMATION FROM MIXTURES OF REALIZATIONS

The square of the magnitude of the frequency components in (8), i.e.,

$$|\bar{u}_{j,n}[k]|^2 \quad (k = 1, \ldots, L), \tag{12}$$

is called the *periodogram*[2] and widely used as an estimate of the PSD $S_n^\star(f)$ on the frequency grid defined in (11). It should be noted that the periodogram needs to be estimated from the mixtures of realizations in (9) for our problem. The periodogram with the DFT shown in Example 2 is an asymptotically unbiased estimator of the PSD under a mild condition [1], [2]. More precisely, since $\bar{u}_{j,n}^{(\text{DFT})}[k]$ can be regarded as a realization of the random variable

$$U_n^\star[k] := \frac{1}{\sqrt{L}} \sum_{\ell=1}^{L} X_n^\star[\ell]e^{-\iota 2\pi f_k(\ell-1)} \quad (k = 1, \ldots, L),$$

---

[2] Strictly speaking, (12) is called the periodogram when $\mathbf{G}$ in (7) is the inverse DFT $\mathbf{F}^{\text{H}}$ in Example 2, and (12) with $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\text{H}}$ is called, e.g., the *windowed periodogram* [1] or the *modified periodogram* [3].

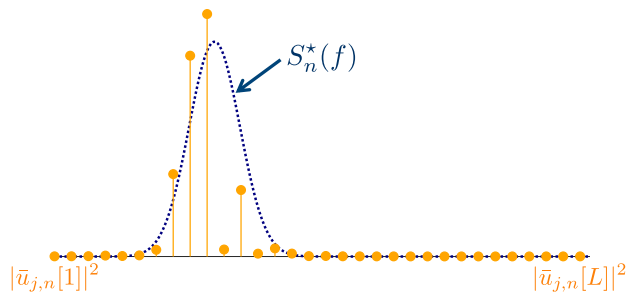**FIGURE 1.** The PSD and the periodogram computed from the realizations for a simulation.

the asymptotic unbiasedness means that

$$\lim_{L \to \infty} E\left[|U_n^\star[k]|^2\right] = S_n^\star(f_k) \quad (k = 1, \dots, L). \quad (13)$$

A simple proof of (13) is provided in [16, Section II-A] under a mild sufficient condition $\sum_{\ell=-\infty}^{\infty} |\ell R_n[\ell]| < \infty$. The periodogram with the windowed DFT in Example 3 is also asymptotically unbiased if the window function is properly designed [3].

The drawback of the periodogram lies in its large variance and often-observed erratic oscillation [1], [2], [3], [4] (see also Fig. 1 for an example from the experiments in Section IV). Indeed, this drawback is theoretically shown for the periodogram with the DFT when $X_n^\star[\ell]$ is a linear combination of i.i.d. Gaussian random variables and $L \to \infty$ [1], [2]. Namely, for this case, the variance is equally large to the square of the PSD, i.e.,

$$\lim_{L \to \infty} E\left[\left(|U_n^\star[k]|^2 - S_n^\star(f_k)\right)^2\right] = (S_n^\star(f_k))^2$$

for $k = 1, \dots, L$, and $U_n^\star[k]$ and $U_n^\star[k']$ ($k \neq k'$) are uncorrelated when $L \to \infty$. These facts validate the often-observed erratic oscillation of the periodogram.

Although several approaches have been developed to reduce the variance of the periodogram, they are not suitable for our problem in which the PSDs have to be estimated from the mixtures of realizations in (9). A simple approach is to exploit the situation that $\bar{u}_{j,n}[k]$ ($j = 1, \dots, J$) are realizations of the common random variable, i.e., to use the ensemble average of the periodograms:

$$\frac{1}{J}\sum_{j=1}^{J}|\bar{u}_{j,n}[k]|^2 \quad (k = 1, \dots, L). \quad (14)$$

Unfortunately, since J is typically very small in weather radar applications (see Remark 1), the ensemble average cannot sufficiently reduce the variance and the erratic oscillation. Since the PSD is usually smooth in weather radar applications [4], another promising approach is to exploit the smoothness of the PSD. However, existing smoothing techniques, e.g., those shown in [1] and [2], are not directly applicable to our problem because these techniques suppose that the frequency components $\bar{u}_{j,n}$ of the realizations are known. Using smoothing techniques as a post-processing step would be sub-optimal because the smoothness of the PSD is not considered in the estimation of the frequency

components. Thus, it remains a major challenge to exploit the smoothness of the PSDs when they need to be estimated from mixtures of realizations.

III. PROPOSED APPROACH

To exploit the sparsity and the smoothness for the PSD estimation from the observed mixtures of realizations in (9), we design a convex model that jointly estimates the frequency components $\bar{u}_{j,n}[k]$ and the PSDs $S_n^\star(f_k)$. In Section III-A, we first apply the optimally structured block-sparse model of [38] for the estimation of the frequency components. Then, in Section III-B, we leverage its latent variable, which is originally introduced for the block structure optimization, to estimate sparse and smooth PSDs.

A. BLOCK-SPARSE ESTIMATION OF FREQUENCY COMPONENTS

We design a block-sparse penalty for the frequency components $\bar{u}_{j,n}$ by applying the optimally structured block-sparse model [38] with the knowledge of the PAWR [16]. For simplicity, we begin by designing a penalty for each $n \in \{1, \dots, N\}$. As demonstrated in [16], the PSD $S_n^\star(f)$ is usually narrow-band for the PAWR, which implies that $\bar{u}_{j,n}$ is block-sparse for each source $n \in \{1, \dots, N\}$ and trial $j \in \{1, \dots, J\}$ due to the relation (13).[3] Moreover, since $\bar{u}_{j,n}$ ($j = 1, \dots, J$) are realizations of the common random variable, suitable block partitions for $\bar{u}_{j,n}$ ($j = 1, \dots, J$) are the same. Thus, using the mixed ℓ_2/ℓ_1 norm that is suitable for the block-sparsity, we introduce a penalty for $\mathbf{u}_n := (\mathbf{u}_{j,n})_{j=1}^J$ as

$$\|\mathbf{u}_n\|_{2,1}^{(\mathcal{B}_{m,n})_{m=1}^{h_n}} := \sum_{m=1}^{h_n} \sqrt{J|\mathcal{B}_{m,n}|} \left\|((u_{j,n}[k])_{j=1}^J)_{k \in \mathcal{B}_{m,n}}\right\|$$

$$= \sum_{m=1}^{h_n} \sqrt{J|\mathcal{B}_{m,n}|} \sqrt{\sum_{j=1}^{J}\sum_{k \in \mathcal{B}_{m,n}}|u_{j,n}[k]|^2},$$

where $\mathcal{B}_{m,n} \subset \{1, \dots, L\}$ ($m = 1, \dots, h_n$) is a block partition in the frequency domain of the n-th source. By suppressing the mixed ℓ_2/ℓ_1 norm, the block-sparsity is promoted because the components $((u_{j,n}[k])_{j=1}^J)_{k \in \mathcal{B}_{m,n}}$ in the same block are forced to be zeros together. The problem in [16] is that an appropriate block partition is unknown a priori because it depends on the unknown Doppler velocity distribution. To solve the problem of unknown block partition, following the approach of [38], we minimize the mixed ℓ_2/ℓ_1 norm over the partition of at most H_n blocks, i.e.,

$$\psi_{H_n}(\mathbf{u}_n) := \min_{h_n \in \{1,\dots,H_n\}} \left[\min_{(\mathcal{B}_{m,n})_{m=1}^{h_n} \in \mathcal{P}_{h_n}} \|\mathbf{u}_n\|_{2,1}^{(\mathcal{B}_{m,n})_{m=1}^{h_n}} \right]. \quad (15)$$

---

[3]Although we present (13) for $\mathbf{G} = \mathbf{F}^H$ in Example 2 for simplicity, $\bar{u}_{j,n}$ is also block-sparse when $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^H$ in Example 3 because the window function in Example 3 is designed to reduce the heights of the sidelobes and slightly increase the width of the mainlobe.
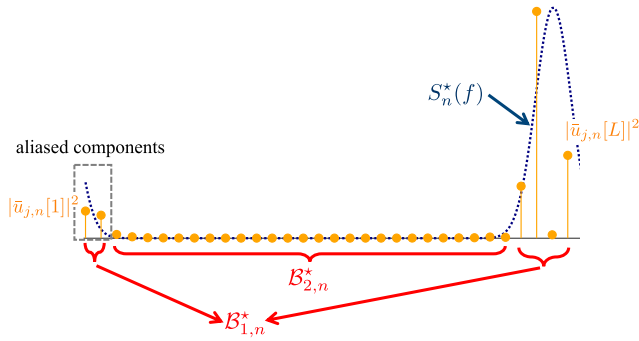
FIGURE 2. The PSD and the squared magnitude of the frequency components when aliasing occurs, and its suitable block partition.

The constraint set $\mathcal{P}_{h_n}$ consists of all h_n block partitions of $\{1, \ldots, L\}$, i.e.,

$$
(\mathcal{B}_{m,n})_{m=1}^{h_n} \in \mathcal{P}_{h_n}
$$

$$
\Leftrightarrow
\begin{cases}
\bigcup_{m=1}^{h_n} \mathcal{B}_{m,n} = \{1, \ldots, L\}, \\
\mathcal{B}_{m,n} \neq \varnothing \quad (m = 1, \ldots, h_n), \\
\mathcal{B}_{m,n} \cap \mathcal{B}_{m',n} = \varnothing \quad (m \neq m'), \\
\mathcal{B}_{m,n} = \mathfrak{M}_L \left(\{\ell \in \mathbb{N} \mid a_{m,n} \leq \ell \leq b_{m,n}\} \right) \\
\quad \text{for some } a_{m,n}, b_{m,n} \in \mathbb{N} \quad (m = 1, \ldots, h_n),
\end{cases}
$$

where

$$
\mathfrak{M}_L(\mathcal{I}) := \left\{ \ell - L \left\lfloor \frac{\ell-1}{L} \right\rfloor \in \{1, \ldots, L\} \;\middle|\; \ell \in \mathcal{I} \right\},
$$

and $\lfloor \cdot \rfloor$ is the floor function. For instance, when $\mathcal{I} = \{L - 1, L, L + 1\}$, $\mathfrak{M}_L(\mathcal{I}) = \{L - 1, L, 1\}$. Differently from the standard design proposed in [38], the present design makes $\mathcal{P}_{h_n}$ includes blocks connected by the first and the last entries, and is suitable for weather radar applications due to the following reason. Since aliasing is not a serious issue in weather radar applications, the anti-aliasing filter is usually not employed (see Example 1 and references [4, Chapter 5] and [16, Section II-A] for detail), and thus aliasing may occur, i.e., some Doppler frequency components may exceed the Nyquist frequency. For instance, in Fig. 2, a part of the PSD that exceeds the Nyquist frequency is aliased, and thus the corresponding frequency components are also aliased. In such cases, aliased nonzero components and non-aliased nonzero components are better to be collected into a single block, as shown in Fig. 2, because they form a single block before the aliasing. To realize such capability, $\mathcal{P}_{h_n}$ is designed to include blocks connected by the first and the last entries. Note that a block connected by the first and the last entries is not always adopted since the block partition is automatically optimized in (15).

Although it is difficult to use $\psi_{H_n}(\mathbf{u}_n)$ directly due to the combinatorial optimization in (15), we can construct a tight convex relaxation of $\psi_{H_n}(\mathbf{u}_n)$ as follows. Let $\phi : \mathbb{C}^J \times \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ be a lower semicontinuous convex function defined

by

$$
\phi(\mathbf{v}, \sigma) :=
\begin{cases}
\dfrac{\|\mathbf{v}\|^2}{2\sigma} + \dfrac{J}{2}\sigma, & \text{if } \sigma > 0; \\
0, & \text{if } \mathbf{v} = \mathbf{0} \text{ and } \sigma = 0; \\
\infty, & \text{otherwise.}
\end{cases}
\tag{16}
$$

Then, similarly to [38, Section II], $\psi_{H_n}(\mathbf{u}_n)$ can be rewritten as

$$
\psi_{H_n}(\mathbf{u}_n) = \min_{\substack{\boldsymbol{\sigma}_n \in \mathbb{R}_+^L \\ \|\mathbf{D}\boldsymbol{\sigma}_n\|_0 \leq H_n}} \sum_{k=1}^{L} \phi\left((u_{j,n}[k])_{j=1}^J, \sigma_n[k] \right), \tag{17}
$$

where $\mathbf{D} \in \mathbb{R}^{L \times L}$ is the first-order difference operator with the periodic boundary condition, i.e., the difference operator on the ring graph [46]. More precisely, $\mathbf{D}$ is defined by

$$
\mathbf{D} := \begin{bmatrix}
-1 & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & -1 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & -1 & 1 \\
1 & 0 & 0 & 0 & \cdots & 0 & -1
\end{bmatrix} \in \mathbb{R}^{L \times L}. \tag{18}
$$

Note that (17) is a slight extension of the result shown in [38] to the case where the blocks are fixed over the trials $j \in \{1, \ldots, J\}$. We can obtain a tight convex relaxation of (17) by replacing the ℓ_0 pseudo-norm in the constraint with its best convex relaxation, i.e., the ℓ_1 norm:

$$
\tilde{\psi}_{\alpha_n}(\mathbf{u}_n) := \min_{\substack{\boldsymbol{\sigma}_n \in \mathbb{R}_+^L \\ \|\mathbf{D}\boldsymbol{\sigma}_n\|_1 \leq \alpha_n}} \sum_{k=1}^{L} \phi\left((u_{j,n}[k])_{j=1}^J, \sigma_n[k] \right),
$$

where $\alpha_n \in \mathbb{R}_+$ is a tuning parameter related to the number of blocks. Although the sum $\sum_{n=1}^N \tilde{\psi}_{\alpha_n}(\mathbf{u}_n)$ can be used for the penalty of $(\mathbf{u}_n)_{n=1}^N$, tuning α_n for each $n \in \{1, \ldots, N\}$ could be troublesome. Thus, to simplify the tuning process, we propose a convex penalty for $\mathbf{u} := (\mathbf{u}_n)_{n=1}^N$ as

$$
\Psi_{\alpha}(\mathbf{u}) := \min_{\substack{(\boldsymbol{\sigma}_n)_{n=1}^N \in \mathbb{R}_+^{NL} \\ \sum_{n=1}^N \|\mathbf{D}\boldsymbol{\sigma}_n\|_1 \leq \alpha}} \sum_{n=1}^{N} \sum_{k=1}^{L} \phi\left((u_{j,n}[k])_{j=1}^J, \sigma_n[k] \right),
$$
$$
\tag{19}
$$

where the single tuning parameter $\alpha \in \mathbb{R}_+$ is related to the number of total blocks for $n = 1, \ldots, N$.

Using the proposed convex penalty $\Psi_{\alpha}(\mathbf{u})$ in (19), we estimate the frequency components by the regularized least squares for the observation model in (9), i.e.,

$$
\underset{\mathbf{u} \in \mathbb{C}^{JNL}}{\text{minimize}} \; \frac{1}{2} \sum_{j=1}^{J} \left\| \mathbf{y}_j - \sum_{n=1}^{N} \mathbf{A}_n \mathbf{G} \mathbf{u}_{j,n} \right\|^2 + \lambda \Psi_{\alpha}(\mathbf{u}), \tag{20}
$$

where $\lambda > 0$ is the regularization parameter that controls the importance of the block-sparsity. Substituting the definition

of $\Psi_\alpha(\mathbf{u})$ in (19) into (20), we can reduce the optimization problem (20) to

$$
\begin{aligned}
\underset{\mathbf{u}\in\mathbb{C}^{JNL},\boldsymbol{\sigma}\in\mathbb{R}_+^{NL}}{\text{minimize}} \quad & \frac{1}{2}\sum_{j=1}^{J}\left\|\mathbf{y}_j - \sum_{n=1}^{N}\mathbf{A}_n\mathbf{G}\mathbf{u}_{j,n}\right\|^2 \\
& + \lambda\sum_{n=1}^{N}\sum_{k=1}^{L}\phi\left((u_{j,n}[k])_{j=1}^{J}, \sigma_n[k]\right) \\
\text{subject to} \quad & \sum_{n=1}^{N}\|\mathbf{D}\boldsymbol{\sigma}_n\|_1 \le \alpha
\end{aligned}
\quad , \quad (21)
$$

where $\boldsymbol{\sigma}$ denotes $(\boldsymbol{\sigma}_n)_{n=1}^{N}$. Although the proposed regularization model (21) is a relatively difficult convex optimization problem due to the discontinuous function ϕ, we can obtain a globally optimal solution of (21) by applying the proximal splitting techniques [47], [48], [49], [50], [51] with the interpretation of ϕ in (16) as a *perspective function* [52], [53], [54]. A concrete algorithm based on the alternating direction method of multipliers (ADMM) and its derivation are provided in Appendix A.

B. LEVERAGING LATENT VARIABLE FOR PSD ESTIMATION

We demonstrate that the solution for the latent variable $\boldsymbol{\sigma}$ of the proposed model (21) is in fact suitable for the PSD estimation. Let $\hat{\mathbf{u}}$ and $\hat{\boldsymbol{\sigma}}$ be the solutions of (21) respectively for the variables $\mathbf{u}$ and $\boldsymbol{\sigma}$. While it is possible to compute the periodogram as in (14) for $\hat{\mathbf{u}}$, $\hat{\boldsymbol{\sigma}}$ is more suitable for the estimation of smooth PSDs. To confirm this, we show that $\hat{\boldsymbol{\sigma}}$ corresponds to the square root of smoothed and averaged periodogram as follows.

i) We begin by considering the case of $\alpha \to \infty$, which is not of our interest but easy to analyze. In this case, since $\boldsymbol{\sigma}$ minimizes

$$
\lambda\sum_{n=1}^{N}\sum_{k=1}^{L}\phi\left((u_{j,n}[k])_{j=1}^{J}, \sigma_n[k]\right) \quad (22)
$$

in (21), the solutions $\hat{\mathbf{u}}$ and $\hat{\boldsymbol{\sigma}}$ satisfy the relationship

$$
\hat{\sigma}_n[k] = \sqrt{\frac{1}{J}\sum_{j=1}^{J}|\hat{u}_{j,n}[k]|^2} \quad (23)
$$

for each $k = 1, \ldots, L$ and $n = 1, \ldots, N$, which can be shown based on [38, Lemma 1]. The relation (23) means that $\hat{\boldsymbol{\sigma}}$ is the square root of the averaged periodogram in (14) computed with $\hat{\mathbf{u}}$ when $\alpha \to \infty$.

ii) Next, we consider the case of our interest, where α is set to a finite value. In this case, the constraint

$$
\sum_{n=1}^{N}\|\mathbf{D}\boldsymbol{\sigma}_n\|_1 \le \alpha
$$

in (21) penalizes the smoothness of $\boldsymbol{\sigma}_n$ ($n = 1\ldots, N$) since $\mathbf{D}$ is the difference operator. Meanwhile, the other part (22) of the proposed model forces $\boldsymbol{\sigma}$ to be the averaged periodogram in (23). Thus,

by the combination of these terms, roughly speaking, $\hat{\boldsymbol{\sigma}}$ is smoothed around the square root of the averaged periodogram in (23).

In addition to being smooth, $\hat{\boldsymbol{\sigma}}$ is block-sparse because $\hat{\mathbf{u}}_{j,n}$ ($j = 1, \ldots, J$) are regularized to have a common block-sparse support. Since the PSDs are smooth and block-sparse for the PAWR, the square of components of $\hat{\boldsymbol{\sigma}}$, i.e.,

$$
\hat{S}_n(f_k) = (\hat{\sigma}_n[k])^2 \quad (k = 1, \ldots, L), \quad (24)
$$

is expected to be a better estimate of the PSDs than the periodogram computed with $\hat{\mathbf{u}}$.

Intuitively, the proposed model is expected to accurately estimate smooth and block-sparse PSDs by the following mechanism. Since $\phi(\mathbf{v}, \sigma)$ is basically $\frac{\|\mathbf{v}\|^2}{2\sigma} + \frac{J}{2}\sigma$ (see (16)), roughly speaking, we can consider that the part (22) acts as

$$
\lambda\sum_{n=1}^{N}\sum_{k=1}^{L}\sum_{j=1}^{J}\left(\frac{|u_{j,n}[k]|^2}{2\sigma_n[k]} + \frac{\sigma_n[k]}{2}\right).
$$

Since $\sigma_n[k]$ estimates the square root of the PSDs,

$$
\frac{|u_{j,n}[k]|^2}{2\sigma_n[k]}
$$

is expected to be an effective regularization for $\mathbf{u}$ because the expectation of the squared magnitude of the realizations $\bar{\mathbf{u}}$ is close to the PSDs (see (13) and Fig. 1). Refining $\mathbf{u}$ leads to an equally refined $\boldsymbol{\sigma}$ because $\boldsymbol{\sigma}$ is smoothed around the value in (23). Thanks to these interactions, the proposed model is expected to effectively estimate the frequency components and the PSDs simultaneously.

While $\|\mathbf{D}\boldsymbol{\sigma}_n\|_1$ with the first difference operator $\mathbf{D}$ in (18) is a good choice for controlling the block structure, more advanced smoothness priors can be incorporated to further improve the estimation accuracy of the PSDs. Thanks to the nonnegativity of $\boldsymbol{\sigma}_n$, many convex smoothness penalties designed for real-valued signals, such as the high-order total variation [39], [40] that uses $\mathbf{D}^r$ ($r \ge 2$) instead of $\mathbf{D}$ and the total generalized variation [41], can be directly applied to $\boldsymbol{\sigma}_n$. For instance, the proposed model with the high-order total variation

$$
\begin{aligned}
\underset{\mathbf{u}\in\mathbb{C}^{JNL},\boldsymbol{\sigma}\in\mathbb{R}_+^{NL}}{\text{minimize}} \quad & \frac{1}{2}\sum_{j=1}^{J}\left\|\mathbf{y}_j - \sum_{n=1}^{N}\mathbf{A}_n\mathbf{G}\mathbf{u}_{j,n}\right\|^2 \\
& + \lambda\sum_{n=1}^{N}\sum_{k=1}^{L}\phi\left((u_{j,n}[k])_{j=1}^{J}, \sigma_n[k]\right) \\
\text{subject to} \quad & \sum_{n=1}^{N}\|\mathbf{D}^r\boldsymbol{\sigma}_n\|_1 \le \alpha
\end{aligned}
\quad (25)
$$

can be solved similarly to the case of (21) by the ADMM-based algorithm shown in Appendix A. In contrast, when these penalties are applied to, e.g., the magnitude of $\mathbf{u}$, their convexity is lost (see, e.g., [31]), which implies that a globally optimal solution is difficult to obtain. Note that the application of these penalties to $\mathbf{u}$, which is complex-valued,

is not a suitable strategy because the magnitude of $\mathbf{u}$ is smooth but the phase of $\mathbf{u}$ is not smooth in most applications.

IV. SIMULATION RESULTS

To demonstrate the effectiveness of the proposed approach, we conduct numerical simulations on the PSD estimation for the PAWR shown in Example 1. Essentially, we follow the simulation setting in [12] and [16]. Uniform elevation angles $\theta_1, \ldots, \theta_N$, ranging between $-15°$ and $30°$ degrees with $N = 110$, are selected. We synthesize the (discrete-time) PSD $S_n^\star(f)$ by

$$S_n^\star(f) = \frac{1}{T} \sum_{m=-\infty}^{\infty} G_n^\star\left(\frac{f-m}{T}\right)$$

for each $n = 1, \ldots, N$, where T is the pulse repetition time, and $G_n^\star(f)$ is a continuous-time Gaussian-shaped PSD

$$G_n^\star(f) = \frac{P_n}{\sqrt{2\pi}\,\varsigma_n} e^{-\frac{(f-\mu_n)^2}{2\varsigma_n^2}},$$

which is an appropriate model when, e.g., the atmospheric turbulence is dominant [4]. The power P_n is set from the actual reflection intensity measured by the PAWR at Osaka University on March 30, 2014. We define the mean Doppler frequency μ_n by the certain sine curve used in [16]. The Doppler frequency width ς_n is converted from the Doppler velocity width, which are chosen randomly from the uniform distribution of $[1, 3]$ [m/s]. Note that this setting is more realistic than that presented in [16] where the Doppler velocity width is merely fixed to 2 [m/s] at every elevation angle. We set $X_n^\star[\ell]$ in (1) to the Gaussian process that has the specified PSD $S_n^\star(f)$, and then generate its realizations $\bar{x}_{j,n}[\ell]$ $(\ell = 1, \ldots, L)$ based on the probability distribution of $X_n^\star[\ell]$ $(\ell = 1, \ldots, L)$, which is computed in the way presented in [4] and [16]. The observation vector $\mathbf{y}_j$ is given by (6), where $\boldsymbol{\varepsilon}_j$ is generated as the white Gaussian noise of the standard deviation $\sqrt{2.5}$. The parameters of the PAWR are set as follows: $M = 128$, $\lambda_{\mathrm{cw}} = 31.8$ [mm], $\Delta = 16.5$ [mm], and $T = 0.4$ [ms]. For the synthesis matrix $\mathbf{G}$ in the observation model (9) in terms of the frequency components, we test both $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$ and $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$ respectively for the standard DFT in Example 2 and the windowed DFT in Example 3. We use the hamming window for the window function $\mathbf{w}$ in Example 3, which is normalized to $\|\mathbf{w}\| = \sqrt{L}$, i.e., to the norm of the rectangular window $(1, 1, \ldots, 1)^\top$ [45].

We compare the proposed approach that jointly estimates the frequency components and the PSDs with the existing approach that first estimates the frequency components and subsequently the PSDs. The proposed approach computes the estimate $\hat{S}_n(f_k)$ of the PSD by (24) with the solution $\hat{\boldsymbol{\sigma}}$ of the proposed model (25) for the variable $\boldsymbol{\sigma}$. For the frequency component estimators used in the existing approach, we employ the mixed ℓ_2/ℓ_1 regularization model using fixed small-size overlapping blocks [16], which is state-of-the-art for the PAWR, and the ℓ_1 regularization

model as a non-structured sparse model. For the mixed ℓ_2/ℓ_1 regularization model, we adopt the formulation based on latent group lasso [29], [30], which selects relevant blocks from the pre-defined overlapping blocks in the mixed ℓ_2/ℓ_1 norm, because its estimation accuracy is (slightly) better than that of the simple overlapping blocks-based formulation in [16]. While the above nonlinear methods outperform the linear methods for the frequency component estimation in [16], we also include the minimum mean square error (MMSE) beamformer [12], which performs best among the linear methods in [16], for comparison. Since the MMSE beamformer is a time-domain method that estimates $\bar{\mathbf{x}}_{j,n}$ from (4), we compute the frequency components from the estimate of $\bar{\mathbf{x}}_{j,n}$ by using the DFT or the windowed DFT shown in Examples 2 and 3 respectively. From the estimated frequency components $\hat{\mathbf{u}}$, the existing approach constructs the estimate of the PSD as the averaged periodogram

$$\hat{S}_n^{(\mathrm{AP})}(f_k) = \frac{1}{J}\sum_{j=1}^{J} |\hat{u}_{j,n}[k]|^2 \quad (k = 1, \ldots, L). \tag{26}$$

We also test the post-smoothing for the existing approach. Specifically, we employ the standard smoothing technique, i.e., *Daniell method* [1], [2]:

$$\hat{S}_n^{(\mathrm{SAP})}(f_k) = \frac{1}{2R+1}\sum_{k'=k-R}^{k+R} \hat{S}_n^{(\mathrm{AP})}(f_{k'}) \quad (k = 1, \ldots, L),$$

where $2R$ neighbor frequency bins[4] are used for the smoothing.

Table 1 shows the normalized mean absolute error (NMAE)

$$\frac{\sum_{n=1}^{N} \sum_{k=1}^{L} \left| S_n^\star(f_k) - \hat{S}_n(f_k) \right|}{\sum_{n=1}^{N} \sum_{k=1}^{L} S_n^\star(f_k)},$$

which is averaged over 100 independent simulations. The tuning parameters of the methods are adjusted in the way that the best NMAE is obtained for each method and setting. Table 2 shows specific settings of the tuning parameters: λ for the importance of the (block-)sparsity, α for the importance of the smoothness in the proposed model, the block-size B for the mixed ℓ_2/ℓ_1 regularization model, R for the post-smoothing, and the standard deviation ς_ε of the noise for the MMSE beamformer. Note that the MMSE beamformer uses the actual standard deviation $\sqrt{2.5}$ in the experiments to achieve the best accuracy. We simply set $r = 2$ in the proposed model, although the tuning of r could improve the estimation accuracy. While the case of $J = 1$ is of particular interest in weather radar applications to keep the frequency resolution of the PSDs (see Remark 1), we also show the results when J is increased to 2, so as to elaborate on the effect of J. In Table 1, the proposed model is shown to achieve the best estimation accuracy for all the settings. The post-smoothing is found to improve the estimation accuracies of

---

[4]When $k' \notin \{1, \ldots, L\}$, we instead use $k' - L\lfloor (k'-1)/L\rfloor$ because the anti-aliasing filter is not employed (see also Example 1).

TABLE 1. A comparison of the methods in terms of the NMAE of the PSDs, where the result is averaged over 100 independent simulations. Values shown in parenthesis for the existing methods are the NMAEs with the post-smoothing.

Settings	MMSE BF	ℓ_1 reg.	Mixed ℓ_2/ℓ_1 reg.	Proposed
$L = 32$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$ (standard DFT in Example 2)	1.1775 (0.9217)	0.8771 (0.6802)	0.7725 (0.6447)	**0.6278**
$L = 32$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$ (windowed DFT in Example 3)	1.1055 (0.9494)	0.8236 (0.7239)	0.7644 (0.7004)	**0.6780**
$L = 128$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$ (standard DFT in Example 2)	1.1808 (0.7396)	0.8161 (0.4874)	0.6843 (0.4906)	**0.4693**
$L = 128$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$ (windowed DFT in Example 3)	1.1767 (0.7975)	0.7899 (0.5428)	0.6970 (0.5244)	**0.5097**
$L = 32$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$ (standard DFT in Example 2)	0.9345 (0.7767)	0.6917 (0.5582)	0.6201 (0.5590)	**0.5580**
$L = 32$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$ (windowed DFT in Example 3)	0.9292 (0.8318)	0.6596 (0.5865)	0.6169 (0.5790)	**0.5718**
$L = 128$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$ (standard DFT in Example 2)	0.9815 (0.7057)	0.6676 (0.4345)	0.5700 (0.4623)	**0.4335**
$L = 128$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$ (windowed DFT in Example 3)	0.9763 (0.7418)	0.6435 (0.4684)	0.5723 (0.4771)	**0.4548**

TABLE 2. Specific settings of the tuning parameters, where ς_ε is the standard deviation of the noise, R is the parameter for the post-smoothing, λ is the weight for the (block-)sparsity, B is the block-size for the mixed ℓ_2/ℓ_1 regularization model, and α is the smoothing parameter for the proposed model.

Settings	MMSE BF	ℓ_1 reg.	Mixed ℓ_2/ℓ_1 reg.	Proposed
$L = 32$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 1$	$\lambda = 0.003\frac{M}{N}$, $R = 1$	$\lambda = 0.03\frac{M}{N}$, $B = 7$, $R = 1$	$\lambda = 0.02\frac{M}{N}$, $\alpha = 60N$
$L = 32$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 1$	$\lambda = 0.003\frac{M}{N}$, $R = 1$	$\lambda = 0.02\frac{M}{N}$, $B = 9$, $R = 1$	$\lambda = 0.01\frac{M}{N}$, $\alpha = 60N$
$L = 128$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 5$	$\lambda = 0.005\frac{M}{N}$, $R = 4$	$\lambda = 0.05\frac{M}{N}$, $B = 36$, $R = 4$	$\lambda = 0.05\frac{M}{N}$, $\alpha = 20N$
$L = 128$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 5$	$\lambda = 0.003\frac{M}{N}$, $R = 5$	$\lambda = 0.02\frac{M}{N}$, $B = 36$, $R = 4$	$\lambda = 0.03\frac{M}{N}$, $\alpha = 20N$
$L = 32$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 1$	$\lambda = 0.01\frac{M}{N}$, $R = 1$	$\lambda = 0.02\frac{M}{N}$, $B = 7$, $R = 1$	$\lambda = 0.03\frac{M}{N}$, $\alpha = 60N$
$L = 32$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 1$	$\lambda = 0.006\frac{M}{N}$, $R = 1$	$\lambda = 0.01\frac{M}{N}$, $B = 6$, $R = 1$	$\lambda = 0.02\frac{M}{N}$, $\alpha = 60N$
$L = 128$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 4$	$\lambda = 0.005\frac{M}{N}$, $R = 4$	$\lambda = 0.02\frac{M}{N}$, $B = 36$, $R = 3$	$\lambda = 0.04\frac{M}{N}$, $\alpha = 20N$
$L = 128$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$	$\varsigma_\varepsilon = \sqrt{2.5}$, $R = 4$	$\lambda = 0.004\frac{M}{N}$, $R = 4$	$\lambda = 0.02\frac{M}{N}$, $B = 36$, $R = 3$	$\lambda = 0.03\frac{M}{N}$, $\alpha = 20N$

the existing models; however, their accuracies remain inferior to those of the proposed model. While the NMSEs of the proposed model and the the existing sparse estimation models combined with the post-smoothing are close for several cases when J is increased to 2, the proposed model yields moderate improvements against them for the cases of $J = 1$. Since the original frequency resolution is preserved when $J = 1$ (see Remark 1), the proposed model has an advantage that it estimates PSDs accurately without sacrificing the frequency resolution. Note that the proposed model also has an advantage that it has fewer tuning parameters than the the mixed ℓ_2/ℓ_1 regularization model with the post-smoothing (see Table 2).

We show the ground-truth and the estimates of the PSDs for examples of simulations: Fig. 3 for $L = 32$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$, Fig. 4 for $L = 32$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$, Fig. 5 for $L = 128$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$, Fig. 6 for $L = 128$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$, Fig. 7 for $L = 32$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$, Fig. 8 for $L = 32$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$, Fig. 9 for $L = 128$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$, Fig. 10 for $L = 128$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$. It can be seen from Figs. 3-10(b)(d)(f) that the estimates of the existing models exhibit erratic oscillation as they do not exploit the smoothness of the PSDs. In Figs. 7-10(b)(d)(f) where J is increased to 2, the erratic oscillation is slightly reduced but still clearly visible, suggesting the limitation of the ensemble average (26) when J is small (see Remark 1 for the reason why J is set to be small in weather radar applications). The post-smoothing is found to reduce the erratic oscillation to

a certain extent, as shown in Figs. 3-10(c)(e)(g). However, the sparsity of the estimate is impaired, i.e., the number of entries of large magnitude that are not present in the ground-truth increases, because the sparsity is not considered in the post-smoothing step. In contrast, in Figs. 3-10(h), the proposed approach obtains the estimates that have both sparsity and smoothness. While the estimates of the mixed ℓ_2/ℓ_1 regularization model after the post-smoothing seem similar to the ground-truth at a glance of Figs. 3-10(g), erroneous spread of the nonzero components are more clearly seen from enlarged views shown in Figs. 11 and 12. From Figs. 11 and 12, we also see that the proposed model estimates the area of nonzero components more accurately than the mixed ℓ_2/ℓ_1 regularization model with the post-smoothing. Since the area of the nonzero components is related to the existence of the corresponding wind velocity components (see Example 1), this is a significant advantage of the proposed approach for weather radar applications.

We also see that the erratic oscillation is still visible for the estimates of the MMSE beamformer and the ℓ_1 regularization model after the post-smoothing in Figs. 3-10(c)(e), suggesting the limitation of the post-smoothing. In particular, while the objective accuracies of the ℓ_1 regularization model after the post-smoothing are close to those of the proposed model when J is increased to 2 and the standard DFT is used, the erratic oscillation is not eliminated as seen in Figs. 7(e) and 9(e). Compared to the ℓ_1 regularization model, the erratic oscillation is considerably reduced in the estimates of the
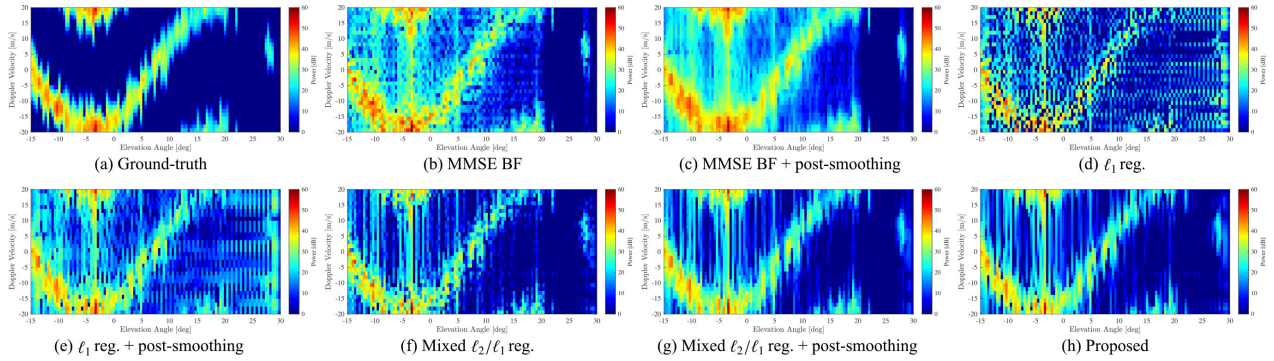
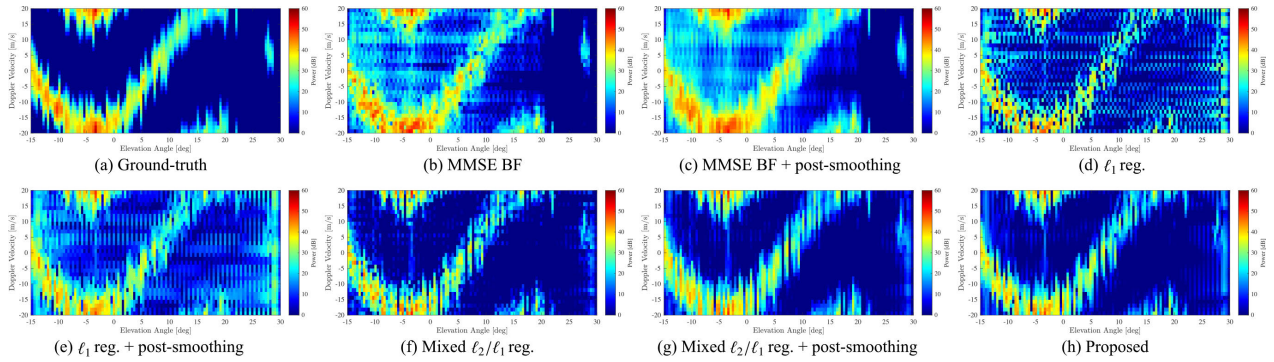**FIGURE 3.** Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 32$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$.

FIGURE 4. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 32$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$.
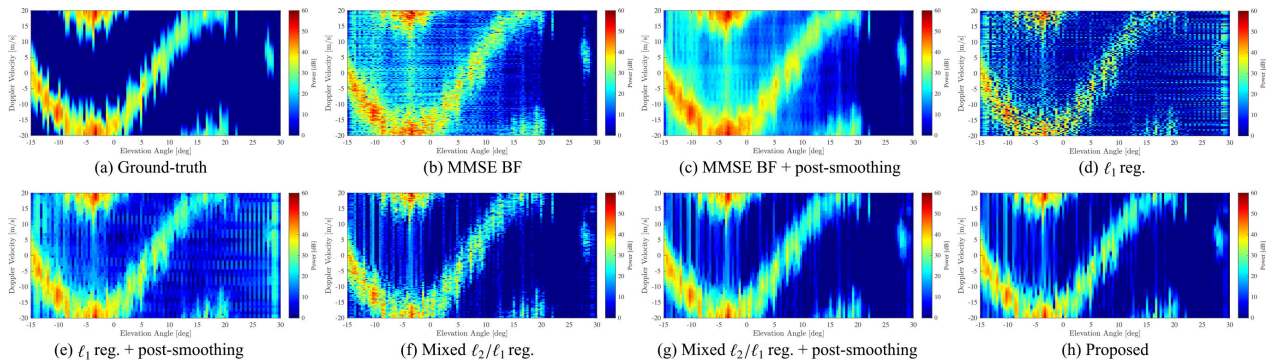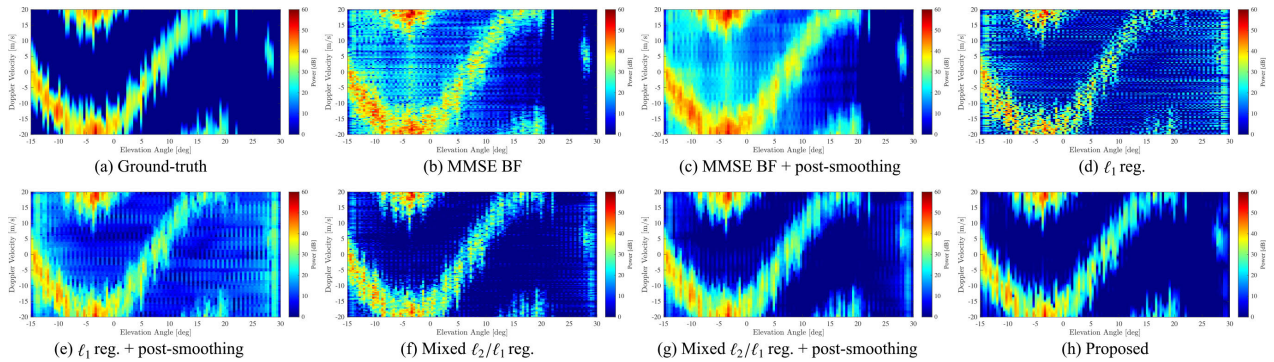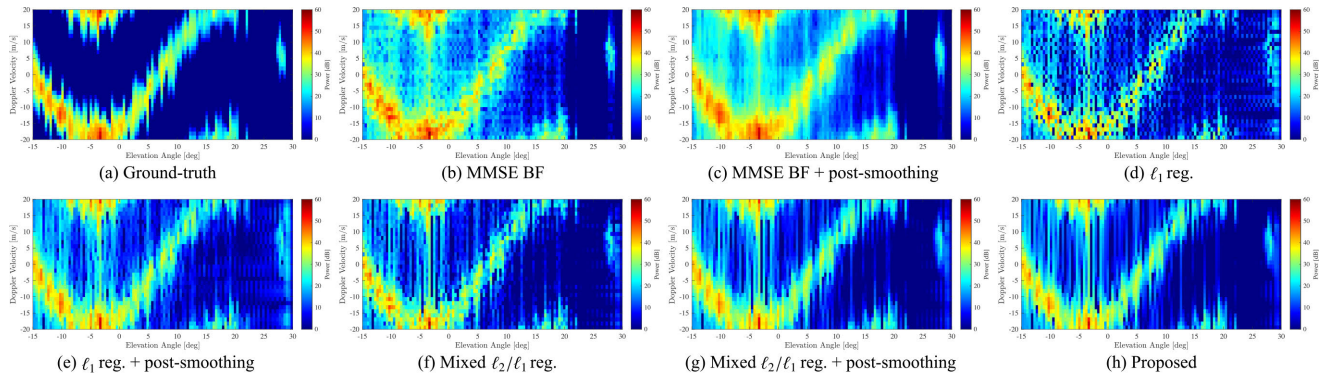

FIGURE 5. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 128$, $J = 1$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$.
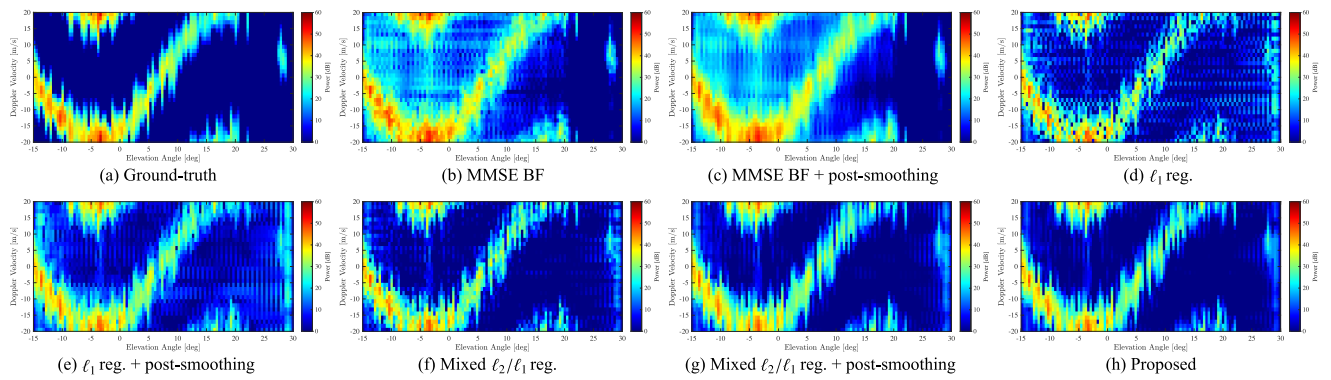

FIGURE 6. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 128$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$.
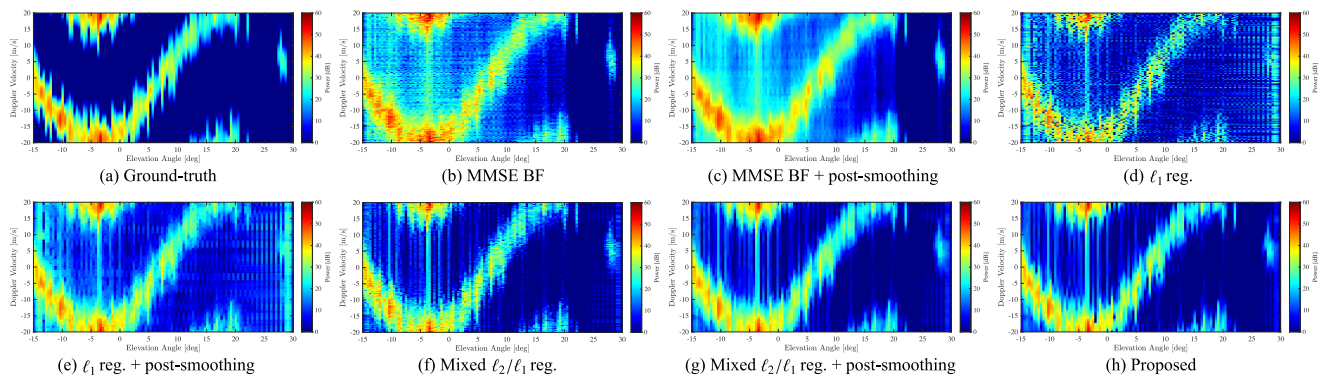
FIGURE 7. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 32$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$.

FIGURE 8. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 32$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$.

FIGURE 9. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 128$, $J = 2$, $\mathbf{G} = \mathbf{F}^{\mathrm{H}}$.
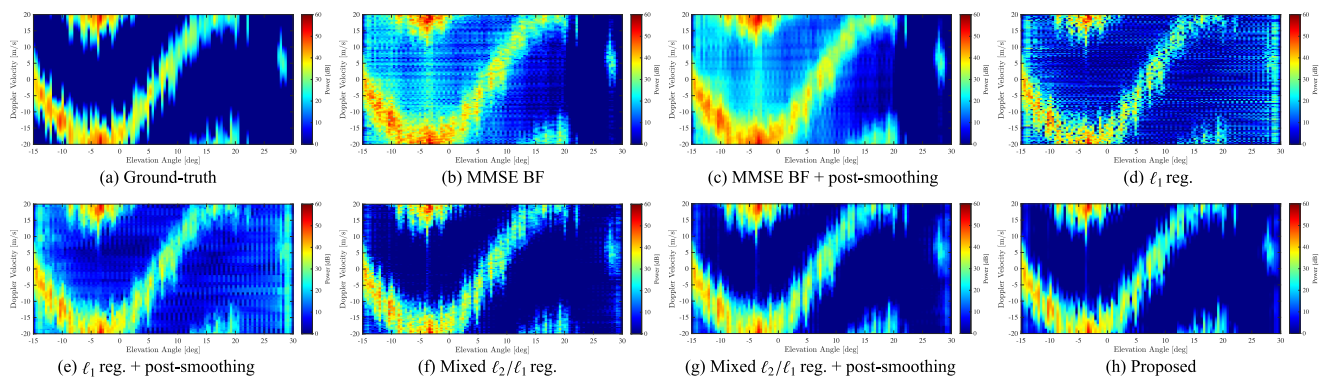

FIGURE 10. Ground-truth of the PSDs and their estimates for a simulation of the PAWR using the following settings: $L = 128$, $J = 2$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^{\mathrm{H}}$.
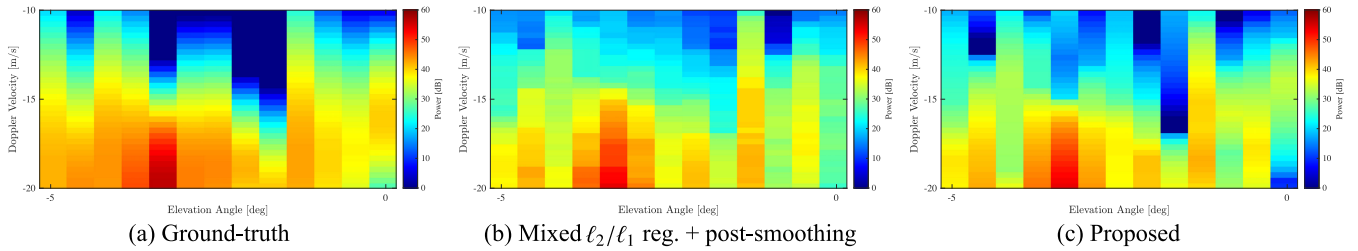
(a) Ground-truth   (b) Mixed ℓ_2/ℓ_1 reg. + post-smoothing   (c) Proposed

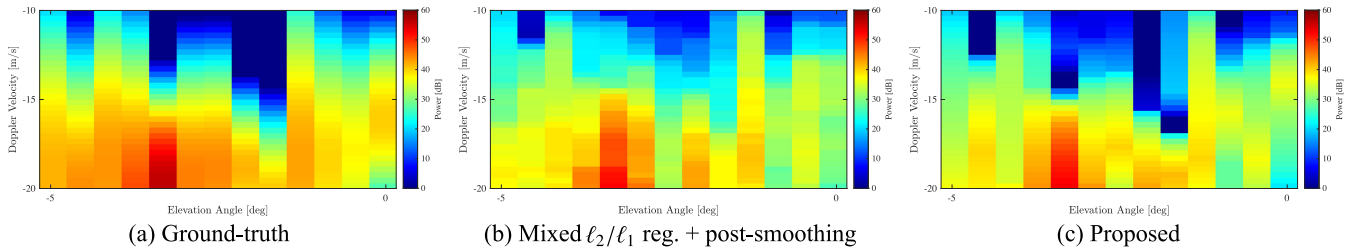**FIGURE 11.** An enlarged view of Ground-truth of the PSDs and their estimates for the following settings: $L = 128$, $J = 1$, $\mathbf{G} = \mathbf{F}^H$.

(a) Ground-truth   (b) Mixed ℓ_2/ℓ_1 reg. + post-smoothing   (c) Proposed

FIGURE 12. An enlarged view of Ground-truth of the PSDs and their estimates for the following settings: $L = 128$, $J = 1$, $\mathbf{G} = \mathbf{W}^{-1}\mathbf{F}^H$.

mixed ℓ_2/ℓ_1 regularization model after the post-smoothing, as shown in Figs. 3-10(g). This could be attributed to the property that the block-sparse model does not necessarily promote the smoothness but force the components in the same block to zeros together, which would be beneficial to promote the smoothness in the post-processing step. Although the estimates of the mixed ℓ_2/ℓ_1 regularization model after the post-smoothing seem smooth enough, from enlarged views in Figs. 11 and 12, we see that they have slight unnatural fluctuation, and the estimates of the proposed model are smoother even compared with them.

The line-like artifacts in the estimates shown in Figs. 3, 5, 7, and 9 are more or less reduced in the estimates shown in Figs. 4, 6, 8, and 10 thanks to the window function that reduces the heights of the sidelobes. Although the estimates obtained with the windowed DFT are visually closer to the ground-truth than those obtained with the standard DFT, the objective accuracy shown in Table 1 is not always improved perhaps because the window function increases the width of the mainlobe. Since the line-like artifacts are caused due to the sidelobes of the window function with finite L, the line-like artifacts are more reduced when $L = 128$. In particular, the line-like artifacts are almost completely eliminated in the proposed estimates when $L = 128$, as shown in Figs. 6(h) and 10(h).

V. CONCLUSION

We presented a convex optimization model for the estimation of sparse and smooth PSDs of complex-valued random processes from noisy mixtures of realizations. While the PSDs are related to the expectation of the magnitude of the frequency components of the realizations, it has been difficult to exploit the smoothness of the PSDs as naive penalties for the difference of the magnitude of the frequency components induce hard nonconvex optimization problems. To resolve this difficulty, we designed the proposed model

that jointly estimates complex-valued frequency components and nonnegative PSDs. More precisely, we first applied the optimally structured block-sparse model of [38] for the frequency component estimation. Then, to estimate the PSDs, we newly leveraged the latent variable of the model, which was originally introduced to optimize the block structure. Namely, we demonstrate that the latent variable is in fact related to the square root of the PSDs, enabling us to exploit the smoothness of the PSDs via convex optimization by penalizing the difference of the nonnegative latent variable. Moreover, to further enhance the smoothness of the PSDs of complex-valued random processes, the proposed framework can readily incorporate many smoothness priors designed for real-valued signals. Numerical experiments on the PAWR showed that the proposed approach achieved better objective accuracy and yielded visually better estimates compared with the existing sparse estimation models, even when they are combined with the post-smoothing.

APPENDIX A
SOLVER FOR PROPOSED REGULARIZATION MODEL

The proposed regularization model (25) can be solved by using the proximal splitting techniques [47], [48], [49], [50], [51] with the closed-form computation of the proximity operator of ϕ. As a concrete example, using the ADMM [48], [49], we provide an iterative solver that is guaranteed to converge to an optimal solution of (25). The ADMM solves the following convex optimization problem

$$\underset{v \in \mathcal{V}, w \in \mathcal{W}}{\text{minimize}} \, F(v) + G(w) \text{ subject to } w = \mathcal{L}v \quad (27)$$

by the iterations

$$
\left|
\begin{array}{l}
v^{(i+1)} \in \underset{v \in \mathcal{V}}{\arg\min} \left[\gamma F(v) + \dfrac{1}{2} \|w^{(i)} - z^{(i)} - \mathcal{L}v\|^2 \right] \\
w^{(i+1)} = \text{prox}_{\gamma G}(\mathcal{L}v^{(i+1)} + z^{(i)}) \\
z^{(i+1)} = z^{(i)} + \mathcal{L}v^{(i+1)} - w^{(i+1)},
\end{array}
\right.
$$

where we suppose that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional Hilbert spaces, F and G are proper lower semicontinuous convex functions, $\mathcal{L}$ is a linear operator, $\mathrm{prox}_{\gamma G}(w) := \arg\min_{\omega \in \mathcal{W}} \left[\gamma G(\omega) + \frac{1}{2} \|w - \omega\|^2 \right]$ is the proximity operator of γG, and $\gamma > 0$.

To apply the ADMM, we rewrite (25) as

$$
\begin{aligned}
\underset{\mathbf{u}, \boldsymbol{\sigma}, \mathbf{x}, \tilde{\mathbf{u}}, \tilde{\boldsymbol{\sigma}}, \boldsymbol{\eta}}{\text{minimize}} \; & F(\mathbf{u}, \boldsymbol{\sigma}) + G(\mathbf{x}, \tilde{\mathbf{u}}, \tilde{\boldsymbol{\sigma}}, \boldsymbol{\eta}) \\
\text{subject to} \; & \mathbf{x}_{j,n} = \mathbf{G}\mathbf{u}_{j,n} \quad (\forall j, n) \\
& \tilde{\mathbf{u}}_{j,n} = \mathbf{u}_{j,n} \quad (\forall j, n) \\
& \tilde{\boldsymbol{\sigma}}_n = \boldsymbol{\sigma}_n \quad (\forall n) \\
& \boldsymbol{\eta}_n = \mathbf{D}^r \boldsymbol{\sigma}_n \quad (\forall n)
\end{aligned}
\right\}, \tag{28}
$$

where we let

$$
F(\mathbf{u}, \boldsymbol{\sigma}) := 0,
$$

$$
\begin{aligned}
G(\mathbf{x}, \tilde{\mathbf{u}}, \tilde{\boldsymbol{\sigma}}, \boldsymbol{\eta}) := & \frac{1}{2} \sum_{j=1}^{J} \left\| \mathbf{y}_j - \sum_{n=1}^{N} \mathbf{A}_n \mathbf{x}_{j,n} \right\|^2 \\
& + \lambda \sum_{n=1}^{N} \sum_{k=1}^{L} \phi\left((\tilde{u}_{j,n}[k])_{j=1}^{J}, \tilde{\sigma}_n[k] \right) + \iota_{B_1^\alpha}(\boldsymbol{\eta}),
\end{aligned}
$$

and $\iota_{B_1^\alpha}(\boldsymbol{\eta})$ is the indicator function of the ℓ_1 ball, i.e.,

$$
\iota_{B_1^\alpha}(\boldsymbol{\eta}) := \begin{cases} 0, & \text{if } \sum_{n=1}^{N} \|\boldsymbol{\eta}_n\|_1 \leq \alpha; \\ \infty, & \text{otherwise.} \end{cases}
$$

Since the constraint of (28) can be expressed in the form of (27), we can therefore apply the ADMM to (28), and obtain the iterative algorithm shown in Algorithm 1. For our formulation, since the minimizer of the first step of the ADMM is unique, the convergence to an optimal solution of (28) follows from [51]. From the equivalence between (25) and (28), $(\mathbf{u}^{(i)})_{i=1}^{\infty}$ and $(\boldsymbol{\sigma}^{(i)})_{i=1}^{\infty}$ generated by Algorithm 1 converges to the solution of (25) for the variables $\mathbf{u}$ and $\boldsymbol{\sigma}$ respectively.

The operators in Algorithm 1 can be computed as follows. Expressing $\phi(\mathbf{v}, \sigma)$ as the sum of the perspective function [52], [53], [54] of $\frac{\|\mathbf{v}\|^2}{2}$ and the linear function $\frac{J}{2}\sigma$, based on [54, Example 2.4], we can compute $\mathrm{prox}_{\kappa\phi}$ for $\kappa = \gamma\lambda$ by

$$
\mathrm{prox}_{\kappa\phi}(\mathbf{v}, \sigma)
$$
$$
= \begin{cases} (\mathbf{0}, 0), & \text{if } 2\kappa\sigma + \|\mathbf{v}\|^2 \leq J\kappa^2; \\ \left(\mathbf{0}, \sigma - \frac{\kappa J}{2}\right), & \text{if } \mathbf{v} = \mathbf{0} \text{ and } 2\sigma > J\kappa; \\ \left(\mathbf{v} - \kappa s \frac{\mathbf{v}}{\|\mathbf{v}\|}, \sigma + \kappa \frac{s^2 - J}{2}\right), & \text{otherwise,} \end{cases}
$$

where $s > 0$ is the unique positive root of

$$
s^3 + \left(\frac{2}{\kappa}\sigma + 2 - J\right)s - \frac{2}{\kappa}\|\mathbf{v}\| = 0,
$$

and can be explicitly given based on Cardano's formula as follows [38] and [55]. Let $p = \frac{2}{\kappa}\sigma + 2 - J$ and $D = -\frac{\|\mathbf{v}\|^2}{\kappa^2} - \frac{p^3}{27}$. Then,

$$
s = \begin{cases} \sqrt[3]{\frac{\|\mathbf{v}\|}{\kappa} + \sqrt{-D}} + \sqrt[3]{\frac{\|\mathbf{v}\|}{\kappa} - \sqrt{-D}}, & \text{if } D < 0; \\ 2\sqrt[3]{\frac{\|\mathbf{v}\|}{\kappa}}, & \text{if } D = 0; \\ 2\sqrt{-\frac{p}{3}} \cos\left(\frac{\arctan(\kappa\sqrt{D}/\|\mathbf{v}\|)}{3}\right), & \text{if } D > 0, \end{cases}
$$

where $\sqrt[3]{\cdot}$ denotes the real cubic root. The ℓ_1 ball projection $P_{B_1^\alpha}$, which is the proximity operator of $\iota_{B_1^\alpha}(\boldsymbol{\eta})$, can be computed in $\mathcal{O}(NL)$ expected complexity, e.g., by the algorithm of [56].

A. IMPLEMENTATION FOR WEATHER RADAR APPLICATIONS

The matrix inversions in Algorithm 1 can be efficiently computed for application to weather radars as follows.

a) To efficiently compute the inversion of $(\mathbf{I}_L + \mathbf{G}^H\mathbf{G})$, we use the property

$$
\mathbf{G}\mathbf{G}^H = \mathrm{diag}(\boldsymbol{v}) \in \mathbb{R}_{++}^L,
$$

which holds for, e.g., the DFT in Example 2 and the windowed DFT in Example 3. More precisely, from this property, we have

$$
\begin{aligned}
(\mathbf{I}_L + \mathbf{G}^H\mathbf{G})^{-1} &= \mathbf{I}_L - \mathbf{G}^H(\mathbf{I}_L + \mathbf{G}\mathbf{G}^H)^{-1}\mathbf{G} \\
&= \mathbf{I}_L - \mathbf{G}^H(\mathbf{I}_L + \mathrm{diag}(\boldsymbol{v}))^{-1}\mathbf{G} \\
&= \mathbf{I}_L - \mathbf{G}^H \mathrm{diag}\left(\left(\frac{1}{1 + v_\ell}\right)_{\ell=1}^{L}\right)\mathbf{G},
\end{aligned}
$$

where the first equality follows from the Sherman-Morrison-Woodbury matrix inversion lemma [57]. Note that the multiplications of $\mathbf{G}$ and $\mathbf{G}^H$ can be computed in $\mathcal{O}(L \log L)$ by the fast Fourier transform (FFT) for Examples 2 and 3.

b) Since $\mathbf{D}$ in (18) is a circulant matrix, the inversion of $(\mathbf{I}_L + (\mathbf{D}^r)^\top \mathbf{D}^r)$ can be computed in $\mathcal{O}(L \log L)$ with the FFT [57].

c) For the PAWR shown in Example 1, the inversion of $\mathbf{I}_{NL} + \mathbf{A}^H\mathbf{A} \in \mathbb{C}^{NL \times NL}$ can be computed in $\mathcal{O}(N^3)$, independently of the value of L, because it can be translated into a block-diagonal matrix after some permutations. We show this explicitly in another way. Notice that $(\mathbf{x}_{j,n}^{(i+1)})_{n=1}^{N}$ in Algorithm 1 is the solution of

$$
\begin{aligned}
\underset{(\mathbf{x}_{j,n})_{n=1}^{N}}{\text{minimize}} \; & \frac{\gamma}{2} \left\| \mathbf{y}_j - \sum_{n=1}^{N} \mathbf{A}_n \mathbf{x}_{j,n} \right\|^2 \\
& + \frac{1}{2} \sum_{n=1}^{N} \|\mathbf{G}\mathbf{u}_{j,n}^{(i+1)} + \mathbf{q}_{j,n}^{(i)} - \mathbf{x}_{j,n}\|^2
\end{aligned}
\right\}. \tag{29}
$$

Algorithm 1 Iterative Solver for the Proposed Regularization Model (25)

Input: $\gamma > 0$, $\mathbf{A} := [\mathbf{A}_1, \ldots, \mathbf{A}_N] \in \mathbb{C}^{d \times NL}$, $\mathbf{x}_{j,n}^{(0)} \in \mathbb{C}^L$, $\tilde{\mathbf{u}}_{j,n}^{(0)} \in \mathbb{C}^L$, $\tilde{\boldsymbol{\sigma}}_n^{(0)} \in \mathbb{R}_+^L$, $\boldsymbol{\eta}_n^{(0)} \in \mathbb{R}^L$, $\mathbf{q}_{j,n}^{(0)} \in \mathbb{C}^L$, $\mathbf{r}_{j,n}^{(0)} \in \mathbb{C}^L$,
$\boldsymbol{\tau}_n^{(0)} \in \mathbb{R}^L$, $\boldsymbol{\zeta}_n^{(0)} \in \mathbb{R}^L$ ($j \in \{1, 2, \ldots, J\}$ and $n \in \{1, 2, \ldots, N\}$).

for $i = 0, 1, 2, \ldots$ **do**
  for $n \in \{1, 2, \ldots, N\}$ **do**
    for $j \in \{1, 2, \ldots, J\}$ **do**
      $\mathbf{u}_{j,n}^{(i+1)} = (\mathbf{I}_L + \mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1} \left(\mathbf{G}^{\mathrm{H}}(\mathbf{x}_{j,n}^{(i)} - \mathbf{q}_{j,n}^{(i)}) + \tilde{\mathbf{u}}_{j,n}^{(i)} - \mathbf{r}_{j,n}^{(i)} \right)$
    $\boldsymbol{\sigma}_n^{(i+1)} = (\mathbf{I}_L + (\mathbf{D}^r)^{\top}\mathbf{D}^r)^{-1} \left((\mathbf{D}^r)^{\top}(\boldsymbol{\eta}_n^{(i)} - \boldsymbol{\zeta}_n^{(i)}) + \tilde{\boldsymbol{\sigma}}_n^{(i)} - \boldsymbol{\tau}_n^{(i)} \right)$
  for $j \in \{1, 2, \ldots, J\}$ **do**
    $\left(\mathbf{x}_{j,n}^{(i+1)} \right)_{n=1}^N = (\mathbf{I}_{NL} + \gamma \mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1} \left(\gamma \mathbf{A}^{\mathrm{H}}\mathbf{y}_j + \left(\mathbf{G}\mathbf{u}_{j,n}^{(i+1)} + \mathbf{q}_{j,n}^{(i)} \right)_{n=1}^N \right)$
  for $(n, k) \in \{1, 2, \ldots, N\} \times \{1, 2, \ldots, L\}$ **do**
    $\left((\tilde{u}_{j,n}^{(i+1)}[k])_{j=1}^J, \tilde{\sigma}_n^{(i+1)}[k] \right) = \mathrm{prox}_{\gamma\lambda\phi} \left(\left(u_{j,n}^{(i+1)}[k] + r_{j,n}^{(i)}[k] \right)_{j=1}^J, \sigma_n^{(i+1)}[k] + \tau_n^{(i)}[k] \right)$
  $\left(\boldsymbol{\eta}_n^{(i+1)} \right)_{n=1}^N = P_{B_1^\alpha} \left(\left(\mathbf{D}^r \boldsymbol{\sigma}_n^{(i+1)} + \boldsymbol{\zeta}_n^{(i)} \right)_{n=1}^N \right)$
  for $n \in \{1, 2, \ldots, N\}$ **do**
    for $j \in \{1, 2, \ldots, J\}$ **do**
      $\mathbf{q}_{j,n}^{(i+1)} = \mathbf{q}_{j,n}^{(i)} + \mathbf{G}\mathbf{u}_{j,n}^{(i+1)} - \mathbf{x}_{j,n}^{(i+1)}$
      $\mathbf{r}_{j,n}^{(i+1)} = \mathbf{r}_{j,n}^{(i)} + \mathbf{u}_{j,n}^{(i+1)} - \tilde{\mathbf{u}}_{j,n}^{(i+1)}$
    $\boldsymbol{\tau}_n^{(i+1)} = \boldsymbol{\tau}_n^{(i)} + \boldsymbol{\sigma}_n^{(i+1)} - \tilde{\boldsymbol{\sigma}}_n^{(i+1)}$
    $\boldsymbol{\zeta}_n^{(i+1)} = \boldsymbol{\zeta}_n^{(i)} + \mathbf{D}^r \boldsymbol{\sigma}_n^{(i+1)} - \boldsymbol{\eta}_n^{(i+1)}$

From the definitions of $\mathbf{y}_j^{(\mathrm{pawr})}$ and $\mathbf{A}_n^{(\mathrm{pawr})}$ in Example 1, we have

$$\left\| \mathbf{y}_j^{(\mathrm{pawr})} - \sum_{n=1}^N \mathbf{A}_n^{(\mathrm{pawr})} \mathbf{x}_{j,n} \right\|^2$$
$$= \left\| \begin{pmatrix} \mathbf{y}_j[1] - \sum_{n=1}^N x_{j,n}[1]\mathbf{a}(\theta_n) \\ \vdots \\ \mathbf{y}_j[L] - \sum_{n=1}^N x_{j,n}[L]\mathbf{a}(\theta_n) \end{pmatrix} \right\|^2$$
$$= \left\| \begin{pmatrix} \mathbf{y}_j[1]^{\top} - \sum_{n=1}^N x_{j,n}[1]\mathbf{a}(\theta_n)^{\top} \\ \vdots \\ \mathbf{y}_j[L]^{\top} - \sum_{n=1}^N x_{j,n}[L]\mathbf{a}(\theta_n)^{\top} \end{pmatrix} \right\|_{\mathrm{fro}}^2$$
$$= \| \mathbf{Y}_j^{\top} - \mathbf{X}_j \mathbf{S}^{\top} \|_{\mathrm{fro}}^2,$$

where we let

$$\mathbf{Y}_j := (\mathbf{y}_j[1], \ldots, \mathbf{y}_j[L]) \in \mathbb{C}^{M \times L},$$
$$\mathbf{X}_j := (\mathbf{x}_{j,1}, \ldots, \mathbf{x}_{j,N}) \in \mathbb{C}^{L \times N},$$
$$\mathbf{S} := (\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_N)) \in \mathbb{C}^{M \times N},$$

and $\| \cdot \|_{\mathrm{fro}}$ denotes the Frobenius norm of the matrix. From this expression, it is clear that the step (29) can be solved in $\mathcal{O}(N^3)$ for the inversion regarding $\mathbf{S}^{\top}$.

Namely, $(\mathbf{x}_{j,n}^{(i+1)})_{n=1}^N$ in Algorithm 1 is obtained by

$$\left(\mathbf{x}_{j,1}^{(i+1)}, \ldots, \mathbf{x}_{j,N}^{(i+1)} \right)$$
$$= \left(\gamma \mathbf{Y}_j^{\top} \mathbf{S}^* + \mathbf{G}\mathbf{U}_j^{(i+1)} + \mathbf{Q}_j^{(i)} \right) (\mathbf{I}_N + \gamma \mathbf{S}^{\top} \mathbf{S}^*)^{-1},$$

where $\mathbf{S}^*$ is the complex conjugate of $\mathbf{S}$, and

$$\mathbf{U}_j^{(i+1)} := \left(\mathbf{u}_{j,1}^{(i+1)}, \ldots, \mathbf{u}_{j,N}^{(i+1)} \right) \in \mathbb{C}^{L \times N},$$
$$\mathbf{Q}_j^{(i)} := \left(\mathbf{q}_{j,1}^{(i)}, \ldots, \mathbf{q}_{j,N}^{(i)} \right) \in \mathbb{C}^{L \times N}.$$

Meanwhile, we remark that the inversions in Algorithm 1 are the same for all the iterations, and thus can be computed in advance and stored in the memory.

REFERENCES

[1] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[2] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Cham, Switzerland: Springer, 2009.

[3] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.

[4] V. N. Bringi and V. Chandrasekar, *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[5] D. S. Zrnic and R. J. Doviak, "Velocity spectra of vortices scanned with a pulse-Doppler radar," *J. Appl. Meteorol.*, vol. 14, no. 8, pp. 1531–1539, Dec. 1975.

[6] L. Janssen and G. Van Der Spek, "The shape of Doppler spectra from precipitation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-21, no. 2, pp. 208–219, Mar. 1985.

[7] Y. Wang, T.-Y. Yu, M. Yeary, A. Shapiro, S. Nemati, M. Foster, D. L. Andra, and M. Jain, "Tornado detection using a neuro–fuzzy system to integrate shear and spectral signatures," *J. Atmos. Ocean. Technol.*, vol. 25, no. 7, pp. 1136–1148, Jul. 2008.

[8] T.-Y. Yu, R. R. Rondinel, and R. D. Palmer, "Investigation of non-Gaussian Doppler spectra observed by weather radar in a tornadic supercell," *J. Atmos. Ocean. Technol.*, vol. 26, no. 3, pp. 444–461, Mar. 2009.

[9] D. S. Zrnic, J. F. Kimpel, D. E. Forsyth, A. Shapiro, G. Crain, R. Ferek, J. Heimmer, W. Benner, F. T. J. McNellis, and R. J. Vogt, "Agile-beam phased array radar for weather observations," *Bull. Amer. Meteorological Soc.*, vol. 88, no. 11, pp. 1753–1766, Nov. 2007.

[10] D. McLaughlin et al., "Short-wavelength technology and the potential for distributed networks of small radar systems," *Bull. Amer. Meteorological Soc.*, vol. 90, no. 12, pp. 1797–1818, Dec. 2009.

[11] B. Isom, R. Palmer, R. Kelley, J. Meier, D. Bodine, M. Yeary, B.-L. Cheong, Y. Zhang, T.-Y. Yu, and M. I. Biggerstaff, "The atmospheric imaging radar: Simultaneous volumetric observations using a phased array weather radar," *J. Atmos. Ocean. Technol.*, vol. 30, no. 4, pp. 655–675, Apr. 2013.

[12] E. Yoshikawa, T. Ushio, Z. Kawasaki, S. Yoshida, T. Morimoto, F. Mizutani, and M. Wada, "MMSE beam forming on fast-scanning phased array weather radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 3077–3088, May 2013.

[13] D. S. Zrnic, V. M. Melnikov, R. J. Doviak, and R. Palmer, "Scanning strategy for the multifunction phased-array radar to satisfy aviation and meteorological needs," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1204–1208, Jun. 2015.

[14] F. Mizutani, T. Ushio, E. Yoshikawa, S. Shimamura, H. Kikuchi, M. Wada, S. Satoh, and T. Iguchi, "Fast-scanning phased-array weather radar with angular imaging technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2664–2673, May 2018.

[15] E. Yoshikawa, T. Ushio, and H. Kikuchi, "A study of comb beam transmission on phased array weather radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6346–6356, Aug. 2021.

[16] D. Kitahara, H. Kuroda, A. Hirabayashi, E. Yoshikawa, H. Kikuchi, and T. Ushio, "Nonlinear beamforming based on group-sparsities of periodograms for phased array weather radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4106819.

[17] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[18] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[19] L. Wang, L. Zhao, G. Bi, C. Wan, L. Zhang, and H. Zhang, "Novel wideband DOA estimation based on sparse Bayesian learning with Dirichlet process priors," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 275–289, Jan. 2016.

[20] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

[21] C. R. Rojas, D. Katselis, and H. Hjalmarsson, "A note on the SPICE method," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4545–4551, Sep. 2013.

[22] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.

[23] X. Shang, J. Li, and P. Stoica, "Weighted SPICE algorithms for range-Doppler imaging using one-bit automotive radar," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 1041–1054, Jun. 2021.

[24] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[25] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[26] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[27] X. Lv, G. Bi, and C. Wan, "The group lasso for stable recovery of block-sparse signal representations," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1371–1382, Apr. 2011.

[28] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4094–4107, Aug. 2012.

[29] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jun. 2009, pp. 433–440.

[30] G. Obozinski, L. Jacob, and J.-P. Vert, "Group lasso with overlaps: The latent group lasso approach," 2011, *arXiv:1110.0413*.

[31] P. Ciuciu, J. Idier, and J.-F. Giovannelli, "Regularized estimation of mixed spectra using a circular Gibbs-Markov model," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2202–2213, Oct. 2001.

[32] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.

[33] L. Wang, T. Gerkmann, and S. Doclo, "Noise power spectral density estimation using MaxNSR blocking matrix," *IEEE/ACM Trans. Audio, Speech, Lang., Process.*, vol. 23, no. 9, pp. 1493–1508, Sep. 2015.

[34] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," *Speech Commun.*, vol. 50, no. 6, pp. 453–466, Jun. 2008.

[35] Y. Zhao, X. Zhao, and B. Wang, "A speech enhancement method based on sparse reconstruction of power spectral density," *Comput. Electr. Eng.*, vol. 40, no. 4, pp. 1080–1089, May 2014.

[36] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[37] F. Li and Z. Xu, "Sparse Bayesian hierarchical prior modeling based cooperative spectrum sensing in wideband cognitive radio networks," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 586–590, May 2014.

[38] H. Kuroda and D. Kitahara, "Block-sparse recovery with optimal block partition," *IEEE Trans. Signal Process.*, vol. 70, pp. 1506–1520, 2022.

[39] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 22, no. 2, pp. 503–516, Jan. 2000.

[40] G. D. Maso, I. Fonseca, G. Leoni, and M. Morini, "A higher order model for image restoration: The one-dimensional case," *SIAM J. Math. Anal.*, vol. 40, no. 6, pp. 2351–2391, Jan. 2009.

[41] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[42] H. Kuroda, D. Kitahara, E. Yoshikawa, H. Kikuchi, and T. Ushio, "Sparsity-smoothness-aware power spectral density estimation with application to phased array weather radar," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[43] H. B. Bluestein, J. G. Ladue, H. Stein, D. Speheger, and W. F. Unruh, "Doppler radar wind spectra of supercell tornadoes," *Monthly Weather Rev.*, vol. 121, no. 8, pp. 2200–2222, Aug. 1993.

[44] E. Yoshikawa, S. Kida, S. Yoshida, T. Morimoto, T. Ushio, and Z. Kawasaki, "Vertical structure of raindrop size distribution in lower atmospheric boundary layer," *Geophys. Res. Lett.*, vol. 37, no. 20, Oct. 2010, Art. no. L20802.

[45] V. Chandrasekar, R. M. Beauchamp, and R. Bechini, *Introduction to Dual Polarization Weather Radar: Fundamentals, Applications, and Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2023.

[46] H. Kuroda and D. Kitahara, "Graph-structured sparse regularization via convex optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 5538–5542.

[47] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, May 2011.

[48] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[49] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, Apr. 1992.

[50] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Cham, Switzerland: Springer, 2011, pp. 185–212.

[51] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, "Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists," *SIAM Rev.*, vol. 65, no. 2, pp. 375–435, May 2023.

[52] P. L. Combettes, "Perspective functions: Properties, constructions, and examples," *Set-Valued Variational Anal.*, vol. 26, no. 2, pp. 247–264, Jun. 2018.

[53] P. L. Combettes and C. L. Müller, "Perspective functions: Proximal calculus and applications in high-dimensional statistics," *J. Math. Anal. Appl.*, vol. 457, no. 2, pp. 1283–1306, Jan. 2018.

[54] P. L. Combettes and C. L. Müller, "Perspective maximum likelihood-type estimation via proximal decomposition," *Electron. J. Statist.*, vol. 14, no. 1, pp. 207–238, Jan. 2020.

[55] H. H. Bauschke, M. Krishan Lal, and X. Wang, "Real roots of real cubics and optimization," 2023, *arXiv:2302.10731*.

[56] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Math. Programm.*, vol. 158, no. 1, pp. 575–585, 2016.

[57] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.

HIROKI KURODA (Member, IEEE) received the B.E. degree in computer science and the M.E. and Ph.D. degrees in information and communications engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 2013, 2015, and 2019, respectively.

In 2019, he was a Postdoctoral Researcher with the National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan. From 2020 to 2022, he was an Assistant Professor with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. Since 2022, he has been an Assistant Professor with the Department of Information and Management Systems Engineering, Nagaoka University of Technology, Niigata, Japan. His research interests include signal processing and its applications, inverse problem, sparse modeling, and optimization theory.

Dr. Kuroda was a recipient of the Young Researcher Award from the IEICE Technical Group on Signal Processing, in 2018; and the Seiichi Tejima Doctoral Dissertation Award from the Tokyo Institute of Technology, in 2020.

DAICHI KITAHARA (Member, IEEE) received the B.E. degree in computer science and the M.E. and Ph.D. degrees in communications and computer engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 2012, 2014, and 2017, respectively.

From 2017 to 2022, he was an Assistant Professor with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. Currently, he is a Researcher with the Graduate School of Engineering, Osaka University. His research interests include signal processing and its applications, inverse problem, optimization theory, and multivariate spline theory.

Dr. Kitahara was a recipient of the Young Researcher Awards from the IEICE Technical Group on Signal Processing, in 2013; the SICE Technical Committee on Remote Sensing, in 2015; and the IEEE Computational Intelligence Society Japan Chapter, in 2016. He was also a recipient of the IEEE Signal Processing Society (SPS) Japan Student Journal Paper Award from the IEEE SPS Tokyo Joint Chapter, in 2016; and the Yasujiro Niwa Outstanding Paper Award from the Tokyo Denki University, in 2018.

EIICHI YOSHIKAWA (Member, IEEE) received the B.E. degree in aerospace engineering from Osaka Prefecture University, Sakai, Japan, in 2005, and the M.E. and Ph.D. degrees from Osaka University, Suita, Japan, in 2008 and 2011, respectively.

In 2011, he was a Postdoctoral Researcher with Osaka University and Colorado State University, Fort Collins, CO, USA; and a Postdoctoral Fellow of the Research Fellowship for Young Scientists, sponsored by the Japan Society for the Promotion of Science (JSPS). In 2012, he joined the Japan Aerospace Exploration Agency (JAXA), Mitaka, Tokyo, Japan, where he is currently an Associate Senior Researcher. He is also a Research Scientist with Colorado State University. His research interests include weather radar remote sensing, radar-based analyses, and applications for general and aviation weather.

HIROSHI KIKUCHI (Member, IEEE) received the B.S. degree from the Department of Engineering, Doshisha University, Kyoto, Japan, in 2008, and the M.S. and Ph.D. degrees from the Division of Electrical, Electronic and Information Engineering, Osaka University, Suita, Japan, in 2010 and 2013, respectively.

He joined the Division of Electrical, Electric and Information Engineering, Osaka University, as a Specially Appointed Researcher, in 2013. In 2017, he was a Research Assistant Professor with Tokyo Metropolitan University. In 2018, he joined The University of Electro-Communications, where he is currently an Associate Professor. His research specialties are the remote sensing for an atmospheric electricity with space-borne platforms, weather radar remote sensing, and a development of the radar systems.

TOMOO USHIO (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Osaka University, Suita, Japan, in 1993, 1995, and 1998, respectively.

He was with the Global Hydrology and Climate Center, Huntsville, AL, USA, as a Postdoctoral Researcher, from 1998 to 2000. In 2000, he joined the Department of Aerospace Engineering, Osaka Prefecture University, Sakai, Japan, as an Assistant Professor. In 2006, he joined the Division of Electrical, Electronic and Information Engineering, Osaka University, as an Associate Professor, where he has been a Professor, since 2019. In 2017, he joined the Department of Aeronautics and Astronautics, Tokyo Metropolitan University, Hino, Tokyo, Japan, as a Professor. His research specialties are radar-based remote sensing, passive and active remote sensing of atmosphere from space-borne platforms, and atmospheric electricity.

. . .