2022 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, AUG. 22–25, 2022, XI'AN, CHINA

EXPLOITING TEMPORAL STRUCTURES OF CYCLOSTATIONARY SIGNALS FOR DATA-DRIVEN SINGLE-CHANNEL SOURCE SEPARATION

Gary C.F. Lee, Amir Weiss, Alejandro Lancho, Jennifer Tang, Yuheng Bu,
Yury Polyanskiy, Gregory W. Wornell
Massachusetts Institute of Technology, Cambridge, MA

ABSTRACT

We study the problem of single-channel source separation (SCSS), and focus on cyclostationary signals, which are particularly suitable in a variety of application domains. Unlike classical SCSS approaches, we consider a setting where only examples of the sources are available rather than their models, inspiring a data-driven approach. For source models with underlying cyclostationary Gaussian constituents, we establish a lower bound on the attainable mean-square-error (MSE) for any separation method, model-based or data-driven. Our analysis further reveals the operation for optimal separation and the associated implementation challenges. As a computationally attractive alternative, we propose a deep learning approach using a U-Net architecture, which is competitive with the minimum MSE estimator. We demonstrate in simulation that, with suitable domain-informed architectural choices, our U-Net method can approach the optimal performance with substantially reduced computational burden.

***Index Terms*—** Source separation, cyclostationary signal processing, deep neural network, supervised learning.

1. INTRODUCTION

Source separation is a well-studied problem with many important applications in radio-frequency systems, wireless communication and biomedical signal monitoring [1–4], to name a few. A common formulation is the blind source separation problem, for which perhaps the most popular framework is independent component analysis [5]. Many approaches rely on some degree of spatial diversity present in multi-channel observations. In contrast, a particularly challenging problem of interest is the *single-channel* (or single-sensor) source separation (SCSS). The goal in SCSS is to separate the latent sources from a single-channel observed mixture. In this regime, the aforementioned algorithms are irrelevant due to the absence of spatial diversity. Instead, other characteristics of the latent sources must be exploited.

Recent efforts demonstrate that machine learning (ML) techniques can serve as powerful tools for source separation, even in the single-channel regime, in image and audio counterparts [6–9]. Solutions typically exploit the inherent structure specific to the signal type. For example, natural images may be separable by color features and local dependencies [8], whereas speech signals are commonly addressed by time-frequency spectrogram masking [10–12]. Furthermore, for time series (1-dimensional) data, if the sources are separable in time and/or frequency, appropriate spectrogram-based masking and classical filtering methods can be adopted, e.g. [13,14]. In contrast, the key challenge is the separation of signals that overlap (partially/fully) in *both* time and frequency; such a setting is not addressed in aforementioned works. Further, in such a regime, it is no longer trivial to identify the properties that are helpful in separating the signals, and what separation performance is attainable.

One particular type of time series data is the class of *cyclostationary* signals, a relevant model adopted in many of the earlier mentioned applications (e.g. [15]). Interestingly, under some conditions, perfect separation of cyclostationary signals could be achieved despite having components with overlapping time-frequency spectra [16, 17]. The challenge, however, is in modeling such cyclostationarities in a way that properly describes the true statistics of the observed mixtures. A more realistic scenario is one where the signal model is unknown, but examples of the signals (by measurements or generation[1]) are available. Further, despite cyclostationarity properties, the available examples are usually finite unsynchronized time segments, meaning that they were extracted at possibly random different "start times", relative to an arbitrarily chosen time instance. In other words, there are additional latent variables—the random time shifts—that are unobservable, adding a layer of difficulty to the problem.

In this paper, we study the SCSS problem involving mixtures of two signals. Specifically, we focus on the scenario

Research was sponsored by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.
The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing HPC resources that have contributed to the research results reported within this paper.
Alejandro Lancho is funded by EU's Horizon 2020 programme under the Marie Sklodowska-Curie grant agreement No. 101024432.

[1]Note that the ability to generate a signal (with some device) does not necessarily mean that one can also analytically characterize its true statistics.

978-1-6654-8547-0/22//$31.00 ©2022 IEEE

where each component is a randomly time-shifted and scaled segment from a cyclostationary complex Gaussian process, a regime in which an analytical form of the optimal estimator, in the sense of minimum mean-square-error (MMSE), can be derived. Nonetheless, we also highlight the challenges in implementing the optimal estimator, thereby motivating less computationally demanding alternatives. We propose a deep learning (DL) approach to the SCSS problem, and demonstrate its performance in two representative examples—one of which is based on real applications in wireless communications. Our simulations show that the performance of our proposed DL strategy is competitive to the optimal estimator.

We recognize that many signals of practical relevance generally depart from the complex Gaussian (otherwise, simply referred to as *Gaussian* in this work) model. However, the analysis involving cyclostationary Gaussian signals allows us to better understand and characterize the performance of DL-based methods. In particular, it focuses our attention to the exploitation of temporal correlation structures of the latent source components, so as to accurately assess the ability of ML approaches in capturing such structures. This would shed light as to how far we are from the optimal performance, and how to close that performance gap.

2. SCSS PROBLEM FORMULATION

We consider the following model of an observed signal of length N, which is a noisy mixture of two latent sources,

$$\boldsymbol{y} = \boldsymbol{s}_{\tau_s} + \underbrace{\kappa\, \boldsymbol{b}_{\tau_b}}_{\text{'interference'}} + \underbrace{\sigma_w \boldsymbol{z}}_{\triangleq \boldsymbol{w}} \in \mathbb{C}^N, \tag{1}$$

where $\boldsymbol{s}_{\tau_s}, \boldsymbol{b}_{\tau_b}$ are the (unobservable) independent components, $\kappa \in \mathbb{R}_+$ is distributed according to some unknown (and for simplicity, discrete) distribution on $\mathcal{K} \subset \mathbb{R}_+$, and $\boldsymbol{w}$ is an additive white Gaussian noise, namely $\boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$, such that σ_w^2 corresponds to the variance of the additive noise component, $\boldsymbol{w}$, where $\boldsymbol{0}$ denotes the all zeros-vector and $\boldsymbol{I}$ denotes the identity matrix (with context-dependent dimensions).

Without loss of generality, for the purposes of our discussion, $\boldsymbol{s}_{\tau_s}$ is termed the "reference" signal and $\boldsymbol{b}_{\tau_b}$ the interference. Additionally, we assume that $\boldsymbol{s}_{\tau_s}$ and $\boldsymbol{b}_{\tau_b}$ have unit average power; hence, κ is related to the inverse square-root of the signal-to-interference ratio (SIR). The goal in SCSS is to produce an estimate $\widehat{\boldsymbol{s}}_{\tau_s}$ based on $\boldsymbol{y}$ so that given some metric d, the value $\mathbb{E}\left[d(\widehat{\boldsymbol{s}}_{\tau_s}, \boldsymbol{s}_{\tau_s})\right]$ is minimized. We focus our attention on the conventional metric $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|_2^2$, leading to the minimum mean square error (MMSE) criterion.

We focus our discussion on signal components being segments extracted from cyclostationary Gaussian processes. We consider two independent, discrete-time, zero-mean circularly-symmetric complex Gaussian processes $\tilde{s}[\cdot]$, $\tilde{b}[\cdot]$, with autocovariance functions satisfying

$$C_{\tilde{s}}[n,l] \triangleq \mathbb{E}\left[\tilde{s}[n]\,\tilde{s}[n+l]\right], \;\; C_{\tilde{s}}[n,l] = C_{\tilde{s}}[n+N_s,l],$$
$$C_{\tilde{b}}[n,l] \triangleq \mathbb{E}\left[\tilde{b}[n]\,\tilde{b}[n+l]\right], \;\; C_{\tilde{b}}[n,l] = C_{\tilde{b}}[n+N_b,l],$$

i.e., cyclostationary with fundamental periods $N_s, N_b > 1$.

We denote the temporal offsets by τ_s, τ_b. Hence,

$$\boldsymbol{s}_{\tau_s} = [s_{\tau_s}[0], \ldots, s_{\tau_s}[N-1]]^{\mathrm{T}} \in \mathbb{C}^N, \; s_{\tau_s}[n] \triangleq \tilde{s}[n+\tau_s],$$

and similarly for $\boldsymbol{b}_{\tau_b}$ and $\tilde{\boldsymbol{b}}$. We consider the case where temporal offsets are random, drawn from a discrete uniform distribution, i.e., $\tau_s \sim \mathcal{U}[0, N_s - 1]$ and $\tau_b \sim \mathcal{U}[0, N_b - 1]$, and assume τ_s, τ_b, $\tilde{s}[\cdot]$, $\tilde{b}[\cdot]$ and κ are statistically independent. Consequently, note that $\boldsymbol{s}_{\tau_s}$ and $\boldsymbol{b}_{\tau_b}$ are Gaussian mixtures.

As mentioned in Section 1, we are particularly interested in the case where we do not have full knowledge of the signal models—the distributions of $\tilde{\boldsymbol{s}}$ and $\tilde{\boldsymbol{b}}$ (and thereby $\boldsymbol{s}_{\tau_s}$ and $\boldsymbol{b}_{\tau_b}$) are unknown. Nevertheless, we assume we have a dataset of M independent, identically distributed (iid) copies of $\{(\boldsymbol{y}^{(i)}, \boldsymbol{s}_{\tau_s}^{(i)})\}_{i=1}^M$, allowing for a data-driven approach. Note that this dataset contains *unsynchronized* examples, wherein the corresponding time offsets $\tau_s^{(i)}$ and $\tau_b^{(i)}$ are unknown. At test time, we assume that given the mixture $\boldsymbol{y}$, we may know the *type* of components $\boldsymbol{s}$ and $\boldsymbol{b}$ present (but not their signal models); hence, the mixture can be associated with a corresponding dataset, and in turn, its estimated statistics and/or learned model.

Under this formulation, an ML approach is to learn a regressor, i.e. to estimate $\boldsymbol{s}_{\tau_s}$ from $\boldsymbol{y}$. Note that the dimension of the regression output is *the same* as that of the input. Thus, it can be thought of as a transformation or filtering (through some parametric model) of $\boldsymbol{y}$ to obtain $\boldsymbol{s}_{\tau_s}$, which is generally more challenging than a conventional single-output regression problem [18]. We will thus investigate the use of deep neural networks (DNNs) to learn such a regressor from data. However, we begin with an in-depth examination of the SCSS problem through approaches to optimal estimation.

3. OPTIMAL MODEL-BASED ESTIMATORS

We now derive, for several cases, optimal estimators that achieve the lower bounds for their respective cases. Specifically, the simplified expression of the MMSE estimator reveals the challenges in attaining it, but also provides valuable intuition that informs and justifies our proposed data-driven methodology for the SCSS problem.

A possible "classical signal processing" approach for this setting, is a two-step process, where one first estimates the model parameters, and then adopts MMSE estimation based on the empirical model. However, estimation of these parameters requires synchronization of the given dataset, which can be a very challenging task in itself. Nevertheless, in this section we assume access to an oracle regarding the signal model, so as to establish a lower bound on MMSE for our problem.

Specifically, in the following we assume oracle knowledge of the signal models—the first and second-order statistics (SOSs) of $\tilde{\boldsymbol{s}}$ and $\tilde{\boldsymbol{b}}$, and the marginal distributions of τ_s, τ_b. We denote the conditional covariance of $\boldsymbol{s}_{\tau_s}$ given τ_s by

$$\boldsymbol{C}_{\boldsymbol{s}}(\tau_s) \triangleq \mathbb{E}\left[\boldsymbol{s}_{\tau_s} \boldsymbol{s}_{\tau_s}^{\mathrm{H}} | \tau_s\right] \in \mathbb{C}^{N \times N},$$

which we highlight that it is a function of τ_s; likewise, $\boldsymbol{C_b}(\tau_b)$ denotes the conditional covariance of $\boldsymbol{b}_{\tau_b}$. We denote the entries of $\boldsymbol{C_s}(\tau_s)$ by $(\boldsymbol{C_s}(\tau_s))_{i,j} = C_{\tilde{s}}[i+\tau_s, i-j]$.

3.1. Case 1: The Optimal Linear MMSE Solution

To obtain the SOS of $\boldsymbol{s}_{\tau_s}$, one must account for the uniform randomness in τ_s. Hence, the covariance of $\boldsymbol{s}_{\tau_s}$ is given by

$$\check{\boldsymbol{C}}_{\boldsymbol{s}} = \mathbb{E}\left[\boldsymbol{s}_{\tau_s}\boldsymbol{s}_{\tau_s}^{\mathrm{H}}\right] = \mathbb{E}_{\tau_s}\left[\mathbb{E}\left[\boldsymbol{s}_{\tau_s}\boldsymbol{s}_{\tau_s}^{\mathrm{H}}\right]|\tau_s\right] = \frac{1}{N_s}\sum_{\tau_s=0}^{N_s-1}\boldsymbol{C_s}(\tau_s).$$

Note that this corresponds to a Toeplitz covariance structure, which is due to the fact that $s_{\tau_s}[\cdot]$ is a wide-sense stationary process (unlike $\tilde{s}[\cdot]$). The same applies for the SOS of $\boldsymbol{b}_{\tau_b}$.

The SOSs of $\boldsymbol{s}_{\tau_s}$ and $\boldsymbol{b}_{\tau_b}$ are sufficient for optimal *linear* estimation. For the sake of practicality, if we restrict our attention to linear operators, the optimal estimator is given by

$$\widehat{\boldsymbol{s}}_{\tau_s,\mathrm{LMMSE}} = \check{\boldsymbol{C}}_{\boldsymbol{s}}\left[\check{\boldsymbol{C}}_{\boldsymbol{s}} + \mathbb{E}\left[\kappa^2\right]\cdot\check{\boldsymbol{C}}_{\boldsymbol{b}} + \sigma_w^2\boldsymbol{I}\right]^{-1}\boldsymbol{y}. \quad (2)$$

This would indeed correspond to the estimator that achieves the MMSE among the family of linear estimators. Further, given a sufficiently large dataset, one can implement an accurate approximation of (2) by replacing the covariance matrices with empirical estimates from the datasets. However, due to the random time offsets, $\boldsymbol{s}_{\tau_s}$ and $\boldsymbol{b}_{\tau_b}$ are not Gaussian (but rather, are Gaussian mixtures), and hence the optimal linear estimator is not the MMSE estimator.

3.2. Case 2: The Oracle MMSE Solution

In this subsection, we develop the optimal solution for the case in which (τ_s, τ_b, κ) are known (or, observable). This would have been the case if we would have at our disposal an oracle that provides side information of perfect synchronization to *both* the reference and interference signals, and the SIR level κ. Under this setting, $\boldsymbol{y}$ and $\boldsymbol{s}_{\tau_s}$ are jointly Gaussian, hence the optimal solution is given by

$$\widehat{\boldsymbol{s}}_{\tau_s}|_{(\tau_s,\tau_b,\kappa)} = \mathbb{E}\left[\boldsymbol{s}_{\tau_s}|\boldsymbol{y},\tau_s,\tau_b,\kappa\right] = \boldsymbol{H}(\tau_s,\tau_b,\kappa)\boldsymbol{y}, \quad (3)$$

where

$$\boldsymbol{H}(\tau_s,\tau_b,\kappa) \triangleq \boldsymbol{C_s}(\tau_s)\left[\boldsymbol{C_s}(\tau_s) + \kappa^2\boldsymbol{C_b}(\tau_b) + \sigma_w^2\boldsymbol{I}\right]^{-1} \quad (4)$$

is the optimal linear (generally) time-varying filter. Since (4) is a function of (τ_s, τ_b, κ), we subsequently refer to (3) as an oracle-synchronized MMSE solution. Nevertheless, it should be noted that (3) is not a realizable estimator, as it is a function of latent, unobservable variables. Also, note that for cyclostationary signals, (3) can also be represented as a frequency-shift filter or a cyclic Wiener filter, i.e., expressed as a linear function of frequency shifted copies of the observation [2,17].

3.3. Case 3: The Optimal MMSE Estimator

In general, the temporal offsets τ_s and τ_b, and the SIR parameter κ, are not known at inference time, and a realizable estimator must account for these random variables, by explicit/implicit (non-linear) estimation. However, upon conditioning on these quantities, i.e., considering (4) as fixed (as if it were not a function of $\boldsymbol{y}$), the resulting optimal MMSE estimator would also be the linear. With this observation in mind, the true, realizable MMSE estimator is expressed as

$$\begin{aligned}\widehat{\boldsymbol{s}}_{\tau_s,\mathrm{MMSE}} &= \mathbb{E}\left[\boldsymbol{s}_{\tau_s}|\boldsymbol{y}\right] = \mathbb{E}_{(\tau_s,\tau_b,\kappa)|\boldsymbol{y}}\left[\mathbb{E}\left[\boldsymbol{s}_{\tau_s}|\boldsymbol{y},\tau_s,\tau_b,\kappa\right]\right] \\ &= \sum_{\tau_s=0}^{N_s-1}\sum_{\tau_b=0}^{N_b-1}\sum_{\kappa\in\mathcal{K}} p(\tau_s,\tau_b,\kappa|\boldsymbol{y})\,\widehat{\boldsymbol{s}}_{\tau_s}|_{(\tau_s,\tau_b,\kappa)}, \quad (5)\end{aligned}$$

which essentially corresponds to the sum of oracle-synchronized MMSE solutions (3) for each set of parameters, weighted by the corresponding posterior probabilities given the observation $\boldsymbol{y}$. We emphasize that (5), unlike (3), is a legitimate estimator, i.e., a function of the observed data only. However, the MMSE achieved by (3) serves as a lower bound of that by (5), by data procecessing inequality of MMSE.

It should be noted that, in cases when the true posterior is a (Kronecker) delta function, (5) in fact *coincides* with (3),

$$\begin{aligned}&p(\tau_s,\tau_b,\kappa|\boldsymbol{y}) = \delta[\tau_s-\tau_s^*,\tau_b-\tau_b^*,\kappa-\kappa^*] \quad (6)\\ &\implies \widehat{\boldsymbol{s}}_{\tau_s,\mathrm{MMSE}} = \widehat{\boldsymbol{s}}_{\tau_s}|_{(\tau_s^*,\tau_b^*,\kappa^*)},\end{aligned}$$

namely the performance of the oracle-aided solution is achievable. Indeed, for the noiseless case ($\sigma_w^2 = 0$), it is easy to construct examples, where perfect separation is attained.

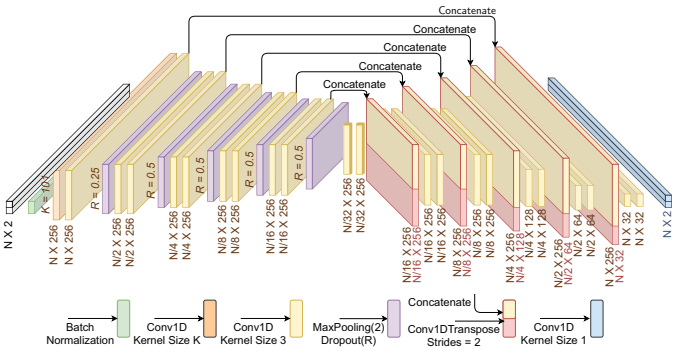

Fig. 1: The U-Net architecture used for SCSS in our simulations.

4. METHODOLOGIES FOR DATA-DRIVEN SCSS

The methods discussed thus far provide an understanding of the functional structure for the MMSE estimator, in terms of the (unknown) time offsets and SIR parameters. Nevertheless, implementing the MMSE estimator may not be possible in practice, due to the challenges outlined below.

4.1. Challenges with the MMSE Estimator

First, as seen in (5), the MMSE estimator involves the computation of a posterior term over the latent variables (τ_s, τ_b, κ). However, obtaining this posterior becomes computationally intensive as the space of parameters increases.

Second, (5) also involves a sum of different (conditionally) linear operators, each corresponding to a set of parameter values. However, each of these operators involves an inversion of a large covariance matrix, which is infeasible in regimes of signals from long observation periods (large N).

Lastly, and importantly, in practice we do not have oracle knowledge of the signal model, corresponding to the first and second-order statistics in the case of the Gaussian models. Instead, we are given many examples, through which we can obtain empirical estimates of the statistics. However, as mentioned, the dataset of the signals is not synchronized. Thus, obtaining the SOSs of the underlying cyclostationary signal, $\boldsymbol{C}_{\tilde{\boldsymbol{s}}}$ and $\boldsymbol{C}_{\tilde{\boldsymbol{b}}}$, requires the estimation of the latent variable τ_s and τ_b, for *each* example in the dataset, corresponding to the synchronization of the entire dataset, which is generally an extremely challenging task of independent interest in itself.

We now show that ML methods can be successfully used to circumvent the aforementioned issues. Specifically, we propose the use of DNNs, trained on *unsynchronized* datasets, to solve the SCSS problem. We use two representative examples to demonstrate our approach, and compare it with the performance of an optimal MMSE solution that utilizes oracle knowledge. The goal henceforth is to establish a practical pipeline, and benchmark it against a theoretical lower bound.

4.2. Supervised Separation with U-Net

Given the formulation in Section 2, a natural approach is to use a DNN to learn a regression model with multivariate output for source separation. We propose to use the so-called "U-Net" architecture for SCSS (architecture shown in Fig. 1).[2] Such DNN was first proposed for biomedical image segmentation [19], but has found its use in other applications, including spectrogram-based RF interference cancellation [1] and audio source separation [6,12]—all of which also corresponding to a multivariate regression setup with the same dimensions on the input and output. Similar to the latter works, we use 1D-convolutional layers to capture features relating to the time series data. The U-Net architecture contains downsampling blocks that operate on successively coarser timescales, and possesses skip connections that combine features at these various timescales with the upsampling blocks.

To technically handle complex-valued signals, borrowing inspiration from widely linear estimation [20], we stack the real and imaginary parts as separate channels to the U-Net.

As these methods are applied to time series signals in practice, the use of domain knowledge to craft an appropriate neural network architecture may be crucial in attaining performance gain, as evident in our experiments and architectural choices. For example, we made the intentional choice of longer kernel sizes on the first convolutional layers. This further reinforces the relevance of this work, that is in identifying and characterizing DNN architectures under study relative to the best possible performance. In our experiments, we observe that kernel sizes that match the effective correlation length (i.e., timescales in which the covariance magnitudes are non-negligible) were required to attain the best performance. This may be an indication to how some partial, though important, information about the signal model is helpful (or even essential) in seeking the appropriate DNN architecture.

[2] https://github.com/RFChallenge/SCSS_CSGaussian

5. SIMULATION RESULTS

We now consider two examples for SCSS. For each setting, we train a U-Net to estimate the corresponding signal $\boldsymbol{s}_{\tau_s}$ from the mixture $\boldsymbol{y}$ in an end-to-end fashion, and with no supervision regarding the time-shifts τ_s and τ_b, and the gain κ.

In the examples below, we describe how long (but finite) segments of the processes $\tilde{s}[\cdot], \tilde{b}[\cdot]$ were generated, from which N-length segments are extracted to create the datasets. The training set is processed as such to yield a labeled dataset of iid copies—mixture and ground-truth reference signal, $\{(\boldsymbol{y}^{(i)}, \boldsymbol{s}_{\tau_s}^{(i)})\}_{i=1}^{M}$—as is done in the supervised learning framework. Our training set comprises $10^4 \times |\mathcal{K}|$ pairs of mixtures $\boldsymbol{y}$ and ground-truth $\boldsymbol{s}_{\tau_s}$, and the validation set comprises $500 \times |\mathcal{K}|$ pairs, where the cardinality $|\mathcal{K}|$ is the total number of levels for κ under consideration. Subsequently, we test the performance across 1000 examples per κ level, reporting the average mean square error (MSE) in dB. Note that varying κ results in different levels of SIR. In our simulations, we assess the separation performance across different SIR levels. We also compare the performance of using such a neural network against that of the optimal model-based estimators. For these optimal estimators, κ is assumed to be known.

Implementation Details: Keras and Tensorflow 2 are used to implement and train the U-Net [21, 22]. For training, we use empirical MSE as the loss function. We also use Adam optimizer [23] and an exponentially decaying learning rate schedule, batch size of 32 with shuffled training samples, and trained for 2000 epochs with early stopping if there is no improvement for 100 epochs on the validation set. We train the neural networks on a computing cluster with Intel Xeon Gold 6248, 192 GB RAM, and a NVidia Volta V100 GPU.

5.1. Signals with Randomly Generated Covariances

We consider (1) with $N = 256$, $N_s = 11$, $N_b = 5$, $\mathcal{K}$ corresponding to 5 equidistant SIR levels, $[-6, 6]$ dB ($|\mathcal{K}| = 5$). The reference and interference signals were generated as

$$\tilde{\boldsymbol{s}} = \boldsymbol{G}_{\boldsymbol{s}}\boldsymbol{a}_1, \;\; \tilde{\boldsymbol{b}} = \boldsymbol{G}_{\boldsymbol{b}}\boldsymbol{a}_2,$$

where $\boldsymbol{a}_1, \boldsymbol{a}_2 \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$, and $\boldsymbol{G}_{\boldsymbol{s}}, \boldsymbol{G}_{\boldsymbol{b}} \in \mathbb{C}^{\tilde{N}\times\tilde{N}}$, with $\tilde{N} = 550$, are block-diagonal matrices with repeating $N_s \times N_s$ and $N_b \times N_b$ blocks respectively. Each entry in the blocks is drawn (once) independently from the Gaussian distribution, and is fixed for the rest of this experiment. Full details on the signal generation are provided in our Github repository.[2] Fig. 2(a) shows the covariance structures of the resulting sources.

Fig. 3 compares the MSE achieved by the U-Net against that obtained by the linear and the "global" MMSE estimators. We also include the oracle-synchronized MMSE (3), which, as evident from the figure, is indistinguishable (in terms of its MSE) from the true, nonoracle MMSE estimator

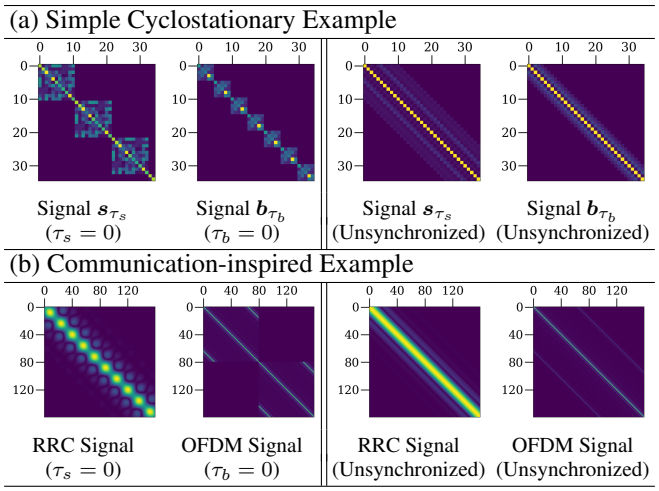

Fig. 2: Visualization of (a sub-matrix of) the covariance matrices for the signals used in the respective examples.

(5). This occurs when all the mass of the posterior is (approximately) concentrated at the point of the true values of (τ_s, τ_b, κ), as in (6).

We observe that a U-Net trained on an *unsynchronized* dataset of signals is capable of obtaining results close to the MMSE performance. This means that the U-Net necessarily learned a significant part of the model, which enables high-quality estimation of the reference signal. We reiterate that the U-Net did not have access to the true statistics of the signal model or any form of synchronization of the signals during training and inference. The slight deterioration performance could be attributed to approximation errors introduced from a DL-based function approximator, or due to the trade-off from lack of access to synchronized dataset. Future work entails identifying factors to close the performance gap.

5.2. Communication-like Waveforms

The problem of SCSS is particularly relevant in the application of wireless communications, where we may be interested in separating multiplex of signals, or extracting a signal-of-interest while mitigating interference [3, 24]. In this example, we consider two types of signals with SOSs properties identical to a single-carrier communication waveform and an orthogonal frequency division multiplexed (OFDM) waveform, respectively. The single-carrier signal is modeled as,

$$\tilde{s}[n] = \sum_{p=-\infty}^{\infty} a_p \, g[n - pN_s], \tag{7}$$

where $a_p \sim \mathcal{CN}(0, 1)$, N_s is the symbol period, and $g[n]$ is the root-raised cosine (RRC) filter. Here, we consider a RRC filter that corresponds to 16 samples per symbol—i.e., $N_s = 16$—and spans 8 symbols, with roll-off factor of 0.5. This also corresponds to an example where a cyclostationary signal's covariance (Fig. 2(b)) is not block-diagonal, which leads to additional computational challenges in the matrix inversion.

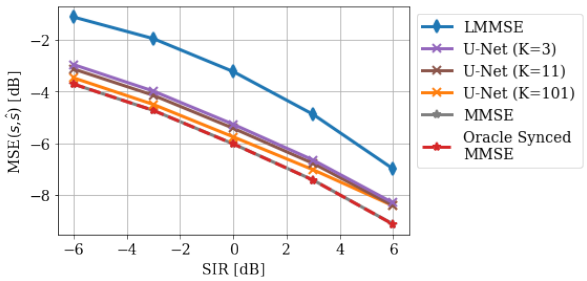

Fig. 3: Separation performance of U-Net separator vs. optimal model-based estimators for waveforms with randomly generated covariance structures.

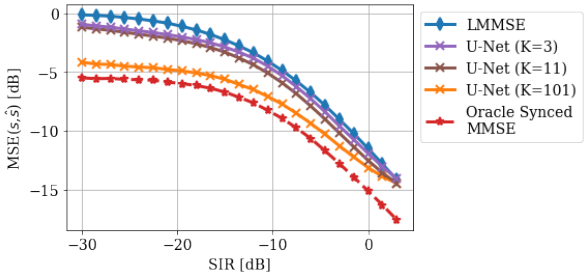

Fig. 4: Separation performance of the U-Net separator vs. optimal model-based estimators for communication-like waveforms.

The second source, an OFDM waveform, is modeled as,

$$\tilde{b}[n] = \frac{1}{\sqrt{N_{sc}}} \sum_{p=-\infty}^{\infty} \sum_{\ell=0}^{N_{sc}-1} a_{p,\ell} \, q[n - pN_b, \ell], \tag{8}$$

$$q[n, \ell] = \mathbb{1}_{\{0 \le n \le N_b - 1\}} \cdot \exp\left(j2\pi\ell \frac{n - N_{cp}}{N_{sc}} \right),$$

where $a_{p,\ell} \sim \mathcal{CN}(0, 1)$ for $\ell \in \mathcal{L}_{sc}$, where $\mathcal{L}_{sc}$ refers to the set of nonzero subcarrier indices, and $a_{p,\ell} = 0$ otherwise, N_{sc} is the number of subcarriers per OFDM symbol, N_{cp} is the cyclic prefix (CP) length, and N_b is the OFDM symbol period, i.e., $N_b = N_{sc} + N_{cp}$. In our specific example, $N_{sc} = 64$, $N_{cp} = 16$, and thus $N_b = 80$. More details on the signal specifications are provided in our Github repository.[2] Fig. 2(b) shows the covariance structures of the resulting sources, whose cyclostationarity is evident.

We consider segments of length $N = 1280$, and $\mathcal{K}$ corresponding to SIR levels $[-30, 3]$ dB at 1.5 dB steps ($|\mathcal{K}| = 23$), namely we focus on the more challenging low SIR regime. We emphasize that the models (7) and (8) are assumed to be *unknown* once we have generated the dataset. Rather, we only have access to a dataset of unsynchronized samples.

Fig. 4 compares the MSE achieved by the linear MMSE and the oracle-synchronized MMSE. We note that, for this example, the true MMSE curve could not be obtained in practice due to the size of parameter space, rendering the computation of the posterior infeasible. Nevertheless, the MSE of the oracle-synchronized MMSE solution—albeit not a realizable estimator, as established earlier—serves as a lower bound.

As observed from Fig. 4, the best performing U-Net, which was trained on unsynchronized data, outperforms the linear MMSE estimator, and is close to the performance of the oracle MMSE (e.g., about 1.2 dB away at SIR levels be-

tween -9 and -30 dB). We also highlight that the choice of U-Net architecture to achieve such a performance benefits from specific domain knowledge. For example, capturing the temporal structures on the order of the signals' effective correlation length yielded significantly improved performance—for which long kernels on the first layer is a way of doing so.[3]

6. CONCLUSION

We studied the SCSS problem when signal models of the underlying latent sources are unknown, and developed a data-driven, DL approach as a prospective solution. Further, we characterized the setup for signals corresponding to segments extracted from cyclostationary Gaussian processes, and presented optimal MMSE estimators. We then outlined practical challenges associated with implementing the MMSE estimator, and through two examples, demonstrate how a trained U-Net with a careful choice of first-layer kernel size was able to achieve comparable performance, motivating DNNs as a competitive, computationally attractive approach.

We reiterate that in this work, the Gaussianity assumption allowed us to have a simplified closed-form expression for the MMSE estimator (5), and in turn a theoretical benchmark to compare against. In practice, signals depart from Gaussianity—for example, digital communication signals possess symbols from discrete sets. In these cases, the MMSE estimator is no longer linear that solely depends on (up to) second-order statistics and higher- statistics can be exploited to attain performance better than what is attained by (5). However, an analytical form of the optimal estimator is less obvious. This strongly motivates learning-based methods that hold the promise of improving upon implementable, optimal model-based estimators by learning those higher-order structures. Characterization and exploration of such performance gains via DL are of high interest for future work.

7. REFERENCES

[1] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," *Astronomy and computing*, vol. 18, pp. 35–39, 2017.

[2] N. Shlezinger and R. Dabora, "Frequency-shift filtering for OFDM signal recovery in narrowband power line communications," *IEEE Trans. Commun.*, vol. 62, no. 4, pp. 1283–1295, 2014.

[3] M. Zhao, X. Yao, J. Wang, Y. Yan, X. Gao, and Y. Fan, "Single-channel blind source separation of spatial aliasing signal based on stacked-LSTM," *Sensors*, vol. 21, no. 14, 2021.

[4] X. Chen, X. Xu, A. Liu, S. Lee, X. Chen, X. Zhang, M. J. McKeown, and Z. J. Wang, "Removal of muscle artifacts from the EEG: A review and recommendations," *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5353–5368, 2019.

[5] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[3]For comparisons to other DNNs for the communication example: `https://github.com/RFChallenge/SCSS_DNN_Comparison`

[6] D. Stoller, S. Ewert, and S. Dixon, "Wave-U-Net: A multi-scale neural network for end-to-end audio source separation," in *Proceedings of the 19th ISMIR*, 2018, pp. 334–340.

[7] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1652–1664, 2016.

[8] Y. Gandelsman, A. Shocher, and M. Irani, ""Double-DIP": Unsupervised image decomposition via coupled deep-image-priors," in *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 11 026–11 035.

[9] Y. Lyu, Z. Cui, S. Li, M. Pollefeys, and B. Shi, "Reflection separation using a pair of unpolarized and polarized images," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 14 559–14 569, 2019.

[10] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Joint optimization of masks and deep recurrent neural networks for monaural source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2136–2147, 2015.

[11] Y. Luo and N. Mesgarani, "Conv-TasNet: Surpassing ideal time–frequency magnitude masking for speech separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 8, pp. 1256–1266, 2019.

[12] E. Tzinis, Z. Wang, and P. Smaragdis, "Sudo RM-RF: Efficient networks for universal audio source separation," in *Proc. of MLSP*, 2020, pp. 1–6.

[13] M. G. Amin, "Interference mitigation in spread spectrum communication systems using time-frequency distributions," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 90–101, 1997.

[14] M. G. Amin, D. Borio, Y. D. Zhang, and L. Galleani, "Time-frequency analysis for GNSSs: From interference mitigation to system monitoring," *IEEE Signal Process. Mag.*, vol. 34, no. 5, pp. 85–95, 2017.

[15] C. M. Spooner and A. N. Mody, "Wideband cyclostationary signal processing using sparse subsets of narrowband subchannels," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 162–176, 2018.

[16] G. B. Giannakis, "Cyclostationary signal analysis," in *Digital Signal Processing Handbook*, V. K. Madisetti and D. Williams, Eds. CRC press, 1998.

[17] W. Gardner, *Cyclostationarity in Communications and Signal Processing*. IEEE Press, 1994.

[18] A. Baietto, J. Boubin, P. Farr, T. J. Bihl, A. M. Jones, and C. Stewart, "Lean neural networks for autonomous radar waveform design," *Sensors*, vol. 22, no. 4, 2022.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015*. Springer, 2015, pp. 234–241.

[20] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, 1995.

[21] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[22] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, tensorflow.org.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. of 3rd Int. Conf. Learn. Represent. ICLR*, 2015.

[24] MIT RLE, "RF Challenge - AI Accelerator," accessed 2021-11-01. [Online]. Available: https://rfchallenge.mit.edu