IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 31, 2021, accepted March 2, 2021, date of publication March 9, 2021, date of current version March 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064558

Deep-Unfolded Sparse CDMA: Multiuser Detector and Sparse Signature Design

SATOSHI TAKABE[1,2], (Member, IEEE), YUKI YAMAUCHI[1], AND TADASHI WADAYAMA[1], (Member, IEEE)

[1]Nagoya Institute of Technology, Nagoya 466-8555, Japan
[2]RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan

Corresponding author: Satoshi Takabe (s_takabe@nitech.ac.jp)

This work was supported in part by the JSPS Grants-in-Aid for Scientific Research (B), under Grant 19H02138 (TW), and in part by the Grant-in-Aid for Young Scientists (Startup), under Grant 17H06758 (ST).

ABSTRACT Sparse code division multiple access (SCDMA) is a promising non-orthogonal multiple access technique for future wireless communications. In SCDMA, transmitted symbols from multiple users are coded by their own sparse signature sequences, and a base station attempts to detect those symbols using the signature sequences. In this paper, we present a new deep-unfolded multiuser detector called a complex sparse trainable projected gradient (C-STPG) detector for SCDMA systems. Deep unfolding is a deep learning method that tunes trainable parameters in iterative algorithms using supervised data and deep learning techniques. The proposed detector provides a much superior detection performance over that of the LMMSE detector. Other advantages of the proposed detector include a low computational complexity in execution and a low training cost owing to the small number of trainable parameters. In addition, we propose a novel joint learning strategy called gradual sparsification for designing sparse signature sequences based on deep unfolding. This method is computationally efficient in optimizing a set of sparse signature sequences. Numerical results show that the gradual sparsification successfully yields sparse signature sequences with a smaller symbol error rate than those of randomly designed sparse signature sequences.

INDEX TERMS SCDMA, deep learning, deep unfolding, multiuser detector, signature design.

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) is one of the key components of the fifth generation (5G) and beyond in wireless communications. NOMA systems can improve the bandwidth efficiency by eliminating orthogonality in conventional orthogonal multiple access (OMA) systems. In addition to the bandwidth efficiency, highly reliable communications in overloaded systems have also been a crucial issue for increasing the areal capacity. Some NOMA techniques such as sparse code multiple access (SCMA) [1] have been proposed to deal with overloaded systems in which the number of users is larger than the number of communication resources.

Code division multiple access (CDMA) [2] is a well-known OMA technique in which multiple users with signature sequences communicate with a base station (BS) simultaneously. Although the orthogonality of signature sequences provides a reasonable multiuser detection performance, the computational complexity of a multiuser detector is relatively high; for example, an MMSE detector needs to compute the inverse of a dense matrix composed of signature sequences. Besides conventional optimization-based methods, some machine learning-based multiuser detectors have been proposed for practical channel models [3]–[5].

By contrast, sparsely spread CDMA (SCDMA) [6] is a variant of NOMA based on sparse signature sequences With SCDMA, each user has a sparse signature sequence that modulates a transmitted signal. Unlike CDMA, SCDMA uses sparse signature sequences with a small number of non-zero elements, which enables us to use a computationally efficient detection algorithm such as belief propagation (BP) [7]. Theoretical analyses show that the multiuser detection performance of SCDMA is comparable to that of CDMA even if a signature sequence contains only a constant number of non-zero elements [6], [8]. However, as discussed later, the computational cost of a BP detector depends

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

exponentially on the signature sparsity, which implies that reducing the computational complexity of a detector remains an important issue.

Signature design, that is, a proper choice of signature sequence, is another crucial issue in SCDMA owing to its sparsity and nonorthogonality. Previous studies have mainly focused on an optimization-based signature design for SCDMA and related systems. For example, Razavi *et al.* proposed a signature design for low-density signature orthogonal frequency-division multiplexing (OFDM) using a BP detector and an extrinsic information transfer chart analysis [9]. Xiao *et al.* presented an optimization of the signature sequences for multicarrier-low-density spreading multiple access using an interior-point algorithm [10]. Song *et al.* proposed a signature design for SCDMA systems by maximizing the code distance [11]. Although these methods exhibit a reasonable performance improvement compared with a random signature design, they are applicable to relatively small systems. Developing a sparse signature design method especially for large systems with hundreds of thousands of users is thus an open and critical problem for realizing a massive SCDMA system accommodating a large number of terminals.

In recent years, deep-learning techniques have been applied to various fields of wireless communication, such as massive MIMO detection [12] and multiuser detection in orthogonal frequency-division multiple access systems [13], [14]. The deep unfolding proposed by Gregor and LeCun [15] has recently been recognized as a promising deep-learning technique for signal processing [16]. Deep unfolding is applied to existing iterative algorithms to improve its convergence performance by unfolding the recursive structures and embedding some trainable parameters. These parameters can be learned using standard deep learning methods such as back propagation and stochastic gradient descent (SGD). One of the notable advantages of deep unfolding is that the number of trainable parameters is smaller than those of conventional deep neural networks, which results in a fast and stable training process. Deep unfolding has been applied to various algorithms in signal processing and wireless communications: sparse signal recovery [15], [17], [18], massive MIMO detection [19]–[21], signal detection for clipped OFDM systems [22], and a trainable decoder for LDPC codes [23].

Inspired by the notion of deep unfolding, the authors [24] proposed a trainable multiuser detector for SCDMA systems with binary phase shift keying (BPSK) modulation. The proposed sparse trainable projected gradient (STPG) detector requires less computational complexity than a standard BP detector, whereas it achieves almost the same detection performance as the BP detector. However, the STPG detector cannot directly deal with practical SCDMA systems with higher-order modulations such as 8PSK. This is because the STPG detector is designed only for real-valued channels.

For higher-order modulations, the computational complexity of a conventional BP detector is prohibitive because the time complexity of a node operation is exponential to the size of the signal constellation. Therefore, a simple detector such as a linear minimum mean squared error (LMMSE) detector is preferable in terms of the computational efficiency, although it exhibits a rather poor detection performance, particularly in overloaded cases. These facts suggest a strong need for an efficient and powerful multiuser detector for SCDMA systems with higher-order modulations.

The goal of this study is to pursue improved *massive* SCDMA systems with hundreds of thousands of users based on the deep unfolding approach. We first propose a deep-unfolded multiuser detector suitable for SCDMA systems with higher-order modulations such as 8PSK. The proposed detector, called a complex sparse trainable projected gradient (C-STPG) detector, is a non-trivial extension of an STPG detector [24] to the complex field. The use of *a complex MMSE function* matched to the signal constellation plays a crucial role in this extension. The dominant part of the iterative computation processes of the C-STPG detector consists of sparse matrix-vector products. This leads to a notable advantage in computational complexity compared with a conventional BP detector.

The second half of this paper is devoted to describing a *sparse signature design through a deep unfolding*. Another contribution of [24] is to clarify the potential of the deep-unfolding approach for a *sparse signature design*. In a nutshell, the idea is as follows. A signature sequence matrix is treated as trainable parameters, and an unfolded graph including the signature matrix is used for training these trainable parameters. It is experimentally shown that such a joint learning strategy for sparse signature sequences and the STPG detector is effective [24]. However, the training process shown in [24] exploits a projection onto a sparse binary matrix, that is, a *support matrix* to sparsify the signature matrix. The detection performance depends heavily on the support matrix, and an appropriate support matrix cannot be easily obtained, which is an unsolved drawback of the joint learning strategy in [24].

In this paper, we propose a new joint learning strategy for complex-valued signature sequences. A novel training strategy called *gradual sparsification* is introduced to circumvent the problem in [24]. In the field of structure learning for deep neural networks, sparsification of the network has recently attracted significant interest because a sparse neural network directly leads to power efficiency in a hardware implementation. The idea of gradual sparsification was inspired from studies such as [25]. For each training iteration, several elements in a trainable signature matrix convert small absolute values into zeros, allowing the signature matrix to gradually become a sparse matrix. In gradual sparsification, no binary support matrix is required for projection.

The outline of this paper is as follows. Section II describes the system model and Sec. III briefly introduces existing SCDMA detectors. In Section IV, we briefly review the STPG detector [24] for BPSK-based SCDMA systems. In Section V, we propose a C-STPG detector and compare its detection performance with that of the LMMSE detector in a

large-scale system. In Section VI, we propose a method for learning the signature matrix using a C-STPG detector and evaluate its performance. Section VII summarizes this paper.

II. MODEL SETTING

The SCDMA system discussed herein is defined as follows. Assume that N users try to send their own symbol $x_i \in \mathcal{X}$ $(i = 1, \ldots, N)$ in the signal constellation $\mathcal{X} \subset \mathbb{C}$ to a BS. Each user has a signature sequence $\boldsymbol{a_i} = (a_{1,i}, \ldots, a_{M,i})^T \in \mathbb{C}^M$ used for encoding the transmitted symbol. Here, we assume that $M \leq N$ holds. In particular, the system is overloaded when $M < N$. Equivalently, the system is overloaded if the overloaded factor $\beta := N/M$ is larger than 1. In addition, we assume that the signature matrix $\boldsymbol{A} := (\boldsymbol{a_1}, \ldots, \boldsymbol{a_N}) \in \mathbb{C}^{M\times N}$ satisfies the following conditions: (i) $\boldsymbol{A}$ is k-row sparse, that is, each row contains exactly k non-zero elements, and (ii) $\boldsymbol{A}$ is normalized such as $\|\boldsymbol{A}\|_F^2 = kM$, where $\|\cdot\|_F$ denotes the Frobenius norm. Under these assumptions, the received signal $\boldsymbol{y} = (y_1, \ldots, y_M)^T$ is given by

$$\boldsymbol{y} = \sqrt{\frac{\mathrm{SNR}}{k}} \sum_{i=1}^{N} \boldsymbol{a_i} x_i + \boldsymbol{w}, \tag{1}$$

where SNR represents the signal-to-noise ratio (SNR) of the system and $\boldsymbol{w} \in \mathbb{C}^M$ is an additive white Gaussian noise vector with zero mean and unit variance [7]. Equivalently, using the signature matrix $\boldsymbol{A}$, the system can be concisely represented by

$$\boldsymbol{y} = \sqrt{\frac{\mathrm{SNR}}{k}} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}. \tag{2}$$

Note that we used the implicit assumption that power and phase control for each active user is perfect to derive the model (2), that is, for simplicity, we will not deal with the near-far problem in this paper. In addition, we assume that complete frame synchronization is available among users. In the case of the real-valued system discussed in [24], all variables including the signal constellation, signature sequences, and noises take real values, and each element of the noise vector $\boldsymbol{w}$ independently follows the normal distribution. By contrast, this study deals with the complex-valued system. Signature sequences take complex values, and $\boldsymbol{w}$ is a complex Gaussian random vector with zero mean and unit variance. In the remaining parts of the paper, we mainly focus on QPSK $\mathcal{X} = \{e^{\mathrm{j}k\pi/4}; k = 0, 1, 2, 3\}$ and 8PSK $\mathcal{X} = \{e^{\mathrm{j}k\pi/8}; k = 0, 1, \ldots, 7\}$ modulations, where $\mathrm{j} := \sqrt{-1}$, although the argument is not limited to QPSK and 8PSK.

III. CONVENTIONAL SCDMA DETECTOR

This section describes well-known detectors for SCDMA systems. These detectors are used for the baselines in the experimental performance, as discussed later.

A. LMMSE DETECTOR

The LMMSE detector is a linear detector for multiple access channels. The detector for the model in (2) is given by the following:

$$\hat{\boldsymbol{x}} = \sqrt{\frac{\mathrm{SNR}}{k}} \boldsymbol{A}^H \left(\boldsymbol{I}_N + \frac{\mathrm{SNR}}{k} \boldsymbol{A}\boldsymbol{A}^H \right)^{-1} \boldsymbol{y}. \tag{3}$$

Although this detector works well in OMA systems, its detection performance significantly degrades when the system is overloaded. Whereas the computational complexity is $O(N^3)$ due to a matrix inversion, the inverse matrix can be pre-computed and used for fast computation within the coherence time.

B. BP DETECTOR

BP is a message-passing algorithm known as an effective detector for SCDMA systems [7]. The update rule of the BP is based on a factor graph whose nodes are variables $\boldsymbol{x}$ and $\boldsymbol{y}$, and whose edges are set according to non-zero elements of $\boldsymbol{A}$. Under the case of BPSK modulation, the message $U_{j\to i}(x)$ $(x \in \{+1, -1\})$ represents a message from a function node y_j to a variable node x_i, and $V_{i\to j}(x)$ is a message from a variable node x_i to a function node y_j. The update rule is then given by the following:

$$V_{i\to j}(x) = Z_{i\to j}^{-1} \prod_{l\in\partial i\setminus j} U_{l\to i}(x), \tag{4}$$

$$U_{j\to i}(x) = Z_{j\to i}^{-1} \sum_{\boldsymbol{x}_{\partial j\setminus i}} \left(\prod_{k\in\partial j\setminus i} V_{k\to j}(x_k) \right) \times \exp\left\{ -\frac{1}{2} \left[y_j - \sqrt{\frac{\mathrm{SNR}}{k}} \left(a_{j,i}x + \sum_{k\in\partial j\setminus i} a_{j,k}x_k \right) \right]^2 \right\}, \tag{5}$$

where $Z_{i\to j}$ and $Z_{j\to i}$ are normalization constants, and $\partial i := \{j \in \{1, \ldots, M\} | a_{j,i} \neq 0\}$ and $\partial j := \{i \in \{1, \ldots, N\} | a_{j,i} \neq 0\}$ are neighboring node sets on the factor graph. After a given number of iterations, the probability that the ith transmitted signal is equal to x is estimated by the following:

$$V_i(x) = Z_i^{-1} \prod_{j\in\partial i} U_{j\to i}(x), \tag{6}$$

where Z_i is a normalization constant. Finally, the ith transmitted signal is detected as $\hat{x}_i = 1$ if $V_i(1) \geq V_i(-1)$, and $\hat{x}_i = -1$ otherwise.

The BP detector exhibits a reasonable detection performance theoretically and practically [7]. In addition, it is computationally efficient in terms of N because an iteration step runs in $O(N)$ when the number of non-zero elements in $\boldsymbol{A}$ is $O(1)$. However, a detailed analysis shows that the computational cost for a BPSK-based system is $O(k^2 2^{k-1} N)$ per iteration because (5) contains a sum over all possible combinations of $k-1$ transmitted signals $\boldsymbol{x}_{\partial j\setminus i}$. This suggests that the cost increases rapidly if the sparsity k increases.

TABLE 1. Number of operations in STPG and BP detectors per iteration, and values for various k when $N = M = 1200$ ($\beta = 1$).

	Number of operations	$k=2$	$k=4$	$k=6$	Big-O notation
STPG additions	$(2\beta^{-1}k+\beta^{-1}+1)N$	7.20×10^3	1.20×10^4	1.68×10^4	$O(kN)$
BP additions	$(k2^k+2)\beta^{-1}kN$	2.40×10^4	3.16×10^5	2.77×10^6	$O(k^2 2^k N)$
STPG multiplications	$(\beta^{-1}k+\beta^{-1}+2)N+1$	6.00×10^3	8.40×10^3	1.08×10^4	$O(kN)$
BP multiplications	$\{(2k+3)2^k+2\beta^{-1}k\}\beta^{-1}kN$	7.68×10^4	8.83×10^5	6.99×10^6	$O(k^2 2^{k-1}N)$

The issue of computational complexity of the BP detector becomes more serious for higher-order modulations. For the PSK modulations, the number of message types is proportional to the constellation size $|\mathcal{X}|$, the computational complexity of an iteration step becomes $O(k^2|\mathcal{X}|^{k-1}N)$. Similarly, for the QAM modulations, the computational complexity of an iteration step is $O(k^2|\mathcal{X}|^{(k-1)/2}N)$ because real and complex part of signals can be divided. These indicate that the time complexity is exponential to the size of the signal constellation. For example, the execution time for $k = 12$ is 256 times longer than that for $k = 6$. Practically, the BP decoder with 30 iterations executes in about 2 seconds on a PC with Intel Core i9-10850K CPU for $(N, M, k) = (1200, 1200, 6)$ with the BPSK modulation ($|\mathcal{X}| = 2$) indicating that the execution time for $(N, M, k) = (1200, 1200, 12)$ is over 8 minutes. Similarly, the execution time for $(N, M, k) = (1200, 1200, 6)$ with the 8PSK modulation and 16QAM modulation is respectively expected to be about 30 and 1 minute(s). These mean that the BP becomes a less attractive solution when we consider a higher-order modulation format owing to its huge time complexity.

IV. STPG DETECTOR

Before presenting the C-STPG detector, we briefly review the STPG detector [24] for BPSK-based SCDMA systems in this section.

A. STRUCTURE OF STPG DETECTOR

The STPG detector is based on an optimization algorithm called *the projected gradient descent*. The recursive relations of the projected gradient descent (PG) is given as follows:

$$\boldsymbol{r}_t = \boldsymbol{s}_t + \gamma\sqrt{\frac{\mathrm{SNR}}{k}}\boldsymbol{A}^T\left(\boldsymbol{y} - \sqrt{\frac{\mathrm{SNR}}{k}}\boldsymbol{A}\boldsymbol{s}_t\right), \tag{7}$$

$$\boldsymbol{s}_{t+1} = \tanh(\alpha\boldsymbol{r}_t), \quad (t = 0, 1, \ldots, T) \tag{8}$$

where $\tanh(\cdot)$ operates in a component-wise manner. In the first step (7) of PG, the estimate is updated by a gradient descent with a step size parameter $\gamma > 0$. In the second step (8), the estimate is mapped to a point close to a BPSK signal $\mathcal{X} = \{1, -1\}$ according to a soft projection function $\tanh(\cdot)$. We call the process (8) a *soft projection*, which is similar to a proximal operation in proximal methods. The parameter α is called a softness parameter that tunes the strength of the soft projection. It is expected that the detection performance depends on two parameters, γ and α. However, the heuristic tuning of these parameters is costly and impractical.

The STPG detector realizes the data-driven tuning of the parameters in the PG using deep-learning techniques. Based on the notion of deep unfolding, we can expand the recursive structure of the PG and embed trainable parameters instead of γ and α. The update rule of an STPG detector is given as follows:

$$\boldsymbol{r}_t = \boldsymbol{s}_t + \gamma_t^2\sqrt{\frac{\mathrm{SNR}}{k}}\boldsymbol{A}^T\left(\boldsymbol{y} - \sqrt{\frac{\mathrm{SNR}}{k}}\boldsymbol{A}\boldsymbol{s}_t\right) \tag{9}$$

$$\boldsymbol{s}_{t+1} = \tanh(\alpha\boldsymbol{r}_t), \tag{10}$$

where $\{\gamma_t\}_{t=1}^T$ and α are trainable parameters. Note that the step size paramter γ_t^2 is squared because the value should be positive. These parameters are learned by using randomly generated supervised data and standard deep-learning techniques such as back propagation and SGD (see IV-C1 for details). The number of trainable parameters is $T + 1$ in T iterations, which results in a fast and stable training process.

B. COMPUTATIONAL COST

The computational cost of the STPG detector is $O(kN)$ because the computationally dominant part of the detector is composed of sparse matrix product operations. Table 1 shows the number of operations per iteration as a function of n, β, and k. In addition, we show the values when $N = M = 1200$ ($\beta = 1$) and $k = 2, 4, 6$ for comparison. It is found that both detectors are linear-time algorithms with respect to N. In particular, the use of a sparse matrix $\boldsymbol{A}^T$ in a gradient step helps the STPG detector reduce its complexity.

We also observe that the STPG detector requires fewer operations than the BP detector in terms of sparsity k. In fact, the STPG detector has $O(kN)$ additions/multiplications in each iteration, whereas the BP detector requires $O(k^2 2^{k-1}N)$ operations for the $|\mathcal{X}|$-PSK modulations, as discussed in Sec. III-B. In general, the constant k should be sufficiently large to ensure a reasonable detection performance, which results in an exponential increase in the BP computational cost.

C. NUMERICAL EXPERIMENTS

In this subsection, we present some numerical experiments of the STPG detector compared with a naive PG detector and a conventional BP detector.

1) EXPERIMENTAL SETUP

In the numerical simulations, we consider the BPSK-based SCDMA system with $N = 1200$ active users. A signature matrix $\boldsymbol{A}$ is randomly generated by an element-wise product

$\boldsymbol{A} = \boldsymbol{H} \odot \boldsymbol{W}$, where $\boldsymbol{H} \in \{0, 1\}^{M \times N}$ is a *support matrix* and $\boldsymbol{W} \in \mathbb{R}^{M \times N}$ is a weight matrix. In numerical simulations, each weight of $\boldsymbol{W}$ is uniformly chosen from $\{1, -1\}$. The support matrix $\boldsymbol{H}$ is also randomly generated by Gallager's construction [29] for low-density parity-check matrix such that its row and column weights are exactly equal to k and $k' = kM/N (\in \mathbb{N})$, respectively.

For the PG and STPG detectors, we set $T = 30$ as a number of iterations, which is sufficient for reasonable convergence. The STPG detector is implemented by using PyTorch 1.4 [27]. The training process is executed using incremental training [18], [21], which avoids a vanishing-gradient problem and provides reasonable results. In incremental training, we begin by learning the trainable parameters γ_1, α, assuming that $T = 1$. This is called the first generation of training. After the first generation is finished, we next train parameters $\gamma_1, \gamma_2, \alpha$ as if $T = 2$ by using the trained values of γ_1, α as their initial values. This process is repeated in an incremental manner until T reaches the desired value. In the following simulations, we use manually tuned parameters $\gamma = 0.01$ and $\alpha = 2$ for the PG detector. For the STPG detector, the initial values of the trainable parameters are set to $\gamma_t^2 = 0.01$ $(t = 1, \ldots, T)$, and $\alpha = 2$. In each generation, we prepare 100 mini batches of size 200 containing a pairing of a transmit signal $\tilde{\boldsymbol{x}}$ and received signal $\tilde{\boldsymbol{y}}$, and train parameters using the Adam optimizer [30] with a learning rate of 5.0×10^{-4} to reduce the mean squared error (MSE) loss between $\tilde{\boldsymbol{x}}$ and $\boldsymbol{s}^{(T)}$. Finally, the SNR of the system is defined as $10 \log_{10}$ SNR [dB]. The training process is executed for each SNR. As another baseline algorithm, we use the BP detector with 30 iterations.

2) ACCELERATION OF CONVERGENCE IN STPG

We first compared the STPG detector to the original PG to demonstrate the advantages of learning parameters by deep unfolding. Figure 1 shows the BER performance of the PG, STPG, and BP detectors with different SNRs. We found that

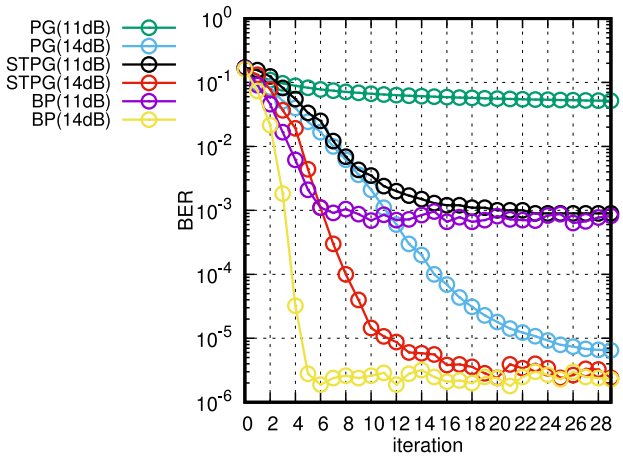


FIGURE 1. The BER of the PG, STPG, and BP detectors as a function of the number of iterations T with different SNRs: $N = 1000$, $M = 1200$ $(\beta = 1.2)$, and $k = 6$. The parameters of the PG are set to $\gamma = 0.01$ and $\alpha = 2$.

the STPG detector exhibits a better performance than the PG. For example, when the SNR is 11 dB, the BER of the STPG detector $(T = 30)$ is approximately 1.0×10^{-3}, whereas that of the PG is approximately 5.1×10^{-2}. In addition, when the SNR $= 14$ dB, the STPG detector shows a fast convergence to a fixed point compared with the PG. These results indicate that training a constant number of parameters in the PG leads to better detection performance and fast convergence to a fixed point. Improvements in the detection performance and convergence acceleration are crucial advantages of deep unfolding, as shown in other signal detection problems [18], [21]. It is also found that the convergent BER performance of the STPG detector is fairly close to the that of the BP detector. Although the BP detector only takes less than 10 iterations for convergence, the time complexity of the BP detector per an iteration is much larger than the STPG detector as discussed in Sec. III-B. This suggests that the STPG detector can reduce execution time without sacrificing the BER performance. This is a crucial advantage of the STPG detector over the BP detector.

3) PERFORMANCE COMPARISON WITH BP DETECTOR

Figure 2 shows the multiuser detection performance with different overloaded factors β when $n = 1200$ and $k = 6$. In the overloaded case, where $\beta = 1.2$ $(M = 1000)$, the two detectors exhibit a similar BER performance. Although the overloaded system suffers from an approximately 1 dB degradation in performance compared with the $\beta = 1$ case, both algorithms successfully detect transmit signals in the high SNR regime, which is an advantage of the SCDMA as a NOMA. In overloaded systems, the computational cost of a detector is still crucial because the signature sparsity k should be as sufficiently large as in the $\beta = 1$ case.

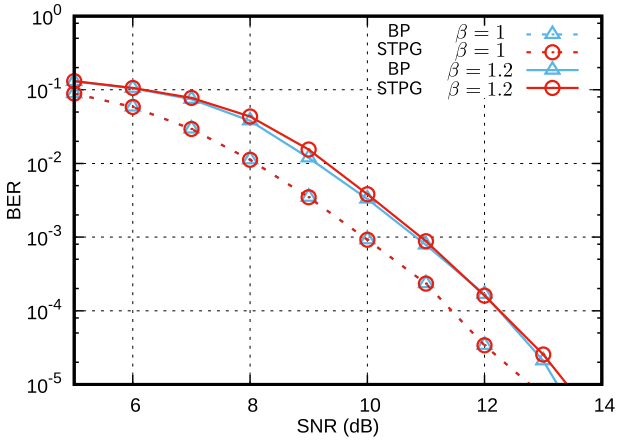


FIGURE 2. The BER performance of the STPG (circles) and BP detectors (triangles) for SCDMA with overloaded factor $\beta = 1$ (dotted line) and 1.2 (solid line); $N = 1200$, $M = N/\beta$, and $k = 6$.

To summarize, the STPG detector shows a fairy close detection performance to the BP detector even in an overloaded case. From the discussion in Sec. IV-B, we can conclude that the STPG detector has an advantage in terms of

the computational cost for a sufficiently large signature sparsity k. The drawback of the STPG detector is that the detector is inapplicable to systems with higher-order modulations such as 8PSK modulations. In the next section, we will resolve this issue by introducing the C-STPG detector.

V. C-STPG DETECTOR

In this section, we propose the C-STPG detector, which is an extension of the STPG detector. We then conducted numerical experiments to demonstrate the detection performance of the STPG detector.

A. STRUCTURE OF C-STPG DETECTOR

As discussed in the previous section, the STPG detector [24] is defined for BPSK modulation. For higher-order modulations, one should extend the detector to the complex field and appropriately replace the soft projection step .

The extended version, called the C-STPG detector, is defined as follows:

$$\boldsymbol{r}_t = \boldsymbol{s}_t + \gamma_t^2 \sqrt{\frac{\mathrm{SNR}}{k}} \boldsymbol{A}^H \left(\boldsymbol{y} - \sqrt{\frac{\mathrm{SNR}}{k}} \boldsymbol{A} \boldsymbol{s}_t \right), \tag{11}$$

$$\boldsymbol{s}_{t+1} = \eta(\boldsymbol{r}_t; \alpha), \tag{12}$$

$$\eta(r; \alpha) := \frac{\sum_{x \in \mathcal{X}} x \exp\left(-\frac{|r-x|^2}{\alpha}\right)}{\sum_{x \in \mathcal{X}} \exp\left(-\frac{|r-x|^2}{\alpha}\right)}, \tag{13}$$

where $\eta : \mathbb{C} \to \mathbb{C}$ is an MMSE function matched to a signal constellation $\mathcal{X}$ with a trainable parameter α, which is described in detail later. The soft projection step (12) consists of an element-wise map of η. The notation $\boldsymbol{A}^H$ represents the Hermitian transpose of $\boldsymbol{A}$. The C-STPG detector starts from $\boldsymbol{s}_0 = \boldsymbol{0}$ and outputs the estimate $\boldsymbol{s}_T$ after T iterations. Eq. (11) corresponds to the gradient descent step, as in that of the STPG detector (9).

Figure 3 shows the block diagram of the C-STPG detector in the Tth iteration (layer). The trainable parameters are α and $\{\gamma_t\}_{t=0}^{T-1}$. Therefore, the number of trainable parameters in T iterations is $T+1$, which is independent of the system size N, M.

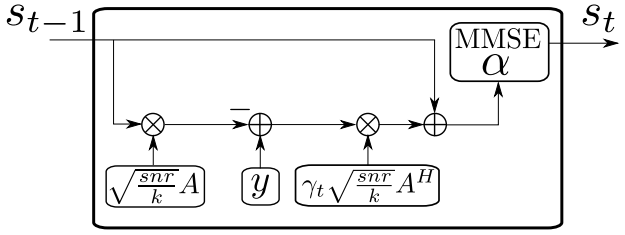


FIGURE 3. Block diagram of the Tth layer of the C-STPG detector. The trainable parameters are $\{\gamma_t, \alpha\}$. It takes inputs s_{t-1}, y and outputs s_t.

The main difference between the C-STPG detector and previous STPG detector is the use of the complex MMSE function η defined in (13). A justification of the use of the complex MMSE function η is given as follows. Let us consider a one-dimensional complex additive white Gaussian noise channel $y = x + w \in \mathbb{C}$, where $x \in \mathcal{X}$ and w is a Gaussian random vector with zero mean and variance α. In an iterative process of the C-STPG detector, it is assumed that an estimation error regarding an element follows an independent Gaussian distribution. For this reason, we introduce a virtual AWGN channel described above. Here, we consider an estimation problem to estimate x from y. It is known that the optimal estimate is given by the MMSE estimator. Then, the complex MMSE function is defined as follows:

$$\begin{aligned} \eta(y; \alpha) &:= E_x[x|y] \\ &= \sum_{x \in \mathcal{X}} x P(x|y) dx. \end{aligned} \tag{14}$$

Using the Bayes' theorem, we have

$$\eta(y; \alpha) = \frac{\sum_{x \in \mathcal{X}} x P(y|x) P(x) dx}{\sum_{x \in \mathcal{X}} P(y|x) P(x) dx}. \tag{15}$$

Assuming that $P(x) = 1/|\mathcal{X}|$, we obtain the complex MMSE function (13) [22]. Figure 4 shows the behavior of the MMSE function η for 8PSK modulation when $\alpha = 0.2$ and 0.6. We can see that the map η shows a soft projection or proximal map-like behavior toward the unit circle. An estimate signal is attracted more strongly to a signal point by the MMSE function as α becomes smaller. It should be noted that the trainable ISTA for sparse signal recovery [18] also successfully employs the MMSE shrinkage function.

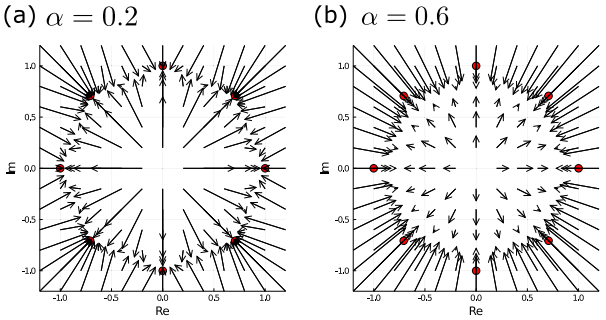


FIGURE 4. Behavior of complex MMSE function η for 8PSK modulation (red points) when (a) $\alpha = 0.2$ and (b) $\alpha = 0.6$. The start and end points of each arrow indicate the input and output of the function η, respectively.

Note that a real MMSE function with Gaussian noise and BPSK modulation corresponds to the soft projection $\tanh(\cdot)$ in the STPG detector. In this sense, the C-STPG detector is a natural extension of the STPG detector to a complex field.

B. COMPUTATIONAL COST

In addition to the STPG-detector, the computational cost of the C-STPG detector is $O(|\mathcal{X}|N + kN)$ for any signal constellation $\mathcal{X}$, whereas the BP and LMMSE detectors respectively take $O(k^2|\mathcal{X}|^{k-1}N)$ and $O(N^3)$. This indicates that the C-STPG detector is more computationally efficient than the BP detector for a high-order modulation format. The C-STPG detector is thus suitable for SCDMA systems with higher order-modulations.

C. NUMERICAL EXPERIMENTS

Here, we demonstrate the SER performances of the C-STPG detector for SCDMA systems with QPSK and 8PSK modulations. Since the BP detector requires a computation cost proportional to $|\mathcal{X}|^k$, we only compare the C-STPG detector with the LMMSE detector.

1) EXPERIMENTAL SETUP

In the following experiments, a transmitted signal is generated using a uniform distribution over $\mathcal{X}$, and the received signal $\boldsymbol{y}$ is generated according to the system model (2).

A random signature matrix $\boldsymbol{A}$ is generated as described in Sec. IV-C1. The weight matrix $\boldsymbol{W}$ is a complex Gaussian random matrix with zero mean and variance $\boldsymbol{I}_N$.

During the training process of the C-STPG detector, the number of iterations is set to $T = 30$. The initial values of the trainable parameters are set to $\gamma_t = 0.1$ $(t = 0, \ldots, T-1)$ and $\alpha = 0.5$. We employ incremental training such as in the training process of the STPG detector. Other learning settings are same as those in Sec. IV-C1. To estimate the symbol error rate (SER), we apply hard thresholding to the output of a detector. Namely, each estimated complex-valued signal is given by the closest signal point in $\mathcal{X}$.

2) EXPERIMENTAL RESULTS

We first show the multiuser detection performance under QPSK modulation. Figure 5 shows the experimental results for the $(N, M, \beta) = (1200, 1200, 1)$ case. In this setting, the C-STPG detector achieves a significantly lower SER than the LMMSE detector when the signature matrix $\boldsymbol{A}$ is dense, i.e., $k = N = 1200$. For the sparse case, $k = 12$, the C-STPG detector achieves a much smaller SER compared with the LMMSE detector as well. Comparing the two cases $k = 12$ and $k = 1200$, the performance degradation of the C-STPG detector with $k = 12$ is only 1 dB when the SER is 10^{-2}. It should be noted, however, that a sparser case such as $k = 6$ results in a significant performance degradation. Sparsity k should be set to a moderately large constant, such as $k = 12$, to maintain the SER performance.

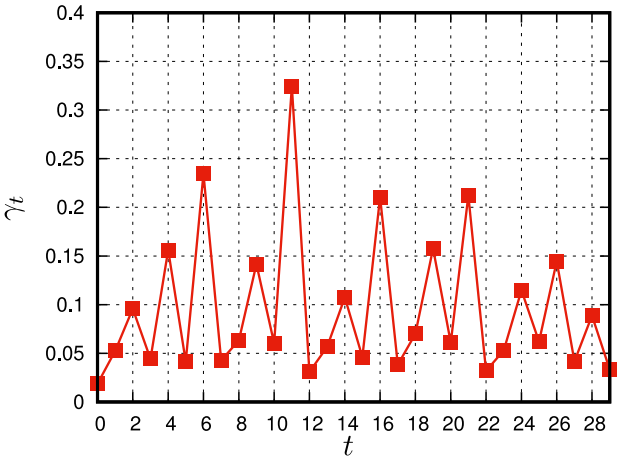


FIGURE 6. Learned $\{\gamma_t\}_{t=0}^{T-1}$ of the C-STPG detector for QPSK modulation when $(N, M, \beta) = (1200, 1200, 1)$ and SNR= 10 dB.

Figure 6 shows the learned $\{\gamma_t\}_{t=0}^{T-1}$ when SNR= 10 dB. We found that $\{\gamma_t\}_{t=0}^{T-1}$ exhibits a zigzag shape, which is similar to the learned parameters observed for other gradient descent-type algorithms such as TISTA [18]. The zigzag shape of step sizes is theoretically analyzed for a naive gradient descent algorithm as Chebyshev steps [28], which can be extended to a part of projection gradient descent algorithms. However, the theory is inapplicable to the C-STPG detector because its fixed point is nontrivial. Further theoretical analyses is a possible future work. Note that the learned α takes 0.366, indicating that a moderately strong soft projection is suitable for detection (see Fig. 4).

Figure 7 shows the results for an overloaded case with $(N, M, \beta) = (1200, 1000, 1.2)$. The C-STPG detector shows a smaller SER than that of the LMMSE detector regardless of the sparsity. Compared with Fig. 5, the performance degradation of the LMMSE detector is apparent, whereas that of the

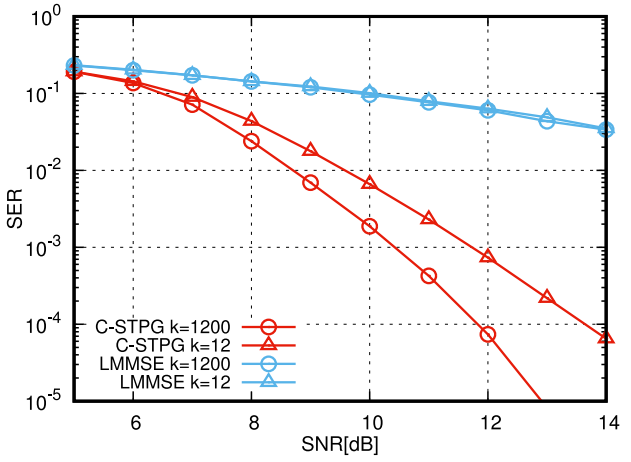


FIGURE 5. SER performance of C-STPG and LMMSE detectors under QPSK modulation with $(N, M, \beta) = (1200, 1200, 1)$.

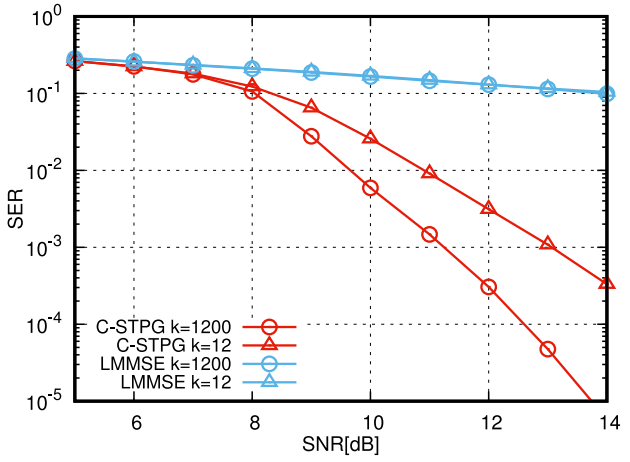


FIGURE 7. SER performance of C-STPG and LMMSE detectors with QPSK modulation in the overloaded case and $(N, M, \beta) = (1200, 1000, 1.2)$.

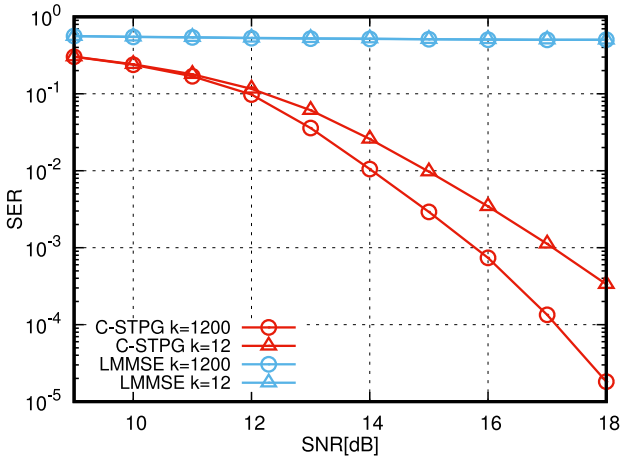


FIGURE 8. SER performance of C-STPG and LMMSE detectors with the 8PSK modulation with $(N, M, \beta) = (1200, 1200, 1)$.

C-STPG detector is not as large. This result implies that the C-STPG detector is also effective for SCDMA systems even under overloaded conditions.

Next, we show the multiuser detection performance of a SCDMA system with 8PSK modulation. Figure 8 shows the results for the $(N, M, \beta) = (1200, 1200, 1)$ case. The superiority of the C-STPG detector is still apparent. In this case, the detection performance of the C-STPG detector decreases by approximately 1 dB because of the sparse signature sequence when the SER is 10^{-2}. Figure 9 shows the results for the overloaded case in which $(N, M, \beta) = (1200, 1000, 1.2)$. We found that the C-STPG detector exhibits a better SER performance than the LMMSE detector, even in this case.

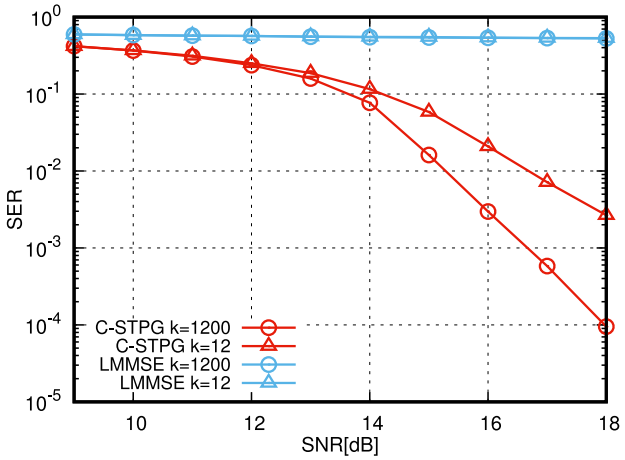


FIGURE 9. SER performance of C-STPG and LMMSE detectors with 8PSK modulation in overladed case and $(N, M, \beta) = (1200, 1000, 1.2)$.

In summary, we confirmed that the proposed C-STPG detector shows an excellent detection performance even for overloaded SCDMA systems with QPSK and 8PSK modulations.

VI. DEEP-UNFOLDING BASED SIGNATURE DESIGN

In the previous section, we proposed the C-STPG detector and demonstrated its SER performance. The detector is highly computationally efficient. and exhibits a superior detection performance compared with the LMMSE detector. However, the results also suggest that the detection performance for SCDMA systems is degraded compared with a dense case in which $k = N$. In a previous study [24], the possibility of deep unfolding for a signature sequence design is shown. In this section, we present a new joint learning strategy called *gradual sparsification* for a signature sequence design.

A. GRADUAL SPARSIFICATION

We can modify the training process of the STPG and C-STPG detector such that not only the trainable parameters of the detector but also the signature matrix $\boldsymbol{A}$ are updated by an optimizer.

The main difficulty of the strategy is how to implement the training process to force matrix $\boldsymbol{A}$ to be sparse, that is, all row weights of trained $\boldsymbol{A}$ should be equal to a given number k. In [24], only a weight matrix $\boldsymbol{W}$ is updated, whereas a binary support matrix $\boldsymbol{H}$ is fixed. Although this ensures the sparsity condition for $\boldsymbol{A} = \boldsymbol{H} \odot \boldsymbol{W}$, the resulting detection performance depends on $\boldsymbol{H}$, which is chosen heuristically.

To eliminate the performance dependence on $\boldsymbol{H}$, we propose a new learning strategy called gradual sparsification, as shown in Fig. 10. Gradual sparsification is closely related to edge pruning methods for deep neural networks, e.g., [25].

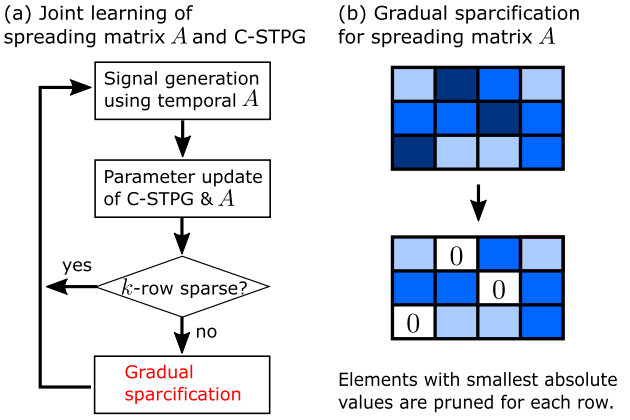


FIGURE 10. (a) Flowchart of the joint learning strategy for the C-STPG detector and signature matrix. (b) Schematic diagram of gradual sparsification to update the signature matrix.

The pseudocode of gradual sparsification is shown in Alg. 1. First, a support matrix is initialized to an all-one matrix. In the pseudocode, Lines 4 and 5 represent the masking and weight normalization of $\boldsymbol{A}$ using a tentative $\boldsymbol{H}$. Lines 6-9 describe the standard incremental training for the C-STPG detector. In Line 10, an optimizer updates the weights of $\boldsymbol{A}$ in addition to the trainable parameters of the detector. The remaining part corresponds to gradual sparsification for the signature matrix. For every I mini-batch,

Algorithm 1 Joint Learning With Gradual Sparsification

Input: M, N, T, k, SNR, mini-batch size bs, number of mini-batches B, sparsification interval I.

Output: Trained params. $\{\gamma_t\}, \alpha, \boldsymbol{A}$

1: Initialize $\{\gamma_t\}$, α, and $\boldsymbol{A}$. Let $\boldsymbol{H}$ be an $M \times N$ support matrix whose elements are 1. Set $l = N$.
2: **for** $t = 1$ to T **do**
3: **for** $b = 1$ to B **do**
 ▷Masking and normalization of $\boldsymbol{A}$.
4: $\boldsymbol{A} := \boldsymbol{A} \odot \boldsymbol{H}$
5: $\boldsymbol{A} := (\sqrt{kM}/\|\boldsymbol{A}\|_F)\boldsymbol{A}$
 ▷Generating training data.
6: Generate $\boldsymbol{x} \in \{1, -1\}^{N \times bs}$ randomly.
7: Generate $\boldsymbol{y}$ by $\boldsymbol{y} = (\sqrt{\mathrm{SNR}/k})\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$.
 ▷Update of training parameters.
8: Calculate an estimate $\hat{\boldsymbol{x}} := \boldsymbol{s}_{t+1}$ of STPG.
9: Calculate MSE loss between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$.
10: Update $\{\gamma_t\}$, α, and $\boldsymbol{A}$ by an optimizer.
 ▷Gradual Sparsification.
11: **if** $l > k$ and $b \bmod I = 0$ **then**
12: **for** $i = 1$ to N **do**
13: Select index j with smallest $|A_{ij}|$ and $H_{ij} \neq 0$.
14: $H_{ij} := 0$.
15: **end for**
16: $l := l - 1$.
17: **end if**
18: **end for**
19: **end for**

the following sparsification process is executed, as shown in Lines 11–17. For each row of $\boldsymbol{A}$, we choose the index j such that $H_{ij} = 1$ and $|A_{ij}|$ takes the smallest absolute values among all elements in the ith row where i represents the row index. Then, the value of H_{ij} is set to zero, which ensures that $A_{ij} = 0$ after this update. Clearly, the row weights of $\boldsymbol{H}$ and $\boldsymbol{A}$ decreases by one for every sparsification process. If the row weights reach a given number k, the sparsification stops.

Gradual sparsification enables us to tune not only the weights but also the locations of non-zero elements in the signature sequences. In addition, the column weights of a learned signature matrix can vary, which may improve the detection performance. The hyperparameter I representing the sparsification interval determines the number of updates of $\boldsymbol{A}$ during sparsification. The value of I should be moderately large because the choice in sparsification depends on the learned weights of $\boldsymbol{A}$.

B. EXPERIMENTAL RESULTS

In the following experiments, we demonstrate the SER performances of an SCDMA system optimized by a gradual sparsification. We compare its SER performance to a learned C-STPG detector with randomly generated sparse signature matrices. The experimental condition basically follows Sec. V-C1. As the number of trainable parameters increases, we change the number of mini-batches per generation to 1000. In addition, we set the sparsification interval I to 10.

Figure 11 shows a transition of a signature matrix representing a gradual sparsification process; (a) an initial random signature matrix of $k = 60$, (b) a signature matrix of $k = 35$, and (c) a trained signature matrix of $k = 12$ when $N = M = 60$ and SNR=10 dB. We found that the number of non-zero elements successfully decreases using gradual

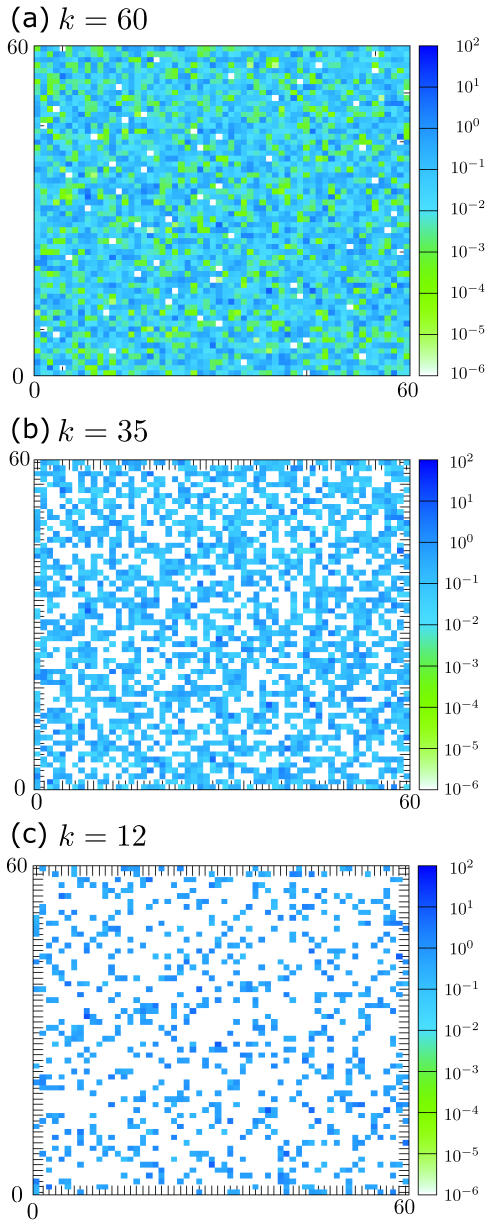


FIGURE 11. Absolute values of elements in a signature matrix in a gradual sparsification process when $N = M = 60$ and the SNR=10 dB. The absolute value of each element in the signature matrix is represented according to the color map at the right side of these figures. (a) $k = 60$ (initial random signature matrix), (b) $k = 35$, and (c) $k = 12$ when the gradual sparsification is completed.

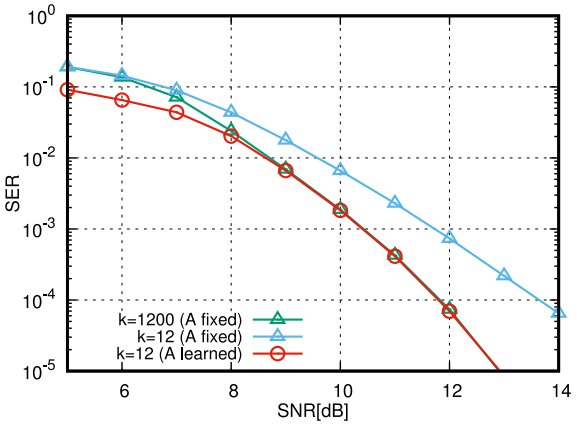


FIGURE 12. SER performance of the QPSK-based SCDMA system using a C-STPG detector with learned signature sequences (A learned) and random signature sequences (A fixed) when $(N, M, \beta) = (1200, 1200, 1)$.

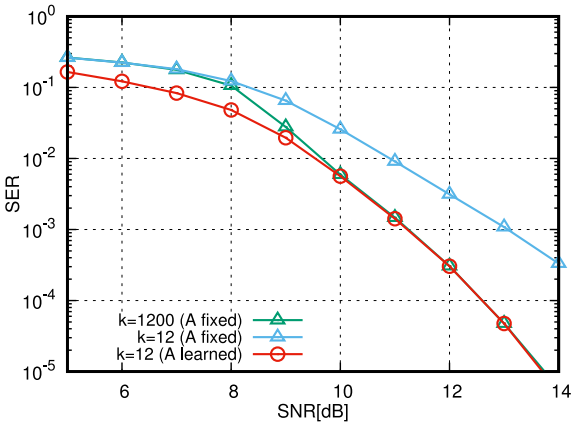


FIGURE 13. SER performance of QPSK-based overloaded SCDMA system using a C-STPG detector with learned signature sequences (A learned) and random signature sequences (A fixed) when $(N, M, \beta) = (1200, 1000, 1.2)$.

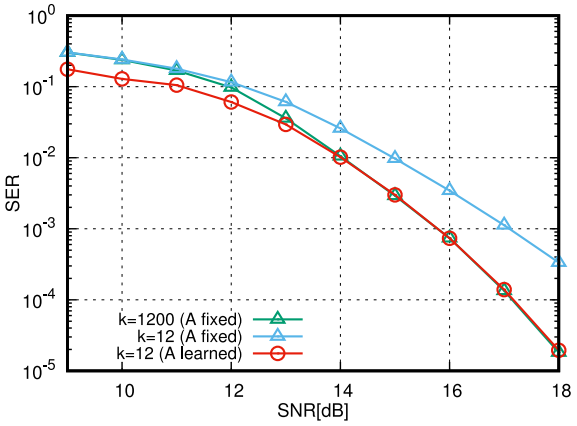


FIGURE 14. SER performance of 8PSK-based SCDMA system using a C-STPG detector with learned signature sequences (A learned) and random signature sequences (A fixed) when $(N, M, \beta) = (1200, 1200, 1)$.

sparsification. The absolute values of the non-zero elements remain relatively large owing to the normalization.

Figures 12-15 show the SER performance of the C-STPG detector with learned signature sequences ($\boldsymbol{A}$ learned) and fixed random signature sequences ($\boldsymbol{A}$ fixed) in various cases.

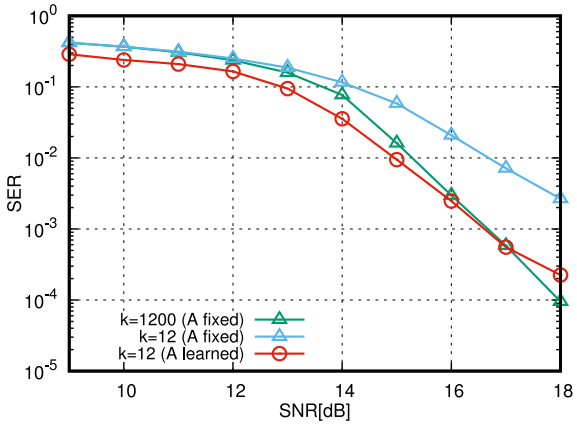


FIGURE 15. SER performance of 8PSK-based overloaded SCDMA system using a C-STPG detector with learned signature sequences (A learned) and random signature sequences (A fixed) when $(N, M, \beta) = (1200, 1000, 1.2)$.

A signature matrix of "$\boldsymbol{A}$ learned" is learned using gradual sparsification (Alg. 1) with $k = 12$. For comparison, we show two SER curves of "$\boldsymbol{A}$ fixed" in the case of sparsity $k = 12$ and $k = N$. We can see that gradual sparsification successfully improves the detection performance of SCDMA systems in all cases. The detection performance of the joint learning is fairly close to that of dense CDMA systems with $k = N$ in a high SNR region. Moreover, the SER performance of the learned signature matrix outperforms that of (S)CDMA with random signature matrices in a low SNR regime. We also find that gradual sparsification is effective even for overloaded systems (see Figs. 13 and 15). These results indicate that gradual sparsification reduces the computational complexity of the CDMA multiuser detector without sacrificing its detection performance. Remarkably, gradual sparsification can provide superior SER performance in the low SNR regime compared with dense randomly generated signature matrices.

Next, we compare gradual sparsification with existing optimal signature design. Although gradual sparsification is effective for large systems as shown above, existing optimal signature design methods mainly focus on small sytems. In such small sytems, we can use a maximum likelihood (ML) detector for mutiuser detection in terms of execution time. We thus compare gradual sparsification with existing optimal signature design in small systems using C-STPG and ML detectors.

Figure 16 shows the SER performance of the C-STPG and ML detector with learned signature sequences and optimal signature sequences in [11] when $(N, M, k) = (6, 4, 3)$. The optimal signature sequence is constructed to maximize the minimum distance among coded transmit signals [11], whereas the learned signature sequence is jointly learned with the C-STPG detector as described above. Although the SER performance of the C-STPG detector is degraded for small systems, gradual sparsification can improve the SER performance. Moreover, by applying the optimal ML detec-

TABLE 2. Example of mutual coherence of signature matrix before and after joint learning.

(N, M, k)	$(1200, 1200, 12)$	$(1200, 1000, 12)$	$(1200, 1200, 1200)$	$(1200, 1000, 1200)$
Before learning (random)	0.4491	0.5756	0.1133	0.1135
After learning (QPSK modulation)	0.3774	0.4575	0.1018	0.0980
After learning (8PSK modulation)	0.3775	0.4647	0.0991	0.0976

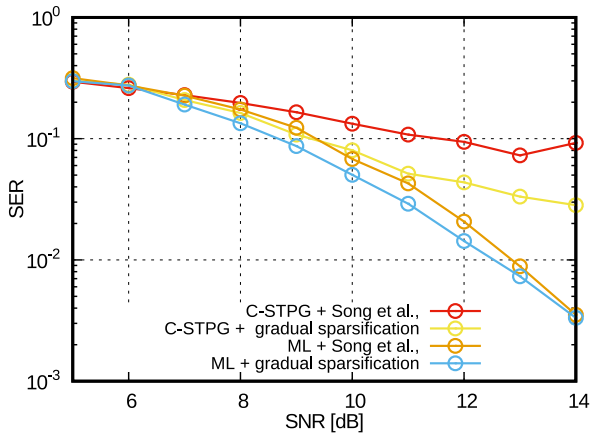


FIGURE 16. SER performance of QPSK-based overloaded SCDMA system using a C-STPG detector and ML detector with learned signature sequences and optimal signature sequences reported by Song *et al.* [11] when $(N, M, k) = (6, 4, 3)$.

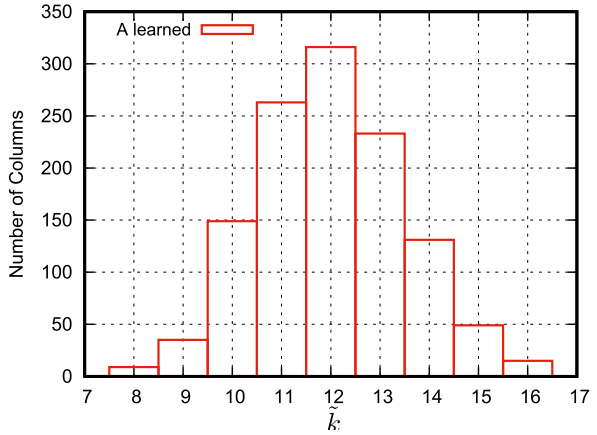


FIGURE 17. Frequency distribution of the number of non-zero elements in each column of a learned signature matrix under QPSK modulation with $(N, M, \beta) = (1200, 1200, 1)$ and SNR=10 dB.

tor, SER performance of both signature sequences improves. It is emphasized that the learned signature sequences using gradual sparsification outperforms existing optimal signature sequence. Note that the maximization of the minimum distance among coded transmit signals is computationally hard in general for large systems. These indicate the effectiveness of gradual sparsification for signature design.

It is conjectured that there are two possible reasons for the performance improvement. The first is the irregular structure of the learned signature matrix $\boldsymbol{A}$. In particular, the column weights $\tilde{k}$ of the learned $\boldsymbol{A}$ fluctuate around its average, as shown in Fig. 17. This differs from random signature matrices used in the experiments. This may improve the resulting detection performance, such as error-correcting performance of irregular low-density parity-check codes [31]. The second is the improvement of correlations called *mutual coherence*. For $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)$, the mutual coherence is given as $\max_{i \neq j} |\boldsymbol{a}_i{}^H \boldsymbol{a}_j|$. It is known that a smaller mutual coherence leads to a higher signal recovery performance in compressed sensing [32]. Table 2 shows the mutual coherence of $\boldsymbol{A}$ before/after joint learning during the experiments. We found that gradual sparsification reduces the mutual coherence in all cases. This suggests that signature sequences with smaller mutual coherence improve the detection performance in SCDMA systems such as compressed sensing. The theoretical justification of these conjectures is an important future problem.

VII. CONCLUDING REMARKS

In this paper, a deep-unfolding approach for improving the SER performance of SCDMA systems is investigated. A trainable SCDMA detector called the C-STPG detector is proposed for higher-order modulations. The detector contains only $T + 1$ trainable parameters in T iterations, which is constant based on the system size. In addition, the computational cost $O(|\mathcal{X}|N + kN)$ per iteration is much smaller than the conventional LMMSE and BP detectors in most cases. In the case of hardware implementation, we can also expect massive parallelism of the C-STPG detector, which may be competitive to known algorithms. Numerical results show that the C-STPG detector exhibits a remarkable multiuser detection performance even for overloaded systems.

In addition, a joint learning strategy for signature sequences based on gradual sparsification is proposed. Gradual sparsification enables us to tune not only weights of a signature matrix but also the location of its non-zero elements. Numerical experiments show that gradual sparsification successfully improves the detection performance, and the resulting improved performance is comparable to dense CDMA systems.

The results obtained in this paper provide insight into the detector design and signature design for SCDMA systems through deep unfolding. The framework shown in this paper may be applicable for improving sparse code multiple access systems, but this remains an open topic for future studies.

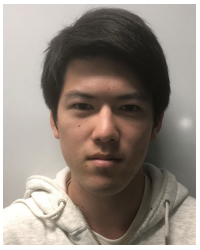
ACKNOWLEDGMENT

This article was presented at the 2020 IEEE International Conference on Communications (ICC) workshop.

REFERENCES

[1] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 332–336.
[2] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag.*, vol. 35, no. 12, pp. 126–133, Dec. 1997.
[3] B. Aazhang, B.-P. Paris, and G. C. Orsak, "Neural networks for multiuser detection in code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 40, no. 7, pp. 1212–1222, Jul. 1992.
[4] J. J. Murillo-Fuentes, S. Caro, and F. Pérez-Cruz, "Gaussian processes for multi-user detection in CDMA receivers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 939–946.
[5] Y. Işık and N. Taşpınar, "Multiuser detection with neural network and PIC in CDMA systems for AWGN and Rayleigh fading asynchronous channels," *Wireless Pers. Commun.*, vol. 43, no. 4, pp. 1185–1194, Nov. 2007.
[6] M. Yoshida and T. Tanaka, "Analysis of sparsely-spread CDMA via statistical mechanics," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, USA, Jul. 2006, pp. 2378–2382.
[7] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.
[8] A. Montanari and D. N. C. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Punta del Este, Uruguay, Mar. 2006, pp. 160–164.
[9] R. Razavi, M. Al-Imari, M. A. Imran, R. Hoshyar, and D. Chen, "On receiver design for uplink low density signature OFDM (LDS-OFDM)," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3499–3508, Nov. 2012.
[10] K. Xiao, B. Xia, Z. Chen, J. Wang, D. Chen, and S. Ma, "On optimizing multicarrier-low-density codebook for GMAC with finite alphabet inputs," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1811–1814, Aug. 2017.
[11] G. Song, X. Wang, and J. Cheng, "Signature design of sparsely spread code division multiple access based on superposed constellation distance analysis," *IEEE Access*, vol. 5, pp. 23809–23821, Nov. 2017.
[12] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.
[13] A. Li, Y. Ma, S. Xue, N. Yi, R. Tafazolli, and T. E. Dodgson, "Unsupervised deep learning for blind multiuser frequency synchronization in OFDMA uplink," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[14] S.-M. Tseng, Y.-F. Chen, C.-S. Tsai, and W.-D. Tsai, "Deep-learning-aided cross-layer resource allocation of OFDMA/NOMA video communication systems," *IEEE Access*, vol. 7, pp. 157730–157740, 2019.
[15] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.
[16] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: A survey and some new directions," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Nanjing, China, Oct. 2019, pp. 266–271.
[17] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[18] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3113–3125, Jun. 2019.
[19] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Anaheim, CA, USA, Nov. 2018, pp. 584–588.
[20] S. Takabe, M. Imanishi, T. Wadayama, and K. Hayashi, "Deep learning-aided projected gradient detector for massive overloaded MIMO channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[21] S. Takabe, M. Imanishi, T. Wadayama, R. Hayakawa, and K. Hayashi, "Trainable projected gradient detector for massive overloaded MIMO channels: Data-driven tuning approach," *IEEE Access*, vol. 7, pp. 93326–93338, 2019.
[22] S. Takabe, T. Wadayama, and Y. C. Eldar, "Complex trainable ista for linear and nonlinear inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 5020–5024.
[23] T. Wadayama and S. Takabe, "Deep learning-aided trainable projected gradient decoding for LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 2444–2448.
[24] S. Takabe, Y. Yamauchi, and T. Wadayama, "Trainable projected gradient detector for sparsely spread code division multiple access," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Dublin, Ireland, Jun. 2020, pp. 1–6.
[25] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 1135–1143.
[26] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
[27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 1–4. [Online]. Available: https://pytorch.org
[28] S. Takabe and T. Wadayama, "Convergence acceleration via Chebyshev step: Plausible interpretation of deep-unfolded gradient descent," 2020, *arXiv:2010.13335*. [Online]. Available: http://arxiv.org/abs/2010.13335
[29] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Ulm, Germany, Jun. 1997, p. 113.
[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[31] M. Yang, W. E. Ryan, and Y. Li, "Design of efficiently encodable moderate-length high-rate irregular LDPC codes," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 564–571, Apr. 2004.
[32] G. D. Mixon, U. W. Bajwa, and R. Calderbank, "Frame coherence and sparse signal processing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint-Petersburg, Russia, Jul. 2011, pp. 663–667.

SATOSHI TAKABE (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in multidisciplinary sciences from The University of Tokyo, Japan, in 2012, 2014, and 2017, respectively. He is currently an Assistant Professor with the Department of Computer Science, Nagoya Institute of Technology, Japan. His research interests include signal processing, information theory, and machine learning. He is a member of IEICE, ORSJ, and JPS. He received the IEEE Information Theory Society Japan Chapter Young Researcher Best Paper Award in 2018.



YUKI YAMAUCHI received the B.E. degree from the Nagoya Institute of Technology, Japan, in 2019. His research interests include signal processing and machine learning.

TADASHI WADAYAMA (Member, IEEE) was born in Kyoto, Japan, on May 9, 1968. He received the B.E., M.E., and D.E. degrees from the Kyoto Institute of Technology, in 1991, 1993, and 1997, respectively. In 1995, he started to work with the Faculty of Computer Science and System Engineering, Okayama Prefectural University, as a Research Associate. From April 1999 to March 2000, he was with the Institute of Experimental Mathematics, Essen University, Germany, as a Visiting Researcher. In 2004, he moved to the Nagoya Institute of Technology, as an Associate Professor. Since 2010, he has been a Full Professor with the Nagoya Institute of Technology. His research interests include coding theory, information theory, and coding and signal processing for digital communication/storage systems. He is a member of IEICE.