Ran and Oh *J Wireless Com Network* (2021) 2021:58
https://doi.org/10.1186/s13638-021-01904-8

EURASIP Journal on Wireless Communications and Networking

RESEARCH

Open Access

Low-complexity sparse-aware multiuser detection for large-scale MIMO systems

Rong Ran[1*] and Hayoung Oh[2*]

*Correspondence: sunnyran@ajou.ac.kr; hyoh79@gmail.com
[1] Department of Electrical and Computer Engineering, Ajou University, World-Cup Road, Suwon, South Korea
[2] SKKU Institute for convergence, Sungkyunkwan University, Sungkyunkwan-ro, Seoul, South Korea

Abstract

Sparse-aware (SA) detectors have attracted a lot attention due to its significant performance and low-complexity, in particular for large-scale multiple-input multiple-output (MIMO) systems. Similar to the conventional multiuser detectors, the nonlinear or compressive sensing based SA detectors provide the better performance but are not appropriate for the overdetermined multiuser MIMO systems in sense of power and time consumption. The linear SA detector provides a more elegant tradeoff between performance and complexity compared to the nonlinear ones. However, the major limitation of the linear SA detector is that, as the zero-forcing or minimum mean square error detector, it was derived by relaxing the finite-alphabet constraints, and therefore its performance is still sub-optimal. In this paper, we propose a novel SA detector, named single-dimensional search-based SA (SDSB-SA) detector, for overdetermined uplink MIMO systems. The proposed SDSB-SA detector adheres to the finite-alphabet constraints so that it outperforms the conventional linear SA detector, in particular, in high SNR regime. Meanwhile, the proposed detector follows a single-dimensional search manner, so it has a very low computational complexity which is feasible for light-ware Internet of Thing devices for ultra-reliable low-latency communication. Numerical results show that the the proposed SDSB-SA detector provides a relatively better tradeoff between the performance and complexity compared with several existing detectors.

Keywords: Multiuser detector, Sparse-aware detector, MIMO systems

1 Introduction

Large-scale MIMO system which uses over tens antennas at both transmitter and receiver, promises to offer high data rates and has been identified as one of key techniques in modern wireless communications [1, 2]. Meanwhile, MIMO is also an emerging technology for supporting communications between enormous devices in IoT environments [3]. However, the multi-antenna interference brings the fundamental limiting characteristic for MIMO systems [5]. Therefore, for downlink multiuser MIMO systems, transmission technologies such as quadrature spatial modulation (QSM) [4] can be considered for reducing the interference for large-scale MIMO systems. While for uplink multiuser MIMO systems, the multiuser detection plays an important role for mitigating the multi-antenna interference.

A review of various detection techniques for uplink multiuser MIMO was provided in [6]. The optimal nonlinear detectors, for example, the maximum a posterior (MAP) or maximum-likelihood (ML) detector, performs an exhaustive search in the whole solution space, and thus is practically prohibitive because its large computational complexity which usually exponentially increases with respect to the number of antennas in MIMO systems. On the contrary, the linear detectors, such as the minimum-mean-square-error (MMSE) and zero-forcing (ZF) detectors have very low complexity but their performances are generally far from the performance bound, in particular for large-scale MIMO systems. Hence, it is desired to develop new multiuser detection methods for achieving a better tradeoff between the performance and complexity for large-scale MIMO systems.

Recently, numerous sparse-aware (SA) detectors in term of sparse signal processing, have been presented for uplink multiuser MIMO systems [7–10]. The SA detectors utilized the hidden sparsity of a residual error vector to refine the results achieved by a low-complexity linear detector (e.g., ZF or MMSE). More specifically, at the receiver with the channel matrix $\mathbf{H}$ being known, the transmitted signal vector $\mathbf{x}$ is first detected by using ZF or MMSE associated with a slicing function. By letting the original received signals ($\mathbf{y} = \mathbf{Hx} + \mathbf{n}$) minus the product of the detected symbol vector $\hat{\mathbf{x}}$ and the channel matrix (i.e., $\hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}} = \mathbf{He} + \mathbf{n}$ where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ and $\mathbf{n}$ represents a noise vector), we can achieve a sparse MIMO system with the residual error vector $\mathbf{e}$ being the inputs. Note that the residual error vector $\mathbf{e}$ is usually sparse because the symbol error rates (SERs) achieved by the underlying linear detector is generally small for practical SNR regimes, and then a SA detector is employed to acquire $\hat{\mathbf{e}}$ which is used in refining $\hat{\mathbf{x}}$. The entire process is illustrated in Fig. 1. Intuitively, we may consider to apply compressive sensing (CS) based techniques for detecting the sparse residual error vector $\mathbf{e}$. For example, in [7], a multipath matching pursuit (MMP) method [14] was considered to detect the residual error vector. [8] first identified the support of the sparse error vector using a CS based method [11]. Then, a very over-determined MIMO system is created by removing the detected supports. In the resultant system, the number of transmit dimensions is much less than the number of receive dimensions, such that a low-complexity linear detector (e.g., ZF or MMSE) is able to detect the remaining non-zero error symbols. Simulation results

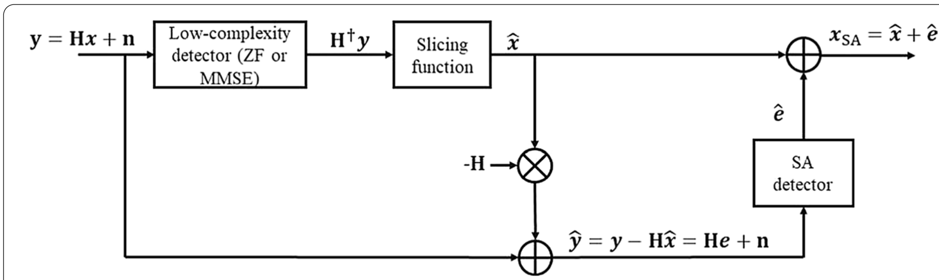

Fig. 1 The generic architecture of SA detectors

demonstrated that those CS-based SA detectors were efficient for improving the performance of the underlying linear detectors. However, most of the CS based methods were originally proposed for under-determined systems where the number of receive dimensions (or antennas) is much less than that of transmit dimensions (or antennas) and an iterative detecting process is usually required [12, 13], but MIMO systems, in particular uplink MIMO systems, are generally overdetermined, which refers that the number of transmit antennas is smaller than that of receive antennas. Therefore, the CS-based SA detectors are not appropriate for the overdetermined MIMO systems in sense of power consumption, execution time etc., especially in ultra-reliable low-latency communication (URLLC) service in 5G communication [15]. Taking the issue into account, a linear sparse-aware detector, named SA-MMSE, was presented in [9, 10], and has achieved attractive performance with a comparable complexity. Leveraging the hidden equality between the pseudo norm ($\|.\|_0$) and the second norm ($\|.\|_2$) when a lower-order modulation BPSK was considered, the conventional non-convex optimization problem on detecting the residual error vector $\mathbf{e}$ was converted into a convex one after relaxing the finite-alphabet constraints. An extension of the SA-MMSE for higher-order modulations was presented in [10]. After relaxing the finite-alphabet constraints of the residual error vector, a closed-form solution on the estimation of the residual error vector was achieved. Then, a well-designed slicing function was employed to map the estimations into the finite alphabet for achieving final decisions. Like the conventional ZF and MMSE detectors, the SA-MMSE detector follows a linear execution manner resulting in low computational complexity. However, this also renders the performance of SA-MMSE is sub-optimal because it relaxed the finite-alphabet constrains as the ZF/MMSE does, and its performance heavily depends on how precise the threshold of the slicing function is. In practical, it is extremely difficult to acquire the exactly accurate threshold due to a lot of unexpected dynamics of wireless channels. As SNR increases, the SA-MMSE detector probably suffers an error-floor. If the SA-MMSE detector takes the finite-alphabet constraints into account, its detection complexity will exponentially increase as the size of the finite alphabet again.

Therefore, in this paper, we aim at designing a novel detection scheme which not only adheres to the finite alphabet constraints for achieving better performance but also costs a comparable computational complexity. In the development of the proposed method, we can consider to utilize the QR decomposition to convert a multiple-dimensional search into a single-dimensional search as MIMO systems are generally overdetermined. Meanwhile, the unfavourable matrix inversion, required in the SA-MMSE detector proposed in [9], can also be avoided. Therefore, we can guarantee low computational complexity of the proposed method. Since we only need to do search in a single dimension, it is not necessary to relax the finite-alphabet constraints. Thus, no information will be lost due to the relaxation. Accordingly, we can expect that the performance of the proposed method should be better than the SA-MMSE detector. We verify through simulations that the proposed method achieves relatively better tradeoff between the complexity and performance compared to other existing detectors. The key contributions of this paper are summarized as follows.

1. We consider the same sparse MIMO system as that defined in [9] after applying a underlying detector (ZF or MMSE) and propose the SDSB-SA detector. Compared with several linear detectors, including the conventional ZF, MMSE and the SA-MMSE proposed in [9], our proposed SDSB-SA detector has superiority in terms of performance while costing a comparable computational complexity.
2. The QR decomposition is utilized to in the development of the SDSB-SA detector, which circumvents the matrix inversion operation required in the SA-MMSE detector and converts joint multi-dimensional search into a single-dimensional search in order to keep low computational complexity. Meanwhile, the single-dimensional search can preserve the finite-alphabet constraints for achieving better performance.
3. In order to avoid the possible error propagation for higher-order modulations, the layered SDSB-SA detector is developed by leveraging the hierarchical structure of the residual error vector.

This rest of paper is organized as follows. Section 2 describes the system model for an uplink MIMO transmission system and briefly overviews the main idea of SA-MMSE. The proposed single-dimensional search-based SA detector is presented in section 3. Simulation results and discussion are illustrated to verify the efficiency of the proposed detection methods in Sect. 4. Conclusion is given in Sect. 5.

1.1 Notation

The uppercase and lowercase boldface letters represent matrices and vectors, respectively. $\mathrm{Im}(x)$ and $\mathrm{Re}\,(x)$ state the imaginary and real parts of a complex-valued vector x, respectively. $\mathbf{X}^T$ represents the complex conjugate transpose of the matrix $\mathbf{X}$. $\mathbb{C}$ and $\mathbb{R}$ represent the entire complex and real domains, respectively.

2 System model

In this section, we first introduce an uplink multiuser MIMO system, and then briefly overview the main idea of SA detection.

2.1 Uplink multiuser MIMO

An uplink multiuser MIMO system is considered, as illustrated in Fig. 2, in which there is one base station (BS) with M antennas and the BS can service K single-antenna users (UEs) simultaneously. We define $\mathbf{x} \in \mathbb{C}^K$ as the symbol vector transmitted by the K UEs with all entries being selected from the same finite constellation set $\mathcal{A}$ (chosen according to a modulation scheme used in the system). At the BS, the received signal vector, denoted by $\bar{\mathbf{y}} \in \mathbb{C}^M$, is expressed as

$$\bar{\mathbf{y}} = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{n}}, \tag{1}$$

where $\bar{\mathbf{H}} \in \mathbb{C}^{M\times K}$ is the channel matrix whose elements are assumed to be independent and identically distributed with zero mean and unit variance over a rich-scattering communication environment as in [16], and $\bar{\mathbf{n}} \in \mathbb{C}^M$ is the additive noise whose entries are drawn from the complex-valued circularly-symmetric Gaussian distribution $\mathbb{CN}(0, \sigma^2)$. In this paper, we assume the channel matrix $\bar{\mathbf{H}}$ is perfectly known at the BS.

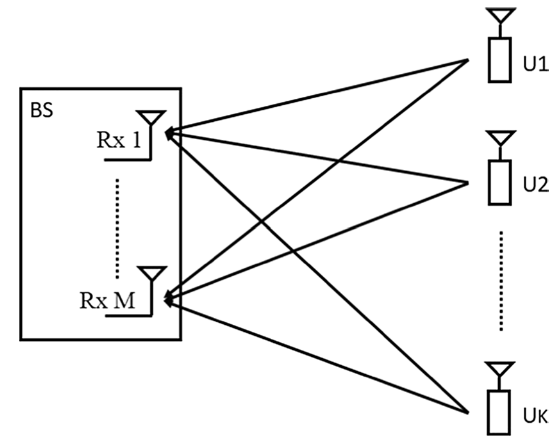

Fig. 2 The diagram of uplink multiuser MIMO systems

For the ease of representation and accommodating to the requirement of the proposed algorithm, we rewrite the complex input-output relationship in (1) into an equivalent real representation as,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathbf{y} = [\mathrm{Re}(\bar{\mathbf{y}})\ \mathrm{Im}(\bar{\mathbf{y}})]^T$, $\mathbf{x} = [\mathrm{Re}(\bar{\mathbf{x}})\ \mathrm{Im}(\bar{\mathbf{x}})]^T$, $\mathbf{n} = [\mathrm{Re}(\bar{\mathbf{n}})\ \mathrm{Im}(\bar{\mathbf{n}})]^T$ and

$$\mathbf{H} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{H}}) & -\mathrm{Im}(\bar{\mathbf{H}}) \\ \mathrm{Im}(\bar{\mathbf{H}}) & \mathrm{Re}(\bar{\mathbf{H}}) \end{bmatrix} \in \mathbb{R}^{2M \times 2K}$$

This real system representation will be used in the sequel.

2.1.1 The sparse-aware detecting

In this section, we briefly overview the methodology of the sparse-aware detectors. Given the received vector $\mathbf{y}$, the low-complexity underlying detector (e.g., ZF or MMSE) is first employed to detect the outputs as

$$\hat{\mathbf{x}} = \mathrm{Q}(\mathbf{W}\mathbf{y}) \tag{3}$$

where Q(.) refers to a slicing function and $\mathbf{W}$ is the weighted matrix of the underlying detector. The sparse MIMO system is achieved by taking the difference from the received signal to the product of the channel matrix and the output vector of the underlying detector, i.e.,

$$\hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}} = \mathbf{H}\mathbf{e} + \mathbf{n} \tag{4}$$

where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$, named a residual error vector (or symbol error vector). Note that the conventional ZF (or MMSE) detector associated with the symbol quantization can successively detect symbols with a certain possibility, namely, the detected symbol vector may be closed to but not always the same as the original transmit vector. Consequently, the residual error vector is sparse since the most of entries in $\mathbf{e}$ should be zeros with an overwhelming possibility in practical communication systems. Therefore, the sparse error vector $\mathbf{e}$ can be detected via the following optimization problem

$$\hat{\mathbf{e}} = \arg \min_{\mathbf{e} \in \hat{\mathcal{A}}^{2K}} \frac{1}{2} \|\hat{\mathbf{y}} - \mathbf{He}\|_2^2 + \lambda \|\mathbf{e}\|_0, \tag{5}$$

where $\hat{\mathcal{A}}$ represents the finite alphabet for symbol errors e_i , (e.g., $\hat{\mathcal{A}} = \{\pm 2, 0\}$ for BPSK), λ is a regularization factor related to the sparsity of the symbol error vector which is determined in term of the bit or symbol error rate achieved by the underlying linear detector. The theoretical analysis of ZF or MMSE on error rates is available in [17, 18] and can be used to calculate the value of λ. Finally, $\|\mathbf{e}\|_0$ states the pseudo-norm of $\mathbf{e}$ returning the number of non-zero entries of $\mathbf{e}$.

It is extremely difficult to tackle (5) since it is a non-convex function. Therefore, [7] and [8] had relaxed the l_0 norm to the l_1 norm and removed the finite-alphabet constraints. Then, a greedy iterative algorithm was employed to estimate the symbol error vector with the attendant high power and time consumption, which in turn preclude their practical implementation. In [9], when BPSK is considered, the mathmatical equality that $\|\mathbf{e}\|_0 = \frac{1}{4}\|\mathbf{e}\|_2^2$ holds. Therefore, the above optimization problem can be equivalently rewritten as

$$\hat{\mathbf{e}} = \arg \min_{\mathbf{e} \in \hat{\mathcal{A}}^{2K}} \frac{1}{2} \|\hat{\mathbf{y}} - \mathbf{He}\|_2^2 + \frac{\lambda}{4} \|\mathbf{e}\|_2^2, \tag{6}$$

By relaxing the finite-alphabet constraints $\mathbf{e} \in \hat{\mathcal{A}}^{2K}$, the estimation of $\mathbf{e}$, named $\tilde{\mathbf{e}}$ can be achieved via

$$\tilde{\mathbf{e}} = \left(\mathbf{H}^T \mathbf{H} + \frac{\lambda}{2} \mathbf{I} \right)^{-1} \mathbf{H}^T \hat{\mathbf{y}} \tag{7}$$

which is named SA-MMSE as it has a similar form as MMSE does except that λ represents the sparsity level of $\mathbf{e}$ rather than the noise variance. Finally, the detection of $\mathbf{e}$ is acquired via $\hat{\mathbf{e}} = \mathcal{Q}_\theta(\tilde{\mathbf{e}})$ where $\mathcal{Q}_\theta(.)$ is a slicing function with θ being a threshold. Similar to the ZF and MMSE detectors, the SA-MMSE detector follows a linear manner guaranteeing low computational complexity. However, it is also sub-optimal in sense of performance as the finite-alphabet constraints are relaxed. Therefore, in this paper, we develop a novel sparse-aware detector which not only adheres to the finite-alphabet constraints for better performance but also has low computational complexity.

3 Method of single-dimensional search-based sparse-aware (SDSB-SA) detector

In this section, we propose the single-dimensional search-based sparse-aware detector for large-scale uplink MIMO systems. First, we consider the case of lower-order modulation schemes. Then, we propose a layered SDSB-SA detector for higher-order modulation schemes in order to mitigate the possible effect of error propagation.

3.1 SDSB-SA for lower-order modulation schemes

Recall the sparse MIMO system model given in (4)

$$\hat{\mathbf{y}} = \mathbf{He} + \mathbf{n}. \tag{8}$$

Note that the channel matrix $\mathbf{H}$ is a tall matrix since the number of receive dimensions is larger than that of transmit dimensions for an uplink multiuser MIMO system. Therefore, the QR decomposition is implementable for the channel matrix $\mathbf{H}$ resulting in $\mathbf{H} = \mathbf{QR}$ where $\mathbf{R}$ is a $2K \times 2K$ upper triangular matrix and $\mathbf{Q}$ is an $2M \times 2K$ unitary matrix [19].

Let (8) be multiplied by $\mathbf{Q}^T$ at both sides, and we can achieve the following expression

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{Q}^T\hat{\mathbf{y}} \\ &= \mathbf{Q}^T\mathbf{He} - \mathbf{Q}^T\mathbf{n} \\ &= \mathbf{Re} + \tilde{\mathbf{n}}\end{aligned} \tag{9}$$

Since the unitary matrix $\mathbf{Q}$ does not change the Gaussian noise distribution, the $\tilde{\mathbf{n}}$ still satisfies Gaussian distribution with zero mean and covariance matrix $\sigma^2\mathbf{I}_M$. Consequently, the independence between the symbol error vector $\mathbf{e}$ and the noise vector $\tilde{\mathbf{n}}$ is preserved. By following the similar processes given in [9], the MAP detector can be equivalently rewritten as

$$\hat{\mathbf{e}} = \arg \min_{\mathbf{e}\in\hat{\mathcal{A}}^{2K}} \frac{1}{2}\|\tilde{\mathbf{y}} - \mathbf{Re}\|_2^2 + \lambda\|\mathbf{e}\|_0. \tag{10}$$

Let $\tilde{y}_k$ and $r_{k,i}$ represent the k-th element of $\tilde{\mathbf{y}}$ and the element at the k-th row and i-th column of $\mathbf{R}$, respectively. Then, by taking use of the triangular structure of $\mathbf{R}$, (10) can be equivalently represented as

$$\hat{\mathbf{e}} = \arg \min_{\mathbf{e}\in\hat{\mathcal{A}}^{2K}} \sum_{k=1}^{K}\left\{\frac{1}{2}\left(\tilde{y}_k - \sum_{i=k}^{K} r_{k,i}e_i\right)^2 + \lambda|e_i|_0\right\}. \tag{11}$$

From the above equation, we can observe that, the upper triangular form of $\mathbf{R}$ enables to successively detect e_k following a reversed order. Therefore, the original detection problem given in (10), which requires a multi-dimensional search in the finite alphabet $\hat{\mathcal{A}}$, can be equivalently decoupled into a set of single-dimensional search-based sub-problems. Specifically, given the detected symbols $\{\hat{e}_i\}_{i=k+1}^{2K}$, the $k-$th symbol can be detected via

$$\hat{e}_k = \arg \min_{e_k\in\hat{\mathcal{A}}} \frac{1}{2}\left(\tilde{y}_k - \sum_{i=k+1}^{K} r_{k,i}\hat{e}_i - r_{k,k}e_k\right)^2 + \lambda|e_k|_0 \tag{12}$$

This minimization problem entails only one scalar variable taking one of $(2 \times 2^n - 1)$ possible elements in $\hat{\mathcal{A}}$. However, since $\mathbf{R}$ usually is not a diagonal matrix, the proposed SDSB algorithm in (12) possibly suffers from error propagation, in particular, when a higher-order modulation scheme is employed. As a compromise, a layered single-dimensional search-based method is developed next, to further improve performance by adhering to the finite-alphabet constraints, at the price of slightly increased complexity compared to (12).

3.2 Layered SDSB-SA for higher-order modulation schemes

We consider a 2^{2n}-order QAM constellation (e.g., 16QAM, 64QAM etc) which results in that the residual error vector takes value from the finite alphabet $\mathcal{A} = \{0, \pm 2, \pm 4, \cdots \pm 2i, \cdots \pm 2(2^n - 1)$. The proposed layered SDSB-SA detector is inspired by the hierarchical structure of the residual error vector, according to which, the residual error vector can be decomposed into a set of orthogonal sub-error vectors, i.e.,

$$\mathbf{e} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_{2^n-1}, \tag{13}$$

where $\mathbf{v}_i \in \mathcal{A}_i^{2K}$ with $\mathcal{A}_i = \{\pm 2i, 0\}^{2K}$ for $i \in \{1, ..., 2^n - 1\}$. Note that $\hat{\mathcal{A}} = \mathcal{A}_1 \cup \mathcal{A}_2 \cdots \cup \mathcal{A}_i \cdots \cup \mathcal{A}_{2^n-1}$, and the sub-symbol error vectors $\mathbf{v}_i$ are orthogonal to each other, i.e., $\mathbf{v}_i \mathbf{v}_j^T = 0$ when $i \neq j$.

Therefore,

$$\arg \min_{\mathbf{e} \in \hat{\mathcal{A}}^{2K}} \frac{1}{2} \|\mathbf{Q}^T \hat{\mathbf{y}} - \mathbf{R}\mathbf{e}\|_2^2 + \lambda \|\mathbf{e}\|_0 \tag{14}$$

$$\begin{aligned} &= \arg \min_{\mathbf{e} \in \hat{\mathcal{A}}^{2K}} \frac{1}{2} \|\tilde{\mathbf{y}} - \sum_{i=1}^{2^n-1} \mathbf{R}\mathbf{v}_i\|_2^2 + \lambda \|\sum_{i=1}^{2^n-1} \mathbf{v}_i\|_0 \\ &=^{(a)} \arg \min_{\mathbf{e} \in \hat{\mathcal{A}}^{2K}} \frac{1}{2} \sum_{i=1}^{2^n-1} \left(\|\tilde{\mathbf{y}}\|_2^2 - \tilde{\mathbf{y}}(\mathbf{R}\mathbf{v}_i)^T - \mathbf{R}\mathbf{v}_i(\tilde{\mathbf{y}})^T + \|\mathbf{R}\mathbf{v}_i\|_2^2 \right) \\ &\quad + \lambda \sum_{i=1}^{2^n-1} \|\mathbf{v}_i\|_0 \\ &= \sum_{i=1}^{2^n-1} \arg \min_{\mathbf{v}_i \in \mathcal{A}_i^{2K}} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{v}_i\|_2^2 + \lambda \|\mathbf{v}_i\|_0 \end{aligned} \tag{15}$$

$^{(a)}$ comes from that $\mathbf{v}_i \mathbf{v}_j^T = 0$ for $i \neq j$ and $\|\tilde{\mathbf{y}}\|_2^2$ is known and regarded as a constant, adding this term to (15) will not affect the final result of (15). Consequently, the detection problem given in (10) is equivalent to

$$\begin{aligned} \hat{\mathbf{e}} &= \arg \min_{\mathbf{v}_1 \in \mathcal{A}_1^{2K}} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{v}_1\|_2^2 + \lambda \|\mathbf{v}_1\|_0 \\ &\quad + \arg \min_{\mathbf{v}_2 \in \mathcal{A}_2^{2K}} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{v}_2\|_2^2 + \lambda \|\mathbf{v}_2\|_0 \\ &\quad \vdots \\ &\quad + \arg \min_{\mathbf{v}_{2^n-1} \in \mathcal{A}_{2^n-1}^{2K}} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{v}_{2^n-1}\|_2^2 + \lambda \|\mathbf{v}_{2^n-1}\|_0 \\ &\qquad \text{subject to } \mathbf{v}_i \mathbf{v}_j^T = 0, i \neq j. \end{aligned}$$

With the relaxation on the orthogonality constraint, the objective function associated with $\mathbf{v}_i$ is given as

$$\hat{\mathbf{v}}_i = \arg \min_{\mathbf{v}_i \in \mathcal{A}_i^{2K}} \frac{1}{2} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{v}_i\|_2^2 + \lambda \|\mathbf{v}_i\|_0. \tag{16}$$

Similarly, the k-th entry of the sub-error vector $\mathbf{v}_i$ can be detected as

$$\hat{v}_k^i = \arg\min_{v_k^i \in \mathcal{A}_i} \frac{1}{2}\left(\tilde{y}_k - \sum_{j=k+1}^{K} r_{k,j}\hat{v}_j^i - r_{k,k}v_k^i\right)^2 + \lambda |v_k^i|_0. \tag{17}$$

Define

$$\bar{v}_k^i = \left(\tilde{y}_k - \sum_{j=k+1}^{2K} R_{k,j}\hat{v}_j^i\right)/R_{k,k}, \tag{18}$$

Then, the optimization problem in (16) can be simplified as

$$\hat{v}_k^i = \arg\min_{v_k^i \in \mathcal{A}_i} \frac{1}{2}\left(\bar{v}_k^i - v_k^i\right)^2 + \frac{\lambda}{r_{k,k}^2}|v_k^i|_0. \tag{19}$$

1 *Remark 1*

The main difference between (12) and (19) is that the former one entails one scalar variable taking one of possible $(2 \times 2^n - 1)$ *values in a full finite alphabet set* $\hat{\mathcal{A}}$*, whereas the later one only considers one of*

three values in a subset $\mathcal{A}_i$.

A two-layer reserved successive execution fashion is required to realize the entire detection processes. The inner-layer successive execution is to detect the sub-symbol errors $\mathbf{v}_i$ via (19) following a reversed order. Once the estimation $\hat{\mathbf{v}}_i$ is achieved, it is used to update the sparse system (4) via

$$\hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}(\hat{\mathbf{x}} + \hat{\mathbf{v}}_i). \tag{20}$$

Then, the resulting sparse system (20) is used in detecting the sub-symbol error vector $\mathbf{v}_{i-1}$ in the outer layer through (19) again. The outer layer *reversed* successive fashion can guarantee that the detected sub-symbol error vector $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{v}}_{i-1}$ can be orthogonal to each other as all previous sub-symbol error vectors do not effect on the detection of $\mathbf{v}_{i-1}$. Moreover, it may reduce the effect of error propagation by letting the sub-symbol error vector with a larger power level be detected first. That is because according to the results on probability of error for M-ary PAM or QAM in [20], the sub-symbol error vector with a larger power level has a smaller error probability than the sub-symbol error vector with a smaller power level. Thus, detecting the sub-symbol error vector with a larger power level first can reduce the effect of error propagation for the following sub-symbol error vector detection.

The proposed layered SDSB-SA detector for MIMO systems is officially summarized in Algorithm I given in Table 1.

Table 1 Layered single-dimensional search-based sparse-aware multiuser detection

Agorithm I	Layered SDSB-SA multiuser detection
Step 1:	Use ZF/MMSE detectors to achieve $\hat{\mathbf{x}}$
Step 2:	Reconstruct a sparse system via $\hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}$
Step 3:	Apply QR decomposition, i.e., $\tilde{\mathbf{y}} = \mathbf{Q}^T\hat{\mathbf{y}}$
Step 4:	Initializations: λ
Step 5:	**for** $i = 2^n - 1, (2^n - 1) - 1, \ldots, 1$ **do**
	for $k = 2K, 2K - 1, \ldots, 1$ **do**
	1. Compute $\bar{v}_k^i$ via (17)
	2. Map $\bar{v}_k^i$ into $\mathcal{A}_i$ via (19)
	end for
	3. Update the sparse system via $\hat{\mathbf{y}} = \bar{\mathbf{y}} - \bar{\mathbf{H}}(\hat{\mathbf{x}} + \hat{\mathbf{v}}_i)$
	end for
Step 6:	Attain the detected symbol error vector $\hat{\mathbf{e}} = \sum_{i=1}^{2^n-1} \hat{\mathbf{v}}_i$
Step 7:	Refine the detected vector through $\hat{\mathbf{x}} = \hat{\mathbf{x}} + \hat{\mathbf{e}}$

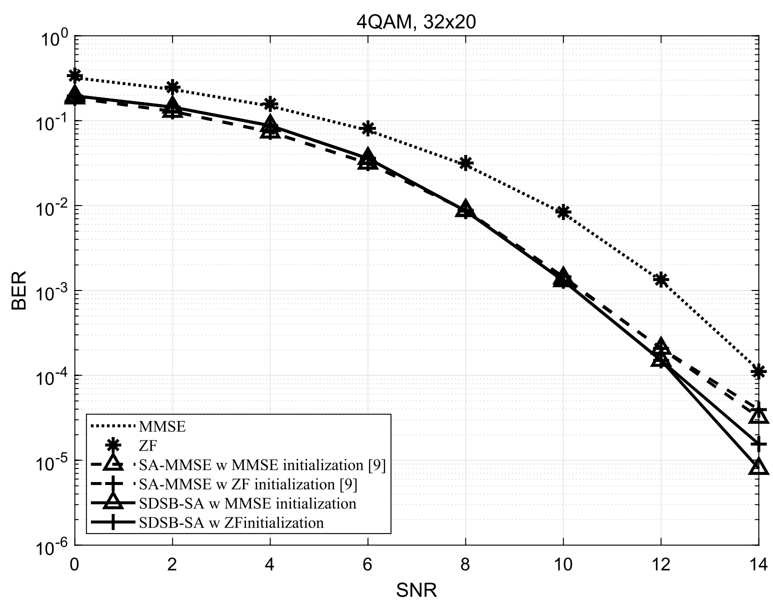

Fig. 3 Performance comparison of MMSE/ZF, MMSE/ZF+SA-MMSE and MMSE/ZF+SDSB-SA for 32x20 MIMO system, QPSK

4 Numerical results and discussion

In this section, the performances of the proposed SDSB-SA detector is evaluated for uplink multiuser MIMO systems. The conventional MMSE and ZF detectors are used as the underlying detectors providing the initialization for the SA-MMSE and the proposed SDSB-SA detectors, and also provides one of baselines for performance comparison. Another baseline is provided by the SA-MMSE detector.

In Fig. 3, we consider an overdetermined uplink MIMO system where $M = 32$, $K = 20$ and 4-QAM is applied.We compared the bit error rates (BERs) of the conventional ZF/MMSE detectors, the SA-MMSE with ZF/MMSE initialization and the proposed SDSB-SA detector with ZF/MMSE initialization. Since the modulation order is small (i.e., $n = 1$), we only consider to employ the proposed SDSB-SA detector given in (12). Simulation results demonstrate that the proposed SDSB-SA detector greatly improves the performance compared with the ZF and MMSE detectors. In particular, at BER $= 10^{-3}$, the SDSB-SA detector using either ZF or MMSE initialization achieves a 2 dB gain over the conventional ZF and MMSE detectors. Furthermore, the proposed SDSB-SA detector with the MMSE initialization performs as the same as the SA-MMSE with the same MMSE initialization in low SNR regimes, but becomes superior to the SA-MMSE detector in high SNR regimes. That is because the proposed SDSB-SA detector counts for the finite-alphabet constraints of error symbols. The similar results are also observed in the comparison with the SA-MMSE with the ZF initialization that the SDSB-SA detector with the ZF initialization performs the same as the SA-MMSE in low SNR regime and becomes better in high SNR regime. When the MIMO system becomes further overdetermined, i.e., $M \gg K$, as illustrated in Fig. 4, where $M = 32$ and $K = 8$, the conventional SA-MMSE detector performs even worse than the conventional MMSE/ZF when SNR is over 10 dB. That is because, for a very overdetermined MIMO system, the underlying MMSE or ZF detector can provide acceptable performances resulting in that the residual error vector is extremely sparse. Therefore, the SA-MMSE detector become more sensitive to the threshold used in the slicing function, and thus might be severally degraded by the sub-optimality caused by the relaxation of finite-alphabet constraints. Whereas, the proposed SDSB-SA detector provides much better performance than either the MMSE detector or the SA-MMSE detector since it adheres to the finite-alphabet constraints. In particular, the proposed SDSB-SA detector with either MMSE or ZF initialization, at BER $= 10^{-3}$, achieves around 1.2 dB gain over the SA-MMSE detector, and around 0.8 dB gain over the conventional ZF/MMSE detectors.

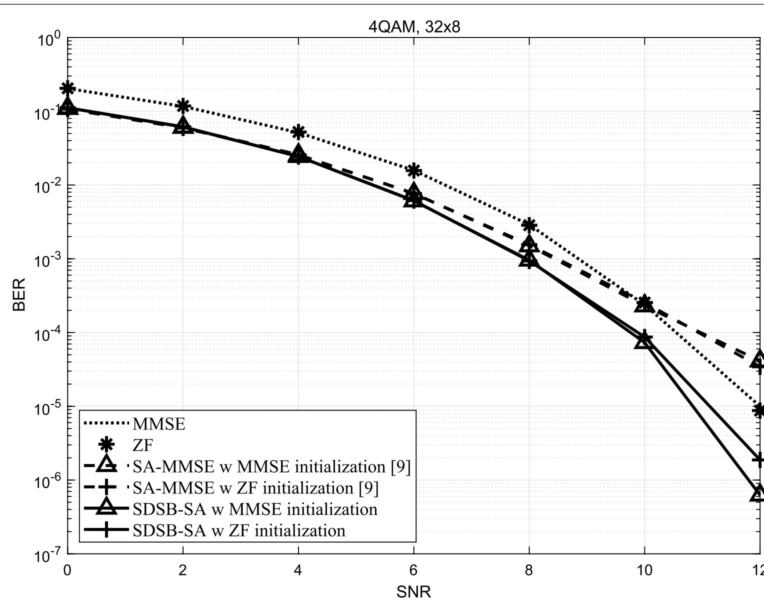

Fig. 4 Performance comparison of MMSE/ZF, MMSE/ZF+SA-MMSE and MMSE/ZF+SDSB-SA for 32 × 8 MIMO system, QPSK

Consider a higher-order modulation scheme, e.g.,16QAM, for the uplink MIMO system with $M = 32$ and $K = 20$. In this case, we only used the MMSE detector for the underlying detector providing the initialization for the SDSB-SA as the MMSE and ZF performs almost identically in an overdetermined MIMO system. We used SERs as the performance measurement for the comparison of the conventional MMSE detector, the SA-MMSE and the proposed SDSB-SA detector. From simulation results illustrated in Fig. 4, we still can observe that the proposed SDSB-SA detector given in (12) achieves better performance than the conventional MMSE detector, but becomes inferior to the SA-MMSE detector as SNR increases. That is because it may suffer from severe error propagation. Whereas, the proposed layered SDSB-SA detector given in (19) performs the best among all the existing detectors since it reduces the effects of error propagation by decomposing the residual error vector into a set of orthogonal sub-error vectors. More specifically, the proposed layered SDSB-SA detector achieves around 1.2 dB gain over the conventional MMSE detector, 0.8 dB over the SA-MMSE detector and 1 dB over the SDSB-SA detector given in (17). Moreover, the total $\mathcal{O}(64 \times 3 \times 3)$ number of complexities are required for the proposed layered SDSB-SA detector in (19) which is slightly increased compared with $\mathcal{O}(64 \times 7)$ required for the non-layered SDSB-SA detector in (12). The increment of computation cost is negligible. For reference, the complexities required by either the conventional MMSE/ZF or SA-MMSE are $\mathcal{O}(64 \times 1600)$.

The similar results can be also observed for a 32×8 MIMO system in Fig. 5. The proposed SDSB-SA detector performs better than the conventional MMSE detector but worse than the SA-MMSE detector due to the possible error propagation. The proposed layered SDSB-SA detector outperforms both the conventional MMSE and SA-MMSE detectors as it utilizes the residual error vector decomposition to mitigate the possible error propagation. Furthermore, its computational complexities ($\mathcal{O}(64 \times 3 \times 3)$) are

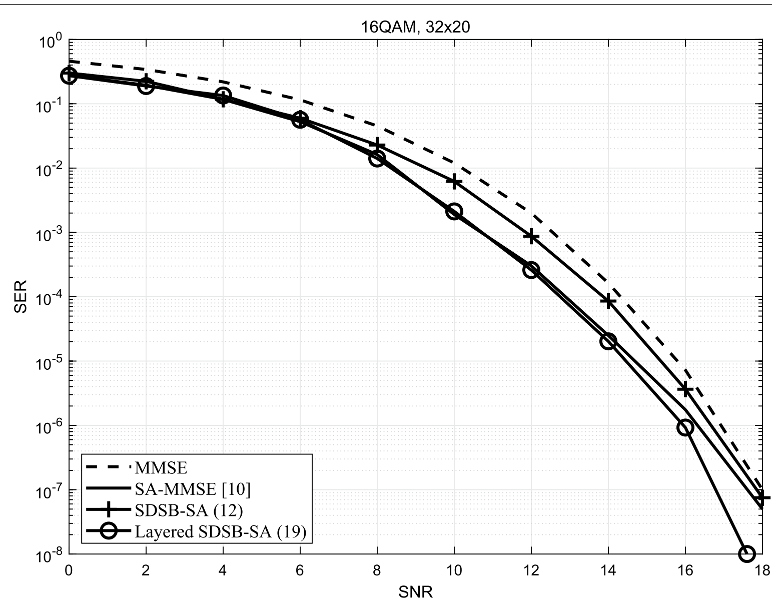

Fig. 5 Performance comparison of MMSE, SA-MMSE and Layered SDSB-SA detector for 32×20 MIMO system, 16QAM

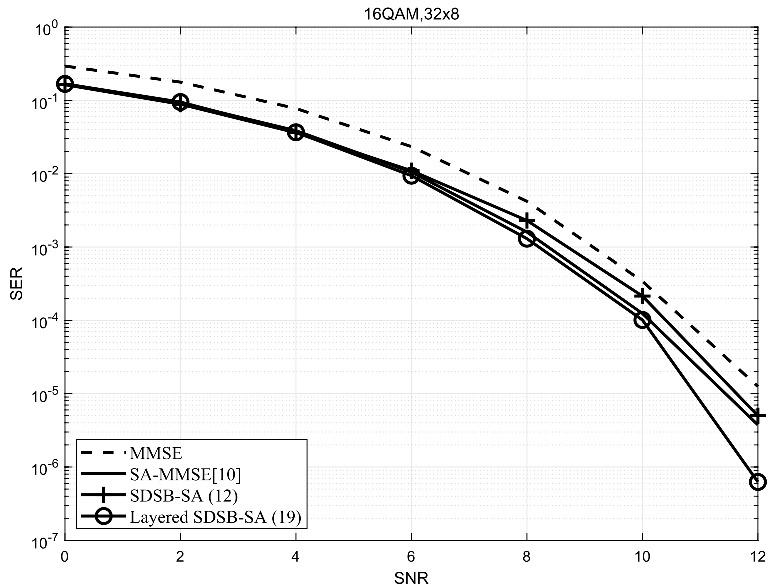

Fig. 6 Performance comparison of MMSE, SA-MMSE and Layered SDSB-SA detector for 32 × 8 MIMO system, 16QAM

Table 2 Computation Complexity Comparison

Detection method	Iteration number	Computation cost per iteration	Total
ML/MAP	1	$\mathcal{O}((2^{2n})^K)$	$\mathcal{O}((2^{2n})^K)$
MMSE/ZF	1	$\mathcal{O}((2M)(2K)^2)$	$\mathcal{O}((2M)(2K)^2)$
SA-MMSE	1	$\mathcal{O}((2M)(2K)^2))$	$\mathcal{O}((2M)(2K)^2))$
Non-layered SDSB-SA (12)	$\mathcal{O}(2K)$	$\mathcal{O}((2 \times 2^n - 1))$	$\mathcal{O}(2K \times (2 \times 2^n - 1))$
Layered SDSB-SA (19)	$\mathcal{O}(2K)$	$\mathcal{O}(3 \times (2^n - 1))$	$\mathcal{O}(2K \times (3 \times (2^n - 1)))$

much less than the complexities ($\mathcal{O}(64 \times 256)$) cost by either the MMSE or SA-MMSE detector (Fig. 6).

Finally, the complexity comparison of the conventional ZF/MMSE, the SA-MMSE and the proposed SDSB-SA detectors are officially summarized in Table 2.

5 Conclusion

In this paper, we exploit the sparsity of the residual error vector to develop a novel SDSB-SA detector for Large-scale MIMO systems. By utilizing the QR decomposition, the unfavourable matrix inversion which is required by the conventional MMSE and SA-MMSE detector is avoided. Meanwhile, the multiple-dimensional search-based detection problem is converted into the single-dimensional search-based one, so the computational complexity of the proposed SDSB-SA is modest compared to that of the MMSE and SA-MMSE detectors. Simulation results show that when lower-order modulations are considered, the proposed SDSB-SA outperforms the MMSA and SA-MMSE detectors because it takes the finite-alphabet constraints into account. When higher-order modulations are considered, the proposed layered SDSB-SA detector achieves better performance than the MMSE and SA-MMSE as it not only adheres to the finite-alphabet constraints and also utilizes the hierarchical structure of the residual error vector to mitigate the possible error propagation.

As an extension of this work, we will consider the more practical factors that channel is imperfect or spatially correlated, and the application of the SDSB-SA detector for downlink multiuser MIMO systems.

Abbreviations

SA: Sparse-aware; MIMO: Multiple-input multiple-output; IoT: Internet of things; ZF: Zero-forcing; SDSB: Single-dimensional search-based; URLLC: Ultra-reliable low-latency communication; ML: Maximum-likelihood; MAP: Maximum a posteriori; MMSE: Minimum-mean-square-error; BER: Bit error rates; SER: Symbol error rate; CS: Compressive sensing; MMP: Multipath matching pursuit; AWGN: Additive white Gaussian noise.

Authors' contributions

RR as the principal investigator takes the primary responsibility for this research. HO analyzed the results. All authors read and approved the final manuscript.

Funding

This research was supported by the Ajou Research Fund, the frame of International Cooperation Program managed by the National Research Foundation of Korea, Grant No. 2017K2A9A2A06016102, and the Basic Research Foundation Program through the National Research Foundation of Korea funded by the Ministry of Education, Grant No. 2017R1D1A1B03035557. This research was partly supported by Sunkyunkwan University Fund, and the frame of Intenatinal Cooperation Program by the National Reseach Fundation of Korea, Grant number 2017K2A9A2A06016102.

Availability of data and materials

The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests

The authors declare that they have no competing interests.

Received: 1 December 2019 Accepted: 19 January 2021
Published online: 18 March 2021

References

1. A. Chockalingam, B.S. Rajan, *Large MIMO Systems* (Cambridge University Press, Cambridge, 2013). https://doi.org/10.1017/CBO9781139208437
2. G.J. Foschini, M.J. Gans, On limits of wireless communications in a fading environment when using multiple antennas. Wirel. Pers. Commun. **6**(3), 311–335 (1998). https://doi.org/10.1023/A:1008889222784
3. L. Liu, W. Yu, Massive connectivity with massive MIMO-Part I: device activity detection and channel estimation. IEEE Trans. Signal Process. **66**(11), 2933–2946 (2018). https://doi.org/10.1109/TSP.2018.2818082
4. F.R. Castillo-Soria, J. Cortez, C.A. Gutiérrez, M. Luna-Rivera, A. Garcia-Barrientos, Extended quadrature spatial modulation for MIMO wireless communications. Phys. Commun. **32**, 88–95 (2019)
5. D. Tse, P. Viswanath, *Fundamental of Wireless Communication* (Cambridge University Press, Cambridge, 2005). https://doi.org/10.1017/CBO9780511807213
6. M.A. Albreem, M. Juntti, S. Shahabuddin, Massive MIMO detection techniques: a survey. IEEE Commun. Surv. Tut. Process. **66**(11), 2933–2946 (2018). https://doi.org/10.1109/COMST.2019.2935810
7. W. Peng, W. Wu, J. Sun, Y. Liu, Sparsity-boosted detection for large MIMO systems. IEEE Commun. Lett. **19**(2), 191–194 (2015). https://doi.org/10.1109/LCOMM.2014.2377215
8. J. Choi, B. Shim, Detection of large-scale wireless systems via sparse error recovery. IEEE Trans. Signal Process. **22**, 6038–6048 (2017). https://doi.org/10.1109/TSP.2017.2749214
9. R. Ran, J. Wang, S. Oh, S. Hong, Sparse aware minimum mean square error detector for MIMO systems. IEEE Commun. Lett. **21**(10), 2214–2217 (2017). https://doi.org/10.1109/LCOMM.2017.2723362
10. R. Ran, G. Park, S. Hong, S. Oh, J. Wang, Generalized sparse aware minimum mean square error detector for large-scale multiuser MIMO systems with higher-order QAM modulation schemes, in *Proceedings of IEEE Conference on Communication*, pp. 1–6 (May 2018). https://doi.org/10.1109/ICC.2018.8423048
11. D. Needell, R. Vershinyn, CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26**(3), 301–321 (2009). https://doi.org/10.1109/ICC.2018.8423048
12. E. Candes, M.B. Waking, An introduction to compressive sampling. IEEE Signal Process. Mag. **25**, 21–30 (2008). https://doi.org/10.1109/MSP.2007.914731
13. R.G. Berger, Z. Wang, J. Huang, S. Zhou, Application of compressive sensing to sparse channel estimation. IEEE Commun. Mag. **48**, 164–174 (2010). https://doi.org/10.1109/MCOM.2010.5621984
14. S. Kwon, J. Wang, B. Shim, Multipath matching pursuit. IEEE Trans. Inf. Theory **60**(5), 2986–3001 (2014). https://doi.org/10.1109/TIT.2014.2310482

15. P. Popovski, J.J. Nielsen, C. Stefanovic, E. de Carvalho, E. Strom, K.F. Trillingsgaard, A.-S. Bana, D.M. Kim, R. Kotaba, J. Park, R.B. Sorensen, Ultra-reliable low-latency communication: Principles and building blocks (2017) (Online). http://arxiv.org/pdf/1708.07862.pdf
16. H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2020). https://doi.org/10.1109/TCOMM.2013.020413.110848
17. Y. Jiang, M.K. Varanasi, J. Li, Performance analysis of ZF and MMSE equalizers for MIMO systems: an in-depth study of the high SNR regime. IEEE Trans. Inf. Theory **57**(4), 2008–2026 (2011). https://doi.org/10.1109/TIT.2011.2112070
18. C. Wang, K.S. Au, R.D. Murch, W.H. Mow, R.S. Cheng, V. Lau, On the performance of MIMO Zero-Forcing receiver in the presence of channel estimation error. IEEE Trans. Wirel. Commun. **6**(3), 805–810 (2007). https://doi.org/10.1109/TWC.2007.05384
19. R.A. Horn, C.R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2012). https://doi.org/10.1017/CBO9780511810817
20. J.G. Proakis, *Digital Communication* (McGram-Hill, London, 2000). https://doi.org/10.1017/CBO9781139208437

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.