Distance Hardening in Large MIMO Systems

Yuan Qi, *Member, IEEE*, and Rongrong Qian, *Member, IEEE*

***Abstract*— Asymptotic analysis plays an important role in the performance evaluation and technology design of large multiple-input multiple-output (MIMO) systems. This paper investigates an asymptotic property of large MIMO systems, called *distance hardening*. It involves distance metrics defined in terms of many factors, such as the channel matrix and the difference matrix of two arbitrary transmitted codewords. The asymptotic property is revealed by the use of the Chernoff bounding method and expressed in an analogous way to the well-known channel hardening. More specifically, we first derive Chernoff bounds on the tail probabilities of distance metrics, in which we are interested. Then, we formulate the asymptotic property based on a generalized definition of the hardening phenomena by referring to a formal definition of the channel hardening. The distance hardening is shown to have implications for space-time code design, MIMO detection, and millimeter wave MIMO systems. Numerical simulations are conducted to verify our theoretical development.**

***Index Terms*— Multiple-input multiple-output (MIMO), large MIMO system, asymptotic analysis.**

I. Introduction

THE emergence of large MIMO systems using large antenna arrays has led to an increased focus on such systems in practical communication systems design. As the demands on the data rate/throughput dramatically increase in practical multi-antenna systems, the number of antennas needs to be scaled up to tens or hundreds to meet performance requirements [1]–[3], [31]. In addition to two obvious benefits of large MIMO systems (increased data rates and enhanced reliability), the large dimensionality of the systems can also result in a host of other advantages that do not come with small and moderate-size systems [1], [31]. For example, as the asymptotics of random matrix theory arise in large MIMO systems [2], [32], some random variables converge to be deterministic [2]. Hence, many metrics that were previously difficult to analyze are clarified by accurate approximations [4], [5]. To illustrate, let us briefly review some previous results.

First, the classical central limit theorem (CLT) establishes that, in most situations, the properly normalized sum of independent random variables tends toward a normal distribution even if the original variables are not normally distributed [35]. This theorem has a number of variants, some of which are relevant to random matrix and large MIMO research [33]. The CLT of linear spectral statistics for large dimensional sample covariance matrices is investigated in [6], which can be used to derive the size-asymptotic, finite-SNR diversity-multiplexing tradeoff for a broad class of fading channels [7], [32]. A CLT for random determinants is developed in [8], which has been extended to study the weak convergence of the capacity random variable [9].

Second, the asymptotic empirical eigenvalue distribution (EED) functions of many random Hermitian matrices have been shown to converge a.s. to non-random limits [10], [11], [34]. The extensive use of this convergence helps to analyze the asymptotic SINR of independent non-identically distributed (IND) channels [12] and the capacity of the sparse family of channels (beamforming and multiplexing channels) [13]. According to [13], although a closed-form expression of the limiting EED in the multiplexing channel may not be available, the capacity formula of this channel is identical to that of the ideal channel with a proper parameter. Note that, the capacity of the ideal channel can be expressed in closed form by using the asymptotic EED function.

Third, the free probability theory [32], [33] provides a sufficient condition (called *almost sure asymptotic freeness* [32]) under which the normalized trace of the product of two random matrices converges almost surely [14]. The power of free probability theory should be evident in both the fresh view it provides on established results and the new properties it unveils on random matrices [32]. The concept of almost sure asymptotic freeness has been used to compute asymptotic pairwise error probabilities (PEPs) with maximum-likelihood (ML) detection and three types of receiver architectures [14].

Fourth, as the number of antennas grows, the MIMO channel quickly "hardens" in the sense that the off-diagonal terms of the matrix $\boldsymbol{H}^H\boldsymbol{H}$ become increasingly weaker compared to the diagonal terms [15] or the mutual information fluctuation rapidly decreases relative to its mean [4]. This is a well-known phenomenon called *channel hardening*. Due to channel hardening, approximate matrix inversion, using series expansion and deterministic approximations from the limiting distribution [31], can be employed to realize low-complexity MIMO detectors with good performance in large dimensions [15]. According to recent studies [4], [16], [17], channel hardening

Manuscript received June 22, 2017; revised October 1, 2017 and December 2, 2017; accepted December 3, 2017. Date of publication December 11, 2017; date of current version April 16, 2018. This work is sponsored by the National Natural Science Foundation of China (Grant No. 61501043) and the Program of Introducing Talents of Discipline to Universities of China (Grant No. B08004). The associate editor coordinating the review of this paper and approving it for publication was V. Raghavan. *(Corresponding author: Rongrong Qian.)*

Y. Qi is with the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: qiyuan@bupt.edu.cn).

R. Qian is with the Automation School, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: rongrongqian@bupt.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2017.2782343

can also affect the scheduling gains, the number of bits needed in rate feedback, and the energy efficiency in MIMO and antenna selection systems.

These results show that the analysis of asymptotic properties (e.g., CLT, asymptotic EED, almost sure asymptotic freeness, and channel hardening) allows us to gain profound insights into a number of important problems in large MIMO systems [2], [4], [5], [7], [9], [13]–[18], [32]. The insights could further facilitate the performance evaluations and technology design for these systems. This is the reason why asymptotic properties of large MIMO systems need to be well investigated. The above arguments motivate theoretical research on exploring new asymptotic properties of large MIMO systems and exploiting more benefits from the known properties.

In this paper, we report on an asymptotic property of large MIMO systems, and attempt to describe this property in a rigorous, intuitive way, by using the Chernoff bounding method [35]. This asymptotic property is named distance hardening, since it is similar to channel hardening [4], [15], [16]. We also discuss implications of this "distance hardening" result for space-time code design, MIMO detection, and mmWave MIMO systems.

Our main contributions and the novelty of this paper can be summarized as follows.

- The Chernoff bound expressions for the tail probabilities of two distance metrics are derived (Theorems 1-3). One metric is used in many areas of space-time MIMO design and the other could appear in spatial multiplexing, beamforming, and precoding systems with uniform linear arrays (ULAs). For the Chernoff bound given in the two most important theorems of this paper (Theorems 1 and 2), several related properties are provided (Lemma 3) and extreme point analysis is performed (Corollary 5) in order to clarify its mathematical characteristics. It is also shown that the expressions of Chernoff bound can be extended to MIMO channels with transmit correlation (Corollary 3).
- The distance hardening properties are formulated (Theorems 4-6) by the use of the obtained Chernoff bound expressions and a generalized definition for the hardening phenomena that originates from the formal definition of the channel hardening introduced in [18]. Based on the distance hardening formulation together with the tail probability analysis, a new space-time decoding strategy and its related criterion of code design are proposed (Corollary 6). The derived results' implications for MIMO detection and mmWave MIMO systems are also discussed (Corollary 7).

The paper is organized as follows. Section II introduces the system models used in our development, which include the widely used analytical model of large MIMO systems and the model available for many realistic large MIMO systems. Section III presents the tail probability analysis that yields the results of Chernoff bound expressions. Section IV formulates the distance hardening and discusses its implications. Numerical results are provided in Section V, and Section VI concludes the paper and gives the outlook.

Notations: Matrices are set in boldface capital letters, and vectors in boldface lowercase letters. We write a_{ij} for the entry in the ith row and jth column of the matrix $\boldsymbol{A}$, and b_i for the ith entry of the vector $\boldsymbol{b}$. $\boldsymbol{A} = \text{diag}\{b_1, b_2, \cdots, b_n\}$ puts the elements $b_1, b_2, \cdots, b_n$ on the main diagonal of matrix $\boldsymbol{A}$. The superscripts T and H stand for the transpose and conjugate transpose, respectively. The Frobenius norm is denoted by $\|\boldsymbol{A}\|_F = \sqrt{\text{Tr}(\boldsymbol{A}^H\boldsymbol{A})} = \sqrt{\text{Tr}(\boldsymbol{A}\boldsymbol{A}^H)}$, where $\text{Tr}(\cdot)$ is the trace of a square matrix. $E[\cdot]$ denotes the expectation operator. We write $\overset{d}{=}$ for equality in distribution and $\xrightarrow{P}$ for convergence *in probability*. The zero-mean complex Gaussian distribution with variance σ^2 is denoted by $\mathcal{CN}(0, \sigma^2)$. If $z_1, \cdots, z_k$ are independent, standard normal random variables, then the sum of their squares, $y = \sum_{i=1}^{k} z_i^2$, is distributed according to *chi-squared distribution* with k degrees of freedom, which is usually denoted as $y \sim \chi_k^2$. Suppose that the sequence $\{x_k\}$ converges to the number a. We say that this sequence converges linearly to a, if there exists a number $r \in (0, 1)$ such that

$$\lim_{k \to +\infty} \frac{|x_{k+1} - a|}{|x_k - a|} = r. \tag{1}$$

Here, r is called the *rate of convergence*. In numerical simulation, if the computation of r is not available, we will be interested in evaluating the *instantaneous rate* defined as $r_k \triangleq \frac{|x_{k+1}-a|}{|x_k-a|}$.

II. System Models

The widely used analytical model of large MIMO channels is the i.i.d. flat fading channel model [7], [20], [31], [32]. This model relies on (i) the antenna elements in the transmitter/receiver being spatially well separated, and (ii) the environment being rich-scattering so that its combined gains can be approximated by a Gaussian random variable owing to the CLT [31]. Here, rich scattering is thought to be common in typical indoor and outdoor urban environments [31] and known to be favorable for the operation of large MIMO [20]. On the other hand, many realistic large MIMO systems are expected to have clustered-scattering channels that can usually be effectively represented by the virtual representation model proposed in [21]. A typical channel that might be clustered-scattering is the channel of millimeter wave (mmWave) communication with a ULA [22], [23]. Taken together, this study will consider two models of Rayleigh fading MIMO systems, as given in Subsections II-A and II-B.

A. Space-Time MIMO System With Rich Scattering (i.i.d. Model) [7], [20], [31], [32]

We consider a Rayleigh-fading MIMO channel with n_T transmit and n_R receive antennas. The channel is quasi-static fading with a quasi-static interval t. The standard discrete-time system model is adopted here,

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}. \tag{2}$$

$\boldsymbol{X} \in \mathbb{C}^{n_T \times t}$ and $\boldsymbol{Y} \in \mathbb{C}^{n_R \times t}$ denote the transmitted and received signals, respectively. We assume that $\boldsymbol{X}$ is a code matrix drawn from a space-time codebook $\mathbf{C}$. $\boldsymbol{H} = [h_{i,j}] \in$

$\mathbb{C}^{n_R \times n_T}$ represents the channel matrix, where $h_{i,j}$ denotes the complex path gain (the fading coefficient) from transmit antenna j to receive antenna i. It is assumed that $h_{i,j}$ are i.i.d. zero mean complex Gaussian random variables with variance ρ_h, i.e., $h_{i,j} \sim \mathcal{CN}(0, \rho_h)$. $\boldsymbol{N} \in \mathbb{C}^{n_R \times t}$ denotes the i.i.d. complex Gaussian noise in the MIMO channel.

Let $\boldsymbol{D}\left(X, \widehat{X}\right) \triangleq X - \widehat{X}$ be the difference matrix of two arbitrary code matrices $X \in \mathbf{C}$ and $\widehat{X} \in \mathbf{C}$. The distance metric, $\left\|\boldsymbol{D}_{H,W}\left(X, \widehat{X}\right)\right\|_F^2$, is of interest in this study, where

$$\boldsymbol{D}_{H,W}\left(X, \widehat{X}\right) \triangleq \boldsymbol{H}\boldsymbol{D}\left(X, \widehat{X}\right) + \boldsymbol{W}. \tag{3}$$

$\boldsymbol{W} = [w_{i,j}] \in \mathbb{C}^{n_R \times t}$ is the circularly symmetric, complex Gaussian noise matrix with i.i.d. entries $w_{i,j} \sim \mathcal{CN}(0, N_0)$. The noise matrix $\boldsymbol{W}$ could be the matrix of channel noise $\boldsymbol{N}$ or an all-zero matrix.

For brevity, in the remainder of the paper, $\boldsymbol{D}_{H,W}$ and $\boldsymbol{D}$ will be used to denote $\boldsymbol{D}_{H,W}\left(X, \widehat{X}\right)$ and $\boldsymbol{D}\left(X, \widehat{X}\right)$, respectively, by dropping $\left(X, \widehat{X}\right)$.

B. Spatial Multiplexing, Beamforming, and Precoding Systems With ULAs (Virtual Representation Model) [13], [22], [23], [26]

We assume that the MIMO channel is Rayleigh-fading with n_T transmit and n_R receive antennas. A unified discrete-time system model available to spatial multiplexing, beamforming, and precoding systems, can be written as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{4}$$

where $\boldsymbol{x} \in \mathbb{C}^{n_T \times 1}$ and $\boldsymbol{y} \in \mathbb{C}^{n_R \times 1}$ are the transmitted and received signals, respectively. Within beamforming systems, $\boldsymbol{x}$ may be given by $\boldsymbol{x} = \boldsymbol{f}c$, where $\boldsymbol{f}$ denotes the beamforming vector and c is the symbol chosen from a constellation for signaling [23]. In precoding systems, $\boldsymbol{x}$ can be given by $\boldsymbol{x} = \boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\boldsymbol{c}_{SV}$, where $\boldsymbol{F}_{RF}$, $\boldsymbol{F}_{BB}$, and $\boldsymbol{c}_{SV}$ are the RF precoder, the baseband precoder, and the symbol vector, respectively [22]. The channel matrix $\boldsymbol{H}$ is expressed by *virtual representation* as [13]

$$\boldsymbol{H} = \boldsymbol{A}_r \boldsymbol{H}_v \boldsymbol{A}_t^H, \tag{5}$$

where $\boldsymbol{A}_t$ and $\boldsymbol{A}_r$ are $n_T \times n_T$ and $n_R \times n_R$ unitary discrete Fourier matrices, respectively. $\boldsymbol{H}_v = [h_{i,j,v}] \in \mathbb{C}^{n_R \times n_T}$ is the virtual channel matrix whose entries are independent zero-mean, complex Gaussian [13]. In other words, $\boldsymbol{H}_v$ has independent non-identically distributed (i.n.d.) entries [12]. The statistics of $\boldsymbol{H}_v$ are characterized by the virtual channel power matrix $\boldsymbol{\Psi} = [\psi_{i,j}] \in \mathbb{C}^{n_R \times n_T}$, where $\psi_{i,j} \triangleq E\left[|h_{i,j,v}|^2\right]$. $\boldsymbol{n} \in \mathbb{C}^{n_R \times 1}$ is the i.i.d. complex Gaussian noise in channel, and its elements are $n_i \sim \mathcal{CN}(0, N_0)$.

Let $\boldsymbol{d}\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right) \triangleq \boldsymbol{x} - \widehat{\boldsymbol{x}}$ be the difference vector of two arbitrary transmitted vectors $\boldsymbol{x} \in \mathbb{C}^{n_T \times 1}$ and $\widehat{\boldsymbol{x}} \in \mathbb{C}^{n_T \times 1}$. Another distance metric of interest in this study, is $\left\|\boldsymbol{d}_{H,n}\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right)\right\|_F^2$, where

$$\boldsymbol{d}_{H,n}\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right) \triangleq \boldsymbol{H}\boldsymbol{d}\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right) + \boldsymbol{n}. \tag{6}$$

For simplicity, by dropping $\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right)$, $\boldsymbol{d}_{H,n}$ and $\boldsymbol{d}$ will be used to denote $\boldsymbol{d}_{H,n}\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right)$ and $\boldsymbol{d}\left(\boldsymbol{x}, \widehat{\boldsymbol{x}}\right)$, respectively. We define

$$\widetilde{\boldsymbol{d}} \triangleq \boldsymbol{A}_t^H \boldsymbol{d} = \left[\widetilde{d}_1, \cdots, \widetilde{d}_{n_T}\right]^T. \tag{7}$$

C. Distance Metrics

The forthcoming analysis will concentrate on $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ and $\left\|\boldsymbol{d}_{H,n}\right\|_F^2$, where

$$\boldsymbol{D}_{H,W} = \boldsymbol{H}\boldsymbol{D} + \boldsymbol{W}, \tag{8}$$

and

$$\boldsymbol{d}_{H,n} = \boldsymbol{H}\boldsymbol{d} + \boldsymbol{n}. \tag{9}$$

Here, it is known that $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ is a crucial metric for many important problems in space-time MIMO systems, such as the evaluation of PEP, MIMO detection, space-time code design, and the analysis of diversity-multiplexing tradeoff (DMT) [14], [27], [31], [36]. The derivation of exact analytical expressions for the statistics of $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ and $\left\|\boldsymbol{d}_{H,n}\right\|_F^2$ can be difficult, so the accurate approximations given by asymptotic analysis will become very helpful, as shown later in the next two sections.

III. Tail Probability Analysis: The Chernoff Bound

The Chernoff bounding method is used as the main analytical tool in this study, and we begin by analyzing the statistics of $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ and $\left\|\boldsymbol{d}_{H,n}\right\|_F^2$ with this method. In analyzing the statistics of a random variable, it is often necessary to determine the area under the tail of the probability distribution function (PDF) [35]. This area is known as the tail probability. In this section, we first derive the Chernoff bound on the tail probability of the following *normalized* distance metric

$$\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2 + tn_RN_0}, \tag{10}$$

which is said to be normalized because $n_R\rho_h\|\boldsymbol{D}\|_F^2 + tn_RN_0 = E_{H,W}\left[\|\boldsymbol{D}_{H,W}\|_F^2\right]$ (see Lemma 1) such that $E_{H,W}\left[\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{\|n_R\rho_h\boldsymbol{D}\|_F^2 + tn_RN_0}\right] = 1$. Then, we perform a similar analysis to another normalized distance metric

$$\frac{\|\boldsymbol{d}_{H,n}\|_F^2}{E_{H,n}\left[\|\boldsymbol{d}_{H,n}\|_F^2\right]}, \tag{11}$$

where $E_{H,n}\left[\|\boldsymbol{d}_{H,n}\|_F^2\right] = \sum_{i=1}^{n_R}\sum_{j=1}^{n_T}\psi_{i,j}\left|\widetilde{d}_j\right|^2 + n_RN_0$. The tail probability analysis will give rise to the fundamental results that enable us to investigate the so-called distance hardening phenomenon in next section.

A. Deriving the Chernoff Bound

Let us begin by developing a simplified equivalent of the distance metric $\|\boldsymbol{D}_{H,W}\|_F^2$, which will be of particular importance in the derivation of the Chernoff bound for the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2 + tn_RN_0}$.

Lemma 1: $\|\boldsymbol{D}_{H,W}\|_F^2$ can be expressed as

$$\|\boldsymbol{D}_{H,W}\|_F^2 = \|\boldsymbol{H}\boldsymbol{D} + \boldsymbol{W}\|_F^2 = \sum_{j=1}^{s} u_j + \sum_{k=s+1}^{t} v_k, \tag{12}$$

where $\left(\frac{\rho_h\sigma_j^2}{2}+\frac{N_0}{2}\right)^{-1} u_j \sim \chi^2_{2\,n_R}$, $\left(\frac{N_0}{2}\right)^{-1} v_k \sim \chi^2_{2\,n_R}$, and $s=\min\{n_T,t\}$. Note that the terms v_k in (12) will disappear if $n_T \geq t$ such that $s=\min\{n_T,t\}=t$. The expectation of the distance metric $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ can be computed in the form

$$E_{H,W}\left[\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right]=n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0. \tag{13}$$

Proof: See Appendix A, where the definitions and properties of u_j and v_k are provided, e.g., (58), Lemma A.1, and Lemma A.2. ■

For simplicity of presentation, let $\xi \triangleq n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0$ such that

$$\xi=E_{H,W}\left[\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right]=n_R\rho_h\left(\sum_{j=1}^{s}\sigma_j^2\right)+tn_RN_0, \tag{14}$$

where σ_j denotes the j-th singular value of $\boldsymbol{D}$.

The exact closed-form expression of the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ is difficult to obtain. Nevertheless, there is an interesting and useful upper bound on this tail probability, which is adequate for the analysis of this study.

Theorem 1: The tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ can be upper-bounded by the Chernoff bound as

$$P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}\geq\theta\right)\leq U_c(\theta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}},N_0) \tag{15}$$

for the cases with $\theta>1$, or

$$P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}\leq\theta\right)\leq U_c(\theta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}},N_0) \tag{16}$$

for the cases with $0<\theta<1$, where

$$U_c(\theta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}},N_0)\triangleq e^{-\overline{\gamma}\theta}\left(1-\frac{\overline{\gamma}N_0}{\xi}\right)^{-(t-s)n_R}\times\prod_{j=1}^{s}\left(1-\frac{\overline{\gamma}\left(\rho_h\sigma_j^2+N_0\right)}{\xi}\right)^{-n_R}, \tag{17}$$

$\widetilde{\boldsymbol{\sigma}}\triangleq[\sigma_1^2\ \cdots\ \sigma_s^2]$, and

$$\frac{t-s}{\left(\frac{\xi}{N_0}-\overline{\gamma}\right)}+\sum_{j=1}^{s}\frac{1}{\left(\frac{\xi}{\rho_h\sigma_j^2+N_0}-\overline{\gamma}\right)}=\frac{\theta}{n_R}. \tag{18}$$

Proof: See Appendix B. ■

The Chernoff bound applied to $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ yields an exponential dependence on n_R, and thus provides a tight upper bound on the tail probability. In Theorem 1, $\overline{\gamma}$ is an important term for the Chernoff bound obtained as (17), which can be computed from θ by the use of the relationship in (18).

Corollary 1: As an implicit equation that gives the relationship between $\overline{\gamma}$ and θ, (18) provides the following results.

1) $\overline{\gamma}$ can be viewed as a function of θ, denoted by $\overline{\gamma}(\theta)$ if needed, which is strictly monotonically increasing with θ.
2) $\overline{\gamma}=0$ if and only if $\theta=1$. This also implies $\overline{\gamma}>0$ when $\theta>1$ and $\overline{\gamma}<0$ when $\theta<1$.
3) $\overline{\gamma}<\min_j\frac{\xi}{\rho_h\sigma_j^2+N_0}$.
4) If $\boldsymbol{D}$, ρ_h, N_0, and θ are fixed, then $\frac{\overline{\gamma}}{n_R}$ will not change as n_R varies.

Proof: Differentiating the implicit function $\overline{\gamma}(\theta)$, defined by (18), yields

$$\frac{d\overline{\gamma}}{d\theta}=\frac{1}{n_R}\left[\frac{t-s}{\left(\frac{\xi}{N_0}-\overline{\gamma}\right)^2}+\sum_{j=1}^{s}\frac{1}{\left(\frac{\xi}{\rho_h\sigma_j^2+N_0}-\overline{\gamma}\right)^2}\right]^{-1}. \tag{19}$$

This implies $\frac{d\overline{\gamma}}{d\theta}>0$; hence, $\overline{\gamma}$ is strictly monotonically increasing as θ grows. Due to (14), it is easy to verify that, if $\theta=1$, then $\overline{\gamma}=0$. According to Appendix B, $\overline{\gamma}$ should be less than $\frac{\xi}{\rho_h\sigma_j^2+N_0}$ with any j. Moreover, (18) can be rewritten as

$$\frac{t-s}{\left(\frac{\xi}{n_RN_0}-\frac{\overline{\gamma}}{n_R}\right)}+\sum_{j=1}^{s}\frac{1}{\left(\frac{\xi}{n_R\rho_h\sigma_j^2+n_RN_0}-\frac{\overline{\gamma}}{n_R}\right)}=\theta. \tag{20}$$

From above, we can find that $\frac{\xi}{n_RN_0}$ and $\frac{\xi}{n_R\rho_h\sigma_j^2+n_RN_0}$, $1\leq j\leq s$, will not change with different n_R, as long as $\boldsymbol{D}$, ρ_h, and N_0 are fixed. Then if θ is also fixed, $\frac{\overline{\gamma}}{n_R}$ in (20) will be unaltered as n_R changes. ■

Furthermore, in case of $t=1$, Theorem 1 takes on a remarkably simple form.

Corollary 2: If $t=1$, there exists the Chernoff bound relation as

$$P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+n_RN_0}\geq\theta\right)\leq\left(\theta e^{1-\theta}\right)^{n_R} \tag{21}$$

for the cases with $\theta>1$, or

$$P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+n_RN_0}\leq\theta\right)\leq\left(\theta e^{1-\theta}\right)^{n_R} \tag{22}$$

for the cases with $0<\theta<1$.

In fact, such a simplified relationship on the Chernoff bound is exactly the main result of the previous study [19]. In this sense, Theorem 1 of this study is a generalization of the main result in [19]. It is easy to find that the expression of Theorem 1 in this study applies to the generic space-time MIMO systems including space-time coding and spatial-multiplexing systems, whereas the corresponding expression obtained by [19] is unavailable to space-time coding system.

It is worth pointing out that the result of Theorem 1 can be extended to space-time MIMO systems with some other channel models. In [2], the channel is modeled as $\boldsymbol{G}=\boldsymbol{H}\boldsymbol{L}^{1/2}$, where $\boldsymbol{H}$ describes the small scale fading. $\boldsymbol{L}^{1/2}$ is a diagonal matrix whose diagonal elements are the large scale fading coefficients that account for path loss and shadow fading. In [24] and [25], the channel that involves the transmit correlation can be written as $\boldsymbol{G}=\boldsymbol{H}\boldsymbol{C}^{1/2}$, where $\boldsymbol{H}$ describes the channel fading and $\boldsymbol{C}$ is a Hermitian Toeplitz deterministic correlation matrix.

Corollary 3: If the channel matrix of the space-time MIMO system becomes $\boldsymbol{G} = \boldsymbol{H}\boldsymbol{L}^{1/2}$ or $\boldsymbol{G} = \boldsymbol{H}\boldsymbol{C}^{1/2}$, where $\boldsymbol{H}$ has i.i.d. zero mean complex Gaussian entries, then the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{L}^{1/2}\boldsymbol{D}\|_F^2+tn_RN_0}$ or $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{C}^{1/2}\boldsymbol{D}\|_F^2+tn_RN_0}$ will have the Chernoff upper bounds as given in Theorem 1, by letting σ_j be the j-th singular value of $\boldsymbol{L}^{1/2}\boldsymbol{D}$ and $\boldsymbol{C}^{1/2}\boldsymbol{D}$, respectively.

However, Theorem 1 cannot be applied to the channel with the receive correlation, denoted by $\boldsymbol{G} = \boldsymbol{C}^{1/2}\boldsymbol{H}$ [24], [25]. In this case, the correlation $\boldsymbol{C}^{1/2}$ is the obstruction to our analytical framework.

The expression of the Chernoff bound in Theorem 1 involves too many parameters such as N_0, σ_j, and ξ, which hinders more insights on the properties of this bound. To be refined, let us define

$$C_e(\lambda) \triangleq \frac{\lambda}{1-\lambda}, \tag{23}$$

so that (18) can be rewritten as

$$(t-s)C_e\left(\frac{\overline{\gamma}N_0}{\xi}\right) + \sum_{j=1}^{s} C_e\left(\frac{\overline{\gamma}\left(\rho_h\sigma_j^2+N_0\right)}{\xi}\right) = \frac{\theta\overline{\gamma}}{n_R}. \tag{24}$$

Moreover, by defining

$$\lambda_0 \triangleq \frac{\overline{\gamma}N_0}{\xi}, \quad \text{and } \lambda_j \triangleq \frac{\overline{\gamma}\left(\rho_h\sigma_j^2+N_0\right)}{\xi}, \tag{25}$$

where $1 \le j \le s$, (24) further becomes

$$(t-s)C_e(\lambda_0) + \sum_{j=1}^{s} C_e(\lambda_j) = \frac{\theta\overline{\gamma}}{n_R}. \tag{26}$$

The Chernoff bound in Theorem 1, $U_c(\theta, \overline{\gamma}, \widetilde{\boldsymbol{\sigma}}, N_0)$, can be rewritten as

$$\begin{aligned} &U_c(\theta, \overline{\gamma}, \widetilde{\boldsymbol{\sigma}}, N_0) \\ &\quad= \left[\left(\frac{1}{1-\lambda_0}\right)\exp\left(-\frac{\lambda_0}{1-\lambda_0}\right)\right]^{(t-s)n_R} \\ &\qquad\times \prod_{1\le j\le s}\left[\left(\frac{1}{1-\lambda_j}\right)\exp\left(-\frac{\lambda_j}{1-\lambda_j}\right)\right]^{n_R}. \end{aligned} \tag{27}$$

From the above equation, we observe that the expression of the Chernoff bound has a tight relationship to an elementary function $\left(\frac{1}{1-\lambda}\right)\exp\left(-\frac{\lambda}{1-\lambda}\right)$. Thus, for simplicity, let us define

$$B_e(\lambda) \triangleq \left(\frac{1}{1-\lambda}\right)\exp\left(-\frac{\lambda}{1-\lambda}\right), \tag{28}$$

where $\lambda < 1$. As a consequence, the expression of Theorem 1 simplifies to the following form.

Theorem 2: The tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ is upper-bounded as

$$P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0} \ge \theta\right) \le U_c(\theta,\boldsymbol{\lambda}) \tag{29}$$

for the cases with $\theta > 1$, or

$$P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0} \le \theta\right) \le U_c(\theta,\boldsymbol{\lambda}) \tag{30}$$

for the cases with $0 < \theta < 1$, where $\boldsymbol{\lambda} = [\lambda_0\ \lambda_1\ \cdots\ \lambda_s]$ with λ_j, $0 \le j \le s$, as defined in (25) and (26), and

$$U_c(\theta,\boldsymbol{\lambda}) \triangleq [B_e(\lambda_0)]^{(t-s)n_R}\prod_{1\le j\le s}\left[B_e(\lambda_j)\right]^{n_R}. \tag{31}$$

Corollary 4: If $\boldsymbol{D}$, ρ_h, N_0, and θ are fixed, then λ_0 and λ_j $(1 \le j \le s)$ will not change as n_R varies.

Proof: Rewriting (25) as

$$\begin{aligned} \lambda_0 &= \frac{\overline{\gamma}}{n_R}\cdot\frac{n_RN_0}{\xi}, \\ \lambda_j &= \frac{\overline{\gamma}}{n_R}\cdot\frac{\left(n_R\rho_h\sigma_j^2+n_RN_0\right)}{\xi}, \quad 1 \le j \le s, \end{aligned} \tag{32}$$

we can derive this corollary from Corollary 1. ∎

Let us turn to deriving the Chernoff bound for the tail probability of $\frac{\|\boldsymbol{d}_{H,n}\|_F^2}{E_{H,n}\left[\|\boldsymbol{d}_{H,n}\|_F^2\right]}$. To attain this goal, we need a simplified equivalent of the distance metric $\|\boldsymbol{d}_{H,n}\|_F^2$.

Lemma 2: $\|\boldsymbol{d}_{H,n}\|_F^2$ can be expressed as

$$\|\boldsymbol{d}_{H,n}\|_F^2 = \|\boldsymbol{H}\boldsymbol{d}+\boldsymbol{n}\|_F^2 = \sum_{i=1}^{n_R}\kappa_i, \tag{33}$$

where $\left(\sum_{j=1}^{n_T}\frac{\psi_{i,j}|\widetilde{d}_j|^2}{2}+\frac{N_0}{2}\right)^{-1}\kappa_i \sim \chi_2^2$. Please refer to (7) for $\widetilde{d}_j$.

Proof: See Appendix C, where the definition and property of κ_i are provided, i.e., Lemma C.1. ∎

Therefore, a result similar to Theorem 1 can be obtained for $\|\boldsymbol{d}_{H,n}\|_F^2$.

Theorem 3: The tail probabilities $P\left(\frac{\|\boldsymbol{d}_{H,n}\|_F^2}{E\left[\|\boldsymbol{d}_{H,n}\|_F^2\right]} \ge \theta\right)$ for $\theta > 1$ and $P\left(\frac{\|\boldsymbol{d}_{H,n}\|_F^2}{E\left[\|\boldsymbol{d}_{H,n}\|_F^2\right]} \le \theta\right)$ for $\theta < 1$ have the Chernoff upper bounds of the same form, as

$$\prod_{i=1}^{n_R} B_e\left(\frac{\overline{\gamma}\left(\sum_{j=1}^{n_T}\psi_{i,j}\left|\widetilde{d}_j\right|^2+N_0\right)}{\eta}\right), \tag{34}$$

where $\eta \triangleq E\left[\left\|\boldsymbol{d}_{H,n}\right\|_F^2\right] = \sum_{i=1}^{n_R}\sum_{j=1}^{n_T}\psi_{i,j}\left|\widetilde{d}_j\right|^2 + n_RN_0$, and $\overline{\gamma}$ can be computed by

$$\sum_{i=1}^{n_R} C_e\left(\frac{\overline{\gamma}\left(\sum_{j=1}^{n_T}\psi_{i,j}\left|\widetilde{d}_j\right|^2+N_0\right)}{\eta}\right) = \theta\overline{\gamma}. \tag{35}$$

Proof: It is similar to the proof of Theorem 1. See Appendix B. ∎

B. Discussion

Up to this point, we have derived the Chernoff bounds on the tail probabilities of the normalized distance metrics $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ and $\frac{\|\boldsymbol{d}_{H,n}\|_F^2}{E\left[\|\boldsymbol{d}_{H,n}\|_F^2\right]}$. We emphasize the following.

1) The Relationship between Theorems 1 and 2: $U_c(\theta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}},N_0)$ in Theorem 1 and $U_c(\theta,\boldsymbol{\lambda})$ in Theorem 2 essentially represent the same bound, and the difference between them is only in the form of an expression.

2) Properties of $B_e(\lambda)$ and $C_e(\lambda)$ defined for Theorems 2 and 3: Although $B_e(\lambda)$ and $C_e(\lambda)$ are very simple in terms of expressions, they play important roles in this study. On the one hand, $U_c(\theta,\boldsymbol{\lambda})$ is expressed in terms of $B_e(\lambda)$ for which one can further have

$$B_e(\lambda)=(1+C_e(\lambda))\exp(-C_e(\lambda)), \tag{36}$$

while $C_e(\lambda)$ simultaneously constructs the relationship in (26). On the other hand, several properties of $B_e(\lambda)$ and $C_e(\lambda)$ will be used in the subsequent analysis.

Lemma 3: The properties of $C_e(\lambda)$ and $B_e(\lambda)$ that prove useful are as follows.

- $C_e(\lambda)$ is a convex function in the range $(-\infty,1)$. That is, $\forall\lambda_1,\lambda_2\in(-\infty,1)$, $\forall\alpha\in[0,1]$, it holds that $C_e(\alpha\lambda_1+(1-\alpha)\lambda_2)\le\alpha C_e(\lambda_1)+(1-\alpha)C_e(\lambda_2)$.
- For $\lambda\in(-\infty,1)$, $B_e(\lambda)\le 1$ with equality if and only if $\lambda=0$.
- The function $\log_e B_e(\lambda)$ is convex in the interval $(-\infty,-1]$ and is concave in the interval $[-1,1)$.

Proof: These properties follow from the development in Appendix D. Moreover, the curves of $B_e(\lambda)$ and $C_e(\lambda)$ appear as shown in Figure 1. From Figure 1, we observe that, within the interval $(-\infty,1)$, the maximum value of $B_e(\lambda)$ is 1 at $\lambda=0$. ∎

3) Extreme point analysis: The value of the Chernoff bound in the form of either $U_c(\theta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}},N_0)$ or $U_c(\theta,\boldsymbol{\lambda})$ depends on many parameters. Sometimes, we may not want such meticulous results and just wish to know the extreme point on the Chernoff bound for a certain MIMO system with static parameters n_T, n_R, and t. Here, a relevant result is presented.

Corollary 5: The Chernoff bound derived for the tail probability of the metric $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ has a strict local minimum $\left(\theta e^{1-\theta}\right)^{tn_R}$, at $\lambda_0^*=\lambda_1^*=\cdots=\lambda_s^*=\frac{\overline{\gamma}(\boldsymbol{\lambda}^*)}{tn_R}=1-\frac{1}{\theta}$ with $\overline{\gamma}(\boldsymbol{\lambda}^*)=tn_R\left(1-\frac{1}{\theta}\right)$. Moreover, such a minimum can be proven to be the global minimum if $0<\theta\le 1/2$.

Proof: The strict local minimum can be verified by means of Lagrangian duality [37], the Karush-Kuhn-Tucker (KKT) condition [37], and the bordered Hessian [38], whereas the validation of the global minimum requires the method of convex optimization [37]. See Appendix E for details. ∎

IV. Distance Hardening

Having performed the tail probability analysis, we now apply the obtained results to large MIMO systems. More specifically, in this section we introduce a generalized definition for the hardening phenomena that includes but is not limited to channel hardening. Then, we reveal a phenomenon named distance hardening for large MIMO systems by the use of the theorems provided in the previous section.

A. Formulating Distance Hardening

In large MIMO systems, channel hardening is the phenomenon that the off-diagonal entries of the matrix $\boldsymbol{H}^H\boldsymbol{H}$ become increasingly weaker compared to the diagonal entries as the number of antennas grows [15], [31]. The mutual information of an MIMO system is defined as $I\triangleq\log\det\left(\boldsymbol{I}+\frac{\rho}{n_T}\boldsymbol{H}^H\boldsymbol{H}\right)$, where $\boldsymbol{I}$ is the identity matrix and ρ is the SNR as physically measured at each receive antenna, which is closely related to $\boldsymbol{H}^H\boldsymbol{H}$. Therefore, channel hardening also refers to the phenomenon where the variance of the mutual information of the MIMO channel grows at a smaller rate relative to the mean or just stays constant in the extreme case [4].

The recent work [18] provides a formal definition of channel hardening. One can say that the propagation offers channel hardening if

$$\frac{\|\boldsymbol{h}_i\|_F^2}{E\left[\|\boldsymbol{h}_i\|_F^2\right]}\xrightarrow{P}1,\quad i=1,\cdots,n_T, \tag{37}$$

as $n_R\to\infty$. By this definition, the channel hardening phenomenon becomes more embodied, which will be thought to occur if the probability that $\frac{\|\boldsymbol{h}_i\|_F^2}{E\left[\|\boldsymbol{h}_i\|_F^2\right]}$ differs from 1 by more than δ ($\delta>0$) approaches zero as n_R approaches infinity. Meanwhile, a simple criterion, based on the Chebyshev inequality, is also proposed to check whether the channel hardens [18].

Then, a natural question is whether a similar hardening phenomenon occurs with metrics other than $\|\boldsymbol{h}_i\|_F^2$. To answer this, we propose a generalized definition as follows.

Definition 1: A metric ϕ in large MIMO systems is said to have the *hardening* property if

$$\mathcal{H}(\phi)\triangleq\frac{\phi}{E[\phi]}\xrightarrow{P}1, \tag{38}$$

in the limit as $n_R\to\infty$, where $\mathcal{H}$ is called the evaluation function for the hardening.

This definition formalizes the phenomenon that certain metrics in MIMO systems, after being normalized by its mean, converge in probability to 1, as the number of receive antennas increases. Then, the relationships in Theorems 1 and 2 immediately allow us to discover a hardening phenomenon related to the distance metric $\|\boldsymbol{D}_{H,W}\|_F^2$ within large MIMO systems.

Theorem 4: If $\boldsymbol{D}$, ρ_h, and N_0 are fixed, the distance metric $\|\boldsymbol{D}_{H,W}\|_F^2$ has the hardening property, i.e., as $n_R\to\infty$,

$$\mathcal{H}\left(\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right)\xrightarrow{P}1. \tag{39}$$

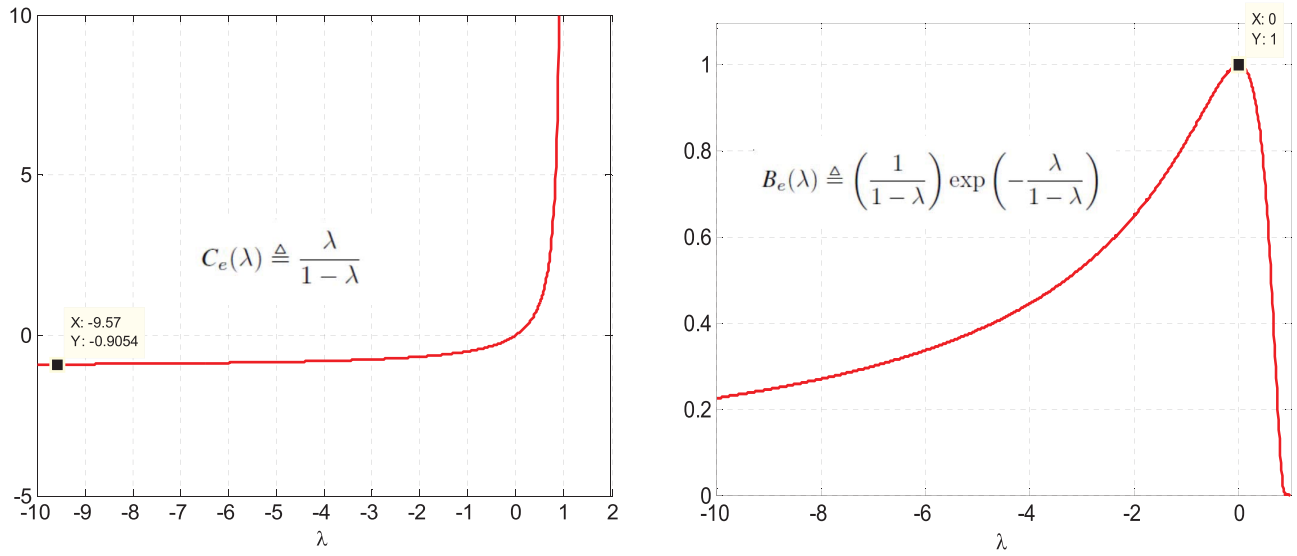

Fig. 1. Curves of C_e and B_e.

Proof: For any $\varepsilon > 0$, referring to Lemma 1 and Theorem 2, we can get

$$\begin{aligned} & P\left(\left|\mathcal{H}\left(\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right)-1\right| \geq \varepsilon\right) \\ & \leq P\left(\frac{\left\|\boldsymbol{D}_{H,W}\right\|_F^2}{n_R \rho_h \|\boldsymbol{D}\|_F^2 + t n_R N_0} \geq 1+\varepsilon\right) \\ & \quad + P\left(\frac{\left\|\boldsymbol{D}_{H,W}\right\|_F^2}{n_R \rho_h \|\boldsymbol{D}\|_F^2 + t n_R N_0} \leq 1-\varepsilon\right) \\ & \leq U_c\left(1+\varepsilon, \boldsymbol{\lambda}_{1+\varepsilon}\right) + U_c\left(1-\varepsilon, \boldsymbol{\lambda}_{1-\varepsilon}\right), \end{aligned} \tag{40}$$

where the term $P\left(\frac{\left\|\boldsymbol{D}_{H,W}\right\|_F^2}{n_R \rho_h \|\boldsymbol{D}\|_F^2 + t n_R N_0} \leq 1-\varepsilon\right) = 0$ if $\varepsilon > 1$. $\boldsymbol{\lambda}_{1+\varepsilon}$ and $\boldsymbol{\lambda}_{1-\varepsilon}$ denote the vectors $\boldsymbol{\lambda}$ which are computed from (25) and (26) with $\theta = 1+\varepsilon$ and $\theta = 1-\varepsilon$, respectively. For a given ε, when $\boldsymbol{D}$, ρ_h, and N_0 are fixed, $\boldsymbol{\lambda}_{1+\varepsilon}$ and $\boldsymbol{\lambda}_{1-\varepsilon}$ should not vary with n_R according to Corollary 4. This means that $U_c\left(1+\varepsilon, \boldsymbol{\lambda}_{1+\varepsilon}\right)$ and $U_c\left(1-\varepsilon, \boldsymbol{\lambda}_{1-\varepsilon}\right)$ exponentially decay as n_R increases, since at least one element of either $\boldsymbol{\lambda}_{1+\varepsilon}$ or $\boldsymbol{\lambda}_{1-\varepsilon}$ is nonzero such that, with respect to Lemma 3, there is $B_e(\lambda_j) < 1$ with some $1 \leq j \leq s$. Thus, for $\varepsilon > 0$, a substitution of (28) and (31) into (40) yields

$$\lim_{n_R \to \infty} P\left(\left|\mathcal{H}\left(\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right)-1\right| \geq \varepsilon\right) = 0. \qquad \blacksquare \tag{41}$$

Theorem 5: With fixed $\boldsymbol{D}$ and N_0, if ρ_h is a decreasing function of n_R and

$$\lim_{n_R \to \infty} \rho_h = 0, \tag{42}$$

the distance metric $\|\boldsymbol{D}_{H,W}\|_F^2$ has the hardening property, i.e., as $n_R \to \infty$,

$$\mathcal{H}\left(\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right) \xrightarrow{P} 1. \tag{43}$$

Proof: When $n_R \to \infty$, $\rho_h \to 0$ such that $\boldsymbol{\lambda}$ tends to the minimum $\boldsymbol{\lambda}^*$. It further yields $U_c(\theta, \boldsymbol{\lambda}) \to \left(\theta e^{1-\theta}\right)^{t n_R}$. Once again, by using (40), we can complete the proof. $\blacksquare$

Theorem 6: If the entries in each column of the virtual channel power matrix $\boldsymbol{\Psi}$ are identical, that is, $\psi_{i_1,j} = \psi_{i_2,j}$, $1 \leq i_1, i_2 \leq n_R$, for every j (note that the entries in each row are not necessarily same), then the distance metric $\|\boldsymbol{d}_{H,n}\|_F^2$ has the hardening property, i.e., as $n_R \to \infty$,

$$\mathcal{H}\left(\left\|\boldsymbol{d}_{H,n}\right\|_F^2\right) \xrightarrow{P} 1. \tag{44}$$

Proof: Since $\psi_{i_1,j} = \psi_{i_2,j}$, $1 \leq i_1, i_2 \leq n_R$, for every j, we can get

$$\frac{\sum_{j=1}^{n_T} \psi_{i,j} \left|\widetilde{d}_j\right|^2 + N_0}{\eta} = \frac{1}{n_R}, \tag{45}$$

and $\overline{\gamma} = n_R\left(1-\frac{1}{\theta}\right)$. The Chernoff bound can be rewritten as $\left[B_e\left(1-\frac{1}{\theta}\right)\right]^{n_R}$ that approaches 0 as $n_R \to \infty$. $\blacksquare$

Theorems 4, 5, and 6 show that the hardening phenomenon is not peculiar to channel propagation, but it also occurs in other type of metrics as well. For example, the normalized distance metric $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{E_{H,W}\left[\|\boldsymbol{D}_{H,W}\|_F^2\right]}$ converges in probability to 1, as the number of receive antennas n_R increases, where $E_{H,W}\left[\|\boldsymbol{D}_{H,W}\|_F^2\right] = n_R \rho_h \|\boldsymbol{D}\|_F^2 + t n_R N_0$. This phenomenon is called distance hardening and formulated by virtue of the evaluation function $\mathcal{H}$. Additionally, the Chernoff bounds given by Theorems 1, 2, and 3 help to quantify the convergence behavior of $P\left(\left|\mathcal{H}\left(\left\|\boldsymbol{D}_{H,W}\right\|_F^2\right)-1\right| \geq \varepsilon\right)$ and $P\left(\left|\mathcal{H}\left(\left\|\boldsymbol{d}_{H,n}\right\|_F^2\right)-1\right| \geq \varepsilon\right)$. Due to space limitations, we neglect the detailed discussion of these aspects.

B. Discussion

The formulation of distance hardening in this section together with the tail probability analysis presented in the

previous section, allow us to gain insights into the statistics of $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ and $\|\boldsymbol{d}_{H,n}\|_F^2$. This could have implications for space-time code design [30], [36], MIMO detection [3], [31], and mmWave MIMO systems [22], [23], [26].

1) Space-Time Code Design: The design criteria for space-time codes are typically based on connecting the PEP behavior of a code for a certain channel under a given decoding algorithm with the specific metric [29], [30], where summing the PEPs gives rise to an upper bound on the average probability of error [27], [29]. We now propose a new space-time decoding strategy and its related criterion of code design. For convenience, in what follows, let $X_1, X_2, X_3 \in \mathbf{C}$ denote three possible transmitted codewords.

Proposed Decoding Strategy: When considering the MIMO system model of (2), the space-time decoding problem can be solved by finding a codeword $\widetilde{X} \in \mathbf{C}$ that satisfies

$$\frac{\left\|\boldsymbol{Y}-\boldsymbol{H}\widetilde{X}\right\|_F^2}{n_R\rho_h d_{\min}^2+tn_RN_0} \le \beta, \tag{46}$$

where $d_{\min}^2 \triangleq \min_{X_1,X_2\in\mathbf{C}} \|X_2-X_1\|_F^2$ and β is an adjustable parameter.

For this strategy, we can get the following result.

Corollary 6: Assume that a space-time code is properly designed such that $\rho_h d_{\min}^2 \ge s$, and its decoding uses the strategy (46) with $\beta = N_0 \log(1/N_0)$, where $\log(\cdot)$ is the logarithm to either base e or base 2. Then, the probability of transmitting X_1 and deciding in favor of X_2 at the decoder (the PEP), denoted by $P(X_1 \to X_2)$, satisfies

$$\lim_{N_0\to 0} \frac{\log P(X_1\to X_2)}{\log N_0} \ge sn_R. \tag{47}$$

Thus a diversity of sn_R can be achieved.

Sketch of Proof: We assume the received signal for X_1 is $\boldsymbol{Y}_1$ $(=\boldsymbol{H}X_1+\boldsymbol{N})$ and $\Delta X \triangleq X_1 - X_2$. It holds that

$$\begin{aligned} P(X_1\to X_2) \le\ & P\left(\frac{\|\boldsymbol{Y}_1-\boldsymbol{H}X_1\|_F^2}{n_R\rho_h d_{\min}^2+tn_RN_0}\ge\beta\right) \\ & + P\left(\frac{\|\boldsymbol{Y}_1-\boldsymbol{H}X_2\|_F^2}{n_R\rho_h d_{\min}^2+tn_RN_0}\le\beta\right). \end{aligned} \tag{48}$$

Here, using Theorem 1, we have

$$\begin{aligned} & P\left(\frac{\|\boldsymbol{Y}_1-\boldsymbol{H}X_1\|_F^2}{n_R\rho_h d_{\min}^2+tn_RN_0}\ge\beta\right) \\ & \le P\left(\frac{\|\boldsymbol{Y}_1-\boldsymbol{H}X_1\|_F^2}{tn_RN_0}\ge\beta\left(\frac{\rho_h d_{\min}^2}{tN_0}\right)\right) \\ & \le U_c\left(\beta\left(\frac{\rho_h d_{\min}^2}{tN_0}\right),\overline{\gamma},\widetilde{\boldsymbol{\sigma}}_{X_1-X_1},N_0\right), \end{aligned} \tag{49}$$

where $\widetilde{\boldsymbol{\sigma}}_{X_1-X_1}$ is an empty vector because $\|\boldsymbol{Y}_1-\boldsymbol{H}X_1\|_F^2 = \|\boldsymbol{H}(X_1-X_1)+\boldsymbol{N}\|_F^2$ in which X_1-X_1 leads to an all-zero matrix $\boldsymbol{D}$ for Theorem 1, and

$$\begin{aligned} & P\left(\frac{\|\boldsymbol{Y}_1-\boldsymbol{H}X_2\|_F^2}{n_R\rho_h d_{\min}^2+tn_RN_0}\le\beta\right) \\ & \le P\left(\frac{\|\boldsymbol{Y}_1-\boldsymbol{H}X_2\|_F^2}{n_R\rho_h\|\Delta X\|_F^2+tn_RN_0}\le\beta\right) \\ & \le U_c\left(\beta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}}_{X_1-X_2},N_0\right), \end{aligned} \tag{50}$$

where the elements of $\widetilde{\boldsymbol{\sigma}}_{X_1-X_2}$ are obtained by setting $\boldsymbol{D} = \Delta X$ in Theorem 1. If $\beta = N_0\log(1/N_0)$, then $\beta < 1$ and $\beta\left(\frac{\rho_h d_{\min}^2}{tN_0}+1\right) > 1$ as $N_0 \to 0$. After some mathematical manipulation, we can obtain

$$\lim_{N_0\to 0} \frac{\log U_c\left(\beta\left(\frac{\rho_h d_{\min}^2}{tN_0}\right),\overline{\gamma},\widetilde{\boldsymbol{\sigma}}_{X_1-X_1},N_0\right)}{\log N_0} \ge n_R\rho_h\|\Delta X\|_F^2, \tag{51}$$

$$\lim_{N_0\to 0} \frac{\log U_c\left(\beta,\overline{\gamma},\widetilde{\boldsymbol{\sigma}}_{X_1-X_2},N_0\right)}{\log N_0} \ge sn_R. \tag{52}$$

Then if $\rho_h d_{\min}^2 \ge s$, (47) follows, because $\|\Delta X\|_F^2 \ge d_{\min}^2$. Note that, being the number of nonzero singular values of ΔX, s is equal to the rank of ΔX. ■

From the preceding analysis, we arrive at the following *proposed design criterion* of space-time code for the decoding strategy (46). To obtain a diversity of sn_R, the minimum rank of ΔX has to be s, and $\rho_h d_{\min}^2 \ge s$, over all distinct pairs of codewords X_1 and X_2.

According to the rank criterion of space-time code design [30], a diversity of sn_R can be achieved by the ML decoder if the minimum rank of all possible ΔX is s. Compared to the ML decoder, the decoding strategy (46) requires an extra condition, $\rho_h d_{\min}^2 \ge s$, to attain the same diversity. The advantage of the strategy (46) is that the decoding immediately terminates once a codeword is found to meet the condition (46), but the ML decoding has to perform a brute-force search until all the codewords in $\mathbf{C}$ have been verified. At first glance, there is some analogy between the relationship in (46) and the idea of sphere decoding. However, sphere decoding still has to search all candidates that lie in a sphere [3] and (46) needs to find just one candidate. This means that we might be able to realize the very low complexity algorithm of space-time decoding with near-ML performance by combining the decoding strategy (46) with some depth-first [3] or local search approaches [31].

To obtain the final solution of the low complexity decoding algorithm based on our proposed decoding strategy and the related space-time code based on the proposed design criterion is still not a simple task. However, (46) and Corollary 6 allow us to exploit some results on the depth-first or local search based decoding algorithms [3], [31] and the space-time codes that have certain Euclidean distance properties, such as spherical codes [29]. This could be an interesting future study.

2) MIMO Detection: According to the classical approach of analyzing any type of modulation or coding schemes, the distance properties of the signals under consideration are

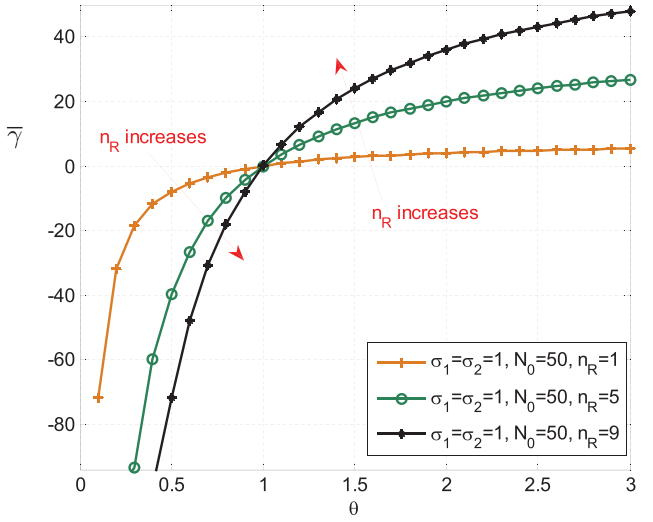

Fig. 2. Plot of $\overline{\gamma}$ as a function of θ for different n_R with fixed σ_j, $1 \le j \le 2$, and N_0.

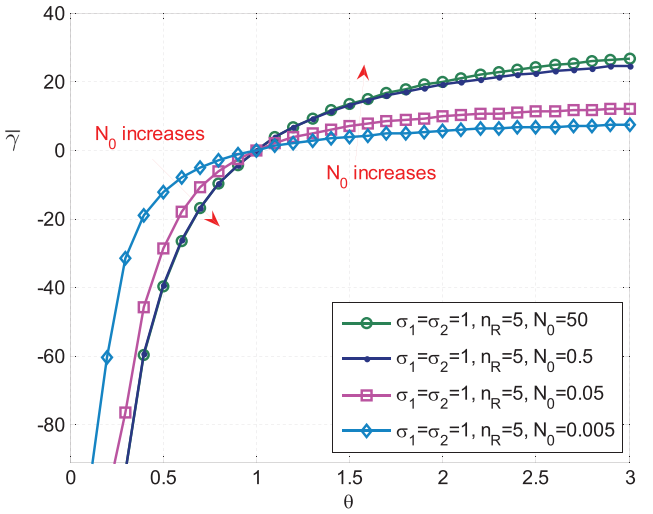

Fig. 3. Plot of $\overline{\gamma}$ as a function of θ for different N_0 with fixed σ_j, $1 \le j \le 2$, and n_R.

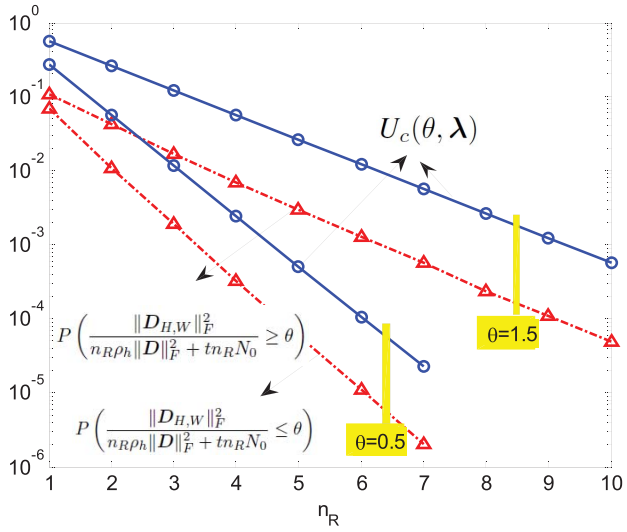

Fig. 4. Tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ vs its Chernoff bound.

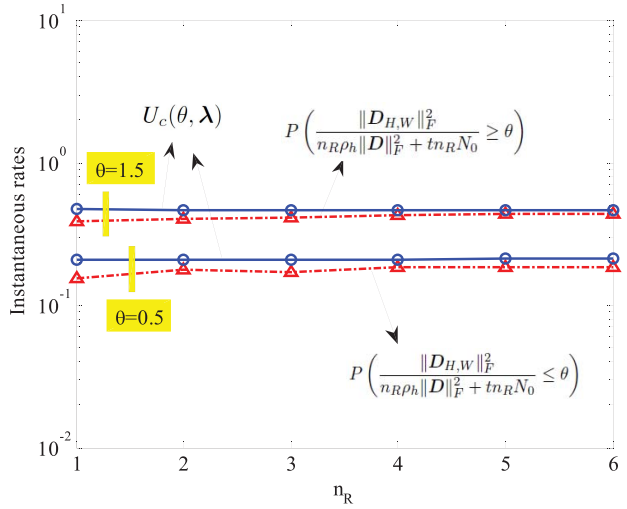

Fig. 5. Instantaneous rates of convergence for the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ and its Chernoff bound.

always worth studying [29]. As far as MIMO encoding is concerned [31], this study reveals a connection between code distances before and after MIMO propagation, i.e., $\|X_1 - X_2\|_F^2$ and $\|\boldsymbol{H}(X_1 - X_2)\|_F^2$ (or $\|\boldsymbol{H}(X_1 - X_2) + \boldsymbol{N}\|_F^2$), respectively. The results obtained might potentially find applications in MIMO detection.

First, we can develop new post-processing functionality for MIMO detection to reduce the computational complexity or improve the performance of the receive links. For example, the theoretical result of Corollary 2 has inspired the design of a functionality that can verify whether a solution obtained by MIMO detection is full-diversity [28].

Second, many local search algorithms, such as likelihood ascent search (LAS) and reactive tabu search (RTS), have been designed in large MIMO detection domains [31]. Typically, a local search starts with an initial solution and then attempts to find better solutions by searching in the neighborhoods defined by certain a neighborhood function [31]. In practice, local search algorithms are known to converge quickly and find high-quality solutions. This study provides an intuitive explanation for the efficiency of local search algorithms. The neighboring codewords of a transmitted code are more likely to become the neighbors of the corresponding received code, which can be observed from a result as follows.

Corollary 7: Let X_1, X_2, and X_3 be three codewords in MIMO encoding. We define $l_{12} \triangleq \|X_1 - X_2\|_F^2$, $l_{13} \triangleq \|X_1 - X_3\|_F^2$, and $\Delta l \triangleq l_{13} - l_{12}$. Assuming that X_2 stays closer to X_1 than X_3 (i.e., $l_{12} < l_{13}$) results in

$$P\left(\|\boldsymbol{H}(X_1 - X_2)\|_F^2 \ge \|\boldsymbol{H}(X_1 - X_3)\|_F^2\right) \le U_c\left(1 + \frac{\Delta l}{2l_{12}}, \boldsymbol{\lambda}_{X_1 - X_2}\right) + U_c\left(1 - \frac{\Delta l}{2l_{13}}, \boldsymbol{\lambda}_{X_1 - X_3}\right), \quad (53)$$

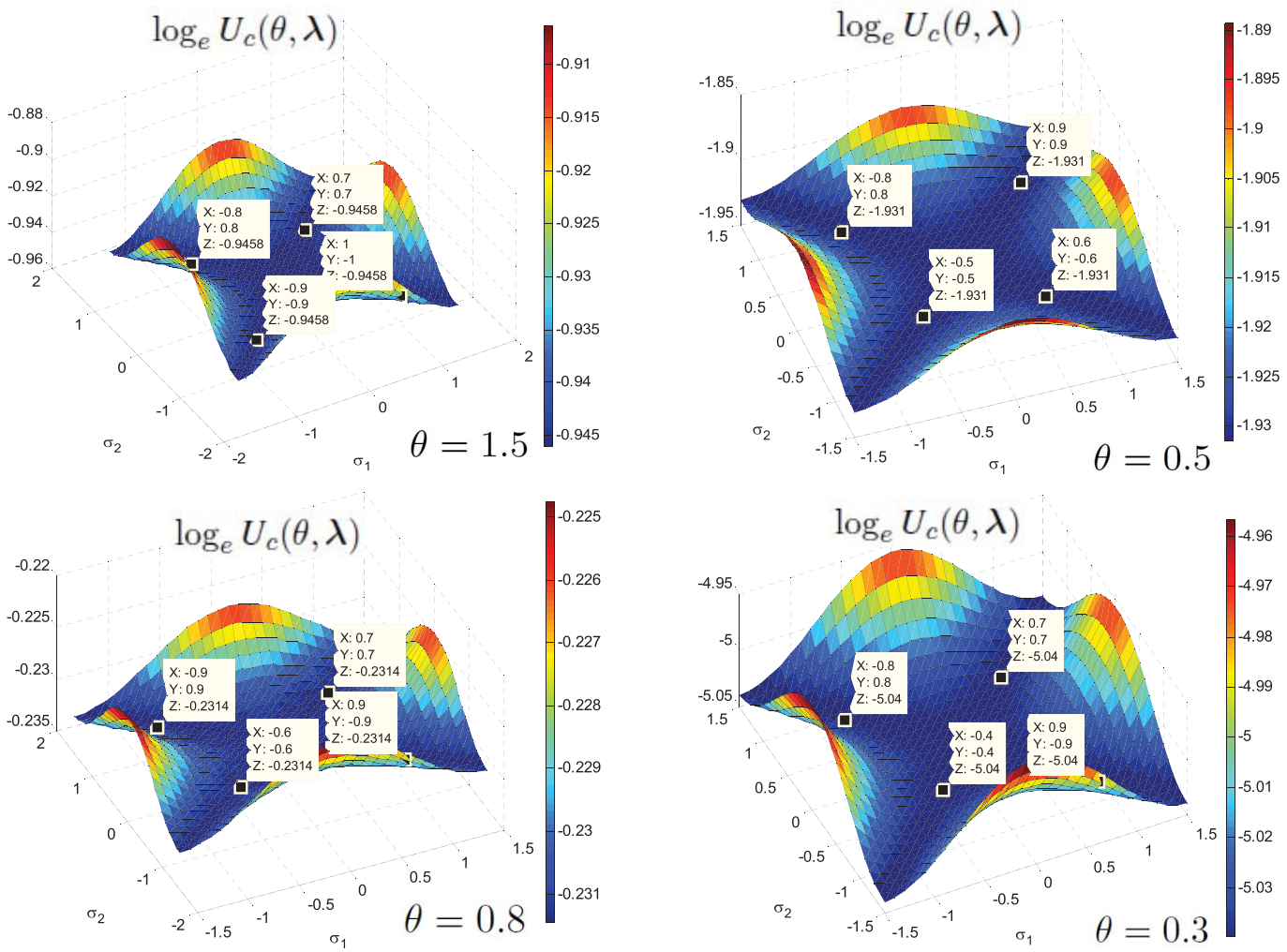

Fig. 6. Numerical results for Corollary 5.

where the elements of $\boldsymbol{\lambda}_{X_1-X_2}$ and $\boldsymbol{\lambda}_{X_1-X_3}$ are obtained by setting $\boldsymbol{D} = X_1 - X_2$ and $\boldsymbol{D} = X_1 - X_3$ in Theorem 2, respectively. Furthermore, the probability $P\left(\|\boldsymbol{H}(X_1 - X_2)\|_F^2 \geq \|\boldsymbol{H}(X_1 - X_3)\|_F^2\right)$ can be arbitrarily small as long as n_R is large enough. The reason is that if X_1, X_2, X_3, ρ_h, and N_0 are fixed, this probability exponentially decays as n_R grows.

Sketch of Proof: First, the event $\|\boldsymbol{H}(X_1 - X_2)\|_F^2 \geq \|\boldsymbol{H}(X_1 - X_3)\|_F^2$ actually belongs to the union of two events: $\|\boldsymbol{H}(X_1 - X_2)\|_F^2 \geq n_R\rho_h\left(l_{12} + \frac{\Delta l}{2}\right)$ and $\|\boldsymbol{H}(X_1 - X_3)\|_F^2 \leq n_R\rho_h\left(l_{13} - \frac{\Delta l}{2}\right)$. Note here that $l_{12} + \frac{\Delta l}{2} = l_{13} - \frac{\Delta l}{2}$. This can lead to (53). Second, according to Corollary 4, if X_1, X_2, X_3, ρ_h, and N_0 are fixed, both $\boldsymbol{\lambda}_{X_1-X_2}$ and $\boldsymbol{\lambda}_{X_1-X_3}$ will not vary with n_R. Hence, from (31), it follows that $U_c\left(1 + \frac{\Delta l}{2\,l_{12}}, \boldsymbol{\lambda}_{X_1-X_2}\right)$ and $U_c\left(1 - \frac{\Delta l}{2\,l_{13}}, \boldsymbol{\lambda}_{X_1-X_3}\right)$ tend to 0 at exponential rates as n_R increases, which concludes the proof. ■

3) mmWave MIMO Systems: The Chernoff bound for the tail probability of $\left\|\boldsymbol{d}_{H,n}\right\|_F^2$, as given in Theorem 3, is applicable to mmWave MIMO systems with the model (4). For these systems, Theorem 6 presents a simple case in which distance hardening can be observed. However, Theorem 6 cannot have a wide significance for clarifying the distance hardening phenomenon in mmWave MIMO systems, since its assumption is quite special and it does not take the sparsity of mmWave MIMO channels into consideration. Recent studies [13], [22], [23] show that mmWave MIMO channels are likely to be sparse, where only a few degrees of freedom are dominant to contribute towards channel capacity. This sparsity makes the formulation of the hardening property complicated. It should be interesting to investigate the distance hardening in mmWave MIMO systems with channel sparsity. We do not present such an investigation in this paper due to space limitation, which we leave for future work.

V. Numerical Studies

This section is devoted to illustrating the analytical results obtained previously via numerical studies.

Figures 2 and 3 illustrate the curves of $\overline{\gamma}$ as a function of θ, with the same parameters as $n_T = 2$ ($\sigma_1 = \sigma_2 = 1$), $\rho_h = 1/n_R$, and $t = 8$ but with different n_R and N_0, respectively. This illustrates the relationship of (18) and confirms the statements in Corollary 1 such as the monotonicity of the implicit function $\overline{\gamma}(\theta)$ and the special point of position,

i.e., $\overline{\gamma} = 0$ at $\theta = 1$. Figures 2 and 3 also show how n_R and N_0 affect the relationship between $\overline{\gamma}$ and θ, respectively. A surprise is that the relationship between $\overline{\gamma}$ and θ tends to be gradually fixed as N_0 becomes sufficiently large since, as we can see, the curves of $\overline{\gamma}$ with $N_0 = 0.5$ and $N_0 = 50$ almost overlap each other in Figure 3. However, this is still reasonable since the relationship between $\overline{\gamma}$ and θ will tend to $\left(\frac{\xi}{N_0} - \overline{\gamma}\right)^{-1} = \frac{\theta}{tn_R}$ if N_0 grows relatively large such that σ_j^2 in (18) can be ignored.

To demonstrate the tightness of the Chernoff bound on the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$, we present the examples in Figures 4 and 5, where the system parameters are set as $n_T = 4$, $\rho_h = 1/n_R$, and $t = 8$. Figure 4 straightforwardly compares the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ with its Chernoff bound, from which we observe that there exists a gap between every tail probability and its bound. Relatively speaking, this gap is not very significant for large MIMO systems because, to reach the same value for curves of the tail probability and its bound, the difference between the corresponding n_R is less than 3. The above gap seems to be roughly constant which implies that the rate of convergence of the Chernoff bound approximately coincides with that of the tail probability if the rate of convergence is defined as (1). Since $P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0} \geq \theta\right)$ for $\theta > 1$, $P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0} \leq \theta\right)$ for $\theta < 1$, and $U_c(\theta, \boldsymbol{\lambda})$ are all related to n_R, we can evaluate their rates of convergence with increasing n_R by replacing k by n_R in (1) and constructing the sequences as $\{\dot{x}_{n_R}\}$, $\{\bar{x}_{n_R}\}$, and $\{\tilde{x}_{n_R}\}$, where

$$\dot{x}_{n_R} = P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2 + tn_RN_0} \geq \theta\right), \tag{54}$$

$$\bar{x}_{n_R} = P\left(\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2 + tn_RN_0} \leq \theta\right), \tag{55}$$

and $\tilde{x}_{n_R} = U_c(\theta, \boldsymbol{\lambda})$.

Computing the rate of convergence r is hard in simulations since it is almost impossible to obtain the values of x_{n_R+1} and x_{n_R} with $n_R = \infty$ as the definition (1) require. Thus, Figure 5 plots the instantaneous rates for $\{\dot{x}_{n_R}\}$, $\{\bar{x}_{n_R}\}$, and $\{\tilde{x}_{n_R}\}$, which can be regarded as the approximation of r. It is clear from Figure 5 that the instantaneous rates of convergence for the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R\rho_h\|\boldsymbol{D}\|_F^2+tn_RN_0}$ and its corresponding Chernoff bound stay close to each other and approach some fixed value.

The numerical results for Corollary 5 that are concerned with the extreme point of Chernoff bound are illustrated by Figure 6. Within the simulation, the parameters are set as $n_T = 2$, $n_R = 5$, $t = 2$, $\rho_h = 1/n_R$, and $N_0 = 1$. Since $n_T = 2$ and $t = 2$ such that $s = \min\{n_T, t\} = 2$, we shall have $\|\boldsymbol{D}\|_F^2 = \sigma_1^2 + \sigma_2^2$. In Figure 6, $\log_e U_c(\theta, \boldsymbol{\lambda})$ is shown as the function of σ_1 and σ_2, where λ_1 and λ_2 that constitute $\boldsymbol{\lambda}$, can be computed from σ_1 and σ_2 according to (25). It can be observed from Figure 6 that $\log_e U_c(\theta, \boldsymbol{\lambda})$ reaches the minimum value -0.9458, -0.2314, -1.931, and -5.04 at $|\sigma_1| = |\sigma_2|$ for $\theta = 1.5$, $\theta = 0.8$, $\theta = 0.5$, and $\theta = 0.3$, respectively. It shall be noted that the relationship $|\sigma_1| = |\sigma_2|$ means that $\lambda_1 = \lambda_2 = 1 - \frac{1}{\theta}$. Thus, these simulation results conform to Corollary 5. Likewise, according to Corollary 5, the solution $\lambda_1 = \lambda_2 = 1 - \frac{1}{\theta}$ is proven to attain the global minimum if $0 < \theta \leq 1/2$, whereas it seems that such a solution can also yield the global minimum for $\theta > 1/2$ as shown in Figure 6.

VI. Conclusion and Future Research Implications

The research on an asymptotic property in large MIMO systems (the distance hardening phenomenon) is presented in this paper. As we have seen, the large number of antennas and the resulting channel hardening make the analysis and design of large MIMO systems more convenient in previous works. The elementary concept of the hardening phenomenon is particularly useful in describing the asymptotic property that we report in this study.

In fact, to clearly reveal and properly describe this property, we first derive the Chernoff bounds on the tail probabilities of the distance metrics of interest (see Theorems 1-3). Then, we formulate the asymptotic property by utilizing a generalized definition for the hardening phenomena that originates from a formal definition of the well-known channel hardening (see Theorems 4-6). For the Chernoff bound given in the two most important theorems of this study (Theorems 1 and 2), several related properties are provided and extreme point analysis is performed in order to clarify its mathematical characteristics. It is also shown that the expressions of Chernoff bound can be extended to MIMO channels with transmit correlation (Corollary 3).

It appears that the asymptotic property reported in this paper may have potential applications in the design of space-time code for large MIMO systems and the design of new post-processing functionality for MIMO detection. First, we propose a new space-time decoding strategy and its related criterion of code design (the most important result therein, Corollary 6, might finally guide us to a promising solution of low complexity decoding algorithms based on our proposed decoding strategy and its related space-time code design). Second, some theoretical result (Corollary 2) has inspired the design of a full-diversity solution identification that can verify whether a solution obtained by MIMO detection is of full-diversity. This study provides an intuitive explanation for the efficiency of local search algorithms, which can be observed from Corollary 7. A more comprehensive analysis and its extension to analyzing the error probability performance of local search algorithms might be an interesting future work, because to the best of the author's knowledge, it still lacks such an analysis. For mmWave MIMO systems, this paper provides a very simple example of the distance hardening phenomenon. While it still lacks an in-depth analysis of the distance hardening for mmWave MIMO systems with channel sparsity, which will be our future work.

Appendix A

Proof of Lemma 1:

By applying the singular value decomposition (SVD), we get $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are $n_T \times n_T$ and $t \times t$

unitary matrices, respectively, $\boldsymbol{\Sigma}$ is a diagonal $n_T \times t$ matrix which is defined as $\boldsymbol{\Sigma} \triangleq \text{diag}\{\sigma_1, \sigma_2, \cdots\}$. Without loss of generality, we assume that $\boldsymbol{D}$ has s singular values such that $\sigma_1, \cdots, \sigma_s \geq 0$, where $s = \min\{n_T, t\}$. This further implies

$$\|\boldsymbol{D}\|_F^2 = \text{Tr}\left(\boldsymbol{\Sigma}^H \boldsymbol{\Sigma}\right) = \sum_{j=1}^{s} \sigma_j^2. \tag{56}$$

For (8), it establishes that

$$\begin{aligned}\|\boldsymbol{HD} + \boldsymbol{W}\|_F^2 &= \left\|\left(\boldsymbol{HU\Sigma} + \boldsymbol{WV}\right)\boldsymbol{V}^H\right\|_F^2 \\ &= \|\boldsymbol{HU\Sigma} + \boldsymbol{WV}\|_F^2 = \left\|\widetilde{\boldsymbol{H}}\boldsymbol{\Sigma} + \widetilde{\boldsymbol{W}}\right\|_F^2, \end{aligned} \tag{57}$$

where $\widetilde{\boldsymbol{H}} \triangleq \boldsymbol{HU}$ and $\widetilde{\boldsymbol{W}} \triangleq \boldsymbol{WV}$ are defined for notational simplicity. From [19, Lemma A.1], it follows that $\widetilde{\boldsymbol{W}} \overset{d}{=} \boldsymbol{W}$, and $\|\widetilde{\boldsymbol{W}}\|_F^2 = \text{Tr}(\widetilde{\boldsymbol{W}}^H \widetilde{\boldsymbol{W}})$ is chi-squared distributed with $2n_R t$ degrees of freedom.

By assuming that $\widetilde{\boldsymbol{H}} = \left[\widetilde{h}_{i,j}\right]$ has columns $\widetilde{\boldsymbol{h}}_1, \cdots, \widetilde{\boldsymbol{h}}_{n_T}$ and $\widetilde{\boldsymbol{W}} = \left[\widetilde{w}_{i,k}\right]$ has columns $\widetilde{\boldsymbol{w}}_1, \cdots, \widetilde{\boldsymbol{w}}_t$, we shall get the below equivalent expression of $\|\boldsymbol{D}_{H,W}\|_F^2$ as

$$\begin{aligned}\|\boldsymbol{D}_{H,W}\|_F^2 &= \|\boldsymbol{HD} + \boldsymbol{W}\|_F^2 = \left\|\widetilde{\boldsymbol{H}}\boldsymbol{\Sigma} + \widetilde{\boldsymbol{W}}\right\|_F^2 \\ &= \sum_{j=1}^{s} u_j + \sum_{k=s+1}^{t} v_k, \end{aligned} \tag{58}$$

where $u_j \triangleq \left(\sigma_j \widetilde{\boldsymbol{h}}_j + \widetilde{\boldsymbol{w}}_j\right)^H \left(\sigma_j \widetilde{\boldsymbol{h}}_j + \widetilde{\boldsymbol{w}}_j\right)$, and $v_k \triangleq (\widetilde{\boldsymbol{w}}_k)^H \widetilde{\boldsymbol{w}}_k$. Note that, the terms v_k will disappear if $n_T \geq t$ such that $s = \min\{n_T, t\} = t$.

Lemma A.1: $\left(\frac{\rho_h \sigma_j^2}{2} + \frac{N_0}{2}\right)^{-1} u_j \sim \chi^2_{2\,n_R}$, and $\left(\frac{N_0}{2}\right)^{-1} v_k \sim \chi^2_{2\,n_R}$ as long as $s < t$.

Proof: The sum of two statistically independent Gaussian random variables is also a Gaussian random variable; hence, we can get $\sigma_j \widetilde{h}_{i,j} + \widetilde{w}_{i,j} \sim \mathcal{CN}\left(0, \rho_h \sigma_j^2 + N_0\right)$ and $\widetilde{w}_{i,k} \sim \mathcal{CN}(0, N_0)$. ∎

Lemma A.2: If $t > n_T$, then

$$\left(\frac{N_0}{2}\right)^{-1} \sum_{k=s+1}^{t} v_k \sim \chi^2_{2(t-s)n_R}. \tag{59}$$

APPENDIX B

Proof of Theorems 1 and 3:

Obviously, the tail probability of $\frac{\|\boldsymbol{D}_{H,W}\|_F^2}{n_R \rho_h \|\boldsymbol{D}\|_F^2 + t n_R N_0}$ is decided by the statistics of $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$. Due to Lemma 1, $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$ can be treated as the random variable: $y = \frac{1}{\varphi}\sum_i^N x_i$, where x_i, $1 \leq i \leq N$, are independent random variables, $\frac{x_i}{\alpha_i^2} \sim \chi_2^2$ is chi-squared distributed with 2 degrees of freedom with $E[x_i] = 2\alpha_i^2$, and $\varphi = \sum_{i=1}^N E[x_i] = 2\sum_{i=1}^N \alpha_i^2$.

For y, the following relations hold:

$$E\left[ye^{\gamma y}\right] = \frac{1}{\varphi}\sum_{i=1}^{N} E\left[x_i \exp\left(\frac{\gamma}{\varphi}x_i\right)\right] E\left[\exp\left(\frac{\gamma}{\varphi}\sum_{j\neq i} x_j\right)\right]$$

in which $E\left[x_i \exp\left(\frac{\gamma}{\varphi}x_i\right)\right] = 2\alpha_i^2\left(1 - \frac{2\gamma\alpha_i^2}{\varphi}\right)^{-2}$, and

$$E\left[e^{\gamma y}\right] = E\left[\exp\left(\frac{\gamma}{\varphi}\sum_{i=1}^{N} x_i\right)\right] = \prod_{i=1}^{N} E\left[\exp\left(\frac{\gamma}{\varphi}x_i\right)\right]$$

in which $E\left[\exp\left(\frac{\gamma}{\varphi}x_i\right)\right] = \left(1 - \frac{2\gamma\alpha_i^2}{\varphi}\right)^{-1}$ where it is required that $\gamma < \frac{\varphi}{2\alpha_i^2}$ such that $\left(1 - \frac{2\gamma\alpha_i^2}{\varphi}\right)^{-1} > 0$.

The Chernoff bound on the upper tail probability $P(y \geq \theta)$ for $\theta > E[y]$ can be obtained as [35]

$$P(y \geq \theta) \leq e^{-\overline{\gamma}\theta} \prod_{i=1}^{N}\left(1 - \frac{2\overline{\gamma}\alpha_i^2}{\varphi}\right)^{-1}, \tag{60}$$

where $\overline{\gamma}$ is the solution of the equation $E\left[ye^{\overline{\gamma}y}\right] = \theta E\left[e^{\overline{\gamma}y}\right]$ which can be simplified to the form as

$$\sum_{i=1}^{N}\left(\frac{\varphi}{2\alpha_i^2} - \overline{\gamma}\right)^{-1} = \theta. \tag{61}$$

In a similar manner, the Chernoff bound on the lower tail probability $P(y \leq \theta)$ for $\theta < E[y]$ can be derived as [35]

$$P(y \leq \theta) \leq e^{-\overline{\gamma}\theta} \prod_{i=1}^{N}\left(1 - \frac{2\overline{\gamma}\alpha_i^2}{\varphi}\right)^{-1}, \tag{62}$$

where $\overline{\gamma}$ is also determined by (61).

Since the relationships in (60) and (62) apply to $\left\|\boldsymbol{D}_{H,W}\right\|_F^2$, Theorem 1 can, therefore, be proved by employing Lemma 1, Lemma A.1, and Lemma A.2. Moreover, using a similar procedure to analyze $\left\|\boldsymbol{d}_{H,n}\right\|_F^2$ can verify Theorem 3.

APPENDIX C

Proof of Lemma 2:

Along similar lines as Lemma 1, we obtain $\|\boldsymbol{d}_{H,n}\|_F^2 = \|\boldsymbol{Hd} + \boldsymbol{n}\|_F^2 = \left\|\boldsymbol{H}_v\widetilde{\boldsymbol{d}} + \widetilde{\boldsymbol{n}}\right\|_F^2$, where $\widetilde{\boldsymbol{n}} \triangleq \boldsymbol{A}_r^H \boldsymbol{n} = \left[\widetilde{n}_1, \cdots, \widetilde{n}_{n_R}\right]^T$. Furthermore, since $\boldsymbol{A}_r$ is unitary, $\widetilde{\boldsymbol{n}} \overset{d}{=} \boldsymbol{n}$ owing to [19, Lemma A.1].

Lemma C.1: Assuming that $\boldsymbol{H}_v = \left[h_{i,j,v}\right]$, then

$$\|\boldsymbol{d}_{H,n}\|_F^2 = \|\boldsymbol{Hd} + \boldsymbol{n}\|_F^2 = \sum_{i=1}^{n_R} \kappa_i, \tag{63}$$

where $\kappa_i = \left|\sum_{j=1}^{n_T} h_{i,j,v}\widetilde{d}_j + \widetilde{n}_i\right|^2$. Furthermore, $h_{i,j,v}$ and $\widetilde{n}_i$ are independent complex Gaussian random variables such that

$$\left(\sum_{j=1}^{n_T} \frac{\psi_{i,j}\left|\widetilde{d}_j\right|^2}{2} + \frac{N_0}{2}\right)^{-1} \kappa_i \sim \chi_2^2. \tag{64}$$

APPENDIX D

Convexity and Concavity Analysis of $C_e(\lambda)$ and $\log_e B_e(\lambda)$:

Lemma D.1: $C_e(\lambda)$ is a convex function in the interval $(-\infty, 1)$.

Proof: If $a+b=1$ with $0 \le a, b \le 1$, for any $x, y < 1$, we have $C_e(ax+by) - (aC_e(x)+bC_e(y)) = \frac{-ab(x-y)^2}{[1-(ax+by)](1-x)(1-y)} < 0$. This completes the proof. ∎

Lemma D.2: The function $\log_e B_e(\lambda)$ is convex in the interval $(-\infty, -1]$, and is concave in the interval $[-1, 1)$.

Proof: Since $C_e'(\lambda) = \frac{1}{(1-\lambda)^2}$ and $C_e''(\lambda) = \frac{2}{(1-\lambda)^3}$, we obtain $\log_e B_e(\lambda) = \log_e(1+C_e(\lambda)) - C_e(\lambda)$, $\left[\log_e B_e(\lambda)\right]' = -\frac{\lambda}{(1-\lambda)^2}$, $\left[\log_e B_e(\lambda)\right]'' = -\frac{1+\lambda}{(1-\lambda)^3}$. The result can be validated because $\lambda < -1$ yields $\left[\log_e B_e(\lambda)\right]'' > 0$ that implies the convexity of $\log_e B_e(\lambda)$, whereas $\lambda > -1$ leads to $\left[\log_e B_e(\lambda)\right]'' < 0$ so that $\log_e B_e(\lambda)$ is concave.

APPENDIX E

Proof of Corollary 5:

Before the analysis is performed, for the sake of simplicity, we define

$$\overline{\gamma}(\boldsymbol{x}) \triangleq \frac{n_R}{\theta}\left[(t-s)C_e(x_0) + \sum_{j=1}^{s} C_e(x_j)\right],$$
$$f_1(\boldsymbol{x}) \triangleq \left[\frac{\overline{\gamma}(\boldsymbol{x})}{n_R} - (t-s)x_0 - \sum_{j=1}^{s} x_j\right], \tag{65}$$

where $\boldsymbol{x} = [x_0\ x_1\ \cdots\ x_s] \in \mathbb{S}^{s+1}$ and $\mathbb{S}^{s+1} \triangleq (-\infty, 1)^{s+1}$.

Although the Chernoff bound can be expressed in either $U_c(\theta, \boldsymbol{\lambda})$ or $U_c(\theta, \overline{\gamma}, \widetilde{\boldsymbol{\sigma}}, N_0)$, our analysis will make use of $U_c(\theta, \boldsymbol{\lambda})$ because $U_c(\theta, \boldsymbol{\lambda})$ is more mathematically tractable than $U_c(\theta, \overline{\gamma}, \widetilde{\boldsymbol{\sigma}}, N_0)$ that has too many parameters such as $\overline{\gamma}$, $\widetilde{\boldsymbol{\sigma}}$, and N_0. In order to find the minimum value of $U_c(\theta, \boldsymbol{\lambda})$ for a given θ, let us consider the optimization problem

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{S}^{s+1}}{\text{minimize}}\ \log_e U_c(\theta, \boldsymbol{x}) \\ &\text{subject to}\ f_1(\boldsymbol{x}) = 0. \end{aligned} \tag{66}$$

where $\mathbb{S}^{s+1}$ is the constraint set, the used objective function is $\log_e U_c(\theta, \boldsymbol{x})$ that is more convenient for mathematical calculation than $U_c(\theta, \boldsymbol{x})$. It is intuitively obvious that minimizing $U_c(\theta, \boldsymbol{x})$ is equivalent to minimizing $\log_e U_c(\theta, \boldsymbol{x})$ because the natural logarithm is a strictly increasing function.

The objective function $\log_e U_c(\theta, \boldsymbol{x})$ is not convex in the set $\mathbb{S}^{s+1}$, and the equality constraint function $f_1(\boldsymbol{x})$ is not *affine*, thus resulting in a nonconvex problem. To the best of our knowledge, such a problem may admit a local minimum, if it can fortunately be handled by means of Lagrangian duality, KKT condition, and bordered Hessian.

Introducing a dual variable $\lambda \in \mathbb{R}$, we can form the Lagrangian function

$$L(\boldsymbol{x}, \lambda) = \log_e U_c(\theta, \boldsymbol{x}) + \lambda f_1(\boldsymbol{x}), \tag{67}$$

where $\log_e U_c(\theta, \boldsymbol{x}) = \sum_{j=1}^{s} n_R \log_e B_e(x_j) + (t-s)\, n_R \log_e B_e(x_0)$.

The KKT condition for a solution of (66) to be the local optimal is written as

$$f_1(\boldsymbol{x}^*) = \frac{\overline{\gamma}(\boldsymbol{x}^*)}{n_R} - (t-s)x_0^* - \sum_{1\le j\le s} x_j^* = 0, \tag{68}$$
$$\lambda^* \ge 0, \tag{69}$$
$$\lambda^* f_1(\boldsymbol{x}^*) = 0, \tag{70}$$
$$\nabla \log_e U_c(\theta, \boldsymbol{x}^*) + \lambda^* \nabla f_1(\boldsymbol{x}^*) = 0. \tag{71}$$

Here the condition (70) is known as *complementary slackness*, which can be interpreted as $\lambda^* > 0 \Rightarrow f_1(\boldsymbol{x}^*) = 0$, or, $f_1(\boldsymbol{x}^*) < 0 \Rightarrow \lambda^* = 0$.

Consequently, the solution to the problem (66), that satisfies the KKT condition (68)-(71), can be obtained as

$$\begin{aligned} x_0^* = x_1^* = \cdots = x_s^* &= 1 - \frac{1}{\theta}, \\ \overline{\gamma}(\boldsymbol{x}^*) &= t n_R \left(1 - \frac{1}{\theta}\right), \\ \lambda^* &= \theta n_R, \end{aligned} \tag{72}$$

for both $\theta > 1$ and $0 < \theta < 1$, in which case

$$\log_e U_c(\theta, \boldsymbol{x}^*) = t n_R \left[\log_e B_e\left(1 - \frac{1}{\theta}\right)\right] = t n_R \log_e\left(\theta e^{1-\theta}\right),$$

or, equivalently, $U_c(\theta, \boldsymbol{x}^*) = \left[B_e\left(1-\frac{1}{\theta}\right)\right]^{t n_R} = \left(\theta e^{1-\theta}\right)^{t n_R}$.

Furthermore, the bordered Hessian satisfies that [38]

$$(-1)\begin{vmatrix} \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_0 \partial x_0} & \cdots & \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_0 \partial x_s} & \frac{\partial f_1(\boldsymbol{x}^*)}{\partial x_0} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_s \partial x_0} & \cdots & \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_s \partial x_s} & \frac{\partial f_1(\boldsymbol{x}^*)}{\partial x_s} \\ \frac{\partial f_1(\boldsymbol{x}^*)}{\partial x_0} & \cdots & \frac{\partial f_1(\boldsymbol{x}^*)}{\partial x_s} & 0 \end{vmatrix} > 0, \tag{73}$$

which can be verified by letting

$$\boldsymbol{A}_{11} = \begin{bmatrix} \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_0 \partial x_0} & \cdots & \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_0 \partial x_s} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_s \partial x_0} & \cdots & \frac{\partial^2 L(\boldsymbol{x}^*, \lambda^*)}{\partial x_s \partial x_s} \end{bmatrix}, \tag{74}$$

$\boldsymbol{A}_{21} = \boldsymbol{A}_{12}^T = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x}^*)}{\partial x_0} & \cdots & \frac{\partial f_1(\boldsymbol{x}^*)}{\partial x_s} \end{bmatrix}$, $\boldsymbol{A}_{22} = \begin{bmatrix} 0 \end{bmatrix}$, and using the following Lemma:

Lemma E.1: Suppose that $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{21} \\ \boldsymbol{A}_{12} & \boldsymbol{A}_{22} \end{bmatrix}$, where $\boldsymbol{A}_{11}$ and $\boldsymbol{A}_{22}$ are $m \times m$ and $n \times n$ square matrices, respectively. If $\boldsymbol{A}_{11}$ is invertible, then $|\boldsymbol{P}| = |\boldsymbol{A}_{11}| \cdot \left|\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right|$.

Proof: Since $\boldsymbol{A}_{11}$ is invertible, the below equality of block matrix multiplication holds:

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{21} \\ \boldsymbol{A}_{12} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{E}_m & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \\ \boldsymbol{O} & \boldsymbol{E}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{O}^T \\ \boldsymbol{A}_{12} & \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{bmatrix}, \tag{75}$$

where $\boldsymbol{E}_m$ and $\boldsymbol{E}_n$ are $m \times m$ and $n \times n$ identity matrices, respectively, and $\boldsymbol{O}$ denotes $n \times m$ matrix of zeros. ■

Thus far, we have validated that the solution $\boldsymbol{x}^*$ in (72) gives the strict local maximum, $\log_e U_c(\theta, \boldsymbol{x}^*)$, see (73) for detailed expression, because the KKT condition (68)-(71) is the necessary condition for the extremum of $\log_e U_c(\theta, \boldsymbol{x})$ and (73) is the sufficient condition for the strict local maximum.

Modifying the constraint set in (66) from $\mathbb{S}^{s+1}$ to $\mathbb{T}^{s+1} \triangleq (-\infty, -1]^{s+1}$, we have the below problem

$$\begin{aligned} &\underset{\boldsymbol{x} \in \mathbb{T}^{s+1}}{\text{minimize}} \ \log_e U_c(\theta, \boldsymbol{x}) \\ &\text{subject to } f_1(\boldsymbol{x}) = 0, \end{aligned} \tag{76}$$

where $\log_e U_c(\theta, \boldsymbol{x})$ is convex function in the convex set $\mathbb{T}^{s+1}$. The equality constraint function $f_1(\boldsymbol{x})$ is not affine. This means (76) is still a nonconvex problem. However, the relaxed form of (76) becomes a convex problem, which is expressed as

$$\begin{aligned} &\underset{\boldsymbol{x} \in \mathbb{T}^{s+1}}{\text{minimize}} \ \log_e U_c(\theta, \boldsymbol{x}) \\ &\text{subject to } f_1(\boldsymbol{x}) \le 0. \end{aligned} \tag{77}$$

The KKT condition for the global optimal solution of (77) is then written as [37]

$$f_1(\boldsymbol{x}^*) = \frac{\overline{\gamma}(\boldsymbol{x}^*)}{n_R} - (t-s)x_0^* - \sum_{1 \le j \le s} x_j^* \le 0, \tag{78}$$

$$\lambda^* \ge 0, \tag{79}$$

$$\lambda^* f_1(\boldsymbol{x}^*) = 0, \tag{80}$$

$$\nabla \log_e U_c(\theta, \boldsymbol{x}^*) + \lambda^* \nabla f_1(\boldsymbol{x}^*) = 0. \tag{81}$$

We find that, the solution $\boldsymbol{x}^*$ and λ^* in (72) can satisfy this KKT condition as well. In case of $0 < \theta \le 1/2$, it holds true that $\boldsymbol{x}^* \in \mathbb{T}^{s+1}$, such that $\boldsymbol{x}^*$ and λ^* give rise to the global minimum, $\log_e U_c(\theta, \boldsymbol{x}^*)$ as expressed in (73), for the convex problem (77) since $\nabla^2 \log_e U_c(\theta, \boldsymbol{x}^*) \ge 0$ in the constraint set $\mathbb{T}^{s+1}$.

It shall be emphasized that, $\{\boldsymbol{x} : f_1(\boldsymbol{x}) = 0\} \subset \{\boldsymbol{x} : f_1(\boldsymbol{x}) \le 0\}$ and the optimal solution $\boldsymbol{x}^*$ of the problem (77) simultaneously satisfies the equality constraint in (76), i.e., $f_1(\boldsymbol{x}^*) = 0$, such that $\boldsymbol{x}^*$ is able to minimize $\log_e U_c(\theta, \boldsymbol{x})$ in the problem (76) without violating the equality constraint, i.e., $\boldsymbol{x}^*$ is the optimal solution of (76) as well; this is the reason for paying our attention to the problem (77). In summary, if $0 < \theta \le 1/2$, we can obtain the global minimum of the problem (76) by directly using that of the relaxed problem (77), i.e., $\log_e U_c(\theta, \boldsymbol{x}^*)$ in (73).

ACKNOWLEDGMENT

The authors would like to thank all the anonymous reviewers and the editor for their constructive suggestions. They also like to thank Chizhong Qian for all his support while writing this paper.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 4th Quart., 2015.

[4] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[5] W. Liu, S. Han, and C. Yang, "Energy efficiency scaling law of massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 107–121, Jan. 2017.

[6] Z. D. Bai and J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, no. 1A, pp. 553–605, 2004.

[7] S. Loyka and G. Levin, "Finite-SNR diversity-multiplexing tradeoff via asymptotic analysis of large MIMO systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4781–4792, Oct. 2010.

[8] V. L. Girko, "The central limit theorem for random determinants," *Theory Probab. Appl.*, vol. 24, no. 4, pp. 729–740, 1981.

[9] V. Raghavan and A. M. Sayeed, "Weak convergence and rate of convergence of MIMO capacity random variable," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3799–3809, Aug. 2006.

[10] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Statist. Sinica*, vol. 9, no. 3, pp. 611–677, 1999.

[11] U. Grenander and J. W. Silverstein, "Spectral analysis of networks with random topologies," *SIAM J. Appl. Math.*, vol. 32, no. 2, pp. 499–519, Mar. 1977.

[12] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[13] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 345–364, Jan. 2011.

[14] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1794–1803, Jul. 2002.

[15] T. L. Narasimhan and A. Chockalingam, "Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 847–860, Oct. 2014.

[16] D. Bai, P. Mitran, S. S. Ghassemzadeh, R. R. Miller, and V. Tarokh, "Rate of channel hardening of antenna selection diversity schemes and its implication on scheduling," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4353–4365, Oct. 2009.

[17] H. Li, L. Song, and M. Debbah, "Energy efficiency of large-scale multiple antenna systems with transmit antenna selection," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 638–647, Feb. 2014.

[18] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, May 2017.

[19] Y. Qi and R. Qian, "Asymptotic analysis of random lattices in high dimensions," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–7.

[20] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[21] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[22] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[23] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[24] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[25] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: Asymptotic evaluation under correlated fading," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 829–838, Jun. 2003.

[26] S. Sun, T. S. Rappaport, R. W. Heath, Jr., A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.
[27] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[28] Y. Qi, R. Qian, and C. Hu, "Identifying the full-diversity solution of MIMO detection: A fine-grained approach," in *Proc. IEEE WCNC*, San Francisco, CA, USA, Mar. 2017, pp. 1–6.
[29] A. E. Geyer, R. Nikjah, S. A. Vorobyov, and N. C. Beaulieu, "Euclidean and space–time block codes: Relationship, optimality, performance analysis revisited," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2912–2923, Aug. 2015.
[30] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[31] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[32] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*. Delft, The Netherlands: Now Publishers, 2004.
[33] R. Couillet and M. Debbah, *Random Matrix Theory Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press. 2011.
[34] V. L. Girko, *Theory of Random Determinants*. Norwell, MA, USA: Kluwer, 1990.
[35] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 1995.
[36] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[38] H. Neudecker and J. R. Magnus, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. New York, NY, USA: Wiley, 1988.

Yuan Qi received the B.S. degree in communication engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecomm (BUPT), Beijing, China, in 2004 and 2009, respectively. Since 2009, she has been a Lecturer with the School of Electronic Engineering, BUPT.

Her research interests include signal processing in wireless communications and industrial electronic.

Rongrong Qian received the B.S. degree in communication engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecomm (BUPT), Beijing, China, in 2004 and 2010, respectively. Since 2010, he has been a Lecturer and an Associate Professor with the Automation School, BUPT. From 2015 to 2016, he was a Visiting Scholar with the College of Engineering, Peking University, Beijing.

His research interests include signal processing in wireless communications and the application of control theory to communication problems.