Novel Matrix Singular Value Inequalities and Their Applications to Uncertain MIMO Channels

Sergey Loyka and Charalambos D. Charalambous

***Abstract*— Novel matrix singular value inequalities are established for a sum/product of three matrices. Their application to the uncertain (compound) multiple-input multiple-output (MIMO) channel subject to normed additive uncertainty establishes the saddle-point property for a wide range of performance metrics monotonic in the channel singular values, including, among others, the mutual information, MMSE, error exponent, and pairwise error probability. This, in turn, implies that the transmission on the eigenmodes of the nominal (or worst case) channel is also optimal for the whole set of channels under a general power constraint and hence achieves the compound channel capacity. The worst case channel turns out to be antiparallel of the nominal one for all these performance metrics. An application of these results to beamforming over compound MIMO channels is discussed. An optimal robust precoder for the uncertain MIMO channel is obtained in a closed-form under the sum-MSE criterion and the total power constraint. The saddle-point property is shown to hold and the optimal strategy is to diagonalize the nominal (or worst case) channel.**

***Index Terms*— Compound channel, channel capacity, multiple-input multiple-output (MIMO), MMSE, optimal precoding, channel uncertainty, singular value inequality.**

I. Introduction

Channel uncertainty plays a significant role in communications and may be caused by several reasons, such as channel dynamics (as in wireless communications), imperfect channel estimation and limitations of feedback link to deliver the channel state information (CSI) to the transmitter (Tx) [1]–[5]. There are several approaches to model channel uncertainty, which can be broadly characterized as stochastic, where the channel is not known but its distribution is known, and deterministic, where a fixed channel realization is known to belong to a certain bounded uncertainty set but is not known otherwise [4], [5]. The latter forms a compound channel model, where one has to develop a coding/transmission technique that would perform well on all members of the channel uncertainty set [1]–[5]. Unless indicated otherwise, we assume below deterministic (compound) channel uncertainty model. From the communication/information-theoretic perspective, the main performance metric is the compound channel capacity, i.e. the maximum achievable rate under arbitrary low error probability for any channel in the uncertainty set using a single codebook, or just an achievable rate or mutual information (MI) under certain coding/transmission strategy [4]–[18]. In many cases, the compound channel capacity can be expressed as the max-min MI, where max is over all possible input distributions and min is over all possible channel realizations in the uncertainty set [4], [5]. Typically, uncertainty is modeled as additive, where the true channel is the nominal (estimated) one plus additive uncertainty bounded in a certain way. Several approaches are used to bound uncertainty, including various norms (Frobenius, spectral, weighted trace etc.) [7], [10]–[12], [23]–[32]. The detailed rationale and practical reasons for using the spectral norm to bound the uncertainty set can be found in [11].

A. Information-Theoretic Perspective

Since the pioneering work in [1]–[3], a large number of studies have appeared that address the channel uncertainty problem from a communication/information-theoretic perspective [4]–[18]; see [4], [5] for an extensive literature review up to 1998. The emphasis has recently shifted towards multiple-input multiple-output (MIMO) systems due to their high spectral efficiency [6]–[14], [16]–[18]. A concise review of recent results on compound MIMO channel capacity can be found in [7] and [11]. Specifically, the optimality of isotropic signaling has been established in [6] under MIMO channel uncertainty when the uncertainty set is isotropic. A code construction achieving the compound channel capacity was proposed in [8]. The optimality of beamforming in terms of ergodic capacity of MIMO channels with limited/imperfect feedback was addressed in [9]. The capacity of compound Gaussian MIMO channel with additive uncertainty under weighted trace constraint has been established in [7] for rank-1 nominal channels, and under the spectral norm constraint in [10] provided that the uncertainty is limited to the singular values only but not to singular vectors, which is unlikely from the physical perspective. This has been extended to uncertainty in singular vectors and no constraint on the rank of the nominal channel in [11], which also includes the case of multiplicative uncertainty. Specifically, a saddle-point property in the behavior of mutual information has been established (min max = max min, where max is over the input distribution and min is over the channel uncertainty), which implies that the compound capacity is equal to the worst-case channel capacity (achievable by a single code for any channel in

Manuscript received May 28, 2014; revised March 17, 2015; accepted July 23, 2015. Date of publication October 2, 2015; date of current version November 18, 2015.

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: sergey.loyka@ieee.org).

C. D. Charalambous is with the ECE Department, University of Cyprus, Nicosia 1678, Cyprus (e-mail: chadcha@ucy.ac.cy).

Communicated by A. Guillen i Fabregas, Associate Editor for Communications.

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/TIT.2015.2485986

the uncertainty set), for the multiplicative uncertainty in full generality [11]. For additive uncertainty, this property has been established for a large number of special cases and, based on extensive evidence, it was conjectured that it holds in full generality [11]. The general additive uncertainty case remained elusive until recently, when a new matrix determinant inequality was established in [12], from which saddle-point and hence the compound capacity follow for the general case. Since all norms are equivalent (in a sense that upper and lower bounds can be constructed with constants which depend only on dimensionality and nothing else) [34], the spectral norm results can be used to construct upper and lower bounds for the compound capacity under any other normed uncertainties [11]. The non-ergodic and ergodic scenarios where neither channel realization nor its statistics is known to the transmitter were considered in [13] and [14] respectively; rather, the Tx knows that the (unknown) true channel distribution is within certain distance (measured by the relative entropy) of the nominal (estimated) channel distribution, which models the imperfections of channel distribution estimation. The robustness of the dirty paper coding under multiplicative uncertainty of interference in a SISO channel has been investigated in [15]. A method of space-time multicasting over a broadcast multi-user MIMO channel was developed in [17], which is based on the joint unitary triangularization [16]. This method can be applied to a finite-state compound channel as well. Unfortunately, its extension to infinite-state channels (e.g. when the uncertainty set is continuous rather than discrete and finite) is not known. Finally, a robust distributed compression algorithm was developed in [18] for cloud radio access under channel uncertainty subject to the spectral norm constraint.

Establishing the compound capacity of a Gaussian MIMO channel involves maximization of the MI over all possible input distributions and minimization over all channel states in the uncertainty set. It is straightforward to show that the optimal input distribution is Gaussian, so it remains to find its covariance matrix, which is a convex problem by itself (since log-det is a concave function). However, the minimization of the MI over channel uncertainty is not a convex problem, so that neither KKT conditions are sufficient for optimality nor von Neumann minimax theorem holds. Thus, a new approach is needed to establish the compound channel capacity in this setting and to show an existence of a saddle-point. Our approach is based on novel matrix singular value inequalities for a product/sum of three matrices established below. Based on these inequalities, not only the conjecture in [11] follows, but also the saddle-point property is established for a large class of MIMO channel performance metrics which are monotonic in the channel singular values (of which the mutual information and sum-MSE are special cases), e.g. error rate performance of space-time codes, error exponent, etc., and also for a general power constraint which includes, as special cases, the total and maximum power constraints. For all these performance metrics and under the general power constraint, the optimal strategy is to diagonalize the nominal (or worst-case) channel and an optimal power allocation depends on a particular performance metric and power constraint.

We note that, on one hand, the new singular value inequalities are stronger than the corresponding determinant inequality in [12] (the latter follows from the former but the converse in not true) under the assumed ordering of singular values. On the other hand, the determinantal inequality in [12] is more general as it does not require any particular ordering, so that these results are complementary to each other. The new singular value inequalities allow one to establish the saddle-point property and also to identify the worst-case channel for a broad class of performance metrics under the general power constraint, which turns out to be "anti-parallel" of the nominal channel, exactly as in [11, Th. 4]. Following the compound channel capacity approach, we also discuss an application of the results to beamforming over compound channels. In particular, we identify the worst-case channel under any beamforming vectors (not only optimal ones).

B. Signal Processing Perspective

From the signal processing perspective, the main metric is the received signal-to-noise ratio (SNR) or mean square error (MSE) [19]–[32]. While the MI and SNR (or MSE) metrics give similar results at low SNR (since the mutual information is proportional to the SNR in this regime, so that maximizing the MI is the same as maximizing the SNR or minimizing the MSE) or in rank-1 channels, they result in different optimal transmission strategies otherwise. A concise review of recent results on robust transmission/processing under channel uncertainty from the signal processing perspective can be found in [32]. In particular, a robust receive (Rx) beamformer was obtained in [19] and [22] that maximizes the worst-case Rx SNR for an uncertain (deterministic) single-input multiple-output (SIMO) channel, which is equivalent to the classical Capon beamformer with diagonal loading [66]. This work was extended to an uncertain MIMO channel subject to the Frobenius (trace) and spectral norm constraints in [24] and [29]. A robust minimax method for signal estimation in a MIMO channel was developed in [20] and [21] under the additive channel uncertainty subject to the spectral norm constraint. A robust precoder design to minimize the zero-forcing equalization MSE was developed in [23] under the Frobenius norm constraint on the additive channel uncertainty. A robust precoder/decoder design to minimize the sum MSE under a statistical model of channel uncertainty (based on a model of channel estimation mechanism) was obtained in [25] and was later extended to a multi-user setting via numerical optimization algorithms in [30] for which, however, global convergence cannot be insured. A robust beamforming (i.e. rank-1 transmission) to maximize the SNR for cognitive radio under deterministic additive MISO channel uncertainty subject to a weighted trace constraint was developed in [26]. A numerical algorithm to minimize the per-user MSEs in a downlink multiuser MIMO system subject to additive channel uncertainty (bounded via the Frobenius norm) was proposed in [27]. A Tx precoder to maximize the worst-case received SNR in a MIMO channel with additive uncertainty bounded via the weighted Frobenius (trace) norm was proposed in [28]. This approach was further extended to several classes of uncertainty sets in [32]. Note however that, since the objective (received SNR) is linear in Tx covariance and quadratic in channel uncertainty, i.e. convex-concave in the right way, von Neumann minimax theorem [58] can be evoked to insure the existence of a saddle-point and Karush-Kuhn-Tucker (KKT) conditions are sufficient for optimality [59].

Neither of these holds for the problem considered in this paper: since the objective (MI or MSE) is not convex in the channel uncertainty, neither KKT conditions are sufficient for optimality nor von Neumann minimax theorem can be used. Therefore, we have to develop a different approach to the problem at hand. This approach is based on novel singular-value inequalities established here and on the tools from majorization theory, which allow us to establish a saddle-point without using the minimax theorem. While the robust designs above [27], [28], [32] optimize the total received SNR over an (deterministic) uncertain MIMO channel, they are optimal in terms of the sum MSE in the low SNR regime only and the general case is still an open problem. While numerical algorithms were proposed to tackle this problem [27], [31], global convergence cannot be insured due to non-convex nature of the problem.

Our approach provides a closed-form analytical solution to this problem under the spectral norm constraint on the channel uncertainty and the total power constraint. It gives both the MI-maximizing and the sum-MSE minimizing designs in a unified way. The saddle-point property is shown to hold and the optimal strategy is to diagonalize the nominal (or worst-case) channel for both metrics.

C. Singular Value Inequalities

There is an extensive literature on eigenvalue/singular value (SV) inequalities of various forms, see e.g. [33]–[57] and references therein, and [33]–[38], [57] for an extensive collection of results and literature review. In particular, a number of inequalities were obtained for the sums and products of two matrices [39]–[47], including Hadamard and Kronecker products and direct sums [44], [45], [48], [54], of their powers and various combinations [42], [47], [49]–[52], [56], including block-partitioned matrices [45], [47], [48], [54], and the conditions for attaining equalities [43]. Majorization theory is one of the main tools in obtaining such inequalities and many inequalities are of the majorization type [38], [39], [42], [52]. Upper and lower bounds for the singular values of the sum $\mathbf{A}+\mathbf{B}$ and product $\mathbf{AB}$ of two matrices $\mathbf{A},\mathbf{B}$ are well-known [33]–[36]:

$$(\sigma_i(\mathbf{A})-\sigma_1(\mathbf{B}))_+ \leq \sigma_i(\mathbf{A}+\mathbf{B}) \leq \sigma_i(\mathbf{A})+\sigma_1(\mathbf{B}) \quad (1)$$

$$\sigma_i(\mathbf{A})\sigma_{\min}(\mathbf{B}) \leq \sigma_i(\mathbf{AB}) \leq \sigma_i(\mathbf{A})\sigma_1(\mathbf{B}) \quad (2)$$

where $\sigma_i(\mathbf{A})$ is i-th largest singular value of $\mathbf{A}$, $\sigma_{\min}$ is the minimum singular value, and $(x)_+ = x$ if $x > 0$ and 0 otherwise. It was shown in [40] that

$$2\sigma_i(\mathbf{A}^+\mathbf{B}) \leq \sigma_i(\mathbf{AA}^+ + \mathbf{BB}^+) \quad (3)$$

where $\mathbf{A}^+$ is Hermitian conjugation of $\mathbf{A}$, which is an extension of arithmetic-geometric mean inequality to matrices. This result was further extended in many directions [44], [47], [49], [50], [53], [56]. It was shown in [48] that

$$\sigma_i(\mathbf{A}+\mathbf{B}) \leq 2\sigma_i(\mathbf{A}\oplus\mathbf{B}),$$

where $\oplus$ denotes direct sum, and that

$$\sigma_i(\mathbf{A}-\mathbf{B}) \leq \sigma_i(\mathbf{A}\oplus\mathbf{B})$$
$$2\sqrt{\sigma_i(\mathbf{AB})} \leq \sigma_i(\mathbf{A}+\mathbf{B})$$

if $\mathbf{A},\mathbf{B}$ are positive semi-definite [45], [53]: $\mathbf{A},\mathbf{B} \geq 0$.

Considerably less is known about singular values of products/sums of three matrices. In particular, [46] considers the product of the form $\mathbf{AZB}$ and relates its singular values to those of the product $\mathbf{ZAB}$ where $\mathbf{A},\mathbf{B}$ are positive semi-definite, $\mathbf{A},\mathbf{B} \geq 0$, and $\mathbf{Z}$ is positive definite, $\mathbf{Z} > 0$. Using (2), it can be shown that

$$\sigma_i(\mathbf{AZB}) \leq \sigma_1(\mathbf{A})\sigma_1(\mathbf{B})\sigma_i(\mathbf{Z})$$

and similarly for lower bound and also for the sum of three matrices; see [42] for more inequalities of this type. (3) can be extended to a product of three matrices using (2), see e.g. [55]:

$$2\sigma_i(\mathbf{A}^+\mathbf{CB}) \leq \sigma_1(\mathbf{C})\sigma_i(\mathbf{AA}^+ + \mathbf{BB}^+) \quad (4)$$

where $\mathbf{C} \geq 0$.

Motivated by applications to transmission over uncertain MIMO channels, we are interested in the singular values of the sum/product of the form $(\mathbf{A}+\mathbf{B})\mathbf{C}$, where $\mathbf{A},\mathbf{B}$ represent channel matrix and its uncertainty (and hence not necessarily Hermitian or square) and $\mathbf{C}$ represents the square root of the transmit covariance matrix (and hence is positive semi-definite). No sharp bounds are known for the singular values of this form. The known bounds for singular values of the sum and product in (1) (2) can be combined to obtain the bounds for the form $(\mathbf{A}+\mathbf{B})\mathbf{C}$:

$$\begin{aligned}\sigma_i((\mathbf{A}+\mathbf{B})\mathbf{C}) &\geq \sigma_i(\mathbf{A}+\mathbf{B})\sigma_{\min}(\mathbf{C}) \\ &\geq (\sigma_i(\mathbf{A})-\sigma_1(\mathbf{B}))_+\sigma_{\min}(\mathbf{C}) \quad (5)\\ \sigma_i((\mathbf{A}+\mathbf{B})\mathbf{C}) &\leq \sigma_i(\mathbf{A}+\mathbf{B})\sigma_1(\mathbf{C}) \\ &\leq (\sigma_i(\mathbf{A})+\sigma_1(\mathbf{B}))\sigma_1(\mathbf{C}) \quad (6)\end{aligned}$$

but those bounds turn out to be not tight enough to establish the result we are targeting. Sharper bounds are need for this latter form, which are established in the present paper using a novel approach. This approach relies on known SV inequalities for products and sums of two matrices, including some majorization-type inequalities, but also includes a non-convex optimization problem as a key step. Even though the problem is non-convex, we obtain its globally-optimum solution in 2 steps: (i) first, we demonstrate that the KKT conditions are necessary for the optimality; then, (ii) we inspect all possible solutions to the KKT conditions, of which there are a finite number, and select the best one. To the best of our knowledge, this approach has never been used before to obtain singular value inequalities.

II. Compound MIMO Channel and Capacity

We consider the standard baseband discrete-time MIMO channel model,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \tag{7}$$

where $\mathbf{x} = [x_1, x_2, \ldots x_m]^T \in \mathbb{C}^{m,1}$ and $\mathbf{y} = [y_1, y_2, \ldots y_n]^T \in \mathbb{C}^{n,1}$ are the vectors representing the Tx and Rx symbols respectively, $\mathbf{x}^T$ denotes transposition, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots \mathbf{h}_m] \in \mathbb{C}^{n,m}$ is the $n \times m$ matrix of the complex channel gains between each Tx and each Rx antenna, where $\mathbf{h}_i$ denotes i-th column of $\mathbf{H}$, n and m are the numbers of Rx and Tx antennas respectively, $\boldsymbol{\xi}$ is the vector of circularly-symmetric additive white Gaussian noise (AWGN), which is independent and identically distributed (i.i.d.) in each receiver. Without loss of generality, we further assume $n \geq m$. The channel is assumed to be deterministic, fixed and frequency-flat, with partial channel state information (CSI) at the Rx and Tx ends, as described below.

The case of multiplicative uncertainty $\mathbf{H} = (\mathbf{I} + \mathbf{E})\mathbf{H}_0$, where $\mathbf{H}_0$ is the nominal channel (without uncertainty) known at the Tx end and, possibly, at the Rx end, and $\mathbf{E} \in \mathbb{C}^{n,n}$ is the multiplicative uncertainty, has been solved in [11] in full generality. Therefore, in this paper we focus our attention on the additive uncertainty, when the nominal channel $\mathbf{H}_0$ experiences an additive perturbation $\Delta\mathbf{H}$,

$$\mathbf{H} = \mathbf{H}_0 + \Delta\mathbf{H} \tag{8}$$

and where $\Delta\mathbf{H}$, and hence $\mathbf{H}$, belong to a bounded uncertainty set S_H,

$$\mathbf{H} \in S_H = \{\mathbf{H} : \|\mathbf{H} - \mathbf{H}_0\|_2 = \sigma_1(\Delta\mathbf{H}) \leq \varepsilon\}, \tag{9}$$

where $\sigma_1(\Delta\mathbf{H}) = \|\Delta\mathbf{H}\|_2 = \max_{|x|=1} |\Delta\mathbf{H}\mathbf{x}|$ is the largest singular value (or spectral norm) of $\Delta\mathbf{H}$, $|\mathbf{x}|^2 = \mathbf{x}^+\mathbf{x} = \sum_i |x_i|^2$ is the vector length squared. A number of advantages provided by the spectral norm as a measure of uncertainty have been pointed out in [11].

For a fixed channel $\mathbf{H}$ and given the covariance of the normalized Tx signal $\mathbf{R} = \overline{\mathbf{x}\mathbf{x}^+}$, where $\overline{\mathbf{x}}$ denotes expectation of $\mathbf{x}$, the mutual information between $\mathbf{x}$ and $\mathbf{y}$ when $\mathbf{x}$ is Gaussian (i.e., capacity-achieving) is given by the celebrated Foschini-Telatar formula,

$$\begin{aligned} I &= \log\left|\mathbf{I} + \gamma\,\mathbf{H}\mathbf{R}\mathbf{H}^+\right| \\ &= \sum_{i=1}^{m} \log\left(1 + \gamma\,\sigma_i^2(\mathbf{H}\mathbf{R}^{1/2})\right) = I(\mathbf{H}, \mathbf{R}) \end{aligned} \tag{10}$$

where γ is the SNR per antenna, $\sigma_i(\mathbf{H}\mathbf{R}^{1/2})$ are singular values of $\mathbf{H}\mathbf{R}^{1/2}$; unless indicated otherwise, we assume $\sigma_1 \geq \sigma_2 \geq \ldots$ If no CSI is available at the Tx end, the popular choice is isotropic signaling, e.g. $\mathbf{R} = \mathbf{I}$. When full CSI is available at the Tx end, the capacity is

$$C(\mathbf{H}) = \max_{\mathbf{R} \in S_R} I(\mathbf{H}, \mathbf{R}) \tag{11}$$

where the maximization is over all possible Tx covariance matrices, $\mathbf{R} \in S_R$ is due to power constraint and S_R is the constraint set of positive-semi definite matrices. In this paper, we consider a general unitary-invariant constraint set, i.e.

$$\mathbf{R} \in S_R \rightarrow \mathbf{U}\mathbf{R}\mathbf{U}^+ \in S_R \tag{12}$$

for any unitary $\mathbf{U}$. Note that this constraint is more general than that in [12] since we do not require $diag\{r_{ii}\} \in S_R$, where r_{ii} is i-th diagonal entry of $\mathbf{R}$, and since S_R is not required to be convex.[1] The constraint in (12) limits the eigenvalues of $\mathbf{R}$ but not its eigenvectors. Particular cases include the popular total power constraint:

$$tr(\mathbf{R}) = \sum_{i=1}^{m} \lambda_i(\mathbf{R}) \leq m \tag{13}$$

where $\lambda_i(\mathbf{R})$ is i-th largest eigenvalue of $\mathbf{R}$, and the maximum per-eigenmode power constraint,

$$\lambda_1(\mathbf{R}) \leq 1 \tag{14}$$

Under the total power constraint (13), The maximum in (11) has the well-known water-filling solution:

$$C(\mathbf{H}) = \sum_{i=1}^{m} \log\left(1 + \gamma\,\sigma_i^2(\mathbf{H})\lambda_i(\mathbf{R}^*)\right) \tag{15}$$

where

$$\lambda_i(\mathbf{R}^*) = \left[\mu - \frac{1}{\gamma\,\sigma_i^2(\mathbf{H})}\right]_+ \tag{16}$$

is i-th eigenvalue of the best (maximizing) covariance matrix $\mathbf{R}^* = \mathbf{V}\Lambda^*\mathbf{V}^+$, $\mathbf{V}$ is a unitary matrix of right singular vectors of $\mathbf{H}$ found from its singular value decomposition (SVD) $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^+$, where $\mathbf{U}$ is a unitary matrix of its left singular vectors and $\boldsymbol{\Sigma}$ is a diagonal matrix of its singular values, $\Lambda^* = diag\{\lambda_i(\mathbf{R}^*)\}$, the constant μ is found from the total power constraint, $\sum_i \lambda_i(\mathbf{R}^*) = m$. We use the compact notation $\mathbf{R}^* = \mathsf{WF}(\mathbf{H})$ to denote the water-filling over the channel $\mathbf{H}$ eigenmodes in (16).

Following the framework developed in [1]–[3], the compound channel capacity C_c is the largest reliable transmission rate achievable by a single code for any channel in the uncertainty set, $\mathbf{H} \in S_H$,

$$C_c = \max_{\mathbf{R} \in S_R} \min_{\mathbf{H} \in S_H} I(\mathbf{H}, \mathbf{R}) \tag{17}$$

while

$$C_w = \min_{\mathbf{H} \in S_H} \max_{\mathbf{R} \in S_R} I(\mathbf{H}, \mathbf{R}) \tag{18}$$

is the capacity of the worst-case channel, and in general $C_w \geq C_c$. The saddle-point property $C_w = C_c$ has been established in [11] under the total power constraint for the case of multiplicative uncertainty in full generality and in the case of additive uncertainty for a wide range of special cases, which was further extended to the general case in [12] using a new deterninantal inequality. In Section IV, we give an alternative

[1]As an example, let $S_R = \{\mathbf{R} : \mathbf{R} = \mathbf{U}\mathbf{E}\mathbf{U}^+\}$, where $\mathbf{U}$ is any $m \times m$ unitary matrix and $\mathbf{E}$ is $m \times m$ all-one matrix, so that all matrices in this set are rank-1 and have eigenvalues $\{m, 0, \ldots, 0\}$. This set is not allowed in [12] since (i) it is not convex, and (ii) it does not include $\mathbf{R} = \mathbf{I}$ (as required by [12, eq. (9)]).

proof under the general power constraint in (12) via the new singular value inequalities established below. This approach is also used to establish an optimal robust signaling under a wide class of performance metrics such as the sum-MSE, the total received SNR, the error exponent, the error rate of space-time codes, and solves the open problem of optimal precoder design under the sum-MSE criterion over a uncertain MIMO channel subject to the spectral norm constraint.

III. New Matrix Singular Value Inequalities

The following Lemma is instrumental in proving the main result.

Lemma 1: Let $a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$, $c_1, c_2, \ldots c_n \geq 0$ *and* $a_1c_1 \geq a_2c_2 \geq \ldots \geq a_nc_n \geq 0$. *Consider the following function,*

$$F(\mathbf{z}) = \left\{ \sqrt{\sum_{i=1}^{n} |a_i c_i z_i|^2} - \varepsilon \sqrt{\sum_{i=1}^{n} |c_i z_i|^2} \right\}_+ \tag{19}$$

where $\varepsilon \geq 0$ *and* z_i *are arbitrary complex numbers,* $\mathbf{z} = [z_1, z_2 \ldots z_n]^T$. *Then,*

$$\min_{|\mathbf{z}|=1} F(\mathbf{z}) = \{a_n - \varepsilon\}_+ c_n \tag{20}$$

Proof: See Appendix. □

The ordering conditions of Lemma 1 are not superfluous: removing anyone of them makes (20) invalid in the general case, as demonstrated by Example 1 below. We are now in a position to establish the new matrix singular value inequalities.

Proposition 1: Let $\mathbf{A}$, $\mathbf{B}$ *be* $n \times m$ *matrices and* $\mathbf{C}$ *be* $m \times p$ *matrix, and let the right singular vectors of* $\mathbf{A}$ *be the same as the left singular vectors of* $\mathbf{C}$ *so that their singular value decompositions (SVD) are* $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}_a\mathbf{V}^+$ *and* $\mathbf{C} = \mathbf{V}\mathbf{\Sigma}_c W^+$, *where* $\mathbf{U}, \mathbf{V}, \mathbf{W}$ *are unitary and* $\mathbf{\Sigma}_a = diag\{\sigma_{ai}\}, \mathbf{\Sigma}_c = diag\{\sigma_{ci}\}$ *are "diagonal" matrices of singular values of* $\mathbf{A}$ *and* $\mathbf{C}$. *Without loss of generality, assume that* $\{\sigma_{ai}\}$ *are in decreasing order,* $\sigma_{a1} \geq \sigma_{a2} \geq \ldots$ *Then, the following inequalities hold*

$$\begin{aligned}(\sigma_i(\mathbf{A}) - \sigma_1(\mathbf{B}))_+\sigma_i(\mathbf{C}) &\overset{(a)}{\leq} \sigma_i((\mathbf{A}+\mathbf{B})\mathbf{C}) \\ &\overset{(b)}{\leq} (\sigma_i(\mathbf{A}) + \sigma_1(\mathbf{B}))\sigma_i(\mathbf{C}),\end{aligned} \tag{21}$$

where $\sigma_i((\mathbf{A}+\mathbf{B})\mathbf{C})$ *are also in decreasing order,* $i = 1 \ldots \min(n, m, p)$, $\sigma_1(\mathbf{B}) = \varepsilon$ *is the largest singular value of* $\mathbf{B}$*; the inequalities in (a) and (b) hold under the conditions in* (22) *and* (23) *respectively:*

$$\{\sigma_{ai}\sigma_{ci}\} \text{ are in decreasing order: } \sigma_{a1}\sigma_{c1} \geq \sigma_{a2}\sigma_{c2} \geq \ldots, \tag{22}$$

$$\{\sigma_{ci}\} \text{ are in decreasing order: } \sigma_{c1} \geq \sigma_{c2} \geq \ldots \tag{23}$$

which cannot be relaxed in general. The bounds are tight for any given $\mathbf{A}$ *and* $\mathbf{C}$*: (a) is achievable with equality by* $\mathbf{B} = -\mathbf{U}\mathbf{\Sigma}_b\mathbf{V}^+$, *where* $\mathbf{\Sigma}_b = diag\{\min(\sigma_i(\mathbf{A}), \varepsilon)\}$, *and (b) is achievable with equality by* $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}_\varepsilon\mathbf{V}^+$, *where* $\mathbf{\Sigma}_\varepsilon = diag\{\varepsilon\}$ *is a diagonal matrix with* ε *as its diagonal entries.*

Proof: To demonstrate the key ideas of the proof, let us begin with (a) in (21) and demonstrate this inequality first for $i = 1$:

$$\begin{aligned}\sigma_1((\mathbf{A}+\mathbf{B})\mathbf{C}) &= \max_{|\mathbf{x}|=1} |(\mathbf{A}+\mathbf{B})\mathbf{C}\mathbf{x}| && (24)\\ &\geq |(\mathbf{A}+\mathbf{B})\mathbf{C}\mathbf{w}_1| && (25)\\ &\geq (|\mathbf{A}\mathbf{C}\mathbf{w}_1| - |\mathbf{B}\mathbf{C}\mathbf{w}_1|)_+ && (26)\\ &\geq (|\mathbf{A}\mathbf{C}\mathbf{w}_1| - \varepsilon\,|\mathbf{C}\mathbf{w}_1|)_+ && (27)\\ &= (\sigma_1(\mathbf{A}) - \varepsilon)_+\sigma_1(\mathbf{C}) && (28)\end{aligned}$$

where $\mathbf{w}_1$ is the right singular vector of $\mathbf{C}$ corresponding to $\sigma_1(\mathbf{C})$; (24) follows from variational characterization of singular values [29], [34], (26) follows from the triangle inequality $|\mathbf{x}+\mathbf{y}| \geq |\mathbf{x}| - |\mathbf{y}|$, where the equality is achievable when $\mathbf{x} = -\alpha\mathbf{y}$, $\alpha \geq 1$, (27) follows from the fact that $|\mathbf{Bz}| \leq \varepsilon\,|\mathbf{z}|$. Let us now demonstrate (a) in (21) for the general case using again the variational characterization of singular values:

$$\begin{aligned}\sigma_k((\mathbf{A}+\mathbf{B})\mathbf{C}) &= \max_{\mathbf{y}^k} \min_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{y}^k\}}} |(\mathbf{A}+\mathbf{B})\mathbf{C}\mathbf{x}| && (29)\\ &\geq \min_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{w}_1\ldots\mathbf{w}_k\}}} |(\mathbf{A}+\mathbf{B})\mathbf{C}\mathbf{x}| && (30)\\ &\geq \min_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{w}_1\ldots\mathbf{w}_k\}}} (|\mathbf{A}\mathbf{C}\mathbf{x}| - |\mathbf{B}\mathbf{C}\mathbf{x}|)_+ && (31)\\ &\geq \min_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{w}_1\ldots\mathbf{w}_k\}}} (|\mathbf{A}\mathbf{C}\mathbf{x}| - \varepsilon\,|\mathbf{C}\mathbf{x}|)_+ && (32)\\ &= \min_{\substack{|\mathbf{z}|=1\\ \mathbf{z}\in span\{\mathbf{e}_1\ldots\mathbf{e}_k\}}} (|\mathbf{\Sigma}_a\mathbf{\Sigma}_c\mathbf{z}| - \varepsilon\,|\mathbf{\Sigma}_c\mathbf{z}|)_+ && (33)\\ &= \min_{|\mathbf{z}|=1} \left\{ \sqrt{\sum_{i=1}^{k} |\sigma_{ai}\sigma_{ci}z_i|^2} - \varepsilon\sqrt{\sum_{i=1}^{k} |\sigma_{ci}z_i|^2} \right\}_+ && (34)\end{aligned}$$

where $\mathbf{y}^k = \{\mathbf{y}_1 \ldots \mathbf{y}_k\}$ is a set of k linearly independent vectors and external maximization is over all such sets, (30) follows from restricting the external maximization to $\{\mathbf{w}_1 \ldots \mathbf{w}_k\}$, the right singular vectors of $\mathbf{C}$, (31) and (32) are as in (26) and (27), (33) and (34) follow from $\mathbf{z} = \mathbf{W}^+\mathbf{x}$, where $\mathbf{e}_i = [0 \ldots 0, 1, 0 \ldots 0]^T$ is a unit vector of all zero entries except entry i. Now applying Lemma 1 to (34), one finally obtains (a) in (21). It is straightforward to see that $\mathbf{B} = -\mathbf{U}\mathbf{\Sigma}_b\mathbf{V}^+$, where $\mathbf{\Sigma}_b = diag\{\min(\sigma_i(\mathbf{A}), \varepsilon)\}$, achieves the lower bound for all i simultaneously and any given $\mathbf{A}$ and $\mathbf{C}$. It can be easily demonstrated by examples that the ordering condition on $\{\sigma_{ai}\sigma_{ci}\}$ cannot be relaxed in general, i.e. this ordering is both sufficient and necessary for the inequality to hold. To demonstrate (b) in (21), use the variational characterization as well [29]:

$$\begin{aligned}\sigma_k((\mathbf{A}+\mathbf{B})\mathbf{C}) &= \min_{\mathbf{y}^{p-k+1}} \max_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{y}^{p-k+1}\}}} |(\mathbf{A}+\mathbf{B})\mathbf{C}\mathbf{x}| && (35)\\ &\leq \max_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{w}_k\ldots\mathbf{w}_p\}}} |(\mathbf{A}+\mathbf{B})\mathbf{C}\mathbf{x}| && (36)\\ &\leq \max_{\substack{|\mathbf{x}|=1\\ \mathbf{x}\in span\{\mathbf{w}_k\ldots\mathbf{w}_p\}}} (|\mathbf{A}\mathbf{C}\mathbf{x}| + \varepsilon\,|\mathbf{C}\mathbf{x}|) && (37)\end{aligned}$$

$$= \max_{\substack{|\mathbf{z}|=1 \\ \mathbf{z}\in span\{\mathbf{e}_k\ldots\mathbf{e}_p\}}} (|\mathbf{\Sigma}_a\mathbf{\Sigma}_c\mathbf{z}| + \varepsilon\,|\mathbf{\Sigma}_c\mathbf{z}|) \tag{38}$$

$$= \max_{|\mathbf{z}|=1}\left\{\sqrt{\sum_{i=k}^{p}|\sigma_{ai}\sigma_{ci}z_i|^2} + \varepsilon\sqrt{\sum_{i=k}^{p}|\sigma_{ci}z_i|^2}\right\} \tag{39}$$

$$\leq (\sigma_{ak} + \varepsilon)\sigma_{ck} \tag{40}$$

where $\mathbf{y}^{p-k+1} = \{\mathbf{y}_1\ldots\mathbf{y}_{p-k+1}\}$; (36)-(39) follow the same logic as in (30)-(34), and (40) follows from the fact that $\{\sigma_{ai}\}$ and $\{\sigma_{ci}\}$ are in decreasing order. The upper bound is achievable by $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}_\varepsilon\mathbf{V}^+$, where $\mathbf{\Sigma}_\varepsilon = diag\{\varepsilon\}$, for any given $\mathbf{A}$ and $\mathbf{C}$. The fact that the ordering of $\{\sigma_{ci}\}$ is necessary can be verified by examples (see Example 1 below). □

Remark 1: It should be noted that the popular singular value inequalities for the sum and product of 2 matrices (from which many others are easily derived) in (1), (2) when applied to $(\mathbf{A}+\mathbf{B})\mathbf{C}$ result in (5), (6), which are *not* tight in general and significantly worse than (21) when $\mathbf{C}$ has a large condition number ($=\sigma_{\max}(\mathbf{C})/\sigma_{\min}(\mathbf{C})$). In particular, (5) is useless when $\sigma_{\min}(\mathbf{C}) = 0$ (which is always the case if $rank(\mathbf{C}) < \min(m, n, p)$), unlike (21) which provides useful information even in this case, albeit at the expense of the assumed ordering as in (22). Therefore, these well-known inequalities cannot provide a solution to the problems at hand in general (see e.g. Theorems 1, 2).

Remark 2: The new matrix determinant inequality reported in [12] (see Lemma 2 there) is equivalent to

$$\prod_i(1+\sigma_i^2((\mathbf{A}+\mathbf{B})\mathbf{C})) \geq \prod_i(1+(\sigma_i(\mathbf{A})-\sigma_1(\mathbf{B}))_+^2\sigma_i^2(\mathbf{C})), \tag{41}$$

which is implied by (21) but the converse is not true, so that Proposition 1 is stronger under the assumed ordering of $\{\sigma_{ai}\sigma_{ci}\}$. On the other hand, (41) is more general since it does not assume any particular ordering of $\{\sigma_{ai}\sigma_{ci}\}$, so that these two inequalities complement each other. The next Section discusses some problems (e.g. robust MMSE receiver) that cannot be solved via (41) while Proposition 1 provides such a solution. In addition, it also provides an upper bound and thus identifies the best channel in the uncertainty set. Note also that since (23) implies (22) but not vice-versa, the lower bound is more general than the upper bound.

Example 1: To see that the required ordering of $\{\sigma_{ai}\sigma_{ci}\}$ in Proposition 1 is not superfluous, let us consider a special case of diagonal 2×2 matrices,

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \tag{42}$$

so that the diagonal entries are also the respective singular values. Fig. 1 shows $\sigma_{\min}((\mathbf{A}+\mathbf{B})\mathbf{C})$ (found via the solution to the problem in (20)) as a function of $\varepsilon = \sigma_1(\mathbf{B})$ ($\mathbf{B}$ is selected to minimize $\sigma_{\min}((\mathbf{A}+\mathbf{B})\mathbf{C})$ so that the lower bound in (29) is attained) as well as the unordered singular values of $(\mathbf{A}-\varepsilon I)\mathbf{C}$, minimum of which corresponds to the lower bound in (21). Note that $\sigma_{\min}((\mathbf{A}-\varepsilon I)\mathbf{C})$ is strictly larger than

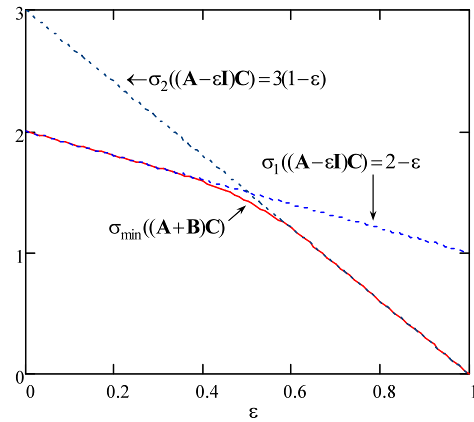

Fig. 1. Unordered singular values of $(\mathbf{A}-\varepsilon\mathbf{I})\mathbf{C}$ and $\sigma_{\min}((\mathbf{A}+\mathbf{B})\mathbf{C})$. The latter is strictly less than $\min\{3(1-\varepsilon), 2-\varepsilon\}$ in the transition region, so that $-\varepsilon\mathbf{I}$ is not the worst-case $\mathbf{B}$.

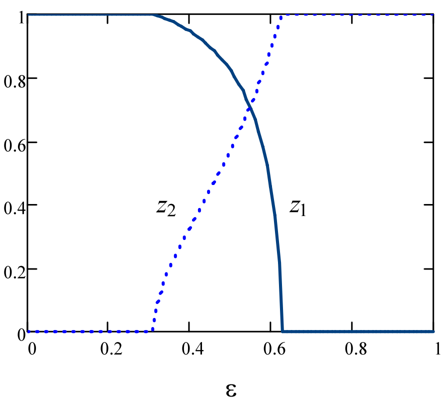

Fig. 2. The right singular vector $z = [z_1, z_2]^T$ corresponding to $\sigma_{\min}((\mathbf{A}+\mathbf{B})\mathbf{C})$. Note that the lower bound in (21) corresponds to either $\mathbf{z} = [1,0]^T$ or $\mathbf{z} = [0,1]^T$, which is not the case in the transition region.

$\sigma_{\min}((\mathbf{A}+\mathbf{B})\mathbf{C})$ for ε in the range

$$\frac{(a_2c_2)^2-(a_1c_1)^2}{a_1(c_2^2-c_1^2)} \approx 0.31 < \varepsilon < \frac{(a_2c_2)^2-(a_1c_1)^2}{a_2(c_2^2-c_1^2)} \approx 0.63, \tag{43}$$

where $\{a_i\}$, $\{c_i\}$ are the diagonal entries of $\mathbf{A}$ and $\mathbf{C}$, so that the lower bound in (21) is not tight in this case, since the required ordering condition is not satisfied. Fig. 2 shows the entries of the right singular vector of $(\mathbf{A}+\mathbf{B})\mathbf{C}$ corresponding to its minimum singular value, which clearly shows that neither $[1,0]^T$ nor $[0,1]^T$ achieve the minimum in the transition region in (43). In a similar way, one can show that the required ordering of $\{\sigma_{ci}\}$ in Proposition 1 is not superfluous.

IV. Performance of Compound MIMO Channels

We are now in a position to characterize the performance of the compound MIMO channel in (8), (9) and to prove the conjecture in [11] under the general power constraint in (12) using the new singular value inequalities. An alternative proof via the determinantal inequality and a less general power constraint can be found in [12].

Theorem 1: The compound capacity C_c of the class of channels in (8), (9) *under the general power constraint in (12) is equal to the worst-case channel capacity C_w,*

$$C_c = C_w = \max_{diag\{\lambda_i\} \in S_R} \sum_{i:\sigma_i(\mathbf{H}_0) > \varepsilon} \log\left(1 + \gamma\left(\sigma_i(\mathbf{H}_0) - \varepsilon\right)^2 \lambda_i\right) \tag{44}$$

so that there is a saddle point $(\mathbf{H}_w, \mathbf{R}_b)$ in $I(\mathbf{H}, \mathbf{R})$ for any admissible $\mathbf{H} \in S_H$ and $\mathbf{R} \in S_R$

$$I(\mathbf{H}_w, \mathbf{R}) \le I(\mathbf{H}_w, \mathbf{R}_b) \le I(\mathbf{H}, \mathbf{R}_b), \tag{45}$$

where $\mathbf{R}_b = \mathbf{V}_0 \Lambda_b \mathbf{V}_0^+$ is the best covariance for $\mathbf{H}_w$, $\Lambda_b = diag\{\lambda_i(\mathbf{R}_b)\}$ is the diagonal matrix of its eigenvalues, which maximize (44)*, and $C_c = C_w = I(\mathbf{H}_w, R_b)$. The worst-case channel $\mathbf{H}_w$ is*

$$\begin{aligned} \mathbf{H}_w &= \mathbf{H}_0 + \Delta\mathbf{H}_w = \mathbf{U}_0 \mathbf{\Sigma}_w \mathbf{V}_0^+, \\ \mathbf{\Sigma}_w &= diag\{(\sigma_i(\mathbf{H}_0) - \varepsilon)_+\}, \end{aligned} \tag{46}$$

$\Delta\mathbf{H}_w = -\mathbf{U}_0 \mathbf{\Sigma}_\varepsilon \mathbf{V}_0^+$ is the worst-case perturbation, $\mathbf{\Sigma}_\varepsilon = diag\{\min(\sigma_i(\mathbf{H}_0), \varepsilon)\}$ is the diagonal matrix of its singular values, $\mathbf{H}_0 = \mathbf{U}_0 \mathbf{\Sigma}_0 \mathbf{V}_0^+$ is the nominal channel SVD.

Proof: First, note that $\{\sigma_i(\mathbf{H}\mathbf{R}^{1/2})\}$ is weakly majorized by $\{\sigma_i(\mathbf{H})\sigma_i(\mathbf{R}^{1/2})\}$ [33], [38]:

$$\sum_{i=1}^{k} \sigma_i(\mathbf{H}\mathbf{R}^{1/2}) \le \sum_{i=1}^{k} \sigma_i(\mathbf{H})\sigma_i(\mathbf{R}^{1/2}), \quad 1 \le k \le m, \tag{47}$$

for any feasible $\mathbf{R}$, and that $f(x) = \log(1 + x^2)$ is increasing on $[0, \infty)$ and $\varphi(t) = f(e^t)$ is convex on $(-\infty, \infty)$ so that, using [33, Th. 3.3.14(c)], one obtains

$$\begin{aligned} I(\mathbf{H}, \mathbf{R}) &= \sum_{i=1}^{m} \log(1 + \gamma\, \sigma_i^2(\mathbf{H}\mathbf{R}^{1/2})) \\ &\le \sum_{i=1}^{m} \log(1 + \gamma\, \sigma_i^2(\mathbf{H})\lambda_i(\mathbf{R})) \end{aligned} \tag{48}$$

where we have used the fact that $\sigma_i^2(\mathbf{R}^{1/2}) = \lambda_i(\mathbf{R})$. Note that this upper bound applies to any $\mathbf{R}$, including optimal one. Also note that, from (12), $\Lambda = diag\{\lambda_i(\mathbf{R})\} \in S_R$ whenever $\mathbf{R} \in S_R$ and that the upper bound in (48) is achieved for given Λ by setting $\mathbf{R} = \mathbf{V}\Lambda\mathbf{V}^+$, i.e. when the eigenvectors of $\mathbf{R}$ are the right singular vectors of $\mathbf{H}$, hence:

$$\max_{\mathbf{R} \in S_{\mathbf{R}}} I(\mathbf{H}, \mathbf{R}) = \max_{diag\{\lambda_i\} \in S_{\mathbf{R}}} \sum_{i=1}^{m} \log(1 + \gamma\, \sigma_i^2(\mathbf{H})\lambda_i) \tag{49}$$

i.e. transmission on right singular vectors of $\mathbf{H}$ is optimal under the general power constraint $\mathbf{R} \in S_R$. Applying (49) to $\mathbf{H} = \mathbf{H}_w$, the best covariance follows,

$$\mathbf{R}_b = \mathbf{V}_0 \Lambda_b \mathbf{V}_0^+ \tag{50}$$

i.e. the optimal signaling is on the right singular vectors of the nominal (or worst-case) channel, where $\Lambda_b = diag\{\lambda_i(\mathbf{R}_b)\} \in S_R$ and $\{\lambda_i(\mathbf{R}_b)\}$ are the best eigenvalues maximizing (49) with $\mathbf{H} = \mathbf{H}_w$. This proves 1^{st} inequality in (45). Further note from (49) with $\mathbf{H} = \mathbf{H}_w$ that $\{\lambda_i(\mathbf{R}_b)\}$, $\{(\sigma_i(\mathbf{H}_0) - \varepsilon)_+\}$ and $\{\sigma_i(\mathbf{H}_0)\}$ are ordered likewise. Therefore, using the lower bound in (21) with $\mathbf{A} = \mathbf{H}_0, \mathbf{B} = \Delta\mathbf{H}, \mathbf{C} = \mathbf{R}_b^{1/2}$, one obtains $\forall(\mathbf{H}_0 + \Delta\mathbf{H}) \in S_{\mathbf{H}}$

$$\sigma_i((\mathbf{H}_0 + \Delta\mathbf{H})\mathbf{R}_b^{1/2}) \ge (\sigma_i(\mathbf{H}_0) - \varepsilon)_+ \sigma_i(\mathbf{R}_b^{1/2}) \tag{51}$$

To prove the upper bound in (45), use (10) with $\mathbf{R} = \mathbf{R}_b$ to obtain:

$$\begin{aligned} I(\mathbf{H}, \mathbf{R}_b) &= \sum_{i=1}^{m} \log\left(1 + \gamma\, \sigma_i^2((\mathbf{H}_0 + \Delta\mathbf{H})\mathbf{R}_b^{1/2})\right) && (52) \\ &\ge \sum_{i=1}^{m} \log\left(1 + \gamma\, (\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2 \sigma_i^2(\mathbf{R}_b^{1/2})\right) && (53) \\ &= \sum_{i=1}^{m} \log\left(1 + \gamma\, (\sigma_i(\mathbf{H}_0) - \varepsilon))_+^2 \lambda_i(\mathbf{R}_b)\right) && (54) \\ &= I(\mathbf{H}_w, \mathbf{R}_b) && (55) \end{aligned}$$

where (53) follows from (51); (54) follows from the fact that $\sigma_i^2(\mathbf{R}_b^{1/2}) = \lambda_i(\mathbf{R}_b)$ since $\mathbf{R}_b$ is positive semi-definite; the last equality is due to (46). This establishes the saddle-point property in (45) from which (44) follows since, using (17) and (18), (44) and (45) are equivalent, see e.g. [58], [59]. □

Examples: under the total power constraint in (13), the compound channel capacity is given by the standard water-filling on the eigenmodes of the worst-case channel $\mathbf{H}_w$ as in (16), $\mathbf{R}^* = \mathsf{WF}(\mathbf{H}_w)$; under the per-eigenmode power constraint in (14), the capacity is achieved by isotropic signaling $\mathbf{R}^* = \mathbf{I}$ and $C_c = \log|\mathbf{I} + \gamma\, \mathbf{H}_w \mathbf{H}_w^+|$. In both cases, S_R is convex. As a non-convex example, consider the following set $S_R = \{\mathbf{R} : \mathbf{R} = \mathbf{U}\mathbf{E}\mathbf{U}^+\}$, where $\mathbf{U}$ is any $m \times m$ unitary matrix and $\mathbf{E}$ is $m \times m$ all-1 matrix, so that all matrices in this set are rank-1 and have eigenvalues $\{m, 0, \ldots, 0\}$; this set is clearly not convex. The optimal signaling is rank-1 (i.e. beamforming) and on the best right singular vector $\mathbf{v}_{10}$ of the nominal channel $\mathbf{H}_0$ (corresponding to its largest singular value): $\mathbf{R}^* = m\mathbf{v}_{10}\mathbf{v}_{10}^+$, and the compound capacity is $C_c = \log(1 + m\gamma\, (\sigma_1(\mathbf{H}_0 - \varepsilon)_+^2)$.

Note that Theorem 4 and the conjecture in [11] are special cases of Theorem 1 (under the total power constraint). While Theorem 1 is limited to the mutual information performance metric, the new singular value inequalities allow one to establish the saddle-point property for a broad class of performance metric monotonic in the channel singular values as shown below.

Theorem 2: Consider the compound MIMO channel in (8), (9) *under the general power constraint in (12), and any performance metric $P(\mathbf{H}, \mathbf{R})$ such that:*

1. It depends on $\{\sigma_i(\mathbf{H}\mathbf{R}^{1/2})\}$ rather than $\mathbf{H}, \mathbf{R}$ individually, $P(\mathbf{H}, \mathbf{R}) = P(\{\sigma_i(\mathbf{H}\mathbf{R}^{1/2})\})$ and is monotonically increasing in each σ_i.

2. The eigenvectors of the performance-optimizing covariance $\mathbf{R}^* = \arg\max_{\mathbf{R}} P(\mathbf{H}, \mathbf{R})$ are the right singular vectors of $\mathbf{H}$; its eigenvalues $\{\lambda_i(\mathbf{R}^*)\}$ are such that $\{\sigma_i^2(\mathbf{H})\lambda_i(\mathbf{R}^*)\}$ and $\{\sigma_i(\mathbf{H})\}$ are ordered likewise. In particular, this holds if $P(\{e^{\sigma_i}\})$ is Schur-convex.

Then, the saddle-point property holds for $P(\mathbf{H}, \mathbf{R})$ and any admissible $\mathbf{H} \in S_H$ and $\mathbf{R} \in S_R$,

$$P(\mathbf{H}_w, \mathbf{R}) \le P(\mathbf{H}_w, \mathbf{R}_b) \le P(\mathbf{H}, \mathbf{R}_b) \tag{56}$$

where the worst-case channel $\mathbf{H}_w$ is as in (46)*, so that the corresponding min-max relation follows:*

$$\max_{\mathbf{R}\in S_R}\min_{\mathbf{H}\in S_H} P(\mathbf{H},\mathbf{R}) = \min_{\mathbf{H}\in S_H}\max_{\mathbf{R}\in S_R} P(\mathbf{H},\mathbf{R}) = P(\mathbf{H}_w,\mathbf{R}_b) \quad (57)$$

i.e. the best covariance $\mathbf{R}_b = \arg\max_{\mathbf{R}} P(\mathbf{H}_w,\mathbf{R})$ that maximizes the performance of the worst-case channel will also maximize the performance of the whole class of channels (universally-optimal transmission).

Proof: Follows in the same way as that of Theorem 1 by noting that it is the conditions 1 and 2 which are essential in the proof of Theorem 1, not a particular functional form of $I(\mathbf{H},\mathbf{R})$. □

We note that Theorem 2 is a considerable extension of Theorem 1 which cannot be obtained using the new determinantal inequality in [12] (since the latter requires the performance metric to be a determinant or its monotonic function). Also note that $P(\mathbf{H},\mathbf{R})$ is not required to be convex in $\mathbf{H}$ and concave in $\mathbf{R}$ and S_R is not necessarily convex, so that Von Neumann minimax theorem does not apply and KKT conditions are not sufficient for optimality either, which rules out the use of convex optimization tools [59]. The ordering in condition 2 is satisfied when $\{\sigma_i(\mathbf{H})\}$ and $\{\lambda_i(\mathbf{R}^*)\}$ are both decreasing sequences, i.e. strong modes get more power; this is inspired by the water-filling in (16) under the total power constraint and also holds under the general power constraint in Theorem 1. In fact, Theorem 2 does not restrict the power constraint set beyond this condition, so it can be more general than in Theorem 1. The robust MMSE precoder below also satisfies the required properties. For all these performance metrics and under the general power constraint, the optimal strategy is to diagonalize the nominal (or worst-case) channel and an optimal power allocation depends on a particular performance metric and power constraint. This extends the corresponding results established for perfect CSI [62] to channels with uncertainty.

Following Theorem 2, the worst-case channel in (46) is indeed worst for a broad class of performance metrics, not just mutual information (i.e. ‘universally worst”). A wealth of performance metrics that satisfy conditions 1, 2 in Theorem 2 can be found in [62]–[65]. Furthermore, these properties are satisfied by such important performance metrics, beyond mutual information, as pair-wise error probability, error exponent and mean square error (MSE) [64]. Proposition 1 also reveals that the best-case channel perturbation is of the form $\Delta\mathbf{H}_b = \mathbf{U}_0\mathbf{\Sigma}_\varepsilon\mathbf{V}_0^+$, where $\mathbf{\Sigma}_\varepsilon = diag\{\varepsilon\}$, so that the best channel is $\mathbf{H}_b = \mathbf{H}_0 + \Delta\mathbf{H}_b$ and (56) can be supplemented by

$$P(\mathbf{H},\mathbf{R}_b) \le P(\mathbf{H}_b,\mathbf{R}_b) \quad (58)$$

provided that $\{\lambda_i(\mathbf{R}_b)\}$ are in decreasing order.

A. Robust MMSE Precoder

In this section, we apply the result above to design a robust MMSE MIMO precoder. Note that the robust designs proposed in [28]–[32] for MIMO channels with uncertainty maximize the worst-case total received SNR, which is not the same as minimizing the sum MSE unless the SNR is low or when the nominal channel is rank-1. Our solution below does not have these limitations: it applies to any SNR and any rank of the nominal channel under the sum-MSE criterion. To the best of our knowledge, no analytical solution is known for the robust MIMO precoder design under the MSE criterion in the general case. While numerical algorithms where developed in [27] and [31] to optimize the sum-MSE, their convergence cannot be insured since the underlining optimization problem is not convex. Our SV inequality-based approach overcomes this difficulty and provides a closed-form solution to this open problem in Proposition 2 under the total power constraint in (13).

The MIMO channel model including precoder $\mathbf{F}$ and receiver (decoder) $\mathbf{W}$ is as follows:

$$\hat{\mathbf{x}} = \mathbf{W}^+\mathbf{H}\mathbf{F}\mathbf{x} + \mathbf{W}^+\boldsymbol{\xi} \quad (59)$$

and the MSE matrix $\mathbf{E} = E\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^+\}$, where $E\{\cdot\}$ denotes statistical expectation over noise and signal distributions; the sum MSE is $tr\mathbf{E}$. The best linear receiver under the sum MSE criterion and the perfect Rx CSI is the Wiener filter $\mathbf{W}^* = (\mathbf{H}\mathbf{R}\mathbf{H}^+ + \mathbf{I})^{-1}\mathbf{H}\mathbf{F}$, for which the MSE matrix $\mathbf{E}$ at its output is [62]–[65]

$$\mathbf{E}^* = (\mathbf{I} + \gamma\,\mathbf{F}^+\mathbf{H}^+\mathbf{H}\mathbf{F})^{-1} \quad (60)$$

where $\mathbf{R} = \mathbf{F}\mathbf{F}^+$ (we assume here that $E\{\mathbf{x}\mathbf{x}^+\} = \mathbf{I}$), and the sum MSE is

$$\mathsf{MSE}(\mathbf{H},\mathbf{R}) = tr\mathbf{E}^* = \sum_i \frac{1}{1+\gamma\,\sigma_i^2(\mathbf{H}\mathbf{R}^{1/2})} \quad (61)$$

which can be further optimized over $\mathbf{F}$ to obtain the best precoder $\mathbf{F}^* = \mathbf{V}\Lambda$ under the full Tx CSI and the total power constraint $tr\mathbf{F}\mathbf{F}^+ = tr\mathbf{R} \le m$, where $\mathbf{V}$ collects the right singular vectors of $\mathbf{H}$ and $\Lambda = diag\{\lambda_{bi}^{1/2}\}$ is the diagonal matrix, λ_{bi} is the power allocated to i-th stream (the i-th eigenvalue of $\mathbf{R}$) found from the water-filling type algorithm:

$$\lambda_{bi} = \left(\frac{\mu}{\sqrt{\gamma}\sigma_i(\mathbf{H})} - \frac{1}{\gamma\,\sigma_i^2(\mathbf{H})}\right)_+ \quad (62)$$

where μ is found from the total power constraint [62]. It can be further shown that channel inversion is the best strategy at high SNR provided that the channel is of full rank. The resulting MMSE is:

$$\mathsf{MMSE} = \sum_{i:\lambda_{bi}>0} \frac{1}{1+\gamma\,\sigma_i^2(\underline{\mathbf{H}})\lambda_{bi}} = \frac{1}{\mu\sqrt{\gamma}} \sum_{i:\mu\sqrt{\gamma}\sigma_i(\mathbf{H})>1} \frac{1}{\sigma_i(\mathbf{H})} \quad (63)$$

Using the singular value inequalities in Proposition 1, this can now be extended to the robust precoder design under the Tx channel uncertainty and perfect Rx CSI:

$$\min_{\mathbf{F}}\max_{\mathbf{H}}\min_{\mathbf{W}} tr\mathbf{E} = \min_{\mathbf{R}}\max_{\mathbf{H}} \mathsf{MSE}(\mathbf{H},\mathbf{R}) \quad (64)$$

where $\min_{\mathbf{F}}$ is subject to the total power constraint $tr\mathbf{F}\mathbf{F}^+ \le m$ and $\max_{\mathbf{H}}$ is over the uncertainty set and subject to the spectral norm constraint $\mathbf{H} \in S_H$ in (9).

Proposition 2: Consider the robust precoder design in (64) *for the compound channel in* (9) *under the total power constraint in (13). The saddle-point property holds:*

$$\min_{\mathbf{R}} \max_{\mathbf{H}} \mathsf{MSE}(\mathbf{H}, \mathbf{R}) = \max_{\mathbf{H}} \min_{\mathbf{R}} \mathsf{MSE}(\mathbf{H}, \mathbf{R}) = \mathsf{MSE}(\mathbf{H}_w, \mathbf{R}^*) \tag{65}$$

where the worst-case channel is $\mathbf{H}_w$ *in* (46) *and the best robust precoder is given by*

$$\mathbf{F}^* = \mathbf{V}_0 \Lambda^* \tag{66}$$

where $\mathbf{V}_0$ *collects the right singular vectors of the nominal (or worst-case) channel* $\mathbf{H}_0$ *and* $\Lambda^* = diag\{\lambda_{bi}^{1/2}\}$ *is the diagonal matrix of optimal power allocation as in* (62) *with* $\mathbf{H} = \mathbf{H}_w$, *for which the robust MMSE can be found from* (63) *with* $\mathbf{H} = \mathbf{H}_w$.

Proof: Take $P(\mathbf{H}, \mathbf{R}) = -\mathsf{MSE}(\mathbf{H}, \mathbf{R})$ and note that 1st condition in Theorem 2 holds. 2nd condition follows from the best precoder and power allocation in (62), for which it can be shown that $\{\lambda_{bi}\sigma_i^2(\mathbf{H})\}$ and $\{\sigma_i^2(\mathbf{H})\}$ are ordered likewise. Therefore, (56) and (57) provide a robust MMSE formulation for the compound channel in (8), (9). □

Note that the worst-case channel is still as in (46) and the best transmission strategy is signaling on the right singular vectors of $\mathbf{H}_0$ with power allocation as in (62) applied to $\mathbf{H} = \mathbf{H}_w$. Because of the saddle-point property in (57), this transmission strategy works for any channel in the uncertainty set, i.e. for the compound channel, and the performance is as good as for the worst-case channel only. Note also that the best transmission strategy is to diagonalize the nominal (or worst-case) channel, which extends the corresponding results obtained for MIMO channels without uncertainty [62].

V. Beamforming Over the Compound MIMO Channel

Let us now consider the case of transmit beamforming, which may be motivated by low implementation complexity. The system model is as in (7) and (8), but the transmit covariance $\mathbf{R}$ is of rank one, $\mathbf{R} = m \cdot \mathbf{w}\mathbf{w}^+$, where $\mathbf{w}$ is its eigenvector corresponding to non-zero eigenvalue, which is also a beamforming vector in the antenna array terminology [66]. When $\mathbf{w}$ is optimized to maximize the mutual information,

$$\mathbf{w}^* = \arg\max_{|w|=1} \log(1 + m\gamma\ |\mathbf{H}\mathbf{w}|^2) \tag{67}$$

it is straightforward to see that the best strategy is the transmission on the strongest eigenmode of $\mathbf{H}$: $\mathbf{w}^* = \mathbf{v}_1$, where $\mathbf{v}_1$ is the right singular vector of $\mathbf{H}$ corresponding to its largest singular value. For the compound channel as in (8) and (9), the best strategy remains the same: transmission on the strongest eigenmode of the nominal channel is optimal,

$$\mathbf{w}_b = \arg\max_{|\mathbf{w}|=1} \min_{\Delta\mathbf{H}\in S_{\mathbf{H}}} \log(1+m\gamma\ |(\mathbf{H}_0+\Delta\mathbf{H})\mathbf{w}|^2) = \mathbf{v}_{10} \tag{68}$$

where $\mathbf{v}_{10}$ is the right singular vector of $\mathbf{H}_0$ corresponding to its largest singular value and the worst channel is still as in (46). This solution is also optimal in terms of the total Rx SNR [29].

However, in many applications, there are some additional constraints on the beamforming vector $\mathbf{w}$ (e.g. placing nulls in certain directions etc.) [66] so that $\mathbf{w} = \mathbf{v}_{10}$ is not feasible. While [24], [26], [29] identify the best beamforming vectors under different performance criteria, they identify worst-case channels, either analytically or numerically, under those "best" beamforming vectors only. A question emerges as to whether the worst-case channel remains the same if the problem is slightly changed, e.g. if extra null constraints are introduced [66], so that the original optimal beamforming vector is not feasible any more. For many metrics (e.g. the MI, MSE), the performance is determined by the received SNR, which is proportional to the channel power gain $|\mathbf{H}\mathbf{w}|^2 = \mathbf{w}^+\mathbf{H}^+\mathbf{H}\mathbf{w}$. Hence, an answer to this question is determined by the properties of channel covariance matrix $\mathbf{H}^+\mathbf{H}$ for the uncertainty set in (9), which is established below.

While the singular values of $\mathbf{H}_w$ and $\mathbf{H}_b$ lower and upper bound the singular values of $\mathbf{H}$, this does not imply in general that $\mathbf{H}_w^+\mathbf{H}_w \leq \mathbf{H}^+\mathbf{H} \leq \mathbf{H}_b^+\mathbf{H}_b$, i.e. the latter are stronger conditions. The proposition below shows that this implication holds when the nominal channel $\mathbf{H}_0$ has identical singular values.

Proposition 3: Consider the compound MIMO channel in (8), (9). *Then,*

$$\mathbf{H}_w^+\mathbf{H}_w \leq \mathbf{H}^+\mathbf{H} \leq \mathbf{H}_b^+\mathbf{H}_b \tag{69}$$

or, equivalently,

$$|\mathbf{H}_w\mathbf{w}| \leq |\mathbf{H}\mathbf{w}| \leq |\mathbf{H}_b\mathbf{w}| \quad \forall\mathbf{w}, \quad \|\Delta\mathbf{H}\|_2 \leq \varepsilon \tag{70}$$

if and only if $\sigma_1(\mathbf{H}_0) = \sigma_2(\mathbf{H}_0) = \ldots = \sigma_{\min(n,m)}(\mathbf{H}_0) = \sigma_0$ *or, for the lower bound,* $\varepsilon \geq \sigma_1(\mathbf{H}_0)$.

Proof: See Appendix. □

Now note that $|\mathbf{H}\mathbf{w}|$ is the channel (voltage) gain in transmit beamforming direction $\mathbf{w}$, assuming $|\mathbf{w}| = 1$, so that (70) states that $\mathbf{H}_w$ is worse than $\mathbf{H}$ for any admissible $\Delta\mathbf{H}$ in any beamforming direction, i.e. designing a beamformer for $\mathbf{H}_w$ is guaranteed to achieve the gain $|\mathbf{H}_w\mathbf{w}|$ for any admissible channel $\mathbf{H}$ and any $\mathbf{w}$ when the nominal channel singular values are identical. This constitutes a robust design under the spectral norm constraint for any beamformer, not just for those that maximize the SNR or MI or minimize the MSE (as in [24], [26], and [29]). In particular, it can handle extra constraints such as nulls in specific directions.

However, when at least two singular values of the nominal channel are distinct, (89) in Appendix implies that $\mathcal{N}(\mathbf{H}) \notin \mathcal{N}(\mathbf{H}_w)$, where $\mathcal{N}(\mathbf{H}) = \{\mathbf{x} : \mathbf{H}\mathbf{x} = 0\}$ is a null space of matrix $\mathbf{H}$, i.e. $\mathbf{H}$ may have null directions different from those of $\mathbf{H}_w$, so that the latter is not universally worst for the transmit beamforming while being universally worst from the information-theoretic viewpoint. This complements the properties of the null space in [11, Proposition 1].

In any case, the worst beamforming gain in transmit direction $\mathbf{w}$, assuming $|\mathbf{w}| = 1$ and using (86) in Appendix, is

$$\min_{\|\Delta\mathbf{H}\|_2\leq\varepsilon} |(\mathbf{H}_0 + \Delta\mathbf{H})\mathbf{w}| = (|\mathbf{H}_0\mathbf{w}| - \varepsilon)_+ \tag{71}$$

i.e. the spectral norm bound ε indeed quantifies the worst-case loss in the beamforming (voltage) gain compared to the nominal channel; this loss is independent of the beamforming direction $\mathbf{w}$ and is determined by the spectral norm of $\Delta\mathbf{H}$. On the other hand, the most favorable channel uncertainty gives

$$\max_{\|\Delta\mathbf{H}\|_2\leq\varepsilon} |(\mathbf{H}_0+\Delta\mathbf{H})\mathbf{w}| = |\mathbf{H}_0\mathbf{w}|+\varepsilon \tag{72}$$

so that under the additive uncertainty with bounded spectral norm in (8), (9), the beamforming gain is bounded, for any $\mathbf{w}$, as

$$(|\mathbf{H}_0\mathbf{w}|-\varepsilon)_+ \leq |\mathbf{H}\mathbf{w}| \leq |\mathbf{H}_0\mathbf{w}|+\varepsilon \tag{73}$$

which underlines the utility of the spectral norm as a measure of channel uncertainty for beamforming applications.

VI. Conclusion

New matrix singular value inequalities for the product/sum of 3 matrices have been established. Based on this, performance of compound MIMO channels has been characterized for a wide range of performance metrics monotonic in the channel singular values including, among others, mutual information, MMSE, error exponent, pairwise error probability, all of which exhibit the saddle-point property so that transmission strategy optimal for the worst-case channel is also optimal for the whole class of channels, and the worst-case channel is anti-parallel of the nominal one. For all these performance metrics and under the general power constraint, the optimal strategy is to diagonalize the nominal (or worst-case) channel and an optimal power allocation depends on a particular performance metric and power constraint. One notable exception is transmit beamforming where the worst-case channel may differ from that under the above-mentioned criteria, unless the nominal channel has identical singular values or if all of them are smaller than the uncertainty spectral bound. The analysis also reveals the best possible channel uncertainty, whose performance upper bounds that of all other channels in the uncertainty set. A robust precoder design has been developed under the sum-MSE criterion and the spectral norm constraint on the channel uncertainty in the general case (not only low SNR or rank-1 channel).

Appendix

A. Proof of Lemma 1

It is straightforward to see that $\min_{|\mathbf{z}|=1} F(\mathbf{z})=0$ if $\varepsilon\geq a_n$. Indeed,

$$F(\mathbf{z}) \geq \{a_n-\varepsilon\}_+ \sqrt{\sum_{i=1}^n |c_i z_i|^2}=0 \tag{74}$$

and the lower bound is achieved by $\mathbf{z}=[0,0\ldots 0,1]^T$. Thus, without loss of generality, we further assume that $\varepsilon<a_n$ and that all z_i are real. In this case,

$$F_1(\mathbf{z}) = \sqrt{\sum_{i=1}^n |a_i c_i z_i|^2} - \varepsilon\sqrt{\sum_{i=1}^n |c_i z_i|^2} \geq 0 \tag{75}$$

so we may drop $\{\}_+$ in (19) and consider the following problem:

$$\min_{\mathbf{z}} F_1(\mathbf{z}) \text{ s.t. } \sum_{i=1}^n z_i^2 = 1 \tag{76}$$

In general, it is a non-convex problem so that KKT conditions are not sufficient for optimality [59]. We solve it using the following 4-step method [61], [60]:

1) Establish an existence of a global solution: since the objective $F_1(\mathbf{z})$ is a continuous function of $\mathbf{z}$ and the constraint set $|\mathbf{z}|=1$ is compact, the existence of a solution follows from Weierstrass theorem.

2) Find necessary conditions: KKT conditions are necessary for optimality (this follows from e.g. [60, Proposition 3.3.8]), so that a global minimum is a solution of KKT conditions.

3) Find all solutions of KKT conditions.

4) By inspection, find the global minimum.

The Lagrangian of the problem is

$$L = F_1(\mathbf{z}) - \lambda\left(\sum_{i=1}^n z_i^2 - 1\right) \tag{77}$$

and the KKT conditions are:

$$\frac{\partial L}{\partial z_i} = \frac{(a_i c_i)^2 z_i}{|\mathbf{D}_a\mathbf{D}_c\mathbf{z}|} - \varepsilon\frac{c_i^2 z_i}{|\mathbf{D}_c\mathbf{z}|} - 2\lambda z_i = 0, \quad i=1\ldots n, \tag{78}$$

$$\sum_{i=1}^n z_i^2 = 1, \tag{79}$$

where $\mathbf{D}_a = diag\{a_i\}$ and likewise for $\mathbf{D}_c$, and λ is the Lagrange multiplier. If $z_i\neq 0$, then from (78),

$$\frac{(a_i c_i)^2}{|\mathbf{D}_a\mathbf{D}_c\mathbf{z}|} - \frac{\varepsilon c_i^2}{|\mathbf{D}_c\mathbf{z}|} = 2\lambda \tag{80}$$

Now take $i=1$ and $i=k\neq 1$ to obtain

$$\frac{(a_1c_1)^2-(a_kc_k)^2}{|\mathbf{D}_a\mathbf{D}_c\mathbf{z}|} = \frac{\varepsilon(c_1^2-c_k^2)}{|\mathbf{D}_c\mathbf{z}|} \tag{81}$$

If $c_1<c_k$, there is no solution, so we further assume that $c_1\geq c_k$. Consider first the case of $c_1>c_k$:

$$\frac{|\mathbf{D}_a\mathbf{D}_c\mathbf{z}|}{|\mathbf{D}_c\mathbf{z}|} = \frac{(a_1c_1)^2-(a_kc_k)^2}{\varepsilon(c_1^2-c_k^2)} \geq \frac{a_1^2}{\varepsilon} > a_1 \tag{82}$$

where the last inequality follows from $a_1>\varepsilon$, which is implied by $a_n>\varepsilon$. On the other hand,

$$\frac{|\mathbf{D}_a\mathbf{D}_c\mathbf{z}|}{|\mathbf{D}_c\mathbf{z}|} \leq a_1 \tag{83}$$

which follows from $\mathbf{D}_a\leq a_1\mathbf{I}$, where $\mathbf{A}\leq\mathbf{B}$ means that $\mathbf{B}-\mathbf{A}$ is positive semi-definite. Clearly, (82) and (83) are in contradiction, so that there is no solution here either. Finally, consider the degenerate case of $c_1=c_k$: there is a solution to (81) only if $a_1=a_k$, in which case $a_1=a_2=\ldots=a_k$ and $a_1c_1=a_2c_2=\ldots=a_kc_k$ so that $c_1=c_2=\ldots=c_k$ and any $\mathbf{z}=[z_1,z_2,\ldots z_k,0\ldots 0]^T$ delivers the same objective

$$F_1(\mathbf{z}) = (a_1-\varepsilon)c_1 \geq (a_n-\varepsilon)c_n \tag{84}$$

which is sub-optimal in general (strict inequality unless $a_1=a_2=\ldots=a_n$ and likewise for c_i, in which case the problem is trivial), since the lower bound in (84) is achievable by $\mathbf{z} = [0, \dots, 0, 1]^T$. Thus, there is no solution here either and we are left with two possibilities: either $z_1 = 0$ or/and $z_k = 0$. If $z_1 \neq 0$, then $z_k = 0 \; \forall k \geq 2$ so that $z_1 = 1$, but this is not optimal as (84) demonstrates. Thus, we conclude that $z_1 = 0$. This reduces the problem from n-dimensional to (n–1)-dimensional and, by induction, we conclude that an optimal solution is $\mathbf{z} = [0, \dots, 0, 1]^T$, which clearly delivers (20). □

B. Proof of Proposition 3

Let us consider the lower bound first. If $\varepsilon \geq \sigma_1(\mathbf{H}_0)$, then $\mathbf{H}_w = 0$ and the inequality is trivial. So, we consider the non-trivial case of $\varepsilon < \sigma_1(\mathbf{H}_0)$ and assume for simplicity of the discussion that $n \geq m$ (with slight modifications, the proof can be applied to the $n < m$ case as well). To prove the if part, observe the following:

$$\begin{aligned} |\mathbf{H}\mathbf{w}| &= |\mathbf{H}_0\mathbf{w} + \Delta\mathbf{H}\mathbf{w}| \\ &= \left|\mathbf{\Sigma}_0\mathbf{z} + \Delta\tilde{\mathbf{H}}\mathbf{z}\right| \\ &\overset{(a)}{\geq} |\mathbf{\Sigma}_0\mathbf{z}| - \left|\Delta\tilde{\mathbf{H}}\mathbf{z}\right| \\ &\overset{(b)}{\geq} |\mathbf{\Sigma}_0\mathbf{z}| - \varepsilon\,|\mathbf{z}|\,(\sigma_0 - \varepsilon)\,|\mathbf{z}|\,|\mathbf{H}_w\mathbf{w}| \end{aligned} \tag{85}$$

where $\mathbf{z} = \mathbf{V}_0^+\mathbf{w}$, $\Delta\tilde{\mathbf{H}} = \mathbf{U}_0^+\Delta\mathbf{H}\mathbf{V}_0$, (a) follows from the triangle inequality, (b) follows from $\left\|\Delta\tilde{\mathbf{H}}\right\|_2 = \left\|\Delta\mathbf{H}\right\|_2 \leq \varepsilon$, and this holds for any $\mathbf{w}$ and any admissible $\Delta\mathbf{H}$.

To prove the only if part, we assume that there are at least 2 different singular values of $\mathbf{H}_0$ and arrange its singular values in decreasing order. Let us consider first the case of $\sigma_m(\mathbf{H}_0) \geq \varepsilon$ and observe that (85) holds all the way up to (b) and in fact can be achieved with equality by using $\Delta\tilde{\mathbf{H}} = -\varepsilon\mathbf{U}\mathbf{\Sigma}_I$, where $\mathbf{\Sigma}_I = diag(1, 1, \dots 1)$ is an $n \times m$ matrix with 1s on the main diagonal and zeros elsewhere and $\mathbf{U}$ is a unitary (rotation) matrix such that $\mathbf{\Sigma}_0\mathbf{z} = \alpha\mathbf{U}\mathbf{\Sigma}_I\mathbf{z}$, $\alpha > 0$, i.e. $\mathbf{\Sigma}_0\mathbf{z}$ and $\mathbf{U}\mathbf{\Sigma}_I\mathbf{z}$ are parallel. Under this choice of $\Delta\tilde{\mathbf{H}}$, one obtains:

$$|\mathbf{H}\mathbf{w}| = |\mathbf{\Sigma}_0\mathbf{z}| - \varepsilon\,|\mathbf{z}| \tag{86}$$

and this can be done for any $\mathbf{\Sigma}_0$ and $\mathbf{z}$. Let us now select $\mathbf{z} = [1, 0, \dots 0, 1]^T$ and observe that

$$\begin{aligned} |\mathbf{H}_w\mathbf{w}| &= |(\mathbf{\Sigma}_0 - \varepsilon\mathbf{\Sigma}_I)\mathbf{z}| \\ &> |\mathbf{\Sigma}_0\mathbf{z}| - \varepsilon\,|\mathbf{\Sigma}_I\mathbf{z}| = |\mathbf{\Sigma}_0\mathbf{z}| - \varepsilon\,|\mathbf{z}| = |\mathbf{H}\mathbf{w}| \end{aligned} \tag{87}$$

where the inequality is due to the fact that $\mathbf{\Sigma}_0\mathbf{z} = [\sigma_1(\mathbf{H}_0), 0, \dots 0, \sigma_m(\mathbf{H}_0), 0, \dots 0]^T$ and $\mathbf{\Sigma}_I\mathbf{z} = [1, 0, \dots 0, 1, 0, \dots 0]^T$ are not parallel so that the equality in the triangle inequality cannot be achieved and the last equality comes from (86). Thus, (69) breaks down if at least two singular values of $\mathbf{H}_0$ are distinct. Let us now consider the case of $\sigma_m(\mathbf{H}_0) < \varepsilon$ and set

$$\begin{aligned} z_1 &= \sqrt{\frac{\varepsilon^2 - \sigma_m^2(\mathbf{H}_0)}{\sigma_1^2(\mathbf{H}_0) - \sigma_m^2(\mathbf{H}_0)}}, \\ z_m &= \sqrt{1 - z_1^2}, \quad z_k = 0 \quad \forall k \neq 1, m \end{aligned} \tag{88}$$

Noting that $0 < z_1 < 1$ and setting $\Delta\tilde{\mathbf{H}}$ to satisfy (86), one obtains

$$|\mathbf{H}\mathbf{w}| = |\mathbf{\Sigma}_0\mathbf{z}| - \varepsilon = 0 < (\sigma_1(\mathbf{H}_0) - \varepsilon)z_1 = |\mathbf{H}_w\mathbf{w}| \tag{89}$$

where 2^{nd} equality is due to the (easy to verify) fact that $|\Sigma_0\mathbf{z}| = \varepsilon$. The upper bound can be proved in a similar way. □

Acknowledgement

The authors would like to thank S. Primak for pointing an earlier version of [12], the authors of [12] for providing an advanced copy of their manuscript, and the reviewers for numerous constructive comments.

References

[1] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radiotech. Electron.*, vol. 4, no. 12, pp. 1951–1956, Dec. 1959.

[2] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, 1959.

[3] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[4] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[5] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[6] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, Jul. 2003.

[7] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Optimization of the MIMO compound capacity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1094–1101, Mar. 2007.

[8] C. Köse and R. D. Wesel, "Universal space-time trellis codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2717–2727, Oct. 2003.

[9] S. Srinivasa and S. A. Jafar, "The optimality of transmit beamforming: A unified view," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1558–1564, Apr. 2007.

[10] S. Z. Denic, C. D. Charalambous, and S. M. Djouadi, "Information theoretic bounds for compound MIMO Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1603–1617, Apr. 2009.

[11] S. Loyka and C. D. Charalambous, "On the compound capacity of a class of MIMO channels subject to normed uncertainty," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2048–2063, Apr. 2012.

[12] Y. Sun, C. E. Koksal, and N. B. Shroff, "Capacity of compound MIMO Gaussian channels with additive uncertainty," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8267–8274, Dec. 2013.

[13] I. Ioannou, C. D. Charalambous, and S. Loyka, "Outage probability under channel distribution uncertainty," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6825–6838, Nov. 2012.

[14] S. Loyka, C. D. Charalambous, and I. Ioannou, "Ergodic capacity under channel distribution uncertainty," in *Proc. 52nd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Oct. 2014, pp. 779–784.

[15] A. Khina and U. Erez, "On the robustness of dirty paper coding," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1437–1446, May 2010.

[16] A. Khina, Y. Kochman, and U. Erez, "Joint unitary triangularization for MIMO networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 326–336, Jan. 2012.

[17] I. Livni, A. Khina, A. Hitron, and U. Erez, "Space-time MIMO multicasting," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Boston, MA, USA, Jul. 2012, pp. 2831–2835.

[18] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 692–703, Feb. 2013.

[19] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[20] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.
[21] Y. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.
[22] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.
[23] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known MIMO communications channels," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2918–2930, Aug. 2005.
[24] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, "Robust transmit eigen beamforming based on imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1596–1609, May 2006.
[25] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.
[26] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.
[27] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.
[28] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.
[29] J. Wang and M. Payaró, "On the robustness of transmit beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5933–5938, Nov. 2010.
[30] T. E. Bogale, B. K. Chalise, and L. Vandendorpe, "Robust transceiver optimization for downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 446–453, Jan. 2011.
[31] J. Wang and M. Bengtsson, "Joint optimization of the worst-case robust MMSE MIMO transceiver," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 295–298, May 2011.
[32] J. Wang, M. Bengtsson, B. Ottersten, and D. P. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3056–3070, Jun. 2013.
[33] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[35] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[36] F. Zhang, *Matrix Theory*. New York, NY, USA: Springer-Verlag, 2011.
[37] X. Zhan, *Matrix Inequalities*. Berlin, Germany: Springer-Verlag, 2002.
[38] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, USA: Springer-Verlag, 2011.
[39] R. C. Thompson, "Singular value inequalities for matrix sums and minors," *Linear Algebra Appl.*, vol. 11, no. 3, pp. 251–269, 1975.
[40] R. Bhatia and F. Kittaneh, "On the singular values of a product of operators," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 2, pp. 272–277, Apr. 1990.
[41] J. F. Queiró and E. M. Sá, "Singular values and invariant factors of matrix sums and products," *Linear Algebra Appl.*, vol. 225, pp. 43–56, Aug. 1995.
[42] B.-Y. Wang and B.-Y. Xi, "Some inequalities for singular values of matrix products," *Linear Algebra Appl.*, vol. 264, pp. 109–115, Oct. 1997.
[43] R. A. Horn, N. H. Rhee, and S. Wasin, "Eigenvalue inequalities and equalities," *Linear Algebra Appl.*, vol. 270, no. 1, pp. 29–44, Feb. 1998.
[44] X. Zhan, "Some research problems on the Hadamard product and singular values of matrices," *Linear Multilinear Algebra*, vol. 47, no. 2, pp. 191–194, 2000.
[45] X. Zhan, "Singular values of differences of positive semidefinite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 3, pp. 819–823, Oct. 2000.
[46] J.-C. Bourin, "Singular values of products of positive operators: *AZB* and *ZAB*," *Linear Algebra Appl.*, vol. 407, pp. 64–70, Sep. 2005.
[47] O. Hirzallah, "Inequalities for sums and products of operators," *Linear Algebra Appl.*, vol. 407, pp. 32–42, Sep. 2005.
[48] O. Hirzallah and F. Kittaneh, "Inequalities for sums and direct sums of Hilbert space operators," *Linear Algebra Appl.*, vol. 424, no. 1, pp. 71–82, Jul. 2007.
[49] Y. Tao, "More results on singular value inequalities of matrices," *Linear Algebra Appl.*, vol. 416, nos. 2–3, pp. 724–729, Jul. 2006.
[50] K. M. R. Audenaert, "A singular value inequality for Heinz means," *Linear Algebra Appl.*, vol. 422, no. 1, pp. 279–283, Apr. 2007.
[51] F. Kittaneh, "Singular value inequalities for commutators of Hilbert space operators," *Linear Algebra Appl.*, vol. 430, nos. 8–9, pp. 2362–2367, Apr. 2009.
[52] K. M. R. Audenaert, "A Lieb–Thirring inequality for singular values," *Linear Algebra Appl.*, vol. 430, nos. 11–12, pp. 3053–3057, Jun. 2009.
[53] S. W. Drury, "On a question of Bhatia and Kittaneh," *Linear Algebra Appl.*, vol. 437, no. 7, pp. 1955–1960, Oct. 2012.
[54] W. Audeh and F. Kittaneh, "Singular value inequalities for compact operators," *Linear Algebra Appl.*, vol. 437, no. 10, pp. 2516–2522, Nov. 2012.
[55] H. Albadawi, "Singular value and arithmetic-geometric mean inequalities for operators," *Ann. Funct. Anal.*, vol. 3, no. 1, pp. 10–18, 2012.
[56] R. Dumitru, R. Levanger, and B. Visinescu, "On singular value inequalities for matrix means," *Linear Algebra Appl.*, vol. 439, no. 8, pp. 2405–2410, Oct. 2013.
[57] A. Mathias, "Singular values and singular value inequalities," in *Handbook of Linear Algebra*, L. Hogben, Ed. Boca Raton, FL, USA: CRC Press, 2014.
[58] E. Zeidler, *Nonlinear Functional Analysis and Its Applications: Fixed-Point Theorems*. New York, NY, USA: Springer-Verlag, 1986.
[59] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[60] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 2008.
[61] J. Brinkhuis and V. Tikhomirov, *Optimization: Insights and Applications*. Princeton, NJ, USA: Princeton Univ. Press, 2005.
[62] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Found. Trends Commun. Inf. Theory*, vol. 3, nos. 4–5, pp. 331–551, Nov. 2006.
[63] E. Jorswieck and H. Boche, "Majorization and matrix-monotone functions in wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 3, no. 6, pp. 553–701, 2006.
[64] E. Biglieri, *MIMO Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2007.
[65] P. P. Vaidyanathan, *Signal Processing and Optimization for Transceiver Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[66] H. L. Van Trees, *Optimum Array Processing*. New York, NY, USA: Wiley, 2002.

Sergey Loyka was born in Minsk, Belarus. He received the Ph.D. degree in Radio Engineering from the Belorussian State University of Informatics and Radioelectronics (BSUIR), Minsk, Belarus in 1995 and the M.S. degree with honors from Minsk Radioengineering Institute, Minsk, Belarus in 1992. Since 2001 he has been a faculty member at the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. Prior to that, he was a research fellow in the Laboratory of Communications and Integrated Microelectronics (LACIME) of Ecole de Technologie Superieure, Montreal, Canada; a senior scientist at the Electromagnetic Compatibility Laboratory of BSUIR, Belarus; an invited scientist at the Laboratory of Electromagnetism and Acoustic (LEMA), Swiss Federal Institute of Technology, Lausanne, Switzerland. His research areas are information theory, wireless communications and networks and, in particular, MIMO systems and security aspects of such systems, in which he has published extensively. He received a number of awards from the URSI, the IEEE, the Swiss, Belarus and former USSR governments, and the Soros Foundation.

Charalambos D. Charalambous received the B.S. degree in electrical engineering, the M.E. degree, and the Ph.D. degree from the Department of Electrical Engineering, Old Dominion University, Norfolk, VA, in 1987, 1988, and 1992, respectively. In 2003, he joined the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus, where he is currently Professor. He was an Associate Professor at the School of Information Technology and Engineering, University of Ottawa, Ottawa, ON, Canada, from 1999 to 2003. He has served on the faculty of the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, as a nontenure faculty member, from 1995 to 1999. From 1993 to 1995, he was a Postdoctoral Fellow at the Engineering Department, Idaho State University, Pocatello. His research group is interested in theoretical and technological developments concerning large scale distributed communication and control systems and networks in science and engineering. Dr. Charalambous is currently an associate editor for Systems and Control Letters, for Mathematics of Control, Signals, and Systems, and the Chair of the IFAC technical committee of Stochastic Systems. He served as associate editor for IEEE COMMUNICATIONS LETTERS, and for IEEE TRANSACTIONS ON AUTOMATIC CONTROL.