Optimal Signaling for Secure Communications Over Gaussian MIMO Wiretap Channels

Sergey Loyka and Charalambos D. Charalambous

***Abstract*— Optimal signaling over the Gaussian multiple-input multiple-output wire-tap channel is studied under the total transmit power constraint. A closed-form solution for an optimal transmit covariance matrix is obtained when the channel is strictly degraded. In combination with the rank-1 solution, this provides the complete characterization of the optimal covariance for the case of two transmit antennas. The cases of weak eavesdropper and high SNR are considered. It is shown that the optimal covariance does not converge to a scaled identity in the high-SNR regime. Necessary optimality conditions and a tight upper bound on the rank of an optimal covariance matrix are established for the general case, along with a lower bound to the secrecy capacity, which is tight in a number of scenarios.**

***Index Terms*— MIMO, wiretap channel, secrecy capacity, optimal signalling.**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) architecture has gained prominence in both academia and industry as a spectrally-efficient approach to wireless communications [1]. With wide deployment of wireless networks, security issues have recently gained additional importance, including information-theoretic approach at the physical layer [2]. The physical-layer security in MIMO systems has been recently under active investigation [3]–[10]. It was demonstrated that Gaussian signaling is optimal over the Gaussian MIMO wire-tap channels (MIMO-WTC) [6]–[10] and the optimal transmit covariance has been found for MISO systems [3], the 2-2-1 system [7], for the parallel channels (where independent signalling is optimal) [11], [12], all under the total power constraint, and in the general MIMO case under the transmit covariance matrix constraint [5]. The high-SNR regime (SNR → ∞) has been studied in [9]. The general case is still an open problem under the total power constraint, since the underlying optimization problem is not convex and explicit solutions are not known, except for some special cases. In fact, an optimal covariance is not known even when the channel is degraded (so that the respective optimization problem is convex), except for the special cases mentioned above.

The main contribution of this paper is a closed-form solution for the optimal covariance when the latter is of full rank under the total power constraint at finite SNR and the conditions for this to be the case in Theorem 1. The optimal covariance is shown to have some properties similar to those of the conventional water-filling, but with a few remarkable differences. In particular, the optimal covariance does not converge to a scaled identity in the high-SNR case and thus isotropic signaling is sub-optimal in this regime. Theorem 1, in combination with the rank-1 solution, provides the complete characterization of the optimal covariance for the case of two transmit antennas (for any channel, degraded or not). The cases of high-SNR and of weak eavesdropper are elaborated in Corollaries 1 and 2. An optimal covariance matrix for the general case (degraded or not) is characterized in Proposition 2, which shows that there is hidden convexity in the respective optimization problem, even when the channel is not degraded.

Proposition 3 gives a necessary condition of optimality for the general case, which is a transmission of the positive directions of the difference channel where the main channel is stronger than the eavesdropper one. This strengthens the earlier result in [13] (transmission on non-negative rather than positive directions). While the proof in [13] is rather straightforward and is based on a singular transformation (multiplication by a matrix that is singular when the covariance matrix is rank-deficient) of the KKT conditions, significantly more effort and a new approach are required to establish the stronger result. It avoids using a singular transformation (since some information about active signalling sub-space is irreversibly lost in the process) but relies on a novel property of positive semi-definite matrices (Lemma 2) and their block-partitioned representation to establish a property of dual variables from which the desired result follow. This result also allows one to establish a tighter bound on the rank of an optimal covariance matrix (Corollary 3) than those available in the literature for the general case.

A lower bound on the secrecy capacity in the general case is established in Proposition 4. While the original problem is non-convex so that all powerful tools of convex optimization [17] cannot be used, the lower bound is expressed via a convex problem and thus can be solved efficiently by a numerical algorithm. This bound is tight (achieved with equality) in a number of cases: when the SNR is low, or when the legitimate and eavesdropper channels have the same right singular vectors, or when the channel is degraded, thus providing an additional insight into optimal signalling.

Manuscript received April 28, 2015; accepted September 13, 2016. Date of publication October 19, 2016; date of current version November 18, 2016. This paper was presented at the 2012 IEEE International Symposium on Information Theory [15] and the 2013 IEEE International Symposium on Information Theory [16]. Communicated by Y. Liang, Associate Editor for Shannon Theory.

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 Electronics and Communication Engineering Department, University of Cyprus, 1678 Nicosia, Cyprus (e-mail: chadcha@ucy.ac.cy).

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.2016.2618857

An upper bound on the rank of an optimal covariance matrix is given in Corollary 3 for the general case: the rank is bounded by the dimensionality of a positive sub-space of the difference channel. This bound is stronger than those in [10] and [13] and can be further used to identify the cases for which an optimal covariance is of rank one (when the difference channel has just one strictly positive eigenvalue). Since the rank-1 structure of optimal covariance is known (unlike the sufficient and necessary conditions under which an optimal covariance is of rank-1, for which only limited knowledge is available), this extends the earlier results in [3], [7], [13], and [14] and provides not only the rank but also an optimal covariance itself in those cases.

II. Gaussian MIMO Wire-Tap Channel Model

Let us consider the standard Gaussian MIMO-WTC model,

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{x} + \boldsymbol{\xi}_1, \quad \mathbf{y}_2 = \mathbf{H}_2\mathbf{x} + \boldsymbol{\xi}_2 \tag{1}$$

where $\mathbf{x} = [x_1, x_2, ...x_m]^T \in \mathcal{C}^{m,1}$ is the transmitted complex-valued signal vector of dimension $m \times 1$, "T" denotes transposition, $\mathbf{y}_{1(2)} \in \mathcal{C}^{n,1}$ are the received vectors at the receiver (eavesdropper), $\boldsymbol{\xi}_{1(2)}$ is the circularly-symmetric additive white Gaussian noise at the receiver (eavesdropper) normalized to unit variance in each dimension, $\mathbf{H}_{1(2)} \in \mathcal{C}^{n_{1(2)},m}$ is the $n_{1(2)} \times m$ matrix of the complex channel gains between each Tx and each receive (eavesdropper) antenna, $n_{1(2)}$ and m are the numbers of Rx (eavesdropper) and Tx antennas respectively. The channels $\mathbf{H}_{1(2)}$ are assumed to be quasistatic (i.e., constant for a sufficiently long period of time so that the standard random coding arguments can be invoked within each coherence block) and frequency-flat, with full channel state information (CSI) at the Rx and Tx ends.

For a given transmit covariance matrix $\mathbf{R} = E\left\{\mathbf{x}\mathbf{x}^+\right\}$, where $E\{\cdot\}$ is statistical expectation, the maximum achievable secure rate between the Tx and Rx (so that the leakage rate between the Tx and eavesdropper converges to zero) is [5]–[10]

$$C(\mathbf{R}) = \ln \frac{|\mathbf{I} + \mathbf{W}_1\mathbf{R}|}{|\mathbf{I} + \mathbf{W}_2\mathbf{R}|} = C_1(\mathbf{R}) - C_2(\mathbf{R}) \tag{2}$$

where negative $C(\mathbf{R})$ is interpreted as zero rate, $\mathbf{W}_i = \mathbf{H}_i^+\mathbf{H}_i$, $()^+$ means Hermitian conjugation, $C_i(\mathbf{R}) = \ln|\mathbf{I} + \mathbf{W}_i\mathbf{R}|$. The secrecy capacity subject to the total Tx power constraint is

$$C_s = \max_{\mathbf{R} \geq 0} C(\mathbf{R}) \text{ s.t. } tr\mathbf{R} \leq P_T \tag{3}$$

where P_T is the total transmit power (also the SNR since the noise is normalized). It is well-known that the problem in (3) is not convex in general and explicit solutions for the optimal Tx covariance are not known except for some special cases (e.g. low-SNR, MISO or parallel channels). It was conjectured in [10] that an optimal transmission in (3) is on the directions where the main channel is stronger than the eavesdropper one (i.e. on the positive directions of the difference channel $\mathbf{W}_1 - \mathbf{W}_2$). A similar conclusion, albeit in a different (indirect) form, has been obtained in [9] using the degraded channel approach.

Theorem 1 below gives an explicit, closed-form solution for the optimal full-rank covariance in (3) at finite SNR. A number of additional insights and properties follow.

III. Closed-Form Solutions

In this section, we consider the problem in (3) and obtain its closed-form solutions. The following theorem establishes the optimal covariance $\mathbf{R}^*$ for the strictly-degraded channel, $\mathbf{W}_1 > \mathbf{W}_2$, where $\mathbf{A} > \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive definite.

Theorem 1: Let $\mathbf{W}_1 > \mathbf{W}_2$ *and* $P_T > P_{T0}$*, where* P_{T0} *is a threshold power given by* (8)*. Then,* $\mathbf{R}^*$ *is of full rank and is given by:*

$$\mathbf{R}^* = \mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^+ - \mathbf{W}_1^{-1} \tag{4}$$

where the columns of the unitary matrix $\mathbf{U}$ *are the eigenvectors of* $\mathbf{Z} = \mathbf{W}_2 + \mathbf{W}_2(\mathbf{W}_1 - \mathbf{W}_2)^{-1}\mathbf{W}_2$, $\boldsymbol{\Lambda}_1 = diag\{\lambda_{1i}\} > \mathbf{0}$ *is a diagonal positive-definite matrix,*

$$\lambda_{1i} = \frac{2}{\lambda}\left(\sqrt{1 + \frac{4\mu_i}{\lambda}} + 1\right)^{-1} \tag{5}$$

and $\mu_i \geq 0$ *are the eigenvalues of* $\mathbf{Z}$*;* $\lambda > 0$ *is found from the total power constraint* $tr\mathbf{R}^* = P_T$ *as a unique solution of the following equation:*

$$\frac{2}{\lambda}\sum_i\left(\sqrt{1 + \frac{4\mu_i}{\lambda}} + 1\right)^{-1} = P_T + tr\mathbf{W}_1^{-1} \tag{6}$$

The corresponding secrecy capacity is

$$\begin{aligned} C_s &= \ln \frac{|\mathbf{W}_1||\boldsymbol{\Lambda}_1|}{|\mathbf{I} - \mathbf{W}_2(\mathbf{W}_1^{-1} - \mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^+)|} \\ &= \ln \frac{|\mathbf{W}_1|}{|\mathbf{W}_2|} + \ln \frac{|\boldsymbol{\Lambda}_1|}{|\boldsymbol{\Lambda}_2|} \end{aligned} \tag{7}$$

where $\boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}_1 + diag\{\mu_i^{-1}\}$ *and 2nd equality holds when* $\mathbf{W}_2 > 0$*.* P_{T0} *can be expressed as follows:*

$$\begin{aligned} P_{T0} = &\frac{2(\mu_1 + \lambda_{min})}{\lambda_{min}^2}\sum_i\left(\sqrt{1 + \frac{4\mu_i(\mu_1 + \lambda_{min})}{\lambda_{min}^2}} + 1\right)^{-1} \\ &- tr\mathbf{W}_1^{-1} \end{aligned} \tag{8}$$

where λ_{min} *is the minimum eigenvalue of* $\mathbf{W}_1$ *and* μ_1 *is the maximum eigenvalue of* $\mathbf{Z}$*.*

Proof: See Appendix. □

It should be pointed out that Theorem 1 gives an exact (not approximate) optimal covariance at finite SNR ($P_T \to \infty$ is not required) since P_{T0} is a finite constant that depends only on $\mathbf{W}_1$ and $\mathbf{W}_2$ and this constant is small in some cases: it follows from (8) that $P_{T0} \to 0$ if $\lambda_{min} \to \infty$, i.e. P_{T0} is small if λ_{min} is large. In particular, P_{T0} can be upper bounded as

$$P_{T0} \leq \frac{m\mu_1}{\lambda_{min}^2} + \frac{m-1}{\lambda_{min}} \tag{9}$$

and if $\lambda_{min} \gg \mu_1$, then

$$P_{T0} \approx \frac{m}{\lambda_{min}} - tr\mathbf{W}_1^{-1} \leq \frac{m-1}{\lambda_{min}} \leq 1 \tag{10}$$

where the last inequality holds if $\lambda_{min} \geq m - 1$. Fig. 1 illustrates this case. On the other hand, when $\mathbf{W}_1 - \mathbf{W}_2$ approaches a singular matrix, it follows that $P_{T0} \to \infty$, so that P_{T0} is large iff $\mathbf{W}_1 - \mathbf{W}_2$ is close to singular.

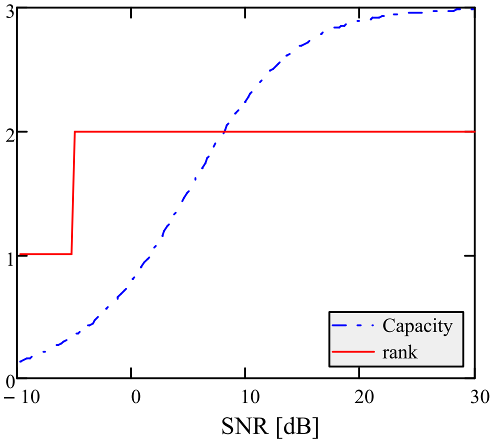

Fig. 1. Secrecy capacity and the rank of $\mathbf{R}^*$ vs. SNR [dB] for the channel in (11). The transition to full-rank covariance takes place at about -6 dB.

Theorem 1, in combination with rank-1 solution in (28), provides the complete solution for the optimal covariance in the $m = 2$ case: if the channel is not strictly degraded or if the SNR is not above the threshold, the rank-1 solution in (28) applies; otherwise, Theorem 1 applies. Fig. 1 illustrates this for the following channel:

$$\mathbf{W}_1 = \begin{bmatrix} 1.5 \ 0.5 \\ 0.5 \ 1.5 \end{bmatrix}, \ \mathbf{W}_2 = \begin{bmatrix} 0.35 \ 0.15 \\ 0.15 \ 0.35 \end{bmatrix} \qquad (11)$$

Note that the transition to full-rank covariance takes place at low SNR of about -6 dB, i.e. P_{T0} is not high at all in this case.

We further observe that 1st term in (7) $C_\infty = \ln \frac{|\mathbf{W}_1|}{|\mathbf{W}_2|}$ is SNR-independent and the 2nd one $\Delta C = \ln \frac{|\mathbf{\Lambda}_1|}{|\mathbf{\Lambda}_2|} < 0$ monotonically increases with the SNR. Furthermore, $C_s \to C_\infty$, $\Delta C \to 0$ as $P_T \to \infty$, in agreement with [9, Th. 2]. This is also clear from Fig. 1.

Note also that the second term in (4) de-emphasizes weak eigenmodes of $\mathbf{W}_1$. Since λ is monotonically decreasing as P_T increases (this follows from (6)), λ_{1i} monotonically increases with P_T, and approaches $\lambda_{1i} \approx 1/\sqrt{\mu_i \lambda}$ at sufficiently high SNR, which is in contrast with the conventional water-filling (WF), where the uniform power allocation is optimal at high SNR. Furthermore, it follows from (5) that λ_{1i} decreases with μ_i, i.e. stronger eigenmodes of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1} = \mathbf{Z}^{-1}$ (which correspond to larger eigenmodes of $\mathbf{W}_1$ and weaker ones of $\mathbf{W}_2$) receive larger power allocation, which follows the same tendency as the conventional WF. It further follows from (4) that when $\mathbf{W}_1$ and $\mathbf{W}_2$ have the same eigenvectors, $\mathbf{R}^*$ also has the same eigenvectors, i.e. the optimal signaling is on the eigenvectors of $\mathbf{W}_{1(2)}$. While the necessary condition for full-rank $\mathbf{R}^*$ ($\mathbf{W}_1 > \mathbf{W}_2$) has been obtained before in [10], no solution was found for $\mathbf{R}^*$, which is given in Theorem 1 here.

The case of singular $\mathbf{W}_1$ can also be included by observing that, under certain conditions, $\mathbf{R}^*$ puts no power on the null space of $\mathbf{W}_1$ so that all matrices can be projected, without loss of generality, on the positive eigenspace of $\mathbf{W}_1$ and Theorem 1 will apply. The following Proposition makes this precise.

Proposition 1: Consider the problem in (3) *when* $\mathcal{N}(\mathbf{W}_1) \in \mathcal{N}(\mathbf{W}_2)$, *where* $\mathcal{N}(\mathbf{W}) = \{\mathbf{x} : \mathbf{W}\mathbf{x} = \mathbf{0}\}$ *is the null space of matrix* $\mathbf{W}$ *[19], and assume that*

$$\mathbf{x}^+(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{x} > 0 \ \forall \mathbf{x} \in \mathcal{N}_\perp, \ \mathbf{x} \neq \mathbf{0}, \qquad (12)$$

where $\mathcal{N}_\perp$ *is orthogonal complement of* $\mathcal{N}(\mathbf{W}_1)$, *i.e.* $\mathbf{W}_1 - \mathbf{W}_2$ *is positive definite on* $\mathcal{N}_\perp$. *When the SNR exceeds a threshold (as in Theorem 1), the optimal covariance in* (3) *is*

$$\mathbf{R}^* = \mathbf{U}_\perp \tilde{\mathbf{R}}^* \mathbf{U}_\perp^+ \qquad (13)$$

where $\tilde{\mathbf{R}}^*$ *is the optimal covariance of Theorem 1 when applied to the projected matrices* $\tilde{\mathbf{W}}_i = \mathbf{U}_\perp^+ \mathbf{W}_i \mathbf{U}_\perp$ *and the columns of semi-unitary matrix* $\mathbf{U}_\perp$ *form an orthonormal basis of* $\mathcal{N}_\perp$. *Furthermore,* $rank(\mathbf{R}^*) = rank(\mathbf{W}_1)$.

Proof: Observe that $\mathbf{W}_i\mathbf{x} = \mathbf{W}_i\mathbf{x}_\perp$, where $\mathbf{x}_\perp = \mathbf{U}_\perp\mathbf{U}_\perp^+\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ on $\mathcal{N}_\perp$, so that

$$\begin{aligned} |\mathbf{I} + \mathbf{W}_i\mathbf{R}| &= |\mathbf{I} + \mathbf{W}_i\mathbf{U}_\perp\mathbf{U}_\perp^+\mathbf{R}\mathbf{U}_\perp\mathbf{U}_\perp^+| \\ &= |\mathbf{I} + \mathbf{U}_\perp^+\mathbf{W}_i\mathbf{U}_\perp\mathbf{U}_\perp^+\mathbf{R}\mathbf{U}_\perp| \qquad (14) \end{aligned}$$

and $tr(\mathbf{U}_\perp^+\mathbf{R}\mathbf{U}_\perp) \leq tr(\mathbf{R})$ so that one can use the projected matrices $\tilde{\mathbf{R}} = \mathbf{U}_\perp^+\mathbf{R}\mathbf{U}_\perp, \tilde{\mathbf{W}}_i = \mathbf{U}_\perp^+\mathbf{W}_i\mathbf{U}_\perp$ in Theorem 1 to obtain the desired solution. (12) insures that $\tilde{\mathbf{W}}_1 - \tilde{\mathbf{W}}_2 > 0$ so that Theorem 1 applies. □

With Proposition 1 in mind, the conditions of Theorem 1 are both sufficient and necessary (except for the power threshold P_{T0} which may be less than in (8)) for an optimal covariance to be of full-rank.

It is instructive to consider the case when the required channel is much stronger than the eavesdropper one, $\mathbf{W}_1 \gg \mathbf{W}_2$, meaning that all eigenvalues of $\mathbf{W}_1$ are much larger than those of $\mathbf{W}_2$.

Corollary 1: Consider the MIMO-WTC in (1) under the conditions of Theorem 1 and when the eavesdropper channel is much weaker than the required one,

$$\lambda_i(\mathbf{W}_2) \ll m(P_T + tr\mathbf{W}_1^{-1})^{-1}/4 \qquad (15)$$

where $\lambda_i(\mathbf{W}_2)$ *is* i*-th eigenvalue of* $\mathbf{W}_2$, *e.g. when* $\mathbf{W}_2 \to \mathbf{0}$ *and fixed* $\mathbf{W}_1$. *Then the optimal covariance in (4) becomes*

$$\mathbf{R}^* \approx \mathbf{U}_1(\lambda^{-1}\mathbf{I} - \mathbf{D}_1^{-1})\mathbf{U}_1^+ - \lambda^{-2}\mathbf{W}_2 \qquad (16)$$

where $\mathbf{W}_1 = \mathbf{U}_1\mathbf{D}_1\mathbf{U}_1^+$ *is the eigenvalue decomposition, so that the columns of* $\mathbf{U}_1$ *are the eigenvectors, and the diagonal entries of* $\mathbf{D}_1$ *are the eigenvalues.*

Proof: See Appendix. □

An interpretation of (16) is immediate: the first term is the standard water-filling on the eigenmodes of $\mathbf{W}_1$ (which is the capacity-achieving strategy for the regular MIMO channel) and the second term is a correction due to the secrecy requirement: those modes that spill over into the eavesdropper channel get less power to accommodate the secrecy requirement.

Let us know consider the high-SNR regime.

Corollary 2: When $\mathbf{W}_2 > 0$, *the optimal covariance* $\mathbf{R}^*$ *in (4) in the high-SNR regime*

$$P_T \gg \mu_m^{-1/2} \sum_i \mu_i^{-1/2} \qquad (17)$$

(e.g. when $P_T \to \infty$*), where* $\mu_m = \min_i \mu_i$, *simplifies to*

$$\mathbf{R}^* \approx \mathbf{U} diag\{d_i\}\mathbf{U}^+, \quad d_i = \frac{P_T \mu_i^{-1/2}}{\sum_i \mu_i^{-1/2}} \qquad (18)$$

The corresponding secrecy capacity is

$$C_s \approx \ln \frac{|\mathbf{W}_1|}{|\mathbf{W}_2|} - \frac{1}{P_T}\left(\sum\nolimits_i \frac{1}{\sqrt{\mu_i}}\right)^2 \tag{19}$$

where we have neglected 2nd and higher order effects in $1/P_T$.

Proof: Follows from Theorem 1 along the same lines as that of Corollary 1. □

Note that the optimal signaling is on the eigenmodes of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1}$ with the optimal power allocation given by $\{d_i\}$. This somewhat resembles the conventional water-filling, but also has a remarkable difference: unlike the conventional WF, the secure WF in (18) does not converge to the uniform one in the high-SNR regime.[1] However, strong eigenmodes of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1}$ (which corresponds to weak modes of $\mathbf{W}_2$ and strong ones of $\mathbf{W}_1$) do get more power, albeit in a form different from that of the conventional WF.

While Theorem 1 gives a closed-form full-rank optimal covariance for the strictly degraded channel, the general case remains an open problem. The proposition below provides a characterization of an optimal covariance for the general case.

Proposition 2: Consider the general Gaussian MIMO-WTC (not necessarily degraded). Let the columns of semi-unitary matrix $\mathbf{U}_a$ span the same subspace as the columns of optimal covariance $\mathbf{R}^*$ in (3)*: $span\{\mathbf{U}_a\} = span\{\mathbf{R}^*\}$. Then, the optimal covariance can be expressed in the following form:*

$$\mathbf{R}^* = \mathbf{U}_a \mathbf{R}' \mathbf{U}_a^+ \tag{20}$$

where $\mathbf{R}'$ is given by Theorem 1 with the substitutions $\mathbf{W}_i \rightarrow \tilde{\mathbf{W}}_i = \mathbf{U}_a^+ \mathbf{W}_i \mathbf{U}_a$ (i.e. applied to the channels projected on $span\{\mathbf{U}_a\}$), and $\tilde{\mathbf{W}}_1 > \tilde{\mathbf{W}}_2$.

Proof: See Appendix. □

Remark 1: Proposition 2 gives a closed-from solution for the general (non-degraded) case provided that the active subspace (i.e. the subspace spanned by the columns or active eigenvectors of $\mathbf{R}^*$) is already known. Note that the knowledge of eigenvectors of $\mathbf{R}^*$ is not required, but only the subspace they span. This in fact splits the entire problem $\mathcal{P}$ into two sub-problems $\mathcal{P}_1$ and $\mathcal{P}_2$:

$$\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2 \tag{21}$$

where $\mathcal{P}_1$ is a non-convex problem of finding the active sub-space (or the active eigenvectors) and $\mathcal{P}_2$ is the convex problem of finding the optimal covariance based on the found active subspace, hence revealing the hidden convexity in the original non-convex problem $\mathcal{P}$. While $\mathcal{P}_2$ is always convex, $\mathcal{P}_1$ and thus $\mathcal{P}$ become convex when the channel is degraded.

IV. Necessary Optimality Conditions and Properties

In this section, we establish the necessary optimality conditions for the problem in (3) and, based on these conditions, some properties of the optimal solutions when the latter are rank-deficient. In particular, we establish an upper bound on the rank of optimal covariance matrix which is tighter than the known bounds. In some cases, this bound results in an explicit closed-form solution for the optimal covariance.

The following Proposition gives a necessary condition of the optimality in (3).

Proposition 3: Let $\mathbf{R}^*$ be an optimal covariance in (3) *and let $\mathbf{U}_{r+}$ be a semi-unitary matrix whose columns are the active eigenvectors $\{\mathbf{u}_{i+}\}$ (i.e. corresponding to positive eigenvalues) of $\mathbf{R}^*$. Then, the following holds:*

$$\mathbf{U}_{r+}^+(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{U}_{r+} > \mathbf{0} \tag{22}$$

so that

$$\mathbf{x}^+(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{x} > \mathbf{0} \ \forall \mathbf{x} \in span\{\mathbf{u}_{i+}\} \tag{23}$$

i.e. a necessary condition for an optimal signaling strategy in (3) *is to transit over the positive directions of $\mathbf{W}_1 - \mathbf{W}_2$ (where the legitimate channel is stronger than the eavesdropper).*[2]

Proof: See the Appendix. □

It was demonstrated in [10] that $rank(\mathbf{R}^*) < m$ unless $\mathbf{W}_1 > \mathbf{W}_2$, i.e. an optimal transmission is of low-rank over a non-degraded channel. The Corollary below gives more precise characterization.

Corollary 3: Let $\mathbf{W}_1 - \mathbf{W}_2 = \mathbf{W}_+ + \mathbf{W}_-$, where $\mathbf{W}_{+(-)}$ collects positive (negative and zero) eigenmodes of $\mathbf{W}_1 - \mathbf{W}_2$ (found from its eigenvalue decomposition). Then,

$$rank(\mathbf{R}^*) \leq rank(\mathbf{W}_+) \leq m, \tag{24}$$

i.e. the rank of an optimal covariance $\mathbf{R}^*$ does not exceed the number of positive eigenvalues of $\mathbf{W}_1 - \mathbf{W}_2$ (the rank of $\mathbf{W}_+$).

Proof: We need the following technical Lemma, which is a direct consequence of [22, Corollary 4.5.11]:

Lemma 1: Let $\mathbf{A}$ be Hermitian and $r_+(\mathbf{A})$ be its number of positive eigenvalues. Then $r_+(\mathbf{S}^+\mathbf{A}\mathbf{S}) \leq r_+(\mathbf{A})$, where $\mathbf{S}$ is any matrix of appropriate size.

Lemma 1 says that applying the transformation $\mathbf{S}^+\mathbf{A}\mathbf{S}$ to $\mathbf{A}$ cannot increase the number of its positive eigenvalues (since $\mathbf{S}$ can be singular; this number stays the same if $\mathbf{S}$ is full rank). Using this Lemma with $\mathbf{S} = \mathbf{R}^{*1/2}$ and $\mathbf{A} = \mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M}$, one obtains:

$$\begin{aligned} r_+(\mathbf{R}^*) &= r_+(\mathbf{R}^{*1/2}(\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M})\mathbf{R}^{*1/2}) && (25)\\ &= r_+(\mathbf{R}^{*1/2}(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{R}^{*1/2}) && (26)\\ &\leq r_+(\mathbf{W}_1 - \mathbf{W}_2) = rank(\mathbf{W}_+) && (27) \end{aligned}$$

where 1st equality follows from the fact that $\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M} > \mathbf{0}$ (which has been established in the proof of Proposition 3), 2nd equality follows from $\mathbf{M}\mathbf{R}^* = 0$, and the inequality follows from Lemma 1. □

[1] The sub-optimality of the isotropic signalling suggested in [9, Th. 2] is hiding in the $o(1)$ term there. 2nd term of Eq. (19) above refines that $o(1)$ term.

[2] After the conference version of this paper has been submitted, we were informed that a weaker result ($\geq$ instead of $>$) was established in [13]. The proof in [13] is based on a singular transformation (multiplication by a singular matrix when $\mathbf{R}$ is singular), so that some information about the active signalling sub-space is lost and strict inequality cannot be established. On the other hand, we avoid using such transformation and base our proof on some novel properties of positive semi-definite matrices (see Lemma 2) and their block-partitioned representation so that the active signaling sub-space can be characterized more precisely and a tighter upper bound on the rank of an optimal covariance can be established.

Note that the rank bound in Corollary 3 is stronger than the corresponding bound in [13], $rank(\mathbf{R}^*) \leq rank(\mathbf{W}_1 - \mathbf{W}_2)$, especially when the difference matrix $\mathbf{W}_1 - \mathbf{W}_2$ has many negative eigenvalues (e.g. when the eigenvalues of $\mathbf{W}_1 - \mathbf{W}_2$ are $\{1, -1, .., -1\}$, the bound in [13] is trivial: $rank(\mathbf{R}^*) \leq m$, while our bound gives the true rank: $rank(\mathbf{R}^*) = 1$).

When $rank(\mathbf{W}_+) = 1$, the optimal covariance $\mathbf{R}^*$ is of rank-1 from Corollary 3 and hence the capacity and the covariance follow from (3)[3]:

$$C_s = \ln \lambda_1, \ \mathbf{R}^* = P_T \mathbf{u}_1 \mathbf{u}_1^+ \tag{28}$$

where λ_1, $\mathbf{u}_1$ are the largest eigenvalue and corresponding eigenvector of $(\mathbf{I} + P_T\mathbf{W}_2)^{-1}(\mathbf{I} + P_T\mathbf{W}_1)$ or, equivalently, the largest generalized eigenvalue and corresponding eigenvector of $(\mathbf{I} + P_T\mathbf{W}_1, \mathbf{I} + P_T\mathbf{W}_2)$, so that transmit beamforming on $\mathbf{u}_1$ is the optimal strategy. Note that this result is more general than those in [3] and [7] as the latter two apply to a single antenna channel (either at the receiver or eavesdropper) while the result above holds for any number of antennas at any end. Furthermore, the signaling in (28) is also optimal for any $rank(\mathbf{W}_+) \geq 1$ at sufficiently small SNR, where λ_1, $\mathbf{u}_1$ become the largest eigenvalue and corresponding eigenvector of the difference channel $\mathbf{W}_1 - \mathbf{W}_2$.

The following Proposition establishes a lower bound to the non-convex problem in (3) via a convex optimization problem (for any channel, degraded or not).

Proposition 4: The secrecy capacity can be lower bounded as follows:

$$C_s \geq \max_{\mathbf{R} \geq 0} C_+(\mathbf{R}) \text{ s.t. } tr\mathbf{R} \leq P_T, \tag{29}$$

where

$$C_+(\mathbf{R}) = \ln \frac{|\mathbf{I} + \mathbf{W}_{1+}\mathbf{R}|}{|\mathbf{I} + \mathbf{W}_{2+}\mathbf{R}|} \tag{30}$$

and $\mathbf{W}_{i+} = \mathbf{P}_+\mathbf{W}_i\mathbf{P}_+$, $\mathbf{P}_+ = \mathbf{U}_+\mathbf{U}_+^+$ *is the projection matrix on the positive eigenspace of* $\mathbf{W}_1 - \mathbf{W}_2$, $\mathbf{U}_+$ *is a semi-unitary matrix whose columns are the eigenvectors of* $\mathbf{W}_1 - \mathbf{W}_2$ *corresponding to its positive eigenvalues:* $\mathbf{W}_+ = \mathbf{U}_+\mathbf{D}_+\mathbf{U}_+^+$, *and* $\mathbf{D}_+$ *is the diagonal matrix of the positive eigenvalues;* $C_+(\mathbf{R})$ *is a non-negative, concave and non-decreasing function of* $\mathbf{R}$ *or strictly positive, concave and increasing when the active eigenmodes of* $\mathbf{R}$ *are in the span of the active eigenmodes of* $\mathbf{W}_+$. *The lower bound is tight (achieved with equality) when the channel is degraded or when* $\mathbf{W}_1$ *and* $\mathbf{W}_2$ *have the same eigenvectors, or in the low-SNR regime.*

Proof: see Appendix. □

The problem in (29) has further significance: while the problem $C_s = \max_{\mathbf{R} \geq 0} C(\mathbf{R})$ is not convex when the channel is not degraded, so that powerful tools of convex optimization [17] cannot be used, the problem $\max_{\mathbf{R} \geq 0} C_+(\mathbf{R})$ is convex for any channel (degraded or not), to which all machinery of convex optimization can be applied and a lower bound (achievable rate) to the secrecy capacity can be evaluated using any standard convex solver.

V. Conclusion

Optimal signalling over the Gaussian MIMO wire-tap channel has been studied under the total power constraint. A closed-form solution is given for the optimal transmit covariance matrix when the channel is strictly degraded. While the optimal signalling has some similarities to the conventional water-filling, it also reveals a number of differences: the optimal signalling does not converge to isotropic at high SNR. The weak eavesdropper and high-SNR regimes are considered, and a tighter upper bound on the rank of the optimal covariance matrix is given for the general case, along with the lower bound to the secrecy capacity, which is tight in a number of cases. While the general case is still an open problem (even when the channel is degraded), a characterization of an optimal covariance based on the active signaling subspace is given, which reveals hidden convexity in the underlying optimization problem.

Appendix

A. Proof of Theorem 1

Using the Lagrange multiplier technique [17], [18], the optimization problem in (3) has the following Lagrangian:

$$L = -\ln|\mathbf{I} + \mathbf{W}_1\mathbf{R}| + \ln|\mathbf{I} + \mathbf{W}_2\mathbf{R}| + \lambda(tr\mathbf{R} - P_T) - tr(\mathbf{M}\mathbf{R}) \tag{31}$$

where $\lambda \geq 0$ is a Lagrange multiplier responsible for the power constraint $tr\mathbf{R} \leq P_T$ and $\mathbf{M} \geq \mathbf{0}$ is a (positive semi-definite) matrix Lagrange multiplier responsible for the constraint $\mathbf{R} \geq \mathbf{0}$. The associated KKT conditions (see e.g. [17]) can be expressed as:

$$\lambda(\mathbf{I} + \mathbf{W}_1\mathbf{R})(\mathbf{I} + \mathbf{R}\mathbf{W}_2) = \mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M} \tag{32}$$

$$\mathbf{M}\mathbf{R} = \mathbf{0}, \ \lambda(tr\mathbf{R} - P_T) = 0, \tag{33}$$

$$\mathbf{R} \geq \mathbf{0}, \ \mathbf{M} \geq \mathbf{0}, \ \lambda \geq 0, \ tr\mathbf{R} \leq P_T \tag{34}$$

where (32) is obtained from $\partial L/\partial \mathbf{R} = \mathbf{0}$,

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{R}} &= (\mathbf{I} + \mathbf{W}_2\mathbf{R})^{-1}\mathbf{W}_2 - (\mathbf{I} + \mathbf{W}_1\mathbf{R})^{-1}\mathbf{W}_1 + \lambda\mathbf{I} - \mathbf{M} \\ &= \mathbf{0}\end{aligned} \tag{35}$$

and the two equalities in (33) are the complementary slackness conditions while (34) are the primal and dual feasibility conditions.

Note that the (affine) constraints $tr\mathbf{R} \leq P_T$, $\mathbf{R} \geq \mathbf{0}$ clearly satisfy the Slater condition [17], [18]. It also follows from Proposition 4 that $C(\mathbf{R})$ is concave when $\mathbf{W}_1 > \mathbf{W}_2$ (no need for projection) so that the problem in (3) is convex and thus the KKT conditions are sufficient for global optimality when the channel is strictly degraded.

Let us consider first the case of $\mathbf{W}_2 > 0$ and extend it to the singular case later. Assuming $\mathbf{R} > \mathbf{0}$ and using $\mathbf{M} = \mathbf{0}$ (which follows from $\mathbf{M}\mathbf{R} = \mathbf{0}$), one obtains from (35),

$$\mathbf{R}_1^{-1} - \mathbf{R}_2^{-1} = \lambda\mathbf{I} \tag{36}$$

where $\mathbf{R}_i = \mathbf{W}_i^{-1} + \mathbf{R}$, $i = 1, 2$. Let $\mathbf{R}_1 = \mathbf{U}\Lambda_1\mathbf{U}^+$ be the eigenvalue decomposition, where the columns of unitary matrix $\mathbf{U}$ are the eigenvectors, and $\mathbf{\Lambda}_1 > \mathbf{0}$ is a diagonal matrix of the corresponding eigenvalues. Using this in (36),

[3] This result has been obtained before, albeit in a different way, in [13]. Note however, that our result here is stronger: it does not require $\mathbf{W}_1 - \mathbf{W}_2$ to be non-singular while [13] does, so that the latter result does not apply when the eigenvalues of $\mathbf{W}_1 - \mathbf{W}_2$ are e.g. $\{1, 0, .., 0, -1, .., -1\}$ while our result does apply to such scenario.

one obtains $\mathbf{\Lambda}_1^{-1} - (\mathbf{U}^+\mathbf{R}_2\mathbf{U})^{-1} = \lambda\mathbf{I}$ and therefore $\mathbf{U}^+\mathbf{R}_2\mathbf{U} = \Lambda_2$ is diagonal, so that $\mathbf{R}_2 = \mathbf{U}\Lambda_2\mathbf{U}^+$ is the eigenvalue decomposition of $\mathbf{R}_2$, from which it follows that $\mathbf{R}_1$ and $\mathbf{R}_2$ have the same eigenvectors. Using this in (36) one obtains

$$\mathbf{\Lambda}_1 = (\lambda\mathbf{I} + \mathbf{\Lambda}_2^{-1})^{-1} \tag{37}$$

Furthermore,

$$\mathbf{R}_2 - \mathbf{R}_1 = \mathbf{W}_2^{-1} - \mathbf{W}_1^{-1} = \mathbf{U}(\mathbf{\Lambda}_2 - \mathbf{\Lambda}_1)\mathbf{U}^+ \tag{38}$$

so that the columns of $\mathbf{U}$ are also the eigenvectors of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1} = \mathbf{Z}^{-1}$ and the diagonal entries of $\mathbf{\Lambda}_2 - \mathbf{\Lambda}_1 = diag\{\mu_i^{-1}\}$ are its eigenvalues. Combining the latter with (37), one obtains after some manipulations (5). (4) follows from $\mathbf{R}_1 = \mathbf{W}_1^{-1} + \mathbf{R}$ and $\mathbf{R}_1 = \mathbf{U}\Lambda_1\mathbf{U}^+$. It is straightforward to see that $\lambda > 0$ (otherwise $\mathbf{W}_1 \le \mathbf{W}_2$), so that transmission with the full power is optimal and (6) follows from the power constraint $tr\mathbf{R} = P_T$. For (4) to be a valid solution, we need $\mathbf{U}\Lambda_1\mathbf{U}^+ > \mathbf{W}_1^{-1}$. This is insured by observing that the left-hand side of (6) is monotonically decreasing in λ, so that the latter is monotonically decreasing as P_T increases and, from (5), λ_{1i} also monotonically increases. Therefore, for sufficiently large P_T, $P_T > P_{T0}$ for some finite P_{T0}, the minimum eigenvalue of $\mathbf{\Lambda}_1$ exceeds the maximum one of $\mathbf{W}_1^{-1}$ and thus the condition $\mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^+ > \mathbf{W}_1^{-1}$ follows. Therefore, (4)-(6) solve the KKT conditions and thus achieve the global optimum. It can be further seen that the solution is unique.

It can be seen that (6) is monotonically decreasing in λ over the interval $(0, \infty)$ when $\lambda \in (0, \infty)$ so that a solution exists and unique for any P_T.

The condition $\mathbf{W}_2 > \mathbf{0}$ can be further removed via the standard continuity argument [19]: use $\mathbf{W}_{2\epsilon} = \mathbf{W}_2 + \epsilon\mathbf{I} > \mathbf{0}$, $\epsilon > 0$, instead of $\mathbf{W}_2$ in Theorem 1 and then take $\epsilon \to 0$. Alternatively, one may observe that $\mathbf{W}$ and $\mathbf{W}^{-1}$ have the same eigenvectors and inverse eigenvalues and use the matrix inversion lemma [19], [22] to obtain:

$$(\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1})^{-1} = \mathbf{W}_2 + \mathbf{W}_2(\mathbf{W}_1 - \mathbf{W}_2)^{-1}\mathbf{W}_2 = \mathbf{Z} \tag{39}$$

Note that $\mathbf{Z}$ is well-defined even for singular $\mathbf{W}_2$ (since $\mathbf{W}_1 > \mathbf{W}_2$), its eigenvectors are those of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1}$ and $\mu_i = \lambda_i(\mathbf{Z})$ so that Theorem 1 applies. Furthermore, $\lambda_i(\mathbf{Z}) = \mathbf{0}$ iff $\lambda_i(\mathbf{W}_2) = 0$, the corresponding eigenvectors are those of $\mathbf{W}_2$ and $\mu_i = 0$ implies $\lambda_{1i} = 1/\lambda$. The equalities in (7) follow by observing that

$$|\mathbf{I} + \mathbf{R}^*\mathbf{W}_1| = |\mathbf{W}_1\mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^+| = |\mathbf{W}_1||\mathbf{\Lambda}_1| \tag{40}$$

and

$$\begin{aligned}|\mathbf{I} + \mathbf{R}^*\mathbf{W}_2| &= |\mathbf{I} - \mathbf{W}_2(\mathbf{W}_1^{-1} - \mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^+)| \\ &= |\mathbf{W}_2||\mathbf{\Lambda}_2| \end{aligned} \tag{41}$$

where 2nd equality holds when $\mathbf{W}_2 > 0$ (1st one allows for singular $\mathbf{W}_2$). Note that $\mathbf{W}_2^{1/2}(\mathbf{W}_1^{-1} - \mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^+)\mathbf{W}_2^{1/2} < \mathbf{I}$ (which follows from $\mathbf{W}_2^{1/2}\mathbf{W}_1^{-1}\mathbf{W}_2^{1/2} < \mathbf{I}$ which in turn is implied by $\mathbf{W}_1 > \mathbf{W}_2$) so that 2nd determinant is indeed strictly positive.

To show (8), observe that $\mathbf{R}^* > 0$. Using (4), this requires $\mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^+ > \mathbf{W}_1^{-1}$, which is insured by $\lambda_{1min}\lambda_{min} > 1$, where $\lambda_{1min} = \min_i\{\lambda_{1i}\}$ and λ_{min} is the minimum eigenvalue of $\mathbf{W}_1$ (this follows from the fact that $\mathbf{W}_1 > \mathbf{W}_2$ is implied by $\lambda_{min}(\mathbf{W}_1) > \lambda_{max}(\mathbf{W}_2)$). Therefore, the threshold power P_{T0} can be found from the boundary condition $\lambda_{1min}(P_{T0}) = 1/\lambda_{min}$, which, after some manipulations, can be expressed as

$$\sqrt{\lambda^2 + 4\mu_1\lambda} = 2\lambda_{min} - \lambda \tag{42}$$

and can be solved for λ:

$$\lambda = \frac{\lambda_{min}^2}{\mu_1 + \lambda_{min}} \tag{43}$$

Substituting this in (6), one finally obtains (8). □

B. Proof of Corollary 1

Using $\sqrt{1+x} \approx 1 + x/2 - x^2/8$ when $x \ll 1$ in (5), one obtains $\lambda_{1i} \approx \lambda^{-1} + \mu_i\lambda^{-2}$, and using this in (6), one obtains $\lambda \approx m(P_T + tr\mathbf{W}_1^{-1})^{-1}$. The condition $x \ll 1$ is equivalent to $\lambda/\mu_i \gg 4$, which in turn is equivalent to (15), and the latter also implies $\min_i \lambda_i(\mathbf{W}_1) \gg \max_i \lambda_i(\mathbf{W}_2)$ (i.e. the eavesdropper channel is indeed much weaker than the main one), from which it follows that $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1} \approx \mathbf{W}_2^{-1}$, and applying these in (4), one obtains (16). □

C. Proof of Proposition 2

Let $\mathbf{R}^*$ be optimal covariance in (3). Observe that

$$\begin{aligned} C_s &= C(\mathbf{R}^*) && (44) \\ &= \ln\frac{|\mathbf{I} + \mathbf{W}_1\mathbf{R}^*|}{|\mathbf{I} + \mathbf{W}_2\mathbf{R}^*|} && (45) \\ &= \ln\frac{|\mathbf{I} + \mathbf{W}_1\mathbf{P}_a\mathbf{R}^*\mathbf{P}_a|}{|\mathbf{I} + \mathbf{W}_2\mathbf{P}_a\mathbf{R}^*\mathbf{P}_a|} && (46) \\ &= \ln\frac{|\mathbf{I} + \tilde{\mathbf{W}}_1\tilde{\mathbf{R}}^*|}{|\mathbf{I} + \tilde{\mathbf{W}}_2\tilde{\mathbf{R}}^*|} && (47) \\ &\le \max_{\tilde{\mathbf{R}}} \ln\frac{|\mathbf{I} + \tilde{\mathbf{W}}_1\tilde{\mathbf{R}}|}{|\mathbf{I} + \tilde{\mathbf{W}}_2\tilde{\mathbf{R}}|} \text{ s.t. } \tilde{\mathbf{R}} \ge 0,\ tr\tilde{\mathbf{R}} \le P_T && (48) \end{aligned}$$

where $\mathbf{P}_a = \mathbf{U}_a\mathbf{U}_a^+$ is the projection matrix on the subspace $span\{\mathbf{U}_a\}$ and $\tilde{\mathbf{R}}^* = \mathbf{U}_a^+\mathbf{R}^*\mathbf{U}_a$; (46) follows from $\mathbf{P}_a\mathbf{R}^*\mathbf{P}_a = \mathbf{R}^*$, (47) follows from

$$|\mathbf{I} + \mathbf{W}_i\mathbf{P}_a\mathbf{R}^*\mathbf{P}_a| = |\mathbf{I} + \mathbf{U}_a^+\mathbf{W}_i\mathbf{U}_a\mathbf{U}_a^+\mathbf{R}^*\mathbf{U}_a| \tag{49}$$

(48) follows from $tr\tilde{\mathbf{R}}^* \le tr\mathbf{R}^* \le P_T$ (since $\mathbf{U}_a$ is semi-unitary). The 1st inequality in (48) holds with equality, as can be proved by contradiction: assume that the inequality is strict so that

$$\begin{aligned} \ln\frac{|\mathbf{I} + \tilde{\mathbf{W}}_1\tilde{\mathbf{R}}^*|}{|\mathbf{I} + \tilde{\mathbf{W}}_2\tilde{\mathbf{R}}^*|} &= \ln\frac{|\mathbf{I} + \mathbf{W}_1\mathbf{U}_a\tilde{\mathbf{R}}^*\mathbf{U}_a^+|}{|\mathbf{I} + \mathbf{W}_2\mathbf{U}_a\tilde{\mathbf{R}}^*\mathbf{U}_a^+|} \\ &= C(\mathbf{R}') \\ &> C(\mathbf{R}^*) \end{aligned} \tag{50}$$

where $\mathbf{R}' = \mathbf{U}_a\tilde{\mathbf{R}}^*\mathbf{U}_a^+$. Now note that $tr\mathbf{R}' = tr\tilde{\mathbf{R}}^* \le P_T$ so that $\mathbf{R}'$ is feasible and hence the strict inequality is impossible. Further note that $\tilde{\mathbf{W}}_1 > \tilde{\mathbf{W}}_2$ (this follows from (22)) and that $\tilde{\mathbf{R}}^*$ is of full rank. Therefore, the problems in (3) and (48) are equivalent and Theorem 1 applies, from which the desired result follows. □

D. Proof of Proposition 3

Observe that the KKT conditions in (31)-(34) are not sufficient for optimality in the general (non-degraded) case since the original problem is not convex (see e.g. [17]). However, since the (affine) constraints $tr\mathbf{R} \leq P_T$, $\mathbf{R} \geq \mathbf{0}$ clearly satisfy the Slater condition [17], [18] and since the maximum is achievable (since the constraint set is compact and the objective function is continuous), the KKT conditions are necessary for optimality [18]. We further need the following technical Lemma.

Lemma 2: Let $\mathbf{A}, \mathbf{B}, \mathbf{C} \geq \mathbf{0}$ *be positive semi-definite matrices and let* $\mathbf{ABC}$ *be Hermitian. Then* $\mathbf{ABC} \geq \mathbf{0}$.

Proof: Since $\mathbf{A}, \mathbf{C} \geq \mathbf{0}$, there exists a non-singular matrix $\mathbf{S}$ such that $\mathbf{SAS}^+ = \mathbf{D}_a \geq \mathbf{0}, \mathbf{SCS}^+ = \mathbf{D}_c \geq \mathbf{0}$ are diagonal [19]. Using the latter,

$$\mathbf{ABC} = \mathbf{SD}_a\overline{\mathbf{B}}\mathbf{D}_c\mathbf{S}^+ \tag{51}$$

where $\overline{\mathbf{B}} = \mathbf{S}^+\mathbf{BS} \geq \mathbf{0}$. Observe further that

$$\begin{aligned} \lambda_i(\mathbf{D}_a\overline{\mathbf{B}}\mathbf{D}_c) &= \lambda_i(\overline{\mathbf{B}}\mathbf{D}_c\mathbf{D}_a) & (52)\\ &= \lambda_i((\mathbf{D}_c\mathbf{D}_a)^{1/2}\overline{\mathbf{B}}(\mathbf{D}_c\mathbf{D}_a)^{1/2}) \geq \mathbf{0} & (53) \end{aligned}$$

since $(\mathbf{D}_c\mathbf{D}_a)^{1/2}\overline{\mathbf{B}}(\mathbf{D}_c\mathbf{D}_a)^{1/2} \geq \mathbf{0}$, where $\lambda_i(\mathbf{B})$ means an eigenvalue of matrix $\mathbf{B}$. Since $\mathbf{D}_a\overline{\mathbf{B}}\mathbf{D}_c$ is Hermitian (because $\mathbf{ABC}$ is) and has non-negative eigenvalues, it is positive semi-definite [19], $\mathbf{D}_a\overline{\mathbf{B}}\mathbf{D}_c \geq \mathbf{0}$. It follows that $\mathbf{ABC} = \mathbf{SD}_a\overline{\mathbf{B}}\mathbf{D}_c\mathbf{S}^+ \geq \mathbf{0}$. □

Note that this Lemma is a generalization of a well known fact: $\mathbf{AB} \geq \mathbf{0}$ if $\mathbf{A}, \mathbf{B} \geq \mathbf{0}$ and $\mathbf{AB}$ is Hermitian [19]. We first prove that $\mathbf{Z} = (\mathbf{I} + \mathbf{W}_1\mathbf{R})(\mathbf{I} + \mathbf{RW}_2) > \mathbf{0}$ when $\mathbf{R} > \mathbf{0}$. In this case, $\mathbf{Z}$ can be expressed as

$$\mathbf{Z} = (\mathbf{R}^{-1} + \mathbf{W}_1)\mathbf{R}^2(\mathbf{R}^{-1} + \mathbf{W}_2) \tag{54}$$

Now identify the right-hand side of (54) with $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and use Lemma 2 to obtain $\mathbf{Z} \geq \mathbf{0}$ (noting that $\mathbf{Z}$ is Hermitian from (32)). Therefore, it follows from (32) that

$$\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M} \geq \mathbf{0} \tag{55}$$

since $\lambda > 0$, as $\lambda = 0$ implies $\mathbf{W}_1 \leq \mathbf{W}_2$ and thus $C_s = 0$ - trivial case not considered here. Since $|(\mathbf{I} + \mathbf{W}_1\mathbf{R})(\mathbf{I} + \mathbf{RW}_2)| > 0$, it further follows that $\mathbf{Z} > \mathbf{0}$ and

$$\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M} > \mathbf{0}. \tag{56}$$

The case of singular $\mathbf{R}$ is somewhat more involved. Let $\mathbf{R} = \mathbf{U\Lambda U}^+$ be the eigenvalue decomposition of $\mathbf{R}$. Consider

$$\begin{aligned} \tilde{\mathbf{Z}} &= \mathbf{U}^+\mathbf{ZU}\\ &= (\mathbf{I} + \tilde{\mathbf{W}}_1\mathbf{\Lambda})(\mathbf{I} + \mathbf{\Lambda}\tilde{\mathbf{W}}_2)\\ &= \tilde{\mathbf{W}}_1 - \tilde{\mathbf{W}}_2 + \mathbf{\Lambda}_M \end{aligned} \tag{57}$$

where $\tilde{\mathbf{W}}_i = \mathbf{U}^+\mathbf{W}_i\mathbf{U}$, $\mathbf{\Lambda}_M = \mathbf{U}^+\mathbf{MU}$, and block-partition $\mathbf{\Lambda}, \tilde{\mathbf{W}}_i$ as follows:

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \ \tilde{\mathbf{W}}_i = \begin{bmatrix} \mathbf{W}_i^{11} & \mathbf{W}_i^{12} \\ \mathbf{W}_i^{21} & \mathbf{W}_i^{22} \end{bmatrix} \tag{58}$$

where $\mathbf{\Lambda}_r$ is a diagonal matrix collecting r positive eigenvalues of $\mathbf{R}$. Using this in (57), one obtains, after some manipulations,

$$\tilde{\mathbf{Z}} = \begin{bmatrix} (\mathbf{W}_1^{11}\mathbf{\Lambda}_r + \mathbf{I}_r)(\mathbf{\Lambda}_r\mathbf{W}_2^{11} + \mathbf{I}_r) & (\mathbf{W}_1^{11}\mathbf{\Lambda}_r + \mathbf{I}_r)\mathbf{\Lambda}_r\mathbf{W}_2^{12} \\ \mathbf{W}_1^{21}\mathbf{\Lambda}_r(\mathbf{\Lambda}_r\mathbf{W}_2^{11} + \mathbf{I}_r) & \mathbf{W}_1^{21}\mathbf{\Lambda}_r^2\mathbf{W}_2^{12} + \mathbf{I}_r \end{bmatrix} \tag{59}$$

where $\mathbf{I}_r$ is $r \times r$ identity matrix. Note that $\tilde{\mathbf{Z}}$ is Hermitian (since $\mathbf{Z}$ is) and use the following fact [19]:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^+ & \mathbf{X} \end{bmatrix} \geq \mathbf{0} \leftrightarrow \mathbf{X} \geq \mathbf{B}^+\mathbf{A}^{-1}\mathbf{B} \tag{60}$$

where $\mathbf{X}, \mathbf{A}$ are Hermitian (and so is the block-partitioned matrix) and ↔ means that the conditions are equivalent. Apply this to (59) to obtain

$$\begin{aligned} \mathbf{B}^+\mathbf{A}^{-1}\mathbf{B} &= \mathbf{W}_1^{21}\mathbf{\Lambda}_r(\mathbf{\Lambda}_r\mathbf{W}_2^{11} + \mathbf{I}_r)((\mathbf{W}_1^{11}\mathbf{\Lambda}_r + \mathbf{I}_r)\\ &\qquad \times(\mathbf{\Lambda}_r\mathbf{W}_2^{11} + \mathbf{I}_r))^{-1}(\mathbf{W}_1^{11}\mathbf{\Lambda}_r + \mathbf{I}_r)\mathbf{\Lambda}_r\mathbf{W}_2^{12}\\ &= \mathbf{W}_1^{21}\mathbf{\Lambda}_r^2\mathbf{W}_2^{12}\\ &\leq \mathbf{W}_1^{21}\mathbf{\Lambda}_r^2\mathbf{W}_2^{12} + \mathbf{I}_r\\ &= \mathbf{X} \end{aligned} \tag{61}$$

so that $\tilde{\mathbf{Z}} \geq \mathbf{0}$ and thus $\mathbf{Z} \geq \mathbf{0}$ follow. Since $|\mathbf{Z}| \neq \mathbf{0}$, it further follows that $\mathbf{Z} > \mathbf{0}$ and thus

$$\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M} > \mathbf{0} \tag{62}$$

To prove (22), note that

$$\mathbf{0} < \mathbf{U}_{r+}^+(\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M})\mathbf{U}_{r+} = \mathbf{U}_{r+}^+(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{U} \tag{63}$$

where the columns of $\mathbf{U}_{r+}$ are the active eigenvectors $\{\mathbf{u}_{i+}\}$. The inequality follows since $\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M} > \mathbf{0}$ and the columns of $\mathbf{U}_{r+}$ being linearly independent:

$$\begin{aligned} &\mathbf{x}^+\mathbf{U}_{r+}^+(\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M})\mathbf{U}_{r+}\mathbf{x}\\ &\quad = \tilde{\mathbf{x}}^+(\mathbf{W}_1 - \mathbf{W}_2 + \mathbf{M})\tilde{\mathbf{x}} > \mathbf{0} \ \ \forall \mathbf{x} \neq \mathbf{0} \end{aligned} \tag{64}$$

where $\tilde{\mathbf{x}} = \mathbf{U}_{r+}\mathbf{x} \neq \mathbf{0}$ since the columns of $\mathbf{U}_{r+}$ are linearly independent. The equality follows since $\mathbf{MR} = \mathbf{0}$ implies $\mathbf{MU}_{r+} = \mathbf{0}$. (23) follows from (22) by expressing $\mathbf{x} = \mathbf{U}_{r+}\mathbf{z}$ for some $\mathbf{z}$.

E. Proof of Proposition 4

We will need the following technical Lemma.

Lemma 3: Consider the function

$$f(\mathbf{X}) = \ln\left|\mathbf{I} - \mathbf{B}(\mathbf{A} + \mathbf{X})^{-1}\mathbf{B}\right|,$$

where $\mathbf{A}, \mathbf{B}, \mathbf{X} \geq \mathbf{0}$ are positive semi-definite matrices, $\mathbf{I}$ is the identity matrix, $\mathbf{BA}^{-1}\mathbf{B} \leq \mathbf{I}$. It has the following properties:

1) $f(\mathbf{X})$ is increasing in $\mathbf{X}$: $\mathbf{X}_1 \leq \mathbf{X}_2 \rightarrow f(\mathbf{X}_1) \leq f(\mathbf{X}_2)$.
2) $f(\mathbf{X})$ is concave in $\mathbf{X}$:

$$f(\alpha\mathbf{X}_1 + \beta\mathbf{X}_2) \geq \alpha f(\mathbf{X}_1) + \beta f(\mathbf{X}_2),$$

for $\alpha + \beta = 1, \ 0 \leq \alpha, \ \beta \leq 1$.

proof: 1st property follows from the (easy to verify) fact that $-\mathbf{B}(\mathbf{A} + \mathbf{X})^{-1}\mathbf{B}$ is increasing in $\mathbf{X}$ (in the matrix positive definite ordering sense [19]). 2nd one is obtained from the following chain argument:

$$\begin{aligned} f(\alpha\mathbf{X}_1+\beta\mathbf{X}_2) &= \ln\left|\mathbf{I}-\mathbf{B}(\mathbf{A}+\alpha\mathbf{X}_1+\beta\mathbf{X}_2)^{-1}\mathbf{B}\right| \\ &\overset{(a)}{\geq} \ln\left|\mathbf{I}-\alpha\mathbf{B}\mathbf{A}_1^{-1}\mathbf{B}-\beta\mathbf{B}\mathbf{A}_2^{-1}\mathbf{B}\right| \\ &\overset{(b)}{\geq} \alpha\ln\left|\mathbf{I}-\mathbf{B}\mathbf{A}_1^{-1}\mathbf{B}\right|+\beta\ln\left|\mathbf{I}-\mathbf{B}\mathbf{A}_2^{-1}\mathbf{B}\right| \\ &= \alpha f(\mathbf{X}_1)+\beta f(\mathbf{X}_2) \end{aligned} \tag{65}$$

where $\mathbf{A}_i = \mathbf{A}+\mathbf{X}_i$; (a) follows from the facts that $F(\mathbf{X}) = \mathbf{X}^{-1}$ is convex in $\mathbf{X}$ and $F(\mathbf{X}) = \ln|\mathbf{X}|$ is increasing [17], [19]; (b) follows from the fact that $F(\mathbf{X}) = \ln|\mathbf{X}|$ is concave [17]. □

We now assume that $\mathbf{W}_{i+} > \mathbf{0}$. The case of singular $\mathbf{W}_{i+}$ will follow from the standard continuity argument [19] (i.e. use $\mathbf{W}_{i\epsilon} = \mathbf{W}_{i+}+\epsilon\mathbf{I}$, $\epsilon > 0$, instead of $\mathbf{W}_{i+}$ and then take $\epsilon \to 0$; see [19, Sec. 2.6] for more details and examples). Observe that

$$\begin{aligned} C_+(\mathbf{R}) &= \ln\frac{|\mathbf{W}_{1+}|}{|\mathbf{W}_{2+}|}+\ln\frac{\left|\mathbf{W}_{1+}^{-1}+\mathbf{R}\right|}{\left|\mathbf{W}_{2+}^{-1}+\mathbf{R}\right|} \\ &= c+\ln\left|\mathbf{I}-\Delta\mathbf{W}(\mathbf{W}_{2+}^{-1}+\mathbf{R})^{-1}\right| \\ &= c+\ln\left|\mathbf{I}-\Delta\mathbf{W}^{1/2}(\mathbf{W}_{2+}^{-1}+\mathbf{R})^{-1}\Delta\mathbf{W}^{1/2}\right| \end{aligned} \tag{66}$$

where $c = \ln|\mathbf{W}_{1+}|-\ln|\mathbf{W}_{2+}|$ and $\Delta\mathbf{W} = \mathbf{W}_{2+}^{-1}-\mathbf{W}_{1+}^{-1}$, and apply Lemma 3 to the last term of the last expression in (66). It is easy to verify that $\mathbf{B}\mathbf{A}^{-1}\mathbf{B} \leq \mathbf{I}$ (since $\mathbf{W}_{2+}^{-1}-\mathbf{W}_{1+}^{-1} \leq \mathbf{W}_{2+}^{-1}$) and that $\mathbf{B} \geq \mathbf{0}$ (since $\mathbf{W}_{1+} \geq \mathbf{W}_{2+}$), so that the properties of $C_+(\mathbf{R})$ follow. To prove the lower bound, note that the problem in (29) limits the optimization to the positive eigenspace of $\mathbf{W}_1-\mathbf{W}_2$ and thus is sub-optimal. To prove the achievability of the lower bound, note that, in the low-SNR regime, one obtains $C(\mathbf{R}) \approx tr(\mathbf{W}_1-\mathbf{W}_2)\mathbf{R}$ so that rank-1 transmission on the largest eigenmode of $\mathbf{W}_1-\mathbf{W}_2$ is optimal. But this eigenmode is in the positive eigenspace of $\mathbf{W}_1-\mathbf{W}_2$ (unless it is negative, in which case the capacity is zero) so that this transmission is also optimal for the projected problem. When eigenvectors of $\mathbf{W}_1$ and $\mathbf{W}_2$ are the same, the achievability follows from the respective result for parallel channels in [11] and [12] (since an optimal covariance also has the same eigenvectors). When the channel is degraded, the projection has no effect since $\mathbf{W}_1-\mathbf{W}_2 \geq 0$ so that the problems in (3) and (29) are identical. □

References

[1] H. Bölcskei, Ed., *Space-Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[2] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[3] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Proc. 41st Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2007, pp. 905–910.

[4] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[5] R. Bustin, R. Liu, H. V. Poor, and S. Shamai (Shitz), "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP J. Wireless Commun. Netw.*, no. 1, Jul. 2009, Art. no. 370970.

[6] T. Liu and S. Shamai (Shitz), "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547–2553, Jun. 2009.

[7] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4033–4039, Sep. 2009.

[8] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2010.

[9] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[10] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[11] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2453–2469, Jun. 2008.

[12] Z. Li *et al.*, "Secrecy capacity of independent parallel channels," in *Securing Wireless Communications at the Physical Layer*, R. Liu and W. Trappe, Eds. New York, NY, USA: Springer, 2010.

[13] J. Li and A. Petropulu. (Sep. 2009). "Transmitter optimization for achieving secrecy capacity in Gaussian MIMO wiretap channels." [Online]. Available: https://arxiv.org/abs/0909.2622

[14] J. Li and A. P. Petropulu, "On beamforming solution for secrecy capacity of MIMO wiretap channels," in *Proc. IEEE Globecom*, Houston, TX, USA, Dec. 2011, pp. 889–892.

[15] S. Loyka and C. D. Charalambous, "On optimal signaling over secure MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 443–447.

[16] S. Loyka and C. D. Charalambous, "Further results on optimal signaling over secure MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2019–2023.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, , 2008.

[19] F. Zhang, *Matrix Theory*. New York, NY, USA: Springer, 1999.

[20] J. Brinkhuis and V. Tikhomirov, *Optimization: Insights and Applications*. Princeton, NJ, USA: Princeton Univ. Press, 2005.

[21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

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.