On Optimal Signaling over Secure MIMO Channels

Sergey Loyka, Charalambos D. Charalambous

***Abstract*—Optimal signalling over the wire-tap MIMO Gaussian channel is studied under the total transmit power constraint. A direct proof of the necessary condition of optimality (signaling on the positive directions of the difference channel) is given using the necessary KKT conditions. Based on it, an explicit, closed-form solution for the optimal transmit covariance matrix is given when the latter is of the full rank. 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 latter regime. A refined estimate of the rank of an optimal covariance matrix is given for the general case.**

I. Introduction

Multiple-input multiple-output (MIMO) architecture has gain 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 wire-tap Gaussian MIMO channels [6]-[10] and the optimal transmit covariance has been found for MISO systems [3] or in the 2-2-1 system [7] under the total power constraint and in the general MIMO case under the transmit covariance matrix constraint [5]. The high-SNR regime (SNR $\to \infty$) 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. The main contribution of this paper is an explicit, closed-form solution for the optimal full-rank covariance under the total power constraint at finite SNR (Theorem 2). Theorem 1 sets the foundation for this giving a direct proof (via the necessary KKT conditions) to a necessary condition of optimality, which is a transmission of the positive directions of the difference channel. The cases of high-SNR and of weak eavesdropper are elaborated in Corollaries 3 and 4. The optimal covariance of Theorem 2 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. Proposition 1 shows that a transmission on the positive eigenspace of the difference channel satisfies the necessary condition and is a convex optimization problem (so that all powerful tools of convex optimization apply [11]). Corollary 1 refines the estimate of the optimal covariance matrix rank given in [10] for the general case.

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

C.D. Charalambous is with the ECE Department, University of Cyprus, 75 Kallipoleos Avenue, P.O. Box 20537, Nicosia, 1678, Cyprus, e-mail: chadcha@ucy.ac.cy

II. Wire-Tap Gaussian MIMO Channel Model

Let us consider the standard wire-tap Gaussian MIMO channel model,

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{x} + \xi_1, \quad \mathbf{y}_2 = \mathbf{H}_2\mathbf{x} + \xi_2 \tag{1}$$

where $\mathbf{x} = [x_1, x_2, ...x_m]^T \in C^{m,1}$ is the transmitted complex-valued signal vector of dimension $m \times 1$, "T" denotes transposition, $\mathbf{y}_{1(2)} \in C^{n,1}$ are the received vectors at the receiver (eavesdropper), $\xi_{1(2)}$ is the circularly-symmetric additive white Gaussian noise at the receiver (eavesdropper) (normalized to unit variance in each dimension), $\mathbf{H}_{1(2)} \in C^{n_{1(2)},m}$ 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 infinite horizon information theory assumption holds) and frequency-flat, with full channel state information (CSI) at the Rx and Tx ends.

For a given transmit covariance matrix $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^+\}$, where $E\{\cdot\}$ is statistical expectation, the maximum achievable secure rate between the Tx and Rx (so that the rate between the Tx and eavesdropper is 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, and 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 or MISO 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 a direct formulation and proof of this fact using the necessary KKT conditions, which is also

instrumental for further development. In particular, Theorem 2 gives an explicit, closed-form solution for the optimal full-rank covariance in (3) at finite SNR. A number of additional insights follow.

III. Optimal Signaling: Solutions and Properties

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

Theorem 1: Let $\mathbf{R}^*$ be an optimal covariance in (3),

$$\mathbf{R}^* = \arg\max_{\mathbf{R}\geq 0} C(\mathbf{R}) \text{ s.t. } tr\mathbf{R} \leq P_T$$

and let $\mathbf{u}_{i+}$ be its active eigenvector (i.e. corresponding to a positive eigenvalue). Then,

$$\mathbf{u}_{i+}^{+}(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{u}_{i+} > 0 \tag{4}$$

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) [1].

Proof: see the Appendix.

It was demonstrated in [10] that $rank(\mathbf{R}^*) < m$ unless $\mathbf{W}_1 > \mathbf{W}_2$ [2], i.e. an optimal transmission is of low-rank over a non-degraded channel. The Corollary below gives more precise characterization.

Corollary 1: 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, $rank(\mathbf{R}^*) \leq rank(\mathbf{W}_+) \leq m$, 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: follows from (4) using the fact that the eigenvectors $\mathbf{u}_{i+}$ are orthogonal to each other.

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

$$C_s = \ln \lambda_1, \ \mathbf{R}^* = P_T \mathbf{u}_1 \mathbf{u}_1^{+} \tag{5}$$

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][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 at (5) is also optimal for any $rank(\mathbf{W}_+) \geq 1$ at low SNR, where λ_1, $\mathbf{u}_1$ become the largest eigenvalue and corresponding eigenvector of $\mathbf{W}_1 - \mathbf{W}_2$.

One way to achieve the necessary condition in (4) is to transmit over the positive eigenspace of $\mathbf{W}_1 - \mathbf{W}_2$, as the following Proposition shows.

[1] After this paper has been submitted, we were informed that a weaker result ($\geq$ instead of $>$) was independently established in [14].

[2] $\mathbf{W}_1 > \mathbf{W}_2$ means that $\mathbf{W}_1 - \mathbf{W}_2$ is positive definite.

[3] This result has been obtained before, albeit in a different way, in [14].

Proposition 1: The following covariance matrix satisfies the necessary condition of Theorem 1:

$$\mathbf{R}' = \arg\max_{\mathbf{R}\geq 0} C_+(\mathbf{R}) \text{ s.t. } tr\mathbf{R} \leq P_T \tag{6}$$

where

$$C_+(\mathbf{R}) = \ln \frac{|\mathbf{I} + \mathbf{W}_{1+}\mathbf{R}|}{|\mathbf{I} + \mathbf{W}_{2+}\mathbf{R}|} \tag{7}$$

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: $\mathbf{W}_+ = \mathbf{U}_+\mathbf{D}_+\mathbf{U}_+^{+} > \mathbf{0}$, and $\mathbf{D}_+$ is the diagonal matrix of the positive eigenvalues. The optimal (maximizing) covariance $\mathbf{R}'$ in (5) satisfies

$$span\{\mathbf{v}_{i+}(\mathbf{R}')\} \in span\{\mathbf{v}_{i+}(\mathbf{W}_1 - \mathbf{W}_2)\} \tag{8}$$

where $\{\mathbf{v}_{i+}(\mathbf{R})\}$ denotes a set of eigenvectors corresponding to positive eigenvalues of matrix $\mathbf{R}$.

Proof: see Appendix.

It follows from (6) that the transmission takes places on the projected channels $\mathbf{W}_{i+} = \mathbf{P}_+\mathbf{W}_i\mathbf{P}_+$. It should be noted that the eigenvectors of the optimal covariance $\mathbf{R}'$ in (5) are not necessarily the same as those of $\mathbf{W}_+$. Rather, they span the same sub-space. In one special case, $\mathbf{R}'$ and $\mathbf{W}_+$ do have the same eigenvectors.

Corollary 2: Consider the secure MIMO channel in (1) such that $rank(\mathbf{W}_+) = 1$. Then

$$\mathbf{R}' = P_T \mathbf{u}_+ \mathbf{u}_+^{+} \tag{9}$$

where $\mathbf{u}_+$ is the only active eigenvector of $\mathbf{W}_+$ (corresponding to a positive eigenvalue), i.e. the optimal transmission is unique and on the positive eigenvector of $\mathbf{W}_+$ with the full available power.

Proof: follows immediately from Proposition 1.

The problem in (6) has further significance: while the problem $C_s = \max_{\mathbf{R}\geq 0} C(\mathbf{R})$ is not convex, so that powerful tools of convex optimization [11] cannot be used, the problem $\max_{\mathbf{R}\geq 0} C_+(\mathbf{R})$ is convex, to which all machinery of convex optimization can be applied. The following proposition makes this precise.

Proposition 2: $C_+(\mathbf{R})$ is a non-negative, concave and non-decreasing function of $\mathbf{R}$ (strictly positive, concave and increasing when the active eigenmodes of $\mathbf{R}$ are in the span of the active eigenmodes of $\mathbf{W}_+$).

Proof: see Appendix.

It follows from Proposition 2 that transmission with the full available power is optimal: $tr\mathbf{R}' = P_T$.

Let us now consider the original problem in (3) and obtain its solution $\mathbf{R}^*$ when the latter is of full rank.

Theorem 2: Consider the case of $\mathbf{W}_1 > \mathbf{W}_2 > \mathbf{0}$ (a degraded full-rank channel) and $P_T > P_{T0}$, where P_{T0} is a certain threshold power (i.e. sufficiently high but finite SNR). Then, $\mathbf{R}^*$ is of full rank and is given by:

$$\mathbf{R}^* = \mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^{+} - \mathbf{W}_1^{-1} \tag{10}$$

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

$$\lambda_{1i} = \frac{\mu_i}{2}\left(\sqrt{1+\frac{4}{\lambda\mu_i}} - 1\right) \tag{11}$$

and $\mu_i > 0$ are the eigenvalues of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1}$; $\lambda > 0$ is found from the total power constraint $tr\mathbf{R}^* = P_T$ as a unique solution of the following equation:

$$\sum_i \frac{\mu_i}{2}\sqrt{1+\frac{4}{\lambda\mu_i}} = P_T + \frac{1}{2}tr(\mathbf{W}_1^{-1} + \mathbf{W}_2^{-1}) \tag{12}$$

The corresponding secrecy capacity is

$$C_s = \ln\frac{|\mathbf{W}_1|}{|\mathbf{W}_2|} + \ln\frac{|\mathbf{\Lambda}_1|}{|\mathbf{\Lambda}_2|} \tag{13}$$

where $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_1 + diag\{\mu_i\}$. P_{T0} can be found as a unique solution of the following equation:

$$\lambda_{1min}(P_{T0})\lambda_{min}(\mathbf{W}_1) = 1$$

where $\lambda_{1min} = \min_i\{\lambda_{1i}\}$ and $\lambda_{min}(\mathbf{W}_1)$ is the minimum eigenvalue of $\mathbf{W}_1$.

Proof: see Appendix.

It should be pointed out that Theorem 2 gives an exact (not approximate) optimal covariance at finite SNR (no $P_T \to \infty$) since P_{T0} is a finite constant that depends only on $\mathbf{W}_1$ and $\mathbf{W}_2$ and can be found numerically (in fact, it can be not high at all, depending on $\mathbf{W}_1$ and $\mathbf{W}_2$). Corollary 4 below makes a more concrete statement. 1st term in (13) $C_\infty = \ln\frac{|\mathbf{W}_1|}{|\mathbf{W}_2|}$ 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 Theorem 2 in [9].

Note also that the second term in (10) de-emphasizes weak eigenmodes of $\mathbf{W}_1$. Since λ is monotonically decreasing as P_T increases (this follows from (12)), λ_{1i} monotonically increases with P_T, and approaches $\lambda_{1i} \approx \sqrt{\mu_i/\lambda}$ 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 (11) that λ_{1i} increases with μ_i, i.e. stronger eigenmodes of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1}$ (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 (9) 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 2 here. The condition $\mathbf{W}_2 > \mathbf{0}$ can be further removed via a limiting transition $\mathbf{W}_{2\epsilon} = \mathbf{W}_2 + \epsilon\mathbf{I} \to \mathbf{W}_2$ as $\epsilon \to 0^+$. The case of singular $\mathbf{W}_1$ can also be included by observing that $\mathbf{R}^*$ 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 2 will apply. With this in mind, the conditions of Theorem 2 are both sufficient and necessary 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 those of $\mathbf{W}_2$.

Corollary 3: Consider the secure MIMO channel in (1) under the conditions of Theorem 2 and when the eavesdropper channel is much weaker than the required one,

$$\lambda_i(\mathbf{W}_2) \ll m(P_T + tr\mathbf{W}_1^{-1})^{-1}/4 \tag{14}$$

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 (10) becomes

$$\mathbf{R}^* \approx \mathbf{U}_1(\lambda^{-1}\mathbf{I} - \mathbf{D}_1^{-1})\mathbf{U}_1^+ - \lambda^{-2}\mathbf{W}_2 \tag{15}$$

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 (15) 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 4: The optimal covariance $\mathbf{R}^*$ in (10) in the high-SNR regime

$$P_T \gg \sqrt{\mu_1}\sum\nolimits_i \sqrt{\mu_i} \tag{16}$$

(e.g. when $P_T \to \infty$), where $\mu_1 = \max_i \mu_i$, simplifies to

$$\mathbf{R}^* \approx \mathbf{U}diag\{d_i\}\mathbf{U}^+, \quad d_i = \frac{P_T\sqrt{\mu_i}}{\sum_i \sqrt{\mu_i}} \tag{17}$$

The corresponding secrecy capacity is

$$C_s \approx \ln\frac{|\mathbf{W}_1|}{|\mathbf{W}_2|} - \frac{1}{P_T}\left(\sum\nolimits_i \sqrt{\mu_i}\right)^2 \tag{18}$$

Proof: follows from Theorem 2 along the same lines as that of Corollary 3.

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 (17) does not converge to the uniform one in the high-SNR regime [4]. 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.

[4]The sub-optimality of the isotropic signalling suggested in Theorem 2 of [9] is hiding in the $o(1)$ term there. 2nd term of Eq. (18) above refines that $o(1)$ term.

IV. Conclusion

Optimal signalling over the wire-tap Gaussian MIMO channel has been studied under the total power constraint. Based on the necessary condition of the optimality, an explicit, closed-form solution is given for the optimal transmit covariance when the latter is of full rank. 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 were considered, and a refined estimate of the rank of the optimal covariance matrix is given for the general case.

References

[1] H. Bolcskei et al (Eds.), Space-Time Wireless Systems: From Array Processing to MIMO Communications, Cambridge University Press, Cambridge, 2006.

[2] Y. Liang, H. V. Poor and S. Shamai(Shitz), Information Theoretic Security, Foundations and Trends in Communications and Information Theory, v. 5, No. 45 (2008), pp. 355-580.

[3] Z. Li, W. Trappe, R. Yates, Secret communication via multi-antenna transmission, Conf. Information Sciences and Systems (CISS), Mar. 2007.

[4] P. K. Gopala, L. Lai, H. El Gamal, On the Secrecy Capacity of Fading Channels, IEEE Trans. Info. Theory, v. 54, No. 10, Oct. 2008.

[5] R. Bustin et al, An MMSE Approach to the Secrecy Capacity of the MIMO GaussianWiretap Channel, EURASIP Journal on Wireless Communications and Networking, 2009, Article ID 370970.

[6] T. Liu, S. Shamai (Shitz), A Note on the Secrecy Capacity of the Multiple-Antenna Wiretap Channel, IEEE Trans. Info. Theory, v. 55, No. 6, June 2009.

[7] S. Shafiee, N. Liu, S. Ulukus, Towards the Secrecy Capacity of the Gaussian MIMO Wire-Tap Channel: The 2-2-1 Channel, IEEE Trans. Info. Theory, v. 55, No. 9, Sep. 2009.

[8] A. Khisti, G.W. Wornell, Secure Transmission With Multiple Antennas—Part I: The MISOME Wiretap Channel, IEEE Trans. Info. Theory, v. 56, No. 7, July 2010.

[9] A. Khisti, G.W. Wornell, Secure Transmission With Multiple Antennas—Part II: The MIMOME Wiretap Channel, IEEE Trans. Info. Theory, v. 56, No. 11, Nov. 2010.

[10] F. Oggier, B. Hassibi, The Secrecy Capacity of the MIMO Wiretap Channel, IEEE Trans. Info. Theory, v. 57, No. 8, Aug. 2011.

[11] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[12] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, 2nd Ed., 2008.

[13] F. Zhang, Matrix Theory, Springer, 1999.

[14] J. Li, A. Petropulu, Transmitter Optimization for Achieving Secrecy Capacity in Gaussian MIMO Wiretap Channels, arXiv:0909.2622v1, Sep 2009.

V. Appendix

Proof of Theorem 1: Using Lagrange multiplier technique [11][12], the optimization problem in (3) has the following Lagrangian:

$$L = -\ln|\mathbf{I}+\mathbf{W}_1\mathbf{R}|+\ln|\mathbf{I}+\mathbf{W}_2\mathbf{R}|+\lambda(tr\mathbf{R}-P_T)-tr(\mathbf{MR}) \tag{19}$$

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. [11]) can be expressed as:

$$\lambda(\mathbf{I}+\mathbf{W}_1\mathbf{R})(\mathbf{I}+\mathbf{R}\mathbf{W}_2) = \mathbf{W}_1-\mathbf{W}_2+\mathbf{M} \tag{20}$$

$$\mathbf{MR}=\mathbf{0},\ \lambda(tr\mathbf{R}-P_T)=0, \qquad \mathbf{R}\geq\mathbf{0},\ \mathbf{M}\geq\mathbf{0},\ \lambda\geq 0,\ tr\mathbf{R}\leq P_T \tag{21}$$

where (20) is obtained from $\partial L/\partial\mathbf{R}=\mathbf{0}$,

$$\partial L/\partial\mathbf{R} = (\mathbf{I}+\mathbf{W}_2\mathbf{R})^{-1}\mathbf{W}_2-(\mathbf{I}+\mathbf{W}_1\mathbf{R})^{-1}\mathbf{W}_1+\lambda\mathbf{I}-\mathbf{M}=\mathbf{0} \tag{22}$$

and the two equalities in (21) are the complementary slackness conditions. Since the original problem is not convex, the KKT conditions are not sufficient for optimality [11]. However, since the (affine) constraints $tr\mathbf{R} \leq P_T$, $\mathbf{R} \geq \mathbf{0}$ clearly satisfy the Slater condition [11][12] 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 [12]. We further need the following technical Lemma.

Lemma 1: 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 [13]. Using the latter,

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

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

$$\lambda_i(\mathbf{D}_a\overline{\mathbf{B}}\mathbf{D}_c)=\lambda_i((\mathbf{D}_c\mathbf{D}_a)^{1/2}\overline{\mathbf{B}}(\mathbf{D}_c\mathbf{D}_a)^{1/2})\geq\mathbf{0}$$

(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 [13], $\mathbf{D}_a\overline{\mathbf{B}}\mathbf{D}_c\geq\mathbf{0}$. It follows that $\mathbf{ABC}=\mathbf{S}\mathbf{D}_a\overline{\mathbf{B}}\mathbf{D}_c\mathbf{S}^+\geq\mathbf{0}$. $Q.E.D.$

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 [13]. We further prove that $\mathbf{Z}=(\mathbf{I}+\mathbf{W}_1\mathbf{R})(\mathbf{I}+\mathbf{R}\mathbf{W}_2)>\mathbf{0}$ when $\mathbf{R}>\mathbf{0}$; the case of singular $\mathbf{R}$ will follow from the standard continuity argument [13]. Assuming $\mathbf{R}>\mathbf{0}$,

$$\mathbf{Z}=(\mathbf{R}^{-1}+\mathbf{W}_1)\mathbf{R}^2(\mathbf{R}^{-1}+\mathbf{W}_2) \tag{23}$$

Now identify the right-hand side of (23) with $\mathbf{A},\mathbf{B},\mathbf{C}$ and use Lemma 1 to obtain $\mathbf{Z}\geq\mathbf{0}$ (noting that $\mathbf{Z}$ is Hermitian from (20)). Therefore, it follows from (20) that $\mathbf{W}_1-\mathbf{W}_2+\mathbf{M}\geq\mathbf{0}$ (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 $|(\mathbf{I}+\mathbf{W}_1\mathbf{R})(\mathbf{I}+\mathbf{R}\mathbf{W}_2)|>0$, it further follows that $\mathbf{Z}>\mathbf{0}$ and $\mathbf{W}_1-\mathbf{W}_2+\mathbf{M}>\mathbf{0}$. Now, let $\mathbf{u}_{i+}$ be an active eigenvector (corresponding to a positive eigenvalue) of $\mathbf{R}^*$. Then,

$$0<\mathbf{u}_{i+}^+(\mathbf{W}_1-\mathbf{W}_2+\mathbf{M})\mathbf{u}_{i+}=\mathbf{u}_{i+}^+(\mathbf{W}_1-\mathbf{W}_2)\mathbf{u}_{i+}$$

where the equality follows from $\mathbf{MR}=\mathbf{0}$. $Q.E.D.$

With more efforts, a stronger statement can be proved:

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

where the columns of $\mathbf{U}_{r+}$ are $\{\mathbf{u}_{i+}\}$.

Proof of Proposition 1: Notice that

$$C_+(\mathbf{R})=\ln\frac{|\mathbf{I}+\mathbf{W}_{1+}\mathbf{R}|}{|\mathbf{I}+\mathbf{W}_{2+}\mathbf{R}|}=\ln\frac{|\mathbf{I}+\mathbf{W}_1\mathbf{R}_+|}{|\mathbf{I}+\mathbf{W}_2\mathbf{R}_+|} \tag{24}$$

where $\mathbf{R}_+ = \mathbf{P}_+\mathbf{R}\mathbf{P}_+$ is the projected covariance. Any component of $\mathbf{R}$ eliminated by the projection will not affect $C_+(\mathbf{R})$ but can possibly increase the total power, since $tr\mathbf{R}_+ \leq tr\mathbf{R}$. Therefore, the optimal covariance $\mathbf{R}'$ in (6) has to satisfy

$$span\{\mathbf{v}_{i+}(\mathbf{R}')\} \in span\{\mathbf{U}_+\} = span\{\mathbf{v}_{i+}(\mathbf{W}_1 - \mathbf{W}_2)\} \tag{25}$$

(in which case $tr\mathbf{R}_+ = tr\mathbf{R}$) and thus clearly satisfies (4). $Q.E.D.$

Proof of Proposition 2: We will need the following technical Lemma.

Lemma 2: 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, $\mathbf{B}\mathbf{A}^{-1}\mathbf{B} \leq \mathbf{I}$. 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 [13]). 2nd one is obtained from the following chain argument:

$$\begin{aligned} f(\alpha\mathbf{X}_1 + \beta\mathbf{X}_2) &= \ln\left|\mathbf{I} - \mathbf{B}(\mathbf{A} + \alpha\mathbf{X}_1 + \beta\mathbf{X}_2)^{-1}\mathbf{B}\right| \\ &\overset{(a)}{\geq} \ln\left|\mathbf{I} - \alpha\mathbf{B}\mathbf{A}_1^{-1}\mathbf{B} - \beta\mathbf{B}\mathbf{A}_2^{-1}\mathbf{B}\right| \\ &\overset{(b)}{\geq} \alpha\ln\left|\mathbf{I} - \mathbf{B}\mathbf{A}_1^{-1}\mathbf{B}\right| + \beta\ln\left|\mathbf{I} - \mathbf{B}\mathbf{A}_2^{-1}\mathbf{B}\right| \\ &= \alpha f(\mathbf{X}_1) + \beta f(\mathbf{X}_2) \end{aligned} \tag{26}$$

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 [11][13]; (b) follows from the fact that $F(\mathbf{X}) = \ln|\mathbf{X}|$ is concave [11]. $Q.E.D.$

We now assume that $\mathbf{W}_{2+} > \mathbf{0}$. The case of singular $\mathbf{W}_{2+}$ will follow from the standard continuity argument [13]. Observe that

$$\begin{aligned} C_+(\mathbf{R}) &= \ln\frac{|\mathbf{W}_{1+}|}{|\mathbf{W}_{2+}|} + \ln\frac{\left|\mathbf{W}_{1+}^{-1} + \mathbf{R}\right|}{\left|\mathbf{W}_{2+}^{-1} + \mathbf{R}\right|} \\ &= c + \ln\left|\mathbf{I} - \boldsymbol{\Delta}\mathbf{W}(\mathbf{W}_{2+}^{-1} + \mathbf{R})^{-1}\right| \\ &= c + \ln\left|\mathbf{I} - \boldsymbol{\Delta}\mathbf{W}^{1/2}(\mathbf{W}_{2+}^{-1} + \mathbf{R})^{-1}\boldsymbol{\Delta}\mathbf{W}^{1/2}\right| \end{aligned} \tag{27}$$

where $c = \ln|\mathbf{W}_{1+}| - \ln|\mathbf{W}_{2+}|$ and $\boldsymbol{\Delta}\mathbf{W} = \mathbf{W}_{2+}^{-1} - \mathbf{W}_{1+}^{-1}$, and apply Lemma 2 to the last term of the last expression in (27). 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 $\mathbf{W}_{1+} \geq \mathbf{W}_{2+}$). Proposition 2 follows. $Q.E.D.$

Proof of Theorem 2: It follows from Proposition 2 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 optimality. Assuming $\mathbf{R} > \mathbf{0}$ and using $\mathbf{M} = \mathbf{0}$ (which follows from $\mathbf{M}\mathbf{R} = \mathbf{0}$), one obtains from (22),

$$\mathbf{R}_1^{-1} - \mathbf{R}_2^{-1} = \lambda\mathbf{I} \tag{28}$$

where $\mathbf{R}_i = \mathbf{W}_i^{-1} + \mathbf{R}$, $i = 1, 2$. Let $\mathbf{R}_1 = \mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^+$ be the eigenvalue decomposition, where the columns of unitary matrix $\mathbf{U}$ are the eigenvectors, and $\boldsymbol{\Lambda}_1 > \mathbf{0}$ is a diagonal matrix of the corresponding eigenvalues. Using this in (28), one obtains $\boldsymbol{\Lambda}_1^{-1} - (\mathbf{U}^+\mathbf{R}_2\mathbf{U})^{-1} = \lambda\mathbf{I}$ and therefore $\mathbf{U}^+\mathbf{R}_2\mathbf{U} = \boldsymbol{\Lambda}_2$ is diagonal, so that $\mathbf{R}_2 = \mathbf{U}\boldsymbol{\Lambda}_2\mathbf{U}^+$ 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 (28) one obtains

$$\boldsymbol{\Lambda}_1 = (\lambda\mathbf{I} + \boldsymbol{\Lambda}_2^{-1})^{-1} \tag{29}$$

Furthermore,

$$\mathbf{R}_2 - \mathbf{R}_1 = \mathbf{W}_2^{-1} - \mathbf{W}_1^{-1} = \mathbf{U}(\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1)\mathbf{U}^+ \tag{30}$$

so that the columns of $\mathbf{U}$ are also the eigenvectors of $\mathbf{W}_2^{-1} - \mathbf{W}_1^{-1}$ and the diagonal entries of $\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1 = diag\{\mu_i\}$ are its eigenvalues. Combining the latter with (29), one obtains after some manipulations (11). (10) follows from $\mathbf{R}_1 = \mathbf{W}_1^{-1} + \mathbf{R}$ and $\mathbf{R}_1 = \mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^+$. It is straightforward to see that $\lambda > 0$ (otherwise $\mathbf{W}_1 \leq \mathbf{W}_2$), so that transmission with the full power is optimal and (12) follows from the power constraint $tr\mathbf{R} = P_T$. For (10) to be a valid solution, we need $\mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^+ > \mathbf{W}_1^{-1}$. This is insured by observing that the left-hand side of (12) is monotonically decreasing in λ, so that the latter is monotonically decreasing as P_T increases and, from (11), λ_{1i} also monotonically increases. Therefore, for sufficiently large P_T, $P_T > P_{T0}$ for some finite P_{T0}, the minimum eigenvalue of $\boldsymbol{\Lambda}_1$ exceeds the maximum one of $\mathbf{W}_1^{-1}$ and thus the condition $\mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^+ > \mathbf{W}_1^{-1}$ follows. Therefore, (10)-(12) solve the KKT conditions and thus achieve the global optimum. It can be further seen that the solution is unique. $Q.E.D.$

Proof of Corollary 3: Using $\sqrt{1+x} \approx 1 + x/2 - x^2/8$ when $x \ll 1$ in (11), one obtains $\lambda_{1i} \approx \lambda^{-1} + (\lambda^2\mu_i)^{-1}$, and using this in (12), 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 (14), 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 (10), one obtains (15). $Q.E.D.$