An Algorithm for Global Maximization of Secrecy Rates in Gaussian MIMO Wiretap Channels

Sergey Loyka and Charalambos D. Charalambous

***Abstract*—Optimal signaling for secrecy rate maximization in Gaussian MIMO wiretap channels is considered. While this channel has attracted a significant attention recently and a number of results have been obtained, including the proof of the optimality of Gaussian signalling, an optimal transmit covariance matrix is known for some special cases only and the general case remains an open problem. An iterative custom-made algorithm to find a globally-optimal transmit covariance matrix in the general case is developed in this paper, with guaranteed convergence to a *global* optimum. While the original optimization problem is not convex and hence difficult to solve, its minimax reformulation can be solved via the convex optimization tools, which is exploited here. The proposed algorithm is based on the barrier method extended to deal with a minimax problem at hand. Its convergence to a global optimum is proved for the general case (degraded or not) and a bound for the optimality gap is given for each step of the barrier method. The performance of the algorithm is demonstrated via numerical examples. In particular, 20 to 40 Newton steps are already sufficient to solve the sufficient optimality conditions with very high precision (up to the machine precision level), even for large systems. Even fewer steps are required if the secrecy capacity is the only quantity of interest. The algorithm can be significantly simplified for the degraded channel case and can also be adopted to include the per-antenna power constraints (instead or in addition to the total power constraint). It also solves the dual problem of minimizing the total power subject to the secrecy rate constraint.**

***Index Terms*—MIMO, secrecy capacity, optimization, wiretap channel.**

I. Introduction

WIDE-SPREAD use of wireless systems has initiated significant interest in their security and related information-theoretic studies [1]. Secrecy capacity has emerged as a key performance metric, which extends the regular channel capacity to accommodate the secrecy requirement. Wyner's wire-tap channel (WTC) [1]–[3] is the most popular model to accommodate secrecy, which was extended to the Gaussian channel [4] and subsequently to the Gaussian multiple-input multiple-output (MIMO) setting [5]–[8]; the reader is referred to [1] for a detailed discussion of this model and extensive literature review. The Gaussian MIMO WTC has been recently a subject of intense study and a number of results have been obtained, including the proof of optimality of Gaussian signaling [1], [5]–[8]. While the functional form of the optimal (capacity-achieving) distribution has been established, significantly less is known about its optimal covariance matrix (the only remaining parameter to completely characterize the distribution since the mean is always zero).

The optimal transmit covariance matrix under the total power constraint has been obtained for some special cases, e.g., low/high SNR, multiple-input single-output (MISO) channels, full-rank, rank-1 or weak eavesdropper cases, or the parallel channel [5]–[19], but the general case remains illusive. The main difficulty lies in the fact that the underlying optimization problem is in general not a convex problem. It was conjectured in [7] and proved in [6] using an indirect approach (via the degraded channel) that the optimal signaling is on the positive directions of the difference channel (where the legitimate channel is stronger than the eavesdropper one). A direct proof based on the necessary Karush-Kuhn-Tucker (KKT) optimality conditions has been obtained in [14]. A weaker form of this result (non-negative instead of positive directions) has been obtained earlier in [9]. In the general case, the rank of an optimal covariance matrix does not exceed the number of positive eigenvalues of the difference channel matrix [14]. An exact full-rank solution for the optimal covariance has been obtained in [14] and its properties have been characterized. In particular, unlike the regular channel (no eavesdropper), the optimal power allocation does not converge to uniform one at high SNR and the latter remains sub-optimal at any finite SNR. In the case of weak eavesdropper (its singular values are much smaller than those of the legitimate channel), the optimal signaling mimics the conventional one (water-filling over the channel eigenmodes) with an adjustment for the eavesdropper channel. The rank-one solution in combination with the full-rank one provides a complete solution for the case of two transmit antennas and any number of receive/eavesdropper antennas. The 2-2-1 case (2 transmit, 2 receive, 1 eavesdropper antenna) has been studied earlier in [10] and the MISO case (single-antenna receiver) has been considered in [11], [12] and settled in [5], [13], for which beamforming is optimal and which is also the case for a MIMO-WTC in the low SNR regime. The case of isotropic eavesdropper is studied in detail in [15], including the optimal signaling in an explicit closed form and its properties. This case is shown to be the worst-case MIMO wire-tap channel. Based on this, lower and upper (tight) capacity bounds have been obtained for the general case, which are achievable by an isotropic eavesdropper. The set of channels for which isotropic signaling is optimal has been fully

Manuscript received August 20, 2014; revised February 5, 2015; accepted March 27, 2015. Date of publication April 17, 2015; date of current version June 12, 2015. The associate editor coordinating the review of this paper and approving it for publication was A. Khisti.

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 Department of Electrical and Computer Engineering (ECE), University of Cyprus, Nicosia 1678, 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/TCOMM.2015.2424235

characterized [15]. It turns out to be more richer than that of the conventional (no eavesdropper) MIMO channel. A closed-form solution was obtained in [16] for the case of weak eavesdropper but otherwise arbitrary channel; its optimal power allocation somewhat resembles the water-filling but is not identical to it. For the case of parallel channels, independent signaling is optimal [17], [18], which implies that the optimal covariance matrix is diagonal; the corresponding optimal power allocation can be found in [18]. This also implies that the eigenvectors of optimal covariance matrix are the same as the right singular vectors of the legitimate or eavesdropper channels when the latter two are the same [16] and the corresponding power allocation is the same as in [18]. The low-SNR regime has been studied in detail in [19]. In particular, signaling on the strongest eigenmode(s) of the difference channel matrix is optimal. Little is known beyond these special cases and the general case is still an open problem.

While numerical algorithms have been proposed in [20], [21] to compute a transmit covariance matrix for the MIMO-WTC, their convergence to a *global* optimum has not been proved. The main difficulty lies in the fact that the underlying optimization problems are not convex and hence KKT conditions are not sufficient for optimality [24]. In particular, while the alternating optimization algorithm in [20] is shown to convergence to a KKT (stationary) point, it is not necessarily a global maximum (due to the above reason); it may, in fact, be a saddle point or a *local* rather than *global* maximum of the secrecy rate[1] and it is not known how far away it is from the global maximum. This remark also applies to the algorithms considered in [21], [22].

The purpose of this paper is to develop a numerical algorithm for computing a *globally*-optimal covariance matrix in the general case, i.e. for the general Gaussian MIMO-WTC (degraded or not), with guaranteed convergence to a *global* optimum, and to prove its convergence. This is a challenging task as the underlying optimization problem is not convex so that standard tools of convex optimization cannot be used; in general, non-convex problems are much harder to solve [23]. We deal with this challenge by using the minimax representation of the secrecy capacity found in [6]. While this representation appears to be more complicated than the standard one (the former involves two conflicting optimizations while the latter—only one), it turns out to be much easier to solve, at least numerically, as we demonstrate using the primal-dual representation of Newton method in combination with the barrier method. The main advantage of this approach is that each of the two problems is convex, the saddle-point property holds and hence the respective KKT conditions are sufficient for global optimality (Slater's condition holds as well). A conceptually-similar approach has been used before for optimizing the transmitter with per-antenna power constraints in the regular (no secrecy) MIMO broadcast channel in [25]. Our custom-made algorithm essentially solves the KKT optimality conditions (see e.g., [23] for a background on these conditions), which are sufficient for the minimax problem at hand, in an iterative way using the primal-dual representation of Newton method in combination with the barrier method (to accommodate inequality constraints) adopted to the MIMO WTC setting, see Section V. A proof of the algorithm's convergence to a *global* optimum is also provided for the general case. While we formulate the algorithm for the total power constraint, it can be easily modified to accommodate other forms of power constraint, e.g. maximum per-antenna constraint (instead or in addition to the total power constraint), and also to solve a dual problem of minimizing the total transmit power under the secrecy rate constraint.

A key part of the convergence proof for our algorithm involves a proof of non-singularity of the KKT matrix[2], so that Newton steps are well-defined for all iterations of the algorithms and they generate a sequence of norm-decreasing residuals and hence converge to a globally-optimal point (i.e., a solution of the KKT conditions which corresponds to zero residual). This is a difficult task since the underlining optimization problems involve both maximization and minimization and the corresponding KKT matrix is indefinite so that the regular tools developed for positive semi-definite matrices [26] do not apply. A block-partitioned factorization of the KKT matrix is used to accomplish it. This is explained in Section V, which also gives a bound on the optimality gap for each step of the barrier method. Numerical examples in Section VII demonstrate fast convergence of the algorithm: 20 to 40 Newton steps are already sufficient to achieve a very high precision (up to the machine precision level), even for large system. Even less steps are required if the secrecy capacity is the only quantity of interest. Section VI demonstrates that significant simplifications in the algorithm are possible for a degraded channel. Section IV gives a brief review of the barrier and Newton methods for inequality-constrained optimization, and presents an algorithm for minimax problems with guaranteed convergence to a global optimum. Section III summarizes the minimax representation of the secrecy capacity on which our algorithm is based. Section II reviews the Gaussian MIMO-WTC model and its secrecy capacity.

II. Wire-Tap Gaussian MIMO Channel Model

Let us consider the standard Gaussian MIMO wire-tap channel model as in Fig. 1,

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

where $\mathbf{x} = [x_1, x_2, \dots x_m]' \in R^{m,1}$ is the (real) transmitted signal vector of dimension $m \times 1$, $'$ denotes transposition, $\mathbf{y}_{1(2)} \in R^{n_{1(2)},1}$ are the (real) received vectors at the receiver (eavesdropper), $\boldsymbol{\xi}_{1(2)}$ is the additive white Gaussian noise at the receiver (eavesdropper) (normalized to unit variance in each dimension), $\mathbf{H}_{1(2)} \in R^{n_{1(2)},m}$ is the $n_{1(2)} \times m$ matrix of the channel gains between each Tx and each receive (eavesdropper) antenna, $n_{1(2)}$ and m are the number 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

[1]For non-convex problems, KKT point can also be a local minimum rather than maximum. This is ruled out in [20] by the non-decreasing nature of the generated sequence of objective values.

[2]A singular KKT matrix would imply that the corresponding Newton step is not defined and thus the algorithm would terminate without converging to a global optimum.

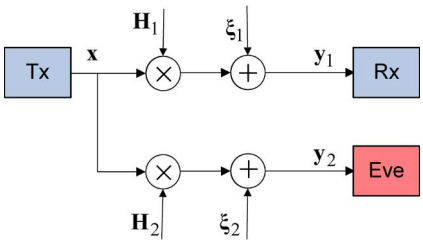

Fig. 1. A block diagram of the Gaussian MIMO wiretap channel. Full channel state information is available at the transmitter. $\mathbf{H}_{1(2)}$ is the channel matrix to the legitimate receiver (eavesdropper); $\mathbf{x}$ is the transmitted signal and $\mathbf{y}_{1(2)}$ is the received (eavesdropper) signal; $\boldsymbol{\xi}_{1(2)}$ is the AWGN at the receiver (eavesdropper). The information leakage to the eavesdropper is required to approach zero asymptotically.

(CSI) at the Rx and Tx ends. A secrecy rate is achievable for this channel if (i) the receiver is able to recover the message with arbitrary low error probability (reliability criterion) and (ii) the information leaked to the eavesdropper approaches zero asymptotically (secrecy criterion) [1].

For a given transmit covariance matrix $\mathbf{R} = E\{\mathbf{x}\mathbf{x}'\}$, where $E\{\cdot\}$ is statistical expectation, the maximum achievable secrecy rate between the Tx and Rx (so that the rate between the Tx and eavesdropper is zero) is [6]–[8]

$$C(\mathbf{R}) = \frac{1}{2}\ln\frac{|\mathbf{I}+\mathbf{W}_1\mathbf{R}|}{|\mathbf{I}+\mathbf{W}_2\mathbf{R}|} = C_1(\mathbf{R}) - C_2(\mathbf{R}) \tag{2}$$

where negative $C(\mathbf{R})$ is interpreted as zero rate, $\mathbf{W}_i = \mathbf{H}_i'\mathbf{H}_i$, 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 and hence very difficult to solve in general and explicit solutions for the optimal Tx covariance is not known for the general case, but only for some special cases, e.g., low/high SNR, MISO channels, full-rank or rank-1 case [5]–[9] or for the parallel channel [17], [18].

Since (3) is not a convex problem in the general case, not only widely-used Karush-Kuhn-Tucker optimality conditions are not sufficient, but also the convergence of a numerical algorithm to a global optimum is very difficult if not impossible to insure since the standard tools of convex optimization fail to work and, in general, non-convex problems are much harder to deal with [23]. Thus, (3) is very difficult to solve either analytically or numerically in the general case. Even when $C(R)$ is concave so that the problem becomes convex (when the channel is degraded, $\mathbf{W}_1 \geq \mathbf{W}_2$), its analytical solution is not known, except for the special cases noted above, and the known convex solvers [30]–[32] are not able to solve the problem, even in this convex setting so that a custom-made algorithm has to be developed.

To go around this difficulty, we use the following minimax representation of the secrecy capacity.

III. Minimax Representation of Secrecy Capacity

A minimax representation of the secrecy capacity was obtained in [6] via a channel enhancement argument and a clever bounding technique, which is instrumental for our algorithm and is summarized below.

Theorem 1 (Theorem 1 in [6]): The secrecy capacity of Gaussian MIMO-WTC channel in (2) can be presented in the following minimax form:

$$C_s = \max_{\mathbf{R}}\min_{\mathbf{K}} f(\mathbf{R},\mathbf{K}) = \min_{\mathbf{K}}\max_{\mathbf{R}} f(\mathbf{R},\mathbf{K}) \tag{4}$$

where

$$f(\mathbf{R},\mathbf{K}) = \frac{1}{2}\ln\frac{|\mathbf{I}+\mathbf{K}^{-1}\mathbf{H}\mathbf{R}\mathbf{H}'|}{|\mathbf{I}+\mathbf{W}_2\mathbf{R}|} \geq C(\mathbf{R}), \tag{5}$$

$$\mathbf{K} = \begin{pmatrix} \mathbf{I} & \mathbf{K}_{21}' \\ \mathbf{K}_{21} & \mathbf{I} \end{pmatrix} \geq \mathbf{0},\ \mathbf{H} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix}, \tag{6}$$

and the optimization is over the set $\mathcal{S}$ of all feasible $\mathbf{R},\ \mathbf{K}$:

$$\mathcal{S} = \{(\mathbf{R},\mathbf{K}) : tr\mathbf{R} \leq P,\ \mathbf{R},\mathbf{K} \geq \mathbf{0},\ \mathbf{K} \text{ is as in (6)}\}. \tag{7}$$

The upper bound in (5) via $f(\mathbf{R},\mathbf{K})$ was obtained from a genie-aided receiver which knows $\mathbf{y}_2$ (in addition to $\mathbf{y}_1$) and $\mathbf{K}$ represents noise covariance between $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. Minimization over $\mathbf{K}$ is due to the fact that the true capacity does not depend on $\mathbf{K}$ while the upper bound does so it's natural to seek the least upper bound. This bound can also be used in a numerical algorithm to evaluate the optimality gap with respect to $\min_{\mathbf{K}}$ for each $\mathbf{R}$. In fact, (4) states that letting the receiver to know $\mathbf{y}_2$ in addition to $\mathbf{y}_1$ does not increase the secrecy capacity under the worst-case noise covariance, which is rather surprising.

Remark 1: 2nd equality in (4) expresses the saddle-point property, which is equivalent to the following inequalities (see e.g. [23], [35]):

$$f(\mathbf{R},\mathbf{K}^*) \leq f(\mathbf{R}^*,\mathbf{K}^*) \leq f(\mathbf{R}^*,\mathbf{K}) \tag{8}$$

which hold for any feasible $\mathbf{R},\mathbf{K}$, where $(\mathbf{R}^*,\mathbf{K}^*)$ is the optimal (saddle) point of (4). These inequalities follow from von Neumann minimax Theorem since $f(\mathbf{K},\mathbf{R})$ is convex in $\mathbf{K}$ for any fixed $\mathbf{R}$ and concave in $\mathbf{R}$ for any fixed $\mathbf{K}$ (and for any channel, degraded or not), and the feasible set in (7) is convex.

Remark 2: It is the convex-concave nature of $f(\mathbf{R},\mathbf{K})$ along with the saddle-point property in (8) and the constraints in (7) that make the respective KKT conditions sufficient for global optimality (see e.g. [23] and [33] for more details; note that Slater's condition holds for these problems). This cannot be said about the original problem in (3). The sufficiency of the KKT conditions is the key for our algorithm and a proof of its convergence to a *global* maximum (rather than just a stationary point).

While the equivalence of (3) and (4) was established in [6], an analytical solution of any one is not known in the general case. In fact, no analytical solution is known for the latter. Despite its more complicated appearance due to two conflicting optimizations, (4) is in fact easier to solve than (3), at least numerically, since both optimizations are convex and the respective KKT conditions are sufficient for global optimality; a proof of convergence of the corresponding numerical algorithm to a *global optimum* is also within reach for *any* channel. While the standard tools developed for single convex optimization [23] do not apply directly here due to two conflicting optimizations involved, their primal-dual reformulation does work, as explained below.

We proceed to solve the minimax problem in (4) via KKT conditions.[3] Subsequently, a numerical algorithm is developed with guaranteed convergence to a global optimum for any channel, degraded or not, which is not possible for (3) due to its non-convex nature in the general case. The Lagrangian for the problem in (4) is

$$L = f(\mathbf{R},\mathbf{K}) - tr\mathbf{M}_1\mathbf{K} + tr\mathbf{M}_2\mathbf{R} - \lambda(tr\mathbf{R} - P) + tr\mathbf{\Lambda}(\mathbf{K}-\mathbf{I}) \quad (9)$$

where $\mathbf{M}_1,\ \mathbf{M}_2 \geq \mathbf{0}$ are (matrix) Lagrange multiplies responsible for the positive semi-definite constraints $\mathbf{K},\mathbf{R} \geq \mathbf{0}, \lambda \geq 0$ is (scalar) Lagrange multiplier responsible for the total power constraint $tr\mathbf{R} \leq P$, and

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix} \quad (10)$$

is a (matrix) Lagrange multiplier responsible for the constraint on $\mathbf{K}$ as in (6). There are two sets of KKT conditions—one per optimization in (4). For the maximization over $\mathbf{R}$, the KKT conditions are (to simplify notations, we have omitted the $\frac{1}{2}$ factor):

$$\begin{aligned} \nabla_R L &= (\mathbf{I}+\mathbf{W}\mathbf{R})^{-1}\mathbf{W} - (\mathbf{I}+\mathbf{W}_2\mathbf{R})^{-1}\mathbf{W}_2 + \mathbf{M}_2 - \lambda\mathbf{I} \\ &= \mathbf{0}, \quad (11)\\ \mathbf{M}_2\mathbf{R} &= \mathbf{0}, \quad (12)\\ tr\mathbf{R} &\leq P,\ \mathbf{R},\mathbf{M}_2 \geq \mathbf{0},\ \lambda \geq 0, \quad (13) \end{aligned}$$

where ∇_R is the gradient (derivative) with respect to $\mathbf{R}$ and $\mathbf{W} = \mathbf{H}'\mathbf{K}^{-1}\mathbf{H}$. The KKT conditions for the minimization over $\mathbf{K}$ are

$$\begin{aligned} \nabla_K L = (\mathbf{K}+\mathbf{Q})^{-1} - \mathbf{K}^{-1} - \mathbf{M}_1 + \mathbf{\Lambda} &= \mathbf{0}, \quad (14)\\ \mathbf{M}_1\mathbf{K} &= \mathbf{0}, \quad (15)\\ \mathbf{K},\mathbf{M}_1 &\geq \mathbf{0}, \quad (16) \end{aligned}$$

and $\mathbf{K},\mathbf{\Lambda}$ are as in (6), (10); $\mathbf{Q} = \mathbf{H}\mathbf{R}\mathbf{H}'$. Here, we implicitly assume that $\mathbf{K} > \mathbf{0}$. While the singular case was treated in a separate way in [6], we do not need a separate treatment here since our numerical algorithm is iterative and, at each step, it produces a non-singular $\mathbf{K}$ which, however, may be arbitrary close to a singular matrix (i.e., may have arbitrary small but positive eigenvalues). This models numerically a case of singular $\mathbf{K}$ and is a standard feature of the barrier method in general, where the boundary of the constraint set can be approached arbitrary closely but never achieved (see e.g. Chapter 11 in [23] for more detail). We remark that negligibly-small eigenvalues can be rounded off to 0 and they also imply that the numerical rank is low.

An optimal point in (4) must satisfy both sets of KKT conditions simultaneously and these conditions are also sufficient for global optimality, as noted above. An analytical solution to these conditions is not known. Our numerical algorithm in Section V solves these two sets of KKT conditions in an iterative way, with guaranteed convergence to a globally-optimal point.

[3]See e.g., [23] for a background on KKT conditions.

IV. Barrier Method for Minimax Optimization

In this section, we first give a brief introduction into Newton and barrier methods for inequality-constrained optimization; the reader is referred to Chapters 9–11 of [23] for more details and background information. These two methods are used as key components to construct an algorithm for minimax optimization. Subsequently, this algorithm is adapted to the secrecy problem in (4) and its guaranteed convergence to a global optimum is proved for any channel (degraded or not) in Section V.

A. Minimax Problem Via Primal-Dual Newton Method

Newton method for an equality-constrained problem essentially transforms the problem into a sequence of quadratic problems for which the sufficient KKT conditions are a system of linear equations [23].

Let us consider the minimax problem of the form[4]

$$\max_{\mathbf{x}} \min_{\mathbf{y}} f(\mathbf{x},\mathbf{y}),\ \text{s.t. } \mathbf{A}_x\mathbf{x} = \mathbf{b}_x,\ \mathbf{A}_y\mathbf{y} = \mathbf{b}_y \quad (17)$$

where vectors $\mathbf{x},\mathbf{y}$ represent optimization variables, the objective $f(\mathbf{x},\mathbf{y})$ is concave in $\mathbf{x}$ and convex in $\mathbf{y}$; given matrices $\mathbf{A}_x,\mathbf{A}_y$ and vectors $\mathbf{b}_x,\mathbf{b}_y$ represent the equality constraints for each variable. The KKT onditions for this problem are

$$\begin{aligned} \nabla_x f + \mathbf{A}_x'\boldsymbol{\lambda}_x = 0,\ \mathbf{A}_x\mathbf{x} - \mathbf{b}_x = 0,\\ \nabla_y f + \mathbf{A}_y'\boldsymbol{\lambda}_y = 0,\ \mathbf{A}_y\mathbf{y} - \mathbf{b}_y = 0, \end{aligned} \quad (18)$$

where $\boldsymbol{\lambda}_x,\boldsymbol{\lambda}_y$ are dual variables, and they are sufficient for global optimality.

While the standard Newton method can be used for both optimizations, a proof of its convergence is challenging since the objective is not monotonous (it decreases in one step and increases at the other). The residual form of the Newton method is preferable since, as it was observed in [23], it reduces the norm of the residual at each step and thus generates a monotonous sequence whose convergence to zero can be guaranteed. To introduce this method, let us aggregate variables, derivatives and parameters as follows:

$$\mathbf{z} = \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix},\ \boldsymbol{\lambda} = \begin{bmatrix}\boldsymbol{\lambda}_x\\ \boldsymbol{\lambda}_y\end{bmatrix},\ \mathbf{b} = \begin{bmatrix}\mathbf{b}_x\\ \mathbf{b}_y\end{bmatrix},\ \mathbf{A} = \begin{bmatrix}\mathbf{A}_x & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_y\end{bmatrix}, \quad (19)$$

$$\nabla f = \begin{bmatrix}\nabla_x f\\ \nabla_y f\end{bmatrix},\ \nabla^2 f = \begin{bmatrix}\nabla^2_{xx} f & \nabla^2_{xy} f\\ \nabla^2_{yx} f & \nabla^2_{yy} f\end{bmatrix}, \quad (20)$$

The KKT conditions in (18) can be cast in a residual form:

$$\mathbf{r} = \left[(\nabla f + \mathbf{A}'\boldsymbol{\lambda})',\ (\mathbf{A}\mathbf{z} - \mathbf{b})'\right]' = \mathbf{0}. \quad (21)$$

The Newton method iteratively solves $\mathbf{r} = \mathbf{0}$ using 1st-order approximation (Newton step):

$$\begin{aligned} \mathbf{r}(\mathbf{w}_0 + \Delta\mathbf{w}) &= \mathbf{r}(\mathbf{w}_0) + D\mathbf{r}\Delta\mathbf{w} + o(\Delta\mathbf{w})\\ &\approx \mathbf{r}(\mathbf{w}_0) + D\mathbf{r}\Delta\mathbf{w} \end{aligned} \quad (22)$$

[4]A similar problem, without equality constraints, have been briefly considered in [23]. More details can be found in [33]. Our development here is tailored to be used for the secrecy problem in (4).

where $\mathbf{w} = [\mathbf{z}', \boldsymbol{\lambda}']$ is the vector of aggregated (primal/dual) variables, $\mathbf{w}_0$ and $\Delta\mathbf{w}$ are its initial value and update, $D\mathbf{r}$ is the derivative of $\mathbf{r}(\mathbf{w})$:

$$D\mathbf{r} = \left[\frac{\partial \mathbf{r}}{\partial \mathbf{z}'}, \frac{\partial \mathbf{r}}{\partial \boldsymbol{\lambda}'}\right] = \begin{bmatrix} \nabla^2 f(\mathbf{z}_0) & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{bmatrix} = \mathbf{T} \tag{23}$$

and $\mathbf{T}$ is the KKT matrix. Now, setting $\mathbf{r}(\mathbf{w}_0 + \Delta\mathbf{w}) = \mathbf{0}$ and solving for $\Delta\mathbf{w}$ from (22) gives the update

$$\Delta\mathbf{w}: \ \mathbf{T}\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w}_0) \tag{24}$$

We further show in Section V that $\mathbf{T}$ is non-singular for our problem so that this system of linear equations is guaranteed to have a unique solution for any set of parameters[5].

Having the steps $\Delta\mathbf{w} = (\Delta\mathbf{z}', \Delta\boldsymbol{\lambda}')'$ computed, the primal/dual variable updates are

$$\mathbf{z} = \mathbf{z}_0 + s\Delta\mathbf{z}, \boldsymbol{\lambda} = \boldsymbol{\lambda}_0 + s\Delta\boldsymbol{\lambda} \tag{25}$$

where the step size s is found via the backtracking line search [23] as in Algorithm 1 below.

Algorithm 1 Backtracking line search

Require: $\mathbf{w}_0, \ 0 < \alpha < 1/2, \ 0 < \beta < 1, \ s = 1.$
while $|\mathbf{r}(\mathbf{w}_0 + s\Delta\mathbf{w})| > (1 - \alpha s)|\mathbf{r}(\mathbf{w}_0)|$ **do** $s := \beta s$
end while

In this Algorithm, α is a % of the linear decrease in the residual one is prepared to accept at each step, and β is a parameter controlling the reduction in step size at each iteration of the algorithm. The Newton method in combination with the backtracking line search is guaranteed to reduce the residual norm $|\mathbf{r}(\mathbf{w})|$ at each step according to the following residual norm-reduction property [23]:

$$\frac{d}{ds}|\mathbf{r}(\mathbf{w}_0 + s\Delta\mathbf{w})| = -|\mathbf{r}(\mathbf{w}_0)| < 0, \tag{26}$$

so that, for sufficiently small s, the residual indeed shrinks at each iteration (unless $|\mathbf{r}(\mathbf{w}_0)| = 0$, which implies that $\mathbf{w}_0$ is optimal). This insures convergence of the algorithm to a global optimum since KKT conditions are sufficient for optimality and any locally-optimal point is automatically globally-optimal as the problem is convex.

Based on this, the Newton method for minimax optimization is as in Algorithm 2. The convergence of this algorithm to a global optimum is insured by the convex/concave nature of the objective, sufficiency of the KKT conditions in (18), non-singularity of the KKT matrix $\mathbf{T}$ at each step (as proved in Section V) and the norm-decreasing residual property in (26), which ensures that the method generates a sequence of sub-optimal solutions with monotonically decreasing residuals, for which the stationary point has zero residual and thus solves the sufficient KKT conditions. While the global optimum point corresponds to zero residual, $|\mathbf{r}| = 0$ (this is equivalent to the KKT conditions in (18)), the practical version $|\mathbf{r}| \leq \epsilon$ of this condition is used in Algorithm 2 as a stopping criterion. This form of the stopping criteria is justified by not only the residual form $|\mathbf{r}| = 0$ of the KKT conditions, but also by the norm-decreasing residual property in (26).

[5]While $\Delta\mathbf{w} = -\mathbf{T}^{-1}\mathbf{r}(\mathbf{w}_0)$ is its analytical solution, it is not computed in practice since computing $\mathbf{T}^{-1}$ is computationally-expensive and may result in loss of accuracy for ill-conditioned $\mathbf{T}$, see e.g., [26].

Algorithm 2 Newton method for minimax optimization

Require: $\mathbf{z}_0, \ \boldsymbol{\lambda}_0, \ \alpha, \ \beta, \ \epsilon$
repeat

1. Find $\Delta\mathbf{z}, \ \Delta\boldsymbol{\lambda}$ using Newton step in (24).
2. Find s using the backtracking line search (Algorithm 1).
3. Update variables: $\mathbf{z}_{k+1} = \mathbf{z}_k + s\Delta\mathbf{z}, \ \boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + s\Delta\boldsymbol{\lambda}$.

until $|\mathbf{r}(\mathbf{z}_{k+1}, \boldsymbol{\lambda}_{k+1})| \leq \epsilon$.

As a side remark, we note that this algorithm can also be used to solve the problem in (17) with max and min interchanged, due to the saddle point property.

B. Barrier Method for Inequality-Constrained Problems

Let us now combine the barrier method and the minimax method above to construct an algorithm for minimax optimization with equality and inequality constraints. Consider the following problem with inequality constraints:

$$\max_{\mathbf{x}} \min_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}), \ \text{s.t.} \ \mathbf{A}_x\mathbf{x} = \mathbf{b}_x, \ \mathbf{A}_y\mathbf{y} = \mathbf{b}_y, \ f_1(\mathbf{x}) \leq 0, \ f_2(\mathbf{y}) \leq 0 \tag{27}$$

where f_1 and f_2 are the constraint functions. The key idea of the barrier method is to use a soft instead of hard constraints by augmenting the objective with the barrier functions responsible for the inequality constraints so that the new objective for the problem in (27) becomes:

$$f_t(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) + \psi_t\left(f_1(\mathbf{x})\right) - \psi_t\left(f_2(\mathbf{y})\right) \tag{28}$$

where we use the logarithmic barrier function:

$$\psi_t(x) = \frac{1}{t}\ln(-x) \tag{29}$$

and where t is the barrier parameter. The barrier method transforms the inequality-constrained problem in (27) into the following problem without inequality constraints:

$$\max_{\mathbf{x}} \min_{\mathbf{y}} f_t(\mathbf{x}, \mathbf{y}), \ \text{s.t.} \ \mathbf{A}_x\mathbf{x} = \mathbf{b}_x, \ \mathbf{A}_y\mathbf{y} = \mathbf{b}_y \tag{30}$$

The optimality gap due to this transformation can be upper bounded as follows.

Proposition 1: The optimality gap of the barrier method in (30) applied to the minimax problem in (27) is as follows:

$$|f\left(\mathbf{x}^*(t), \mathbf{y}^*(t)\right) - p^*| \leq 1/t \tag{31}$$

where p^* is an optimal value of the original problem in (27) and $(\mathbf{x}^*(t), \mathbf{y}^*(t))$ is an optimal point for the modified problem in (30).

Proof: This is a special case of Proposition 3 below with $m = n_1 = n_2 = 1$. □

Thus, by selecting sufficiently high t, one can obtain arbitrary small gap. Newton method is used to solve the modified problem with any desired accuracy.

In practice, the modified problem is solved in an iterative way by selecting first a moderately-large value of t, solving the problem, increasing t and using the previous solution as a starting point for a new one. In this way, the total number of Newton steps required to achieve certain accuracy is minimized [23]. The algorithm is as follows.

V. Barrier Method for Secrecy Rate Maximization

In this section, we use the minimax barrier method above to solve the optimal covariance problem in (4) iteratively with guaranteed convergence to a global optimum, which is also optimal for (3).

A. Choice of Variables

Since the original variables are positive semi-definite matrices $\mathbf{R}, \mathbf{K}$ and the barrier method above requires vectors, we have two options:

1. Use all entries of $\mathbf{R}, \mathbf{K}$ as independent variables via $\mathbf{x} = vec(\mathbf{R}),\ \mathbf{y} = vec(\mathbf{K})$, where operator vec stacks all columns into a single vector. Enforce the symmetry constraints $\mathbf{R}' = \mathbf{R},\ \mathbf{K}' = \mathbf{K}$ and the equality constraint on $\mathbf{K}$ in (6) via extra equality constraints.
2. Use only lower-triangular entries of $\mathbf{R}$ as independent variables via $\mathbf{x} = vech(\mathbf{R})$, where $vech$ stacks column-wise all lower-triangular entries into a single column vector, and use only $\mathbf{K}_{21}$: $\mathbf{y} = vec(\mathbf{K}_{21})$.

It can be shown that these two options are mathematically equivalent, i.e. produce exactly the same solutions at each step of Newton method. Option 2 is a preferable choice for implementation since the number of variables and constraints is reduced so that it is more efficient. Therefore, we use Option 2 for further exposition. Gradient and Hessian can be evaluated either numerically (in a standard way) or analytically as given below. We find the analytical evaluation to be preferable as numerical one entails a loss of precision while approaching an optimal point (this is especially pronounced at high SNR, large t and for large systems).

Since the algorithm requires initial point to begin with, we use the following point:

$$\mathbf{R}_0 = \frac{P}{m}\mathbf{I} \to \mathbf{x}_0 = vech(\mathbf{R}_0), \tag{32}$$

$$\mathbf{K}_0 = \mathbf{I} \to \mathbf{y}_0 = \mathbf{0}, \tag{33}$$

$$\boldsymbol{\lambda}_0 = \mathbf{0} \tag{34}$$

As can be easily verified, the initial point above is feasible (i.e. satisfies the constraints). The choice of $\mathbf{R}_0$ is motivated by the fact that isotropic signalling does not prefer any direction and thus is equally good a priori for any channel. $\mathbf{K}_0$ corresponds to isotropic noise and is motivated by the same reason. It should be emphasized that the algorithm converges for any (feasible) initial point, due to the convex nature of the problem, to a global optimum; the difference is in how fast.

To account for the positive semi-definite constraints $\mathbf{R}, \mathbf{K} \geq \mathbf{0}$, the following barrier function is used

$$\psi_t(\mathbf{R}) = \frac{1}{t}\ln|\mathbf{R}| \tag{35}$$

so that the modified objective f_t is

$$f_t(\mathbf{R}, \mathbf{K}) = f(\mathbf{R}, \mathbf{K}) + \psi_t(\mathbf{R}) - \psi_t(\mathbf{K}) \tag{36}$$

Note that this requires $\mathbf{K}, \mathbf{R} > \mathbf{0}$, i.e. they are strictly inside of the feasible set but can approach the boundary arbitrary closely as t increases, so that some eigenvalues may become arbitrary close to zero (and the numerical rank may be deficient); this models numerically the case of singular $\mathbf{R}$ and/or $\mathbf{K}$ and is a standard feature of the barrier method in general [23]. The inequality in (53) makes sure that the optimality gap due to this can be made as small as desired. In a practical implementation, one can round off negligibly-small eigenvalues of $\mathbf{R}$ to zero to simplify implementation.

After some manipulations (see Appendix for details), the gradients and Hessians can be expressed as:

$$\nabla_x f_t = \mathbf{D}_m' vec(\nabla_R f_t),\ \nabla_y f_t = \widetilde{\mathbf{D}}_n' vec(\nabla_K f_t), \tag{37}$$

$$\nabla^2_{xx} f_t = -\mathbf{D}_m'(\mathbf{Z}_1 \otimes \mathbf{Z}_1 - \mathbf{Z}_2 \otimes \mathbf{Z}_2 + t^{-1}\mathbf{R}^{-1} \otimes \mathbf{R}^{-1})\mathbf{D}_m, \tag{38}$$

$$\nabla^2_{yy} f_t = \widetilde{\mathbf{D}}_n' \left(-(\mathbf{K}+\mathbf{Q})^{-1} \otimes (\mathbf{K}+\mathbf{Q})^{-1} + (1+t^{-1})\mathbf{K}^{-1} \otimes \mathbf{K}^{-1}\right) \widetilde{\mathbf{D}}_n, \tag{39}$$

$$\nabla^2_{xy} f_t = -\mathbf{D}_m' \left(\mathbf{H}'(\mathbf{K}+\mathbf{Q})^{-1} \otimes \mathbf{H}'(\mathbf{K}+\mathbf{Q})^{-1}\right) \widetilde{\mathbf{D}}_n, \tag{40}$$

where

$$\nabla_R f_t = \mathbf{Z}_1 - \mathbf{Z}_2 + t^{-1}\mathbf{R}^{-1}, \tag{41}$$

$$\nabla_K f_t = (\mathbf{K}+\mathbf{Q})^{-1} - (1+t^{-1})\mathbf{K}^{-1}, \tag{42}$$

$$\mathbf{Z}_1 = (\mathbf{I} + \mathbf{W}\mathbf{R})^{-1}\mathbf{W}, \tag{43}$$

$$\mathbf{Z}_2 = (\mathbf{I} + \mathbf{W}_2\mathbf{R})^{-1}\mathbf{W}_2, \tag{44}$$

and $\otimes$ is a Kronecker product, $\mathbf{D}_m$ is a $m^2 \times m(m+1)/2$ duplication matrix defined from $vec(\mathbf{R}) = \mathbf{D}_m vech(\mathbf{R})$ [27], [28], $\widetilde{\mathbf{D}}_n$ is a $n^2 \times n_1 n_2$ reduced duplication matrix defined from $d\mathbf{k} = \widetilde{\mathbf{D}}_n d\tilde{\mathbf{k}}$, where

$$d\mathbf{k} = vec(d\mathbf{K}),\ d\tilde{\mathbf{k}} = vec(d\mathbf{K}_{21}),$$

$$d\mathbf{K} = \begin{pmatrix} \mathbf{0} & d\mathbf{K}_{21}' \\ d\mathbf{K}_{21} & \mathbf{0} \end{pmatrix} \tag{45}$$

and $n = n_1 + n_2$. It can be obtained from $\mathbf{D}_n$ by removing its columns corresponding to all entries of $\mathbf{K}$ but those in $\mathbf{K}_{21}$.

It can be shown (see e.g. [14]) that using the full available power is optimal. Therefore, one can use the equality constraint $tr\mathbf{R} = P$ instead of the inequality $tr\mathbf{R} \leq P$. The equality constraint matrix $\mathbf{A}$ and vector $\mathbf{b}$ take the following form:

$$\mathbf{A} = [\mathbf{a}', \mathbf{0}'],\ \mathbf{b} = P \tag{46}$$

where $\mathbf{I}_m$ is $m \times m$ identity matrix, $\mathbf{a} = vech(\mathbf{I}_m)$, and $\mathbf{0}$ is $n_1 n_2 \times 1$ zero vector, i.e. $\mathbf{A}$ is a row vector and $\mathbf{b}$ is a scalar in our setting.

With this choice of variables and initial points, Algorithm 3, in combinations with Algorithms 1 and 2, can now be used to solve numerically the minimax problem in (4).

Algorithm 3 Barrier Method

Require: $\mathbf{z},\ \boldsymbol{\lambda},\ \epsilon > 0,\ t > 0,\ \mu > 1$

repeat

1. Solve the problem in (30) using Newton method (Algorithm 2) starting at $\mathbf{z},\ \boldsymbol{\lambda}$.
2. Update variables: $\mathbf{z} := \mathbf{z}^*(t),\ \boldsymbol{\lambda} := \boldsymbol{\lambda}^*(t),\ t := \mu t$.

until $1/t < \epsilon$.

B. Convergence of the Algorithm

Here, we provide a proof of convergence of the proposed algorithm to a global optimum. First, one has to insure that Newton step is well defined for all $t, \mathbf{R}, \mathbf{K} > \mathbf{0}$. This, in turn, insures that the Newton method produces a sequence of decreasing-norm residuals (according to (26)), which converge to zero for each t. Consequently, the minimax barrier method applied to our problem generates a sequence of sub-optimal points $\mathbf{z}^*(t)$ that converges to a global optimum (a solution of the sufficient KKT conditions in (11)–(16)) as t increases, since $f_t(\mathbf{R}, \mathbf{K})$ is convex in $\mathbf{K}$ and concave in $\mathbf{R}$ and also twice continuously differentiable for each $\mathbf{R} > \mathbf{0},\ \mathbf{K} > \mathbf{0}$ (more details can be found in [23]).

To make sure that Newton step is well defined for each $t, \mathbf{R}, \mathbf{K} > \mathbf{0}$, we demonstrate that the KKT matrix for the modified objective f_t is non-singular, so that the Newton equations have a well-defined solution as in (24).

Proposition 2: Consider the minimax problem in (17) for the objective in (36) under the equality constraint parameters as in (46). Its KKT matrix

$$\mathbf{T} = \begin{bmatrix} \nabla^2 f_t & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \tag{47}$$

is non-singular for each $t > 0,\ \mathbf{R}, \mathbf{K} > \mathbf{0}$.

Proof: The proof is based on the following three Lemmas.

Lemma 1: The Hessian

$$\nabla^2 f_t = \breve{\mathbf{H}} = \begin{bmatrix} -\mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \tag{48}$$

is non-singular if partial Hessians $\mathbf{H}_{11},\ \mathbf{H}_{22}$ are non-singular, i.e. if $\mathbf{H}_{11},\ \mathbf{H}_{22} > \mathbf{0}$, where $\mathbf{H}_{11} = -\nabla^2_{xx} f_t,\ \mathbf{H}_{12} = \nabla^2_{xy} f_t,\ \mathbf{H}_{21} = \mathbf{H}'_{12} = \nabla^2_{yx} f_t,\ \mathbf{H}_{22} = \nabla^2_{yy} f_t$. Furthermore, block (1,1) $[\breve{\mathbf{H}}^{-1}]_{11}$ of the inverse $\breve{\mathbf{H}}^{-1}$ is also non-singular.

Proof: The proof is complicated by the fact that $\nabla^2 f_t$ is indefinite matrix, since f_t is concave in $\mathbf{x}$ and convex in $\mathbf{y}$ (i.e. $\nabla^2_{xx} f_t \le \mathbf{0},\ \nabla^2_{yy} f_t \ge \mathbf{0}$), so that the standard proofs tailored for positive definite matrices [26] do not apply here. However, since $\mathbf{H}_{11},\ \mathbf{H}_{22} > \mathbf{0}$, it follows that

$$\begin{aligned} \mathbf{S}_{22} &= -\mathbf{H}_{11} - \mathbf{H}'_{21}\mathbf{H}_{22}^{-1}\mathbf{H}_{21} < \mathbf{0}, \\ \mathbf{S}_{11} &= \mathbf{H}_{22} + \mathbf{H}_{21}\mathbf{H}_{11}^{-1}\mathbf{H}'_{21} > \mathbf{0}, \end{aligned} \tag{49}$$

where $\mathbf{S}_{11(22)}$ is Schur complement of $-\mathbf{H}_{11}(\mathbf{H}_{22})$, so that the matrix inversion Lemma in Proposition 2.8.7 of [34] applies and one can invert $\breve{\mathbf{H}}$ as follows[6]

$$\begin{aligned} \breve{\mathbf{H}}^{-1} &= \begin{bmatrix} -\mathbf{H}_{11} & \mathbf{H}'_{21} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{S}_{22}^{-1} & -\mathbf{S}_{22}^{-1}\mathbf{H}'_{21}\mathbf{H}_{22}^{-1} \\ \mathbf{S}_{11}^{-1}\mathbf{H}_{21}\mathbf{H}_{11}^{-1} & \mathbf{S}_{11}^{-1} \end{bmatrix} \end{aligned} \tag{50}$$

which implies that $\breve{\mathbf{H}}$ is non-singular and that $[\breve{\mathbf{H}}^{-1}]_{11} = \mathbf{S}_{22}^{-1} < \mathbf{0}$. □

Lemma 2: The KKT matrix in Proposition 2 is non-singular under the conditions of Lemma 1.

Proof: We proceed as follows. Since the Hessian $\nabla^2 f_t = \breve{\mathbf{H}}$ is non-singular (under conditions of Lemma 1), let us apply the following transformation that preserves the determinant of $\mathbf{T}$:

$$\begin{aligned} \widetilde{\mathbf{T}} &= \begin{bmatrix} \breve{\mathbf{H}} & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\breve{\mathbf{H}}^{-1}\mathbf{A}' \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \breve{\mathbf{H}} & \mathbf{0} \\ \mathbf{A} & -\mathbf{A}\breve{\mathbf{H}}^{-1}\mathbf{A}' \end{bmatrix}, \end{aligned} \tag{51}$$

and observe that

$$|\widetilde{\mathbf{T}}| = |\mathbf{T}| = |\breve{\mathbf{H}}|(-\mathbf{A}\breve{\mathbf{H}}^{-1}\mathbf{A}') \tag{52}$$

(this follows from the properties of block-partitioned matrices and their determinants, see e.g., [29]). From Lemma 1, $|\breve{\mathbf{H}}| \neq 0$. Further notice that $\mathbf{A}\breve{\mathbf{H}}^{-1}\mathbf{A}' = \mathbf{a}'[\breve{\mathbf{H}}^{-1}]_{11}\mathbf{a} < 0$, since $[\breve{\mathbf{H}}^{-1}]_{11} < \mathbf{0}$ from Lemma 1 and $\mathbf{a} \neq 0$. Using (52), $|\mathbf{T}| = |\breve{\mathbf{H}}|(-\mathbf{A}\breve{\mathbf{H}}^{-1}\mathbf{A}') \neq 0$ so that the KKT matrix $\mathbf{T}$ is non-singular. □

Thus, Lemmas 1 and 2 establish the non-singularity of KKT matrix provided that partial Hessians $\nabla^2_{xx} f_t,\ \nabla^2_{yy} f_t$ are non-singular. This is indeed the case as Lemma 3 below shows.

Lemma 3: Partial Hessian $\nabla^2_{xx} f_t,\ \nabla^2_{yy} f_t$ in (38) and (39) are non-singular for each $t > 0,\ \mathbf{R}, \mathbf{K} > \mathbf{0}$.

Proof: See Appendix. □

Combining Lemmas 1–3, Proposition 2 follows. □

Thus, Proposition 2 insures that Newton step is always well-defined and hence generates a sequence of decreasing-norm residuals (according to (26)) which converges to zero for each $t > 0$. The next proposition specifies the optimality gap of the minimax barrier method for a given t.

Proposition 3: For each $t > 0$, the optimality gap of the barrier method applied to the minimax problem in (4) can be upper bounded as follows:

$$|f(\mathbf{R}^*(t), \mathbf{K}^*(t)) - C_s| \le \max(m, n_1 + n_2)/t \tag{53}$$

[6]This idea of the proof was suggested by a reviewer.

where $\mathbf{R}^*(t), \mathbf{K}^*(t)$ are the optimal signal and noise covariance matrices returned by the barrier method for a given t.

Proof: Using the bounds for the minimax problem in [33] and adopting them to the problem in (4), one obtains

$$\begin{aligned} \max_R f\left(\mathbf{R}, \mathbf{K}^*(t)\right) - m/t &\le f\left(\mathbf{R}^*(t), \mathbf{K}^*(t)\right) \qquad (54)\\ &\le \min_K f\left(\mathbf{R}^*(t), \mathbf{K}\right) + (n_1+n_2)/t \end{aligned}$$

so that

$$\begin{aligned} f\left(\mathbf{R}^*(t), \mathbf{K}^*(t)\right) &\le \min_K f\left(\mathbf{R}^*(t), \mathbf{K}\right) + (n_1+n_2)/t\\ &\le \max_R \min_K f(\mathbf{R}, \mathbf{K}) + (n_1+n_2)/t\\ &= C_s + (n_1+n_2)/t, \qquad (55)\\ f\left(\mathbf{R}^*(t), \mathbf{K}^*(t)\right) &\ge \max_R f\left(\mathbf{R}, \mathbf{K}^*(t)\right) - m/t \qquad (56)\\ &\ge \min_K \max_R f(\mathbf{R}, \mathbf{K}) - m/t = C_s - m/t \end{aligned}$$

from which (53) follows. □

Therefore, using sufficiently large barrier parameter t insures any desired accuracy, and $f(\mathbf{R}^*(t), \mathbf{K}^*(t)) \to C_s$ as $t \to \infty$. If desired accuracy is ϵ, then the stopping criterion in Algorithm 3 should be $\max(m, n_1+n_2)/t < \epsilon$ (assuming that the Newton method produces sufficiently-accurate solution, which is always the case in practice due to its quadratic convergence, see [23]).

C. Dual Problem

While the algorithm above is designed to maximize the secrecy rate, its optimal covariance also solves the dual problem of minimizing the total transmit power subject to the secrecy rate constraint $C(\mathbf{R}) \ge R_s$, i.e.

$$\min\ tr\mathbf{R}\ \text{ s.t. }\ C(\mathbf{R}) \ge R_s,\ \mathbf{R} \ge 0 \qquad (57)$$

This can be easily shown by contradiction and observing that 1st inequality in (57) always holds with equality, or by comparing the respective KKT conditions (which are necessary for optimality in both problems), both under the condition $R_s = C_s$.

D. Per-Antenna Power Constraints

Different forms of power constraint can also be incorporated into the proposed algorithm in a straightforward way. In particular, the per-antenna power constraint $r_{ii} \le P_i$, where r_{ii} is i-th diagonal entry of $\mathbf{R}$ (power in antenna i) and P_i is the maximum power of i-th antenna, can be adopted by eliminating matrix $\mathbf{A}$ from the KKT equations and adding m extra barrier terms $t^{-1}\ln(P_i - r_{ii})$ representing new power constraints in (36). As a starting point, one can use e.g. $r_{ii} = P_i/2$.

In fact, these new constraints can be added to the existing ones as well, representing the scenario where not only the total power budget is limited but also the per-antenna powers are limited due to e.g. limited dynamic range of power amplifiers.

The convergence of this modified algorithm to a global optimum can be proved in the same way as above (with minor modifications). In particular, one can observe that the new barrier terms preserve the non-singularity of the KKT matrix and the convex nature of the problem.

VI. Degraded Channel

If the channel is degraded, $\mathbf{W}_1 \ge \mathbf{W}_2$, then $C(\mathbf{R})$ is concave and the corresponding optimization problem in (3) is convex. Therefore, the barrier method can be applied directly to this problem with guaranteed convergence to a global optimum. This reduces the problem complexity since there is no minimization over $\mathbf{K}$ so that the number of variables reduces from $m(m+1)/2 + n_1 n_2$ to $m(m+1)/2$, which is a significant improvement when $n_1 n_2$ is large.

The modified objective (with the barrier term) becomes

$$f_t(\mathbf{R}) = C(\mathbf{R}) + \psi_t(\mathbf{R}), \qquad (58)$$

the variables are $\mathbf{z} = \mathbf{x} = vech(\mathbf{R})$ (no $\mathbf{y}$) and the equality constraint parameters are

$$\mathbf{A} = \mathbf{a}' = vech(\mathbf{I}),\ \mathbf{b} = P, \qquad (59)$$

Non-singularity of the KKT matrix, which guarantees well-defined Newton steps, can be established following the lines of the analysis in Section V. In particular, one observes that Lemmas 1–3 hold. Lemma 3 holds since

$$\nabla^2_{xx} f_t < \mathbf{0} \qquad (60)$$

Lemma 1 holds since the Hessian in this case is $\breve{\mathbf{H}} = \nabla^2_{xx} f_t$. Lemma 2 holds since

$$\mathbf{a}'\breve{\mathbf{H}}^{-1}\mathbf{a} < 0 \qquad (61)$$

so that the KKT matrix is non-singular and thus KKT conditions have a well-defined solution at each step of the barrier method.

The optimality gap in this case becomes

$$\left|C\left(\mathbf{R}^*(t)\right) - C_s\right| \le m/t \qquad (62)$$

where $\mathbf{R}^*(t)$ is an optimal $\mathbf{R}$ returned by the Newton method for a given t, i.e. it is smaller for the same t than in the non-degraded case (53), which is an extra advantage (in addition to having less variables). For desired accuracy ϵ, the stopping criterion in Algorithm 3 is $m/t < \epsilon$.

As a side remark, we note that even though the problem is convex in this case, existing convex solvers (see e.g. [30]–[32]) cannot be used to solve it directly since they do not allow difference of logarithms or matrix powers in objective/constraint functions, while the algorithm above solves it with guaranteed convergence to a global optimum.

VII. Numerical Experiments

To validate the algorithm and analysis and to demonstrate the performance of the algorithm, extensive numerical experiments have been carried out. Some of the representative results are shown below.

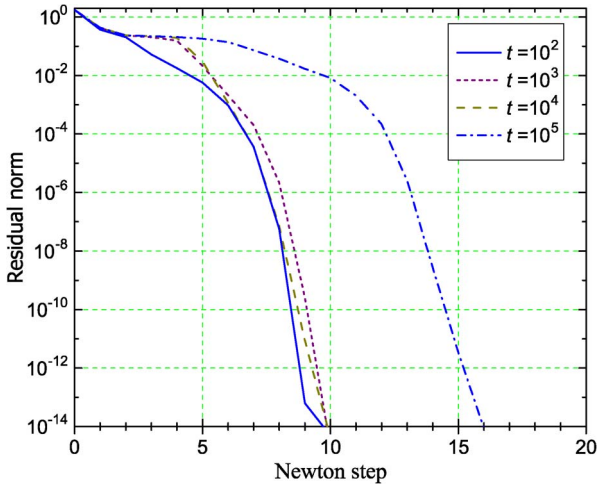

Fig. 2. Convergence of the Newton method for different values of $t; m = 2, P = 10, \alpha = 0.3, \beta = 0.5, \mathbf{H}_1, \mathbf{H}_2$ as in (63). Note the presence of two convergence phases: linear and quadratic. It takes only about 10 to 20 Newton steps to reach the machine precision level.

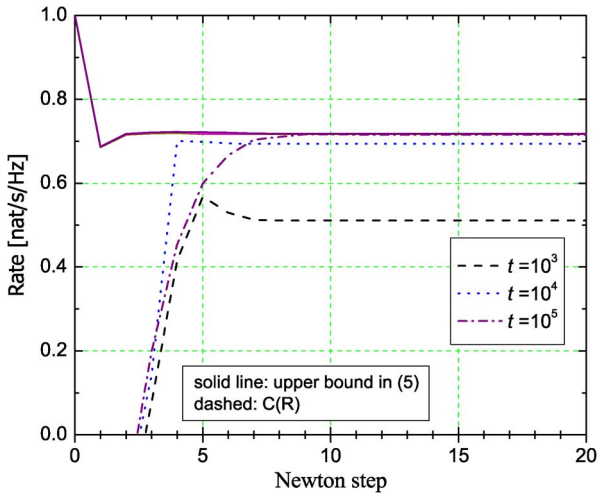

Fig. 3. Secrecy rates for the same setting as in Fig. 2. Solid line—via the upper bound in (5) (the lines coincide for different t), dashed—via $C(\mathbf{R})$ in (2).

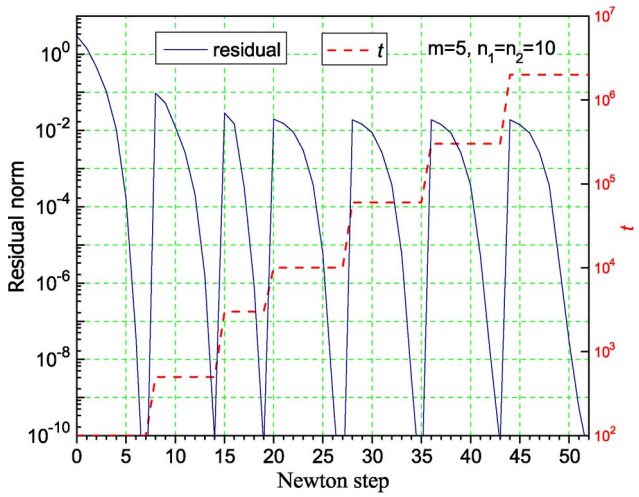

Fig. 4. Convergence of the barrier method (incrementally increasing t); $m = 5, n_1 = n_2 = 10, P = 10, \alpha = 0.3, \beta = 0.5, \mu = 5, \mathbf{H}_1, \mathbf{H}_2$ are randomly generated (i.i.d. Gaussian entries of zero mean and unit variance). It takes about 5 to 10 steps to reduce the residual to a very low value of 10^{-10} for each t.

Convergence of the Newton method for different values of the barrier parameter t is demonstrated in Fig. 2 for

$$\mathbf{H}_1 = \begin{bmatrix} 0.77 & -0.30 \\ -0.32 & -0.64 \end{bmatrix}, \quad \mathbf{H}_2 = \begin{bmatrix} 0.54 & -0.11 \\ -0.93 & -1.71 \end{bmatrix}, \tag{63}$$

which shows the residual $\mathbf{r}$ Euclidian norm versus Newton steps. Even though this channel is not degraded, since the eigenvalues of $\mathbf{W}_1 - \mathbf{W}_2$ are $\{0.395, -3.293\}$, the algorithm does find the global optimum (this particular channel was selected because it is "difficult" for optimization). Note the presence of two convergence phases: linear and quadratic, which is typical for Newton method in general. After the quadratic phase is reached, the convergence is very fast (water-fall region). It takes about 10–20 Newton steps to reach very low residual (at the level of machine precision). This is in agreement with the observations in [23] (although obtained for different problems).

Fig. 3 shows the corresponding secrecy rate evaluated via the upper bound in (5) and the actual achievable rate via $C(\mathbf{R}(t))$ in (2), where $\mathbf{R}(t)$ is an optimal covariance at a particular step of the Newton method and for a given t. As the algorithm converges, they become almost equal if t is sufficiently large (in this case, about $10^4 \ldots 10^5$). While t has negligible impact on the upper bound, it does affect significantly the corresponding $C(\mathbf{R}(t))$ (since the optimal covariance $\mathbf{R}(t)$ returned by the barrier method depends on t and $C(\mathbf{R})$ is sensitive to $\mathbf{R}$), so that the choice of t is not critical if the secrecy capacity is the only quantity of interest (since the upper bound is quite tight even for moderate t). However, if a transmitter is implemented with the optimal covariance $\mathbf{R}(t)$ returned by the algorithm, it is $C(\mathbf{R}(t))$ that determines the achievable rate and this choice is important. We attribute this fact to higher sensitivity of $C(\mathbf{R})$ to $\mathbf{R}$ compared to that of $f(\mathbf{R}, \mathbf{K})$. Similar observations apply to the number of Newton steps required to achieve a certain performance: if C_s is the quantity of interest, the upper bound converges to it in about 3–5 steps. However, if implementing $\mathbf{R}$ is involved, one should use $C(\mathbf{R})$ and, in addition to proper choice of t, it takes about $5 \ldots 10$ steps to achieve the convergence. Note that, in both cases, the number of steps is not large and the execution time is small (a few seconds). In general, larger t and m, n_1, n_2 require more steps to achieve the same accuracy. As expected, the behavior of upper bound is not monotonic while the residual norm does decrease monotonically in each step.

Figs. 4 and 5 demonstrate the convergence of the minimax barrier method (incrementally increasing t) for a larger system $(m = 5, n_1 = n_2 = 10)$. Note that a very low residual value of 10^{-10} is achieved after about 7 Newton steps for each value of t. Using incrementally-increasing t as opposed to a fixed large value results in a smaller number of the total Newton steps required to achieve a given residual value and is less sensitive to system parameters and size. Also observe from Fig. 5 that while the upper bound converges quite fast

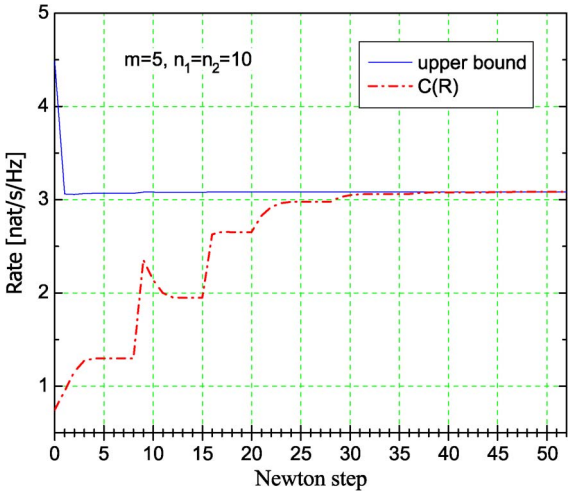

Fig. 5. Secrecy rates for the same setting as in Fig. 4. Solid line—via the upper bound in (5), dashed—via $C(\mathbf{R})$ in (2). Note that while the capacity value evaluated via the upper bound converges very fast, significantly more iterations are required for convergence of the secrecy rate $C(\mathbf{R})$. We attribute this to the fact that $C(\mathbf{R})$ is more sensitive to $\mathbf{R}$ than $f(\mathbf{R},\mathbf{K})$ is. Also note the significantly non-monotonique behavior of the former.

(in a few Newton steps), it takes significantly more steps for $C(\mathbf{R})$ to converge and the convergence process is significantly non-monotonic.

To demonstrate the convergence performance for different channel realizations, Figs. 6 and 7 show the distribution (histograms) of the number of steps required to achieve the residual of 10^{-10} and 10^{-8} for 100 randomly-generated channels (with i.i.d. Gaussian entries of zero mean and unit variance) for $m=4, n_1=n_2=3$ and $m=5, n_1=n_2=10$ systems. While the actual number of required steps depends on a particular channel realization, 20 to 40 steps are sufficient in most cases. We attribute this to the two-phase behaviour of the algorithm's convergence: once the quadratic (water-fall) phase is reached, it takes just a few steps to reduce the residual to a very low value (which is consistent with similar observations in [23], albeit for different problems). Different channel realizations result in a different number of required steps for the linear phase, before the quadratic phase is reached, but do not affect much the latter.

VIII. Conclusion

Global secrecy rate maximization for (non-degraded) Gaussian MIMO-WTC has been discussed. The problem is challenging due to its non-convex nature and no analytical solution is known for this setting. While the known numerical algorithms converge to a stationary point (which may be a local rather than global maximum or just a saddle point), the algorithm proposed herein is guaranteed to converge to a *global* rather than *local* maximum. The algorithm is based on the minimax reformulation of the secrecy capacity problem (to insure global convergence) and the primal-dual reformulation of the Newton method in combination with the barrier method. A proof of its global convergence is also given. Numerical experiments indicate that 20 to 40 Newton steps are sufficient for convergence with high precision (up to the machine precision

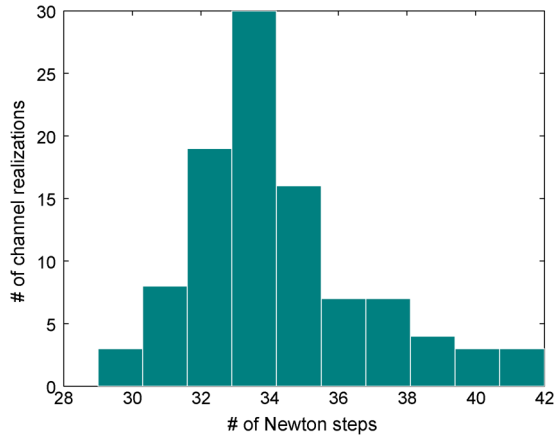

Fig. 6. A histogram showing the distribution of the number of Newton steps needed to achieve the residual of 10^{-10} via the minimax barrier method for 100 randomly generated channels (i.i.d. Gaussian entries of zero mean and unit variance); $P=10, \alpha=0.3, \beta=0.5, m=4, n_1=n_2=3, t_0=100, t_{max}=10^5, \mu=10$.

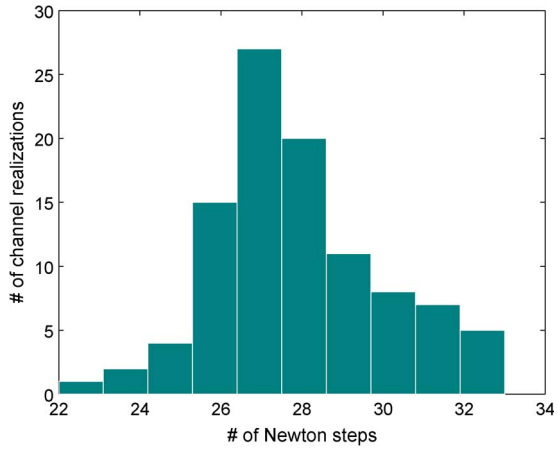

Fig. 7. A histogram showing the distribution of the number of Newton steps needed to achieve the residual of 10^{-8} for 100 randomly generated channels (i.i.d. Gaussian entries of zero mean and unit variance); $m=5, n_1=n_2=10, t_0=100, t_{max}=10^5, \mu=10, P=10, \alpha=0.3, \beta=0.5$.

level). Extra power constraints (e.g., maximum per-antenna power) can be easily incorporated in the algorithm. The dual problem of total power minimization subject to the secrecy rate constraint can also be solved.

Appendix

A. Gradients and Hessians

To derive the gradient and Hessian expressions, we use the tools of matrix differential calculus [27], [28]. Let us consider $f(\mathbf{X})=\ln|\mathbf{X}|$, where $\mathbf{X}>\mathbf{0}$ is $n\times n$ positive definite matrix. Using the perturbation method,

$$\begin{aligned} f(\mathbf{X}+d\mathbf{X}) &= \ln|\mathbf{X}|+\ln|\mathbf{I}+\mathbf{X}^{-1}d\mathbf{X}| \\ &= f(\mathbf{X})+\sum_i \lambda_i(\mathbf{X}^{-1}d\mathbf{X}) \\ &\quad -\frac{1}{2}\lambda_i^2(\mathbf{X}^{-1}d\mathbf{X})+o\left(\lambda_i^2\right) \\ &= f(\mathbf{X})+tr(\mathbf{X}^{-1}d\mathbf{X}) \\ &\quad -\frac{1}{2}tr(\mathbf{X}^{-1}d\mathbf{X}\mathbf{X}^{-1}d\mathbf{X})+o\left(\left\{\lambda_i^2\right\}\right) \end{aligned} \tag{64}$$

Using

$$tr(\mathbf{X}^{-1}d\mathbf{X}) = vec(d\mathbf{X})'vec(\mathbf{X}^{-1}) = d\mathbf{x}'\mathbf{D}_n'vec(\mathbf{X}^{-1}) \quad (65)$$

where $d\mathbf{x} = vech(d\mathbf{X})$, one obtains the gradient $\nabla_x f = \mathbf{D}_n'vec(\mathbf{X}^{-1})$. Applying this to

$$f(\mathbf{K}) = \ln|\mathbf{I} + \mathbf{K}^{-1}\mathbf{Q}| = \ln|\mathbf{K} + \mathbf{Q}| - \ln|\mathbf{K}|, \quad (66)$$

$\nabla_y f_t$ follows. Using

$$tr(\mathbf{X}^{-1}d\mathbf{X}\mathbf{X}^{-1}d\mathbf{X}) = vec(d\mathbf{X})'(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1})vec(d\mathbf{X}) = d\mathbf{x}'D_n'(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1})\mathbf{D}_n d\mathbf{x} \quad (67)$$

the Hessian $\nabla_{xx}^2 f$ can be identified as

$$\nabla_{xx}^2 f = -\mathbf{D}_n'(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1})\mathbf{D}_n \quad (68)$$

Applying this to $f(\mathbf{K}), \nabla_y^2 f_t$ follows.

To derive $\nabla_x f_t$ and $\nabla_x^2 f_t$, use a modification of (64) for $f(\mathbf{R}) = \ln|\mathbf{I} + \mathbf{W}\mathbf{R}|$:

$$f(\mathbf{R} + d\mathbf{R}) = f(\mathbf{R}) + tr(\mathbf{Z}d\mathbf{R}) - \frac{1}{2}tr(\mathbf{Z}d\mathbf{R}\mathbf{Z}d\mathbf{R}) + \sum_i o\left(\lambda_i^2\right) \quad (69)$$

where $\mathbf{Z} = (\mathbf{I} + \mathbf{W}\mathbf{R})^{-1}\mathbf{W}$, so that

$$\nabla_r f_t = \mathbf{D}_m'vec(\mathbf{Z}) \quad (70)$$

where $\mathbf{r} = vech(\mathbf{R})$, and

$$\nabla_r^2 f_t = -\mathbf{D}_m'(\mathbf{Z}\otimes\mathbf{Z})\mathbf{D}_m \quad (71)$$

from which (37), (38) follow, where we have used the following identities [27]:

$$tr(\mathbf{A}\mathbf{B}) = vec(\mathbf{A}')'vec(\mathbf{B}),$$
$$tr(\mathbf{A}\mathbf{B}\mathbf{C}\mathbf{D}) = (vec\mathbf{D})'(\mathbf{A}\otimes\mathbf{C}')vec(\mathbf{B}') \quad (72)$$

and the fact that $\mathbf{Z}$ is symmetric, $\mathbf{Z}' = \mathbf{Z}$. To derive $\nabla_{xy}^2 f_t$, observe that

$$\nabla_{kr}^2 f(\mathbf{R},\mathbf{K}) = \nabla_{kr}^2 \ln|\mathbf{K} + \mathbf{H}\mathbf{R}\mathbf{H}'| \quad (73)$$

where $d\mathbf{k} = vec(d\mathbf{K})$, so that one needs to consider only

$$\tilde{f}(\mathbf{R},\mathbf{K}) = \ln|\mathbf{K} + \mathbf{H}\mathbf{R}\mathbf{H}'| \quad (74)$$

for which the perturbation method gives

$$\tilde{f}(\mathbf{R} + d\mathbf{R}, \mathbf{K} + d\mathbf{K}) = \tilde{f}(\mathbf{R},\mathbf{K}) - tr\left(\mathbf{H}'(\mathbf{K}+\mathbf{Q})^{-1}d\mathbf{K}(\mathbf{K}+\mathbf{Q})^{-1}\mathbf{H}d\mathbf{R}\right) + \Delta\tilde{f} \quad (75)$$

where $\Delta\tilde{f}$ denotes all other terms (which do not affect the mixed derivatives), from which (40) follows by using vec operator inside the trace.

B. Proof of Lemma 3

Observe that $\mathbf{Q} \geq \mathbf{0}$ so that $(\mathbf{K}+\mathbf{Q})^{-1} \leq \mathbf{K}^{-1}$ and thus

$$\mathbf{K}^{-1}\otimes\mathbf{K}^{-1} - (\mathbf{K}+\mathbf{Q})^{-1}\otimes(\mathbf{K}+\mathbf{Q})^{-1} \geq \mathbf{0} \quad (76)$$

(this follows from the properties of Kronecker products, see e.g., [29]) and

$$(1+t^{-1})\mathbf{K}^{-1}\otimes\mathbf{K}^{-1} - (\mathbf{K}+\mathbf{Q})^{-1}\otimes(\mathbf{K}+\mathbf{Q})^{-1} \geq t^{-1}\mathbf{K}^{-1}\otimes\mathbf{K}^{-1} > \mathbf{0} \quad (77)$$

Now consider the following quadratic form for any $\mathbf{y} \neq \mathbf{0}$:

$$\mathbf{y}'\nabla_{yy}^2 f_t\mathbf{y} = \tilde{\mathbf{y}}'\left((1+t^{-1})\mathbf{K}^{-1}\otimes\mathbf{K}^{-1} - (\mathbf{K}+\mathbf{Q})^{-1}\otimes(\mathbf{K}+\mathbf{Q})^{-1}\right)\tilde{\mathbf{y}} > 0 \quad (78)$$

since $\tilde{\mathbf{y}} = \widetilde{\mathbf{D}}_n\mathbf{y} \neq 0$ (this follows from the fact that all columns of $\widetilde{\mathbf{D}}_n$ are linearly independent, which in turn is implied by linear independence of columns of $\mathbf{D}_n$ since it has a full column rank [27]). Therefore, $\nabla_{yy}^2 f_t > \mathbf{0}$. Non-singularity of $\nabla_{xx}^2 f_t$ can be proved in a similar way. First, one observes that $\mathbf{W} \geq \mathbf{W}_2$:

$$\mathbf{W} = \mathbf{H}'\mathbf{K}^{-1}\mathbf{H} \quad (79)$$
$$= [\mathbf{H}_1'\mathbf{H}_2']\begin{bmatrix}\mathbf{I} & \mathbf{K}_{21}'\\ \mathbf{K}_{21} & \mathbf{I}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\end{bmatrix} \quad (80)$$
$$= \mathbf{H}_2'\mathbf{H}_2 + (\mathbf{H}_1 - \mathbf{K}_{21}'\mathbf{H}_2)'(\mathbf{I} - \mathbf{K}_{21}'\mathbf{K}_{21})^{-1} \times (\mathbf{H}_1 - \mathbf{K}_{21}'\mathbf{H}_2) \quad (81)$$
$$\geq \mathbf{H}_2'\mathbf{H}_2 = \mathbf{W}_2 \quad (82)$$

since 2nd term in (81) is positive semi-definite, where we have used the matrix inversion Lemma:

$$\mathbf{K}^{-1} = \begin{bmatrix}\mathbf{I} & \mathbf{K}_{21}'\\ \mathbf{K}_{21} & \mathbf{I}\end{bmatrix}^{-1} = \begin{bmatrix}(\mathbf{I}-\mathbf{K}_{21}'\mathbf{K}_{21})^{-1} & \mathbf{K}_{21}'(\mathbf{K}_{21}\mathbf{K}_{21}'-\mathbf{I})^{-1}\\ (\mathbf{K}_{21}\mathbf{K}_{21}'-\mathbf{I})^{-1}\mathbf{K}_{21} & (\mathbf{I}-\mathbf{K}_{21}\mathbf{K}_{21}')^{-1}\end{bmatrix} \quad (83)$$

and the fact that $\mathbf{K}_{21}'\mathbf{K}_{21} < \mathbf{I}, \mathbf{K}_{21}\mathbf{K}_{21}' < \mathbf{I}$, which follows from $\mathbf{K} > \mathbf{0}$ (since this implies $|\mathbf{K}_{21}|_2 < 1$, where $|\cdot|_2$ is the spectral norm, see e.g., [29]). Therefore, $\mathbf{Z}_1 \geq \mathbf{Z}_2$, which follows from the following argument when $\mathbf{W}$, $\mathbf{W}_2$ are non-singular:

$$\mathbf{W} \geq \mathbf{W}_2 \Rightarrow \mathbf{W}^{-1} \leq \mathbf{W}_2^{-1} \quad (84)$$
$$\Rightarrow \mathbf{W}^{-1} + \mathbf{R} \leq \mathbf{W}_2^{-1} + \mathbf{R} \quad (85)$$
$$\Rightarrow \mathbf{Z}_1 = (\mathbf{W}^{-1}+\mathbf{R})^{-1} \geq (\mathbf{W}_2^{-1}+\mathbf{R})^{-1} = \mathbf{Z}_2 \quad (86)$$

When $\mathbf{W}$ and/or $\mathbf{W}_2$ are singular, one can use the continuity argument [29]: use $\mathbf{W}_\epsilon = \mathbf{W} + \epsilon\mathbf{I} > \mathbf{0}, \mathbf{W}_{2\epsilon} = \mathbf{W}_2 + \epsilon\mathbf{I} > \mathbf{0}$ with $\epsilon > 0$, instead of $\mathbf{W}$, $\mathbf{W}_2$ and then take $\epsilon \to 0$; since both sides of the inequality are continuous functions, the result follows. Since $\mathbf{Z}_1 \geq \mathbf{Z}_2$, it follows that $\mathbf{Z}_1\otimes\mathbf{Z}_1 \geq \mathbf{Z}_2\otimes\mathbf{Z}_2$ and thus

$$\mathbf{Z}_1\otimes\mathbf{Z}_1 - \mathbf{Z}_2\otimes\mathbf{Z}_2 + t^{-1}\mathbf{R}^{-1}\otimes\mathbf{R}^{-1} > \mathbf{0} \quad (87)$$

(since $\mathbf{R}^{-1}\otimes\mathbf{R}^{-1} > \mathbf{0}$) from which it follows that $\nabla_{xx}^2 f_t < \mathbf{0}$.

ACKNOWLEDGMENT

The authors would like to thank A.Khisti for numerous stimulating and insightful discussions, and the reviewers for constructive comments and suggestions.

REFERENCES

[1] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[2] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Jan. 1975.
[3] J. L. Massey, "A simplified treatment of Wyner's wire-tap channel," *Proc. 21st Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 1983, pp. 268–276.
[4] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.
[5] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part I: The MISOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2010.
[6] 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.
[7] 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.
[8] T. Liu and S. Shamai, "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.
[9] J. Li and A. Petropulu, Transmitter optimization for achieving secrecy capacity in Gaussian MIMO wiretap channels, *arXiv:0909.2622v1*, Sep. 2009.
[10] 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.
[11] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Proc. 41st Annu. CISS*, Mar. 2007, pp. 905–910.
[12] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. IEEE ISIT*, Nice, France, Jun. 2007, pp. 2466–2470.
[13] A. Khisti *et al.*, "On the Gaussian MIMO wiretap channel," in *Proc. IEEE ISIT*, Nice, France, Jun. 2007, pp. 2471–2475.
[14] S. Loyka and C. D. Charalambous, "On optimal signaling over secure MIMO channels," in *Proc. IEEE ISIT*, Cambridge, MA, USA, Jul. 2012, pp. 443–447.
[15] S. Loyka and C. D. Charalambous, "Further results on optimal signaling over secure MIMO channels," in *Proc. IEEE ISIT*, Istanbul, Turkey, Jul. 2013, pp. 2019–2023.
[16] S. Loyka and C. D. Charalambous, "Rank-deficient solutions for optimal signaling over secure MIMO channels," in *Proc. IEEE ISIT*, Honolulu, HI, USA, Jul. 2014, pp. 201–205.
[17] 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.
[18] Z. Li, R. Yates, and W. Trappe, "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-Verlag, 2010.
[19] M. C. Gursoy, "Secure communication in the low-SNR regime," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1114–1123, Apr. 2012.
[20] Q. Li *et al.*, "Transmit solutions for MIMO wiretap channels using alternating optimization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1714–1727, Sep. 2013.
[21] J. Steinwandt, S. A. Vorobyov, and M. Haardt, "Secrecy rate maximization for MIMO Gaussian wiretap channels with multiple eavesdroppers via alternating matrix POTDC," in *Proc. IEEE ICASSP*, May 4–9, 2014, Florence, Italy, pp. 5686–5690.
[22] A. Alvarado, G. Scutari, and J. S. Pang, "A new decomposition method for multiuser DC-programming and its applications," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2984–2998, Jun. 2014.
[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[24] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, CA, USA: Athena Scientific, 2008.
[25] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power costraint," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.
[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[27] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications to Statistics and Econometrics*. Hoboken, NJ, USA: Wiley, 1999.
[28] D. A. Harville, *Matrix Algebra From a Statistcian's Perspective*. New York, NY, USA: Springer-Verlag, 1997.
[29] F. Zhang, *Matrix Theory: Basic Results and Techniques*. New York, NY, USA: Springer, 1999.
[30] M. Grant, S. Boyd, and Y. Ye, "Disciplined convex programming," in *Global Optimization: From Theory to Implementation*, L. Liberti and N. Maculan, Eds. New York, NY, USA: Springer-Verlag, 2006, pp. 155–210.
[31] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control (tribute to M. Vidyasagar)*, V. Blondel, S. Boyd, and H. Kimura, Eds. London, U.K.: Springer-Verlag, 2008, pp. 95–110.
[32] CVX: Matlab Software for Disciplined Convex Programming, Apr. 2014. [Online]. Available: http://cvxr.com/cvx/
[33] A. Ghosh and S. Boyd, Minimax and Convex-Concave Games, May 2014. [Online]. Available: http://www.stanford.edu/class/ee364b/lectures.html
[34] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ, USA: Princeton Univ. Press, 2005.
[35] E. Zeidler. *Nonlinear Functional Analysis and Its Applications, Vol. I: Fixed-Point Theorems*. New York, NY, USA: Springer-Verlag, 1986.

Sergey Loyka was born in Minsk, Belarus. He received the M.S. degree with honors from Minsk Radioengineering Institute, Minsk, Belarus, in 1992 and the Ph.D. degree in radio engineering from the Belorussian State University of Informatics and Radioelectronics (BSUIR), Minsk, Belarus, in 1995. 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 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.