A Low-Complexity Algorithm for Achieving Secrecy Capacity in MIMO Wiretap Channels

Thang Van Nguyen*, Quang-Doanh Vu†, Markku Juntti†, and Le-Nam Tran*
*School of Electrical and Electronic Engineering, University College Dublin, Ireland.
Email: {thang.nguyen, nam.tran}@ucd.ie
†Centre for Wireless Communications, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland
Email: {doanh.vu, markku.juntti}@oulu.fi

***Abstract*—We consider a secure transmission including a transmitter, a receiver and an eavesdropper, each being equipped with multiple antennas. The aim is to develop a low-complexity and scalable method to find a globally optimal solution to the problem of secrecy rate maximization under a total power constraint at the transmitter. In principle, the original formulation of the problem is nonconvex. However, it can be equivalently translated into finding a saddle point of a minimax convex-concave program. An existing approach finds the saddle point using the Newton method, whose computational cost increases quickly with the number of transmit antennas, making it unsuitable for large scale antenna systems. To this end, we propose an iterative algorithm based on alternating optimization, which is guaranteed to converge to a saddle point, and thus achieves a globally optimal solution to the considered problem. In particular, each subproblem of the proposed iterative method admits a closed-form solution. We analytically show that the iteration cost of our proposed method is much cheaper than that of the known solution. As a result, numerical results demonstrate that the proposed method remarkably outperforms the existing one in terms of the overall run time.**

I. Introduction

Wireless communications is expected to support every aspects of daily life. Many applications require that confidential information such as bank transfer data or personal information be transmitted securely over wireless medium. Thus secure communications has become one of the primary concerns in both academia and industry. To this end, wireless physical layer security has been receiving growing attention in recent years. This is because, compared to the conventional security approaches using cryptographic techniques, physical layer security provides a more efficient secure communications method via protecting the communication phase [1], [2].

The secrecy capacity of multiple-input multiple-output (MIMO) wiretap channels has been investigated intensively due to the broad applications of multiantenna systems [3]–[5]. A main result is that the secrecy capacity can be achieved by Gaussian wiretap codes [3], which leads to the problem of optimizing the transmit covariance matrix for maximizing the capacity. However, the established secrecy capacity function is nonconvex with the input covariance matrix in general, and thus solving the secrecy capacity maximization problem is difficult. Interestingly, the secrecy capacity maximization problem can be equivalently expressed as a minimax program. And the secrecy capacity equals the objective value of the minimax problem at a saddle point [3], [4].

The equivalent reformulation of the secrecy capacity as a minimax program allows for efficient solutions. In principle, a saddle point for such a program can be found using the Newton method [6]. In fact, this is exactly the idea of the method presented in [5] to compute the secrecy capacity. Although specific structure of the problem is exploited to reduce the number of independent variables, the complexity of the solution proposed in [5] is still high, since the computational cost for each Newton step is $\mathcal{O}(n_{\mathrm{T}}^6)$, where n_{T} is the number of transmit antennas.

Motivated by the above discussions, our aim in this paper is to derive a low-complexity and scalable method for finding the optimal transmit covariance matrix of the secrecy rate maximization problem under a total transmit power constraint. Suggested by the structure of the equivalent minimax problem, we apply the concept of alternating optimization (AO), but in a novel way, to find a saddle point [7], [8]. We remark that the proposed method is entirely different from the one introduced in [9]. It is also worth noting that purely applying AO to the considered problem might result in divergence [7]. Herein, to ensure the convergence, we exploit the specific structure of the problem to derive an upper bound for the minimization subproblem, based on which the convergence to a saddle point of the proposed method is proved. Importantly, we derive closed-form solution to each convex subproblem of the proposed iterative method. Analytical and numerical results are provided which show that our method is superior to the state-of-the-art in terms of computational efficiency.

Notation: Bold lower and upper case letters represent vectors and matrices, respectively. $\mathbf{X}^H$ and $\mathbf{X}^T$ denotes Hermitian and normal transpose of $\mathbf{X}$, respectively. $|\mathbf{X}|$ is the determinant of $\mathbf{X}$. $\mathrm{Tr}(\mathbf{X})$ is trace of $\mathbf{X}$. $\mathbb{C}^{a\times b}$ denote the the space complex matrices of size a×b. $\mathbf{X}\succeq\mathbf{0}$ means $\mathbf{X}$ is positive semidefinite. $\mathbf{I}$ defines an identity matrix. $[x]_+$ denotes $\max\{0,x\}$; $\lceil x\rceil$ denotes the nearest integer greater than or equal to x. Notation $\mathrm{diag}\,(\mathbf{x})$ denotes the square diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal.

II. Background

A. Wiretap MIMO Channel Model

We consider a communication system including a transmitter, an intended receiver, and an eavesdropper, each is

equipped with multiple antennas [3]–[5]. Let $n_{\rm T}$, $n_{\rm R}$, and $n_{\rm E}$ denote the number of antennas at the transmitter, receiver, and eavesdropper, respectively. Let us denote by $\mathbf{H} \in \mathbb{C}^{n_{\rm R}\times n_{\rm T}}$ and $\mathbf{G} \in \mathbb{C}^{n_{\rm R}\times n_{\rm T}}$ the channel matrices corresponding to the receiver, and eavesdropper, respectively, and by $\mathbf{x} \in \mathbb{C}^{n_{\rm T}\times 1}$ the transmit signal. The received signal at the receiver and eavesdropper are, respectively,

$$\mathbf{y}_{\rm R} = \mathbf{H}\mathbf{x} + \mathbf{z}_{\rm R},\ \mathbf{y}_{\rm E} = \mathbf{G}\mathbf{x} + \mathbf{z}_{\rm E} \tag{1}$$

where $\mathbf{z}_{\rm R} \sim \mathcal{CN}(0, \mathbf{I})$ and $\mathbf{z}_{\rm E} \sim \mathcal{CN}(0, \mathbf{I})$ are the additive white Gaussian noise. Suppose that the perfect channel state information is available, and the coding scheme follows $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{S})$, the problem of maximizing secrecy rate is given by [3]

$$\underset{\mathbf{S}\succeq\mathbf{0}}{\text{maximize}}\ C(\mathbf{S}) \triangleq \log\frac{|\mathbf{I}+\mathbf{H}\mathbf{S}\mathbf{H}^H|}{|\mathbf{I}+\mathbf{G}\mathbf{S}\mathbf{G}^H|}\ \text{subject to}\ \mathrm{Tr}(\mathbf{S}) \le P \tag{2}$$

where P is the maximum transmit power. Problem (2) is convex if the channel is degraded, i.e. $\mathbf{H}^H\mathbf{H} \succeq \mathbf{G}^H\mathbf{G}$. For other cases, problem (2) is nonconvex [5], which is difficult to solve optimally. Fortunately, (2) possesses hidden tractability, which is presented next.

B. Minimax Reformulation and Existing Solution

Let us define $\bar{\mathbf{H}} = [\mathbf{H}^T, \mathbf{G}^T]^T$, and $\bar{\mathbf{\Omega}} \in \mathbb{C}^{n_{\rm R}\times n_{\rm E}}$. Suppose $\mathbf{S}^\star$ is an optimal solution to (2), then we have [3], [4]

$$C(\mathbf{S}^\star) = \min_{\mathbf{\Omega}\in\Omega}\max_{\mathbf{S}\in\mathcal{S}} f(\mathbf{\Omega}, \mathbf{S}) \triangleq \log\frac{|\mathbf{I}+\mathbf{\Omega}^{-1}\bar{\mathbf{H}}\mathbf{S}\bar{\mathbf{H}}^H|}{|\mathbf{I}+\mathbf{G}\mathbf{S}\mathbf{G}^H|} \tag{3}$$

where

$$\Omega \triangleq \left\{\mathbf{\Omega}|\ \mathbf{\Omega} = \begin{bmatrix} \mathbf{I}_{n_{\rm R}} & \bar{\mathbf{\Omega}} \\ \bar{\mathbf{\Omega}}^H & \mathbf{I}_{n_{\rm T}} \end{bmatrix} \succeq \mathbf{0}\right\} \tag{4}$$

and

$$\mathcal{S} \triangleq \{\mathbf{S}|\mathbf{S}\succeq\mathbf{0}, \mathrm{Tr}(\mathbf{S}) \le P\}. \tag{5}$$

We note that Ω and $\mathcal{S}$ are compact and convex. In addition, $f(\mathbf{\Omega}, \mathbf{S})$ is an upper bound of $C(\mathbf{S})$, i.e. $f(\mathbf{\Omega}, \mathbf{S}) \ge C(\mathbf{S})$ for any feasible point $(\mathbf{\Omega}, \mathbf{S})$.

The equivalence between (2) and (3) means that we can alternatively find a saddle point of (3) to obtain a globally optimal solution to (2). It is important to note that $f(\mathbf{\Omega}, \mathbf{S})$ is convex with $\mathbf{\Omega}$ for any fixed $\mathbf{S}$, and concave with $\mathbf{S}$ for any fixed $\mathbf{\Omega}$ [4], $f(\mathbf{\Omega}, \mathbf{S})$ is twice differentiable, and the feasible set of (3) is convex. Thus, there exists a saddle point $(\mathbf{\Omega}^\star, \mathbf{S}^\star)$ of (3), i.e. $f(\mathbf{\Omega}^\star, \mathbf{S}) \le f(\mathbf{\Omega}^\star, \mathbf{S}^\star) \le f(\mathbf{\Omega}, \mathbf{S}^\star)$ for any feasible point $(\mathbf{\Omega}, \mathbf{S})$ [4].

An approach finding a saddle point of a minimax convex-concave program is to use the Newton method [6], which is an iterative procedure where each iteration requires calculating a Newton step. In fact, such an algorithm was proposed in [5]. Therein, in order to reduce the computation cost, only $\bar{\mathbf{\Omega}}$ and the the lower triangular portion of $\mathbf{S}$ are considered as independent variables. As a result, the total number of variables is $\mathcal{O}\left(n_{\rm T}^2 + n_{\rm R}n_{\rm E}\right)$, and the iteration cost is $\mathcal{O}(n_{\rm T}^6)$, assuming $n_{\rm T}^2 > n_{\rm R}n_{\rm E}$ [10, Chapter 10].

III. Proposed Low-Complexity Algorithm

We now propose a low-complexity algorithm to obtain a saddle point of (3) based on combining successive convex approximation (SCA) and AO. The idea of the proposed method is as follows:

- At step t, for a given $\mathbf{S} = \mathbf{S}_t$, we consider an approximation of $f(\mathbf{\Omega}, \mathbf{S})$ at $\mathbf{S} = \mathbf{S}_t$ given by

$$f_t(\mathbf{\Omega}, \mathbf{S}) \triangleq \log|\mathbf{\Omega} + \bar{\mathbf{H}}\mathbf{S}\bar{\mathbf{H}}^H| - \log|\mathbf{\Omega}| - \mathrm{Tr}\big(\mathbf{Q}_t\big(\mathbf{S}-\mathbf{S}_t\big)\big) - \log|\mathbf{I}+\mathbf{G}\mathbf{S}_t\mathbf{G}^H| \tag{6}$$

  where $\mathbf{Q}_t = \mathbf{G}^H(\mathbf{I}+\mathbf{G}\mathbf{S}_t\mathbf{G}^H)^{-1}\mathbf{G}$. We note that $f_t(\mathbf{\Omega}, \mathbf{S})$ is attained by linearizing the term $\log|\mathbf{I}+\mathbf{G}\mathbf{S}\mathbf{G}^H|$ around $\mathbf{S}_t$, and thus $f_t(\mathbf{\Omega}, \mathbf{S}) \le f(\mathbf{\Omega}, \mathbf{S})$ where the equality holds at $\mathbf{S} = \mathbf{S}_t$, and $\nabla f_t(\mathbf{\Omega}, \mathbf{S}) = \nabla f(\mathbf{\Omega}, \mathbf{S})$ at $\mathbf{S} = \mathbf{S}_t$. These properties are critical for the proposed method to achieve a saddle point of (3), as shown in Subsection III-D.
- At step $t+1$, we find $\mathbf{S}_{t+1}$ as $\mathbf{S}_{t+1} = \mathbf{W}^\star$, where $\mathbf{W}^\star$ is obtained from a saddle point of the following problem

$$(\mathbf{K}^\star, \mathbf{W}^\star) = \arg\min_{\mathbf{K}\in\Omega}\ \arg\max_{\mathbf{W}\in\mathcal{S}}\ f_t(\mathbf{K}, \mathbf{W}). \tag{7}$$

The two steps are repeated until a convergence criterion is met.

It is obvious that solving (7) is the key to the implementation of the proposed method. In the following we describe an iterative procedure which can solve (7) efficiently, i.e., by closed-form expressions. Let us denote by $(\mathbf{W}_n, \mathbf{K}_n)$ the iterate at iteration n of the iterative procedure. The idea for solving (7) is to alternatively optimize $\mathbf{K}$ and $\mathbf{W}$: finding $\mathbf{W}_n$ for fixed $\mathbf{K}_n$, and then finding $\mathbf{K}_{n+1}$ for fixed $\mathbf{W}_n$. The details are presented next.

A. Finding $\mathbf{W}_n$ for Fixed $\mathbf{K}_n$

For fixed $\mathbf{K}_n$, $\mathbf{W}_n$ is found to be the optimal solution to the following problem

$$\underset{\mathbf{W}\succeq\mathbf{0}}{\text{maximize}}\ \log|\mathbf{K}_n + \bar{\mathbf{H}}\mathbf{W}\bar{\mathbf{H}}^H| - \mathrm{Tr}(\mathbf{Q}_t\mathbf{W}) \tag{8a}$$

$$\text{subject to}\ \mathrm{Tr}(\mathbf{W}) \le P. \tag{8b}$$

The objective in (8) is simply obtained from $f_t(\mathbf{K}, \mathbf{W})$ by ignoring the constants independent of $\mathbf{W}$. We show that problem (8) admits a closed-form solution. To proceed, let us consider the partial Lagrangian function given by

$$\mathcal{L}(\mathbf{W}, \mu) = \log|\mathbf{K}_n + \bar{\mathbf{H}}\mathbf{W}\bar{\mathbf{H}}^H| - \mathrm{Tr}\left(\left(\mu\mathbf{I}_{n_{\rm T}} + \mathbf{Q}_t\right)\mathbf{W}\right) + \mu P \tag{9}$$

where $\mu \ge 0$ is the Lagrangian multiplier. Let $\hat{\mathbf{W}} = \mathbf{M}_\mu^{1/2}\mathbf{W}\mathbf{M}_\mu^{1/2}$ where $\mathbf{M}_\mu = \mu\mathbf{I}_{n_{\rm T}} + \mathbf{Q}_t$, then finding $\mathbf{W}$ to maximize $\mathcal{L}(\mathbf{W}, \mu)$ is equivalent to

$$\underset{\hat{\mathbf{W}}\succeq\mathbf{0}}{\text{maximize}}\ \log|\mathbf{K}_n + \bar{\mathbf{H}}\mathbf{M}_\mu^{-1/2}\hat{\mathbf{W}}\mathbf{M}_\mu^{-1/2}\bar{\mathbf{H}}^H| - \mathrm{Tr}(\hat{\mathbf{W}}). \tag{10}$$

Let $\mathbf{V}\mathbf{\Sigma}\mathbf{V}^H$ be the eigenvalue decomposition (EVD) of $\mathbf{M}_\mu^{-1/2}\bar{\mathbf{H}}^H\mathbf{K}_n^{-1}\bar{\mathbf{H}}\mathbf{M}_\mu^{-1/2}$ with $\mathbf{V} \in \mathbb{C}^{n_{\rm T}\times n_{\rm T}}$ being a unitary

Algorithm 1 Finding optimal solution to (8)

1: **Initialization**: Set $\mu_{\min} = 0$, $\mu_{\max} = \frac{n_{\mathrm{T}}}{P}$, $\mu_{\mathrm{l}} \coloneqq \mu_{\min}$, $\mu_{\mathrm{u}} \coloneqq \mu_{\max}$, tolerance $\epsilon > 0$,
2: **repeat**
3: $\mu \coloneqq (\mu_{\mathrm{l}} + \mu_{\mathrm{u}})/2$, $\mathbf{M}_\mu = \mu \mathbf{I}_{n_{\mathrm{T}}} + \mathbf{Q}_t$
4: Compute EVD of $\mathbf{M}_\mu^{-1/2} \bar{\mathbf{H}}^H \mathbf{K}_n^{-1} \bar{\mathbf{H}} \mathbf{M}_\mu^{-1/2}$ as $\mathbf{V}\mathbf{\Sigma}\mathbf{V}^H$ with positive eigenvalues σ_i, $i = 1, 2, \ldots, r$.
5: Update $\mathbf{W}_\mu$ as (12)
6: If $\mathrm{Tr}(\mathbf{W}_\mu) > P$, $\mu_{\mathrm{l}} \coloneqq \mu$, otherwise $\mu_{\mathrm{u}} \coloneqq \mu$
7: **until** $\mu_{\mathrm{u}} - \mu_{\mathrm{l}} < \epsilon$
8: **Output**: $\mathbf{W} \coloneqq \mathbf{W}_\mu$

matrix, and $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, \mathbf{0}_{n_{\mathrm{T}}-r})$ where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, and $r \leq n_{\mathrm{T}}$ is the rank of $\bar{\mathbf{H}}$. For some fixed $\mu \geq 0$, (10) admits a water-filling solution as follows

$$\hat{\mathbf{W}} = \mathbf{V}\mathbf{\Phi}\mathbf{V}^H \tag{11}$$

where $\mathbf{\Phi} \triangleq \mathrm{diag}(\left[1 - \sigma_1^{-1}\right]_+, \ldots, \left[1 - \sigma_r^{-1}\right]_+, \mathbf{0}_{n_{\mathrm{T}}-r})$. Consequently, we arrive at

$$\mathbf{W} = \mathbf{M}_\mu^{-1/2} \mathbf{V}\mathbf{\Phi}\mathbf{V}^H \mathbf{M}_\mu^{-1/2}. \tag{12}$$

To find the optimal μ, we state the following lemma.

Lemma 1. *Let $\mathbf{W}_\mu$ be the matrix given in* (12) *corresponding to μ. Then, $\mathrm{Tr}(\mathbf{W}_\mu)$ is a decreasing function of μ for $\mu \geq 0$.*

We omit the proof of the above lemma for the sake of brevity. In summary, the procedure finding the optimal solution to (8) is outlined in Algorithm 1.

B. Finding $\mathbf{K}_{n+1}$ for Fixed $\mathbf{W}_n$

After obtaining $\mathbf{W}_n$ we need to update $\mathbf{K}$, i.e. computing $\mathbf{K}_{n+1}$ for fixed $\mathbf{W}_n$. To do this we do not minimize $f_t(\mathbf{K}, \mathbf{W})$ directly, since doing so does not guarantee the convergence. Instead, we consider an upper bound of the objective which is obtained using the following inequality

$$\log|\mathbf{K} + \bar{\mathbf{H}}\mathbf{W}_n\bar{\mathbf{H}}^H| \leq \log|\mathbf{T}_n^{-1}| + \mathrm{Tr}(\mathbf{T}_n(\mathbf{K} - \mathbf{K}_n)) \tag{13}$$

where $\mathbf{T}_n = (\mathbf{K}_n + \bar{\mathbf{H}}\mathbf{W}_n\bar{\mathbf{H}}^H)^{-1}$. We remark that the same idea was also used in [8] in a different context. Then $\mathbf{K}_{n+1}$ is an optimal solution to the following convex problem

$$\underset{\mathbf{K}}{\text{minimize}} \ \mathrm{Tr}(\mathbf{T}_n\mathbf{K}) - \log|\mathbf{K}| \tag{14a}$$

$$\text{subject to } \mathbf{K} = \begin{bmatrix} \mathbf{I}_{n_{\mathrm{R}}} & \bar{\mathbf{K}}^H \\ \bar{\mathbf{K}} & \mathbf{I}_{n_{\mathrm{E}}} \end{bmatrix} \succeq \mathbf{0} \tag{14b}$$

We now show that the above problem can also be solved in closed-form. To begin with, let us write $\mathbf{T}_n$ as $\mathbf{T}_n = \begin{bmatrix} \bar{-} & \bar{\mathbf{T}}_n^H \\ \bar{\mathbf{T}}_n & \bar{-} \end{bmatrix}$ where $\bar{\mathbf{T}}_n \in \mathbb{C}^{n_{\mathrm{E}} \times n_{\mathrm{R}}}$. (Note that $-$ denotes matrices which have no effect on solving (14).) Then it is easy to see that (14) is equivalent to

$$\text{minimize } \{g(\bar{\mathbf{K}}) \mid \mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H \succeq \mathbf{0}\} \tag{15}$$

where $g(\bar{\mathbf{K}}) \triangleq \mathrm{Tr}(\bar{\mathbf{T}}_n\bar{\mathbf{K}}^H) + \mathrm{Tr}(\bar{\mathbf{T}}_n^H\bar{\mathbf{K}}) - \log|\mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H|$. As far as the closed-form solution to (15) is concerned, the following lemma is in order.

Algorithm 2 The proposed algorithm to find a globally optimal solution of (2)

1: **Initialization**: Set $t \coloneqq 0$, $\mathbf{S}_0 = \mathbf{0}$, $\mathbf{\Omega}_0 = \mathbf{I}$
2: **repeat**
3: Compute $\mathbf{Q}_t$ as given below (6) and set $n \coloneqq 0$
4: Set $\mathbf{K}_0 = \mathbf{\Omega}_t$
5: **repeat**
6: Solve (8) using Algorithm 1 and assign the optimal solution to $\mathbf{W}_n$
7: Compute $\mathbf{T}_n$ as shown below (13) and use (16) to obtain $\mathbf{K}_{n+1}$, the optimal solution to (14)
8: $n \to n + 1$
9: **until** convergence criterion is met
10: Update $\mathbf{S}_{t+1} = \mathbf{W}_t^\star$, $\mathbf{\Omega}_{t+1} = \mathbf{K}_t^\star$
11: $t \to t + 1$
12: **until** convergence criterion is met
13: **Output**: $\mathbf{S}_t$

Lemma 2. *The optimal solution to problem* (15) *is given by*

$$\bar{\mathbf{K}} = -\mathbf{U}\mathbf{\Delta}\mathbf{U}^H\bar{\mathbf{T}}_n \tag{16}$$

where $\mathbf{U}$ is a unitary matrix obtained from of the EVD of $\bar{\mathbf{T}}_n\bar{\mathbf{T}}_n^H = \mathbf{U}diag(\rho_1, \ldots, \rho_{n_E})\mathbf{U}^H$ and $\mathbf{\Delta} = 2diag\left(\frac{1}{1+\sqrt{1+4\rho_1}}, \ldots, \frac{1}{1+\sqrt{1+4\rho_{n_E}}}\right)$.

The proof is given in Appendix A.

C. The Proposed Algorithm

Our proposed algorithm for solving (2) is outlined in Algorithm 2, which is basically a two-stage iterative method. The outer stage is to update the lower bound of the objective $f(\mathbf{\Omega}, \mathbf{S})$, which follows the notion of SCA. The inner stage is to find a saddle point of the lower bound, i.e. (7), which is denoted by $(\mathbf{K}_t^\star, \mathbf{W}_t^\star)$ (at Step 10).

D. Convergence Analysis

We now prove that Algorithm 2 is guaranteed to achieve a globally optimal solution to (2). In particular we state the following two lemmas, one concerned with the convergence of the inner iterative process and one concerned with the convergence of the outer one.

Lemma 3. *For an outer iteration t the sequence $\{f_t(\mathbf{K}_n, \mathbf{W}_n)\}_n$ is decreasing and thus convergent. Moreover, there exists at least a convergent subsequence $\{\mathbf{K}_{[n]}, \mathbf{W}_{[n]}\}_{n=0}^\infty$ of the sequence $\{\mathbf{K}_n, \mathbf{W}_n\}_{n=0}^\infty$ whose limit point is a saddle point of* (7).

Lemma 4. *Let $(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star)$ be the saddle point solution of* (7) *for each t. Then, the sequence $\{f(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star)\}_t$ is increasing and convergent. Moreover, there exists at least a convergent subsequence $\{(\mathbf{\Omega}_{[t]}^\star, \mathbf{S}_{[t]}^\star)\}_{t=0}^\infty$ from the sequence $\{(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star)\}_{t=0}^\infty$ whose limit point is a saddle point of* (3).

The proof of Lemmas 3 and 4 are given in Appendices B and C, respectively.

E. Complexity Analysis

We now discuss the complexity of Algorithm 2. To shorten the complexity comparison we mainly use big-O notation of the floating-point operations (flops) and assume $n_{\mathrm{T}} \geq \max(n_{\mathrm{E}}, n_{\mathrm{R}})$. It is clear that the main complexity of Algorithm 1 is due to computing $\mathbf{M}_{\mu}^{-1/2} \in \mathbb{C}^{n_{\mathrm{T}} \times n_{\mathrm{T}}}$ and the SVD of $\mathbf{K}_n^{-1/2} \bar{\mathbf{H}} \mathbf{M}_{\mu}^{-1/2} \in \mathbb{C}^{(n_{\mathrm{E}}+n_{\mathrm{R}}) \times n_{\mathrm{T}}}$. The complexities of these two operations are $\mathcal{O}(n_{\mathrm{T}}^3)$. To find $\mathbf{K}_{n+1}$ we need to compute the SVD of $\bar{\mathbf{T}}_n$ which also requires $\mathcal{O}(n_{\mathrm{T}}^3)$ flops. At the outer stage, updating $\mathbf{Q}_t$ requires the inversion of the matrix $\mathbf{I} + \mathbf{G}\mathbf{S}_t\mathbf{G}^H$ which needs $\mathcal{O}(n_{\mathrm{E}}^3)$ flops. If $n_{\mathrm{T}} < n_{\mathrm{E}}$ then the complexity of this step can be reduced to $\mathcal{O}(n_{\mathrm{E}}^3)$ flops by applying the matrix inversion lemma. In summary, the iteration complexity of our proposed method (i.e. Algorithm 2) is $\mathcal{O}(n_{\mathrm{T}}^3)$, which is significantly smaller compared to that of the customized Newton method presented in [5].

IV. Numerical Results

We now numerically evaluate the performance of the proposed method. The channel matrices $\mathbf{H}$ and $\mathbf{G}$ are randomly generated following the independent and identically distributed zero mean and unit variance Gaussian distribution. Unless otherwise stated, we set the parameters as follows. The total transmit power is taken as $P = 10$ dB. The error tolerance for bisection search in Algorithm 1 is 10^{-12}. The inner and outer stages of Algorithm 2 stops when the difference of the objective function in each stage between two successive iterations is smaller than 10^{-6}.

We consider the solution proposed in [5] based on the Newton method as the baseline. The parameters of this method are taken following those in [5, Figure 7]. All the algorithms are written on MATLAB.

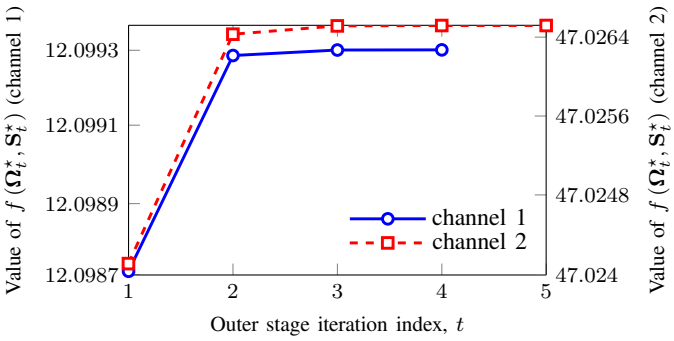

Figure 1. Convergence performance of the proposed algorithm.

In Figure 1, we study the convergence behavior of the proposed algorithm over two random channel realizations. We take $(n_{\mathrm{T}}, n_{\mathrm{R}}, n_{\mathrm{E}}) = (16, 4, 4)$ for channel 1, and $(n_{\mathrm{T}}, n_{\mathrm{R}}, n_{\mathrm{E}}) = (50, 16, 16)$ for channel 2. The figure plots the achieved value of objective function $f(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star)$ over outer iteration index. We can see that $f(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star)$ is increasing and convergent with outer stage iteration index. This result numerically confirms Lemma 4.

Fig. 2 plots the achieved secrecy rate of Algorithm 2 and the method in [5] as the functions of run time with the two

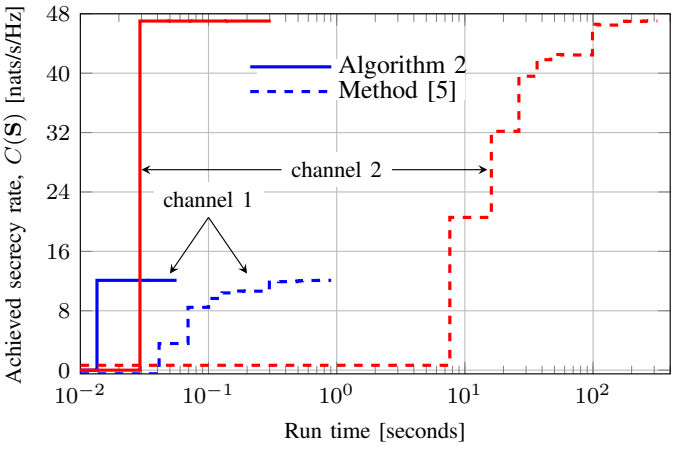

Figure 2. Achieved secrecy rate over run time (in second) of the considered methods with the two random channel realizations considered in Fig. 1.

random channel realizations considered in Figure 1. We can see from the figure that the proposed method needs only 0.06s for channel 1 and 0.3s for channel 2, while these numbers of method in [5] are 0.9s and 318s, respectively. This means the proposed method not only is significantly small but also scales slowly with the problem size compared with the existing method. The results confirm the theoretical predictions discussed in the previous section that the proposed method is computationally cheap and scalable. In the next set of experiments, we extensively investigate the computational cost of the proposed method with different system settings.

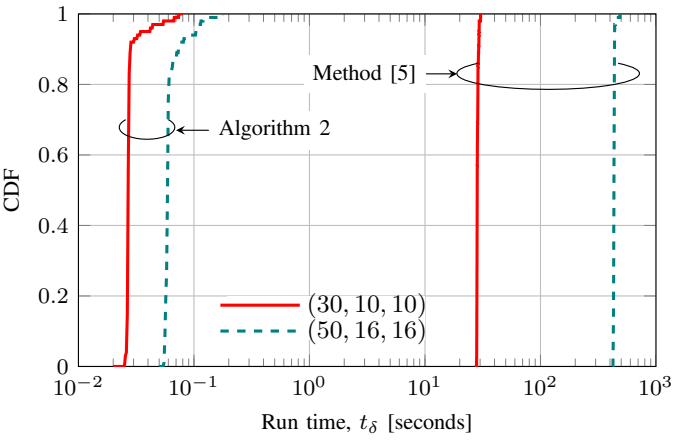

Figure 3. CDF of the run time of the considered methods needed to achieve accuracy level δ with two different settings of $(n_{\mathrm{T}}, n_{\mathrm{R}}, n_{\mathrm{E}})$. We take $\delta = 10^{-2}$ (nats/s/Hz).

To demonstrate the computational efficiency of the proposed method compared to the method in [5], we consider two different system settings of $(n_{\mathrm{T}}, n_{\mathrm{R}}, n_{\mathrm{E}})$, each is performed over 100 random channels. Because the convergence criterion of the two methods are different, we report the run time t_δ, the time needed to achieve the relatively small gap $\delta \triangleq C(\mathbf{S}^\star) - C(\mathbf{S})$. Here, we take $\delta = 10^{-2}$ (nats/s/Hz). Figure 3 shows the cumulative distribution function (CDF) of the run time t_δ

of the considered methods. The results clearly point out that our proposed method is much more computationally efficient compared to the existing one. In addition, the complexity of the proposed method is less sensitive to the system size, which is consistent with the theoretical analysis in Section III-E. This numerically confirms the scalability of the proposed method.

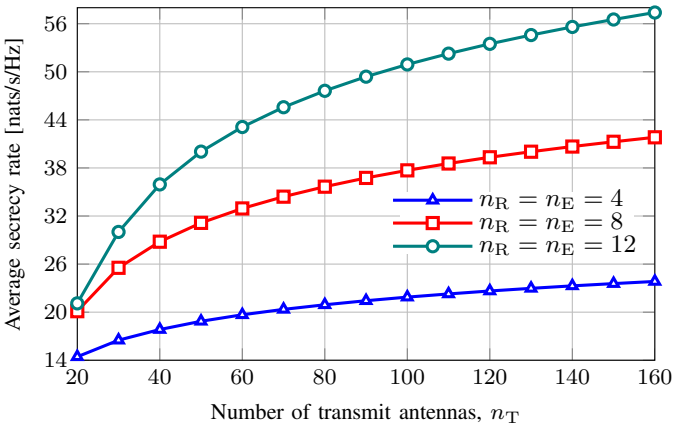

Figure 4. Average secrecy rate versus number of transmit antennas with different number of antennas at the receiver and eavesdropper.

Figure 4 plots the average secrecy rate as a function of number of transmit antennas. We take different number of antennas at the receiver and eavesdropper. An important observation from the figure is that the secrecy rate increases when n_{T} increases. In addition, the increase rate is faster with the larger number of receive antennas. The results imply that that we can achieve performance improvement with large scale systems.

V. Conclusion

We have introduced a new approach to obtain the globally optimal transmit covariance matrix for secrecy rate maximization in MIMO wiretap channels. The proposed iterative algorithm has been developed based on the AO, where the solutions admit closed-forms. We have proven that the algorithm is guaranteed to converge, and achieve global optimality. The complexity analysis has shown that the computational cost of the proposed approach scales with order n_{T}^3 in the number of transmit antennas. Finally, the extensive numerical results have demonstrated the computational effectiveness of the proposed algorithm.

Appendix A
Proof of Lemma 2

Proof: Since $g(\bar{\mathbf{K}})$ is convex in $\bar{\mathbf{K}}$ and $\{\bar{\mathbf{K}}|\mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H \succ \mathbf{0}\}$ is the effective domain of $g(\bar{\mathbf{K}})$, it is sufficient to find the stationary point of $g(\bar{\mathbf{K}})$ in its effective domain for minimizing (15). To this end, the gradient of $g(\bar{\mathbf{K}})$ is given by

$$\nabla g(\bar{\mathbf{K}}) = 2\bar{\mathbf{T}}_n + 2(\mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H)^{-1}\bar{\mathbf{K}}. \tag{17}$$

Thus, a stationary point of $g(\bar{\mathbf{K}})$ must satisfy

$$\bar{\mathbf{K}} = -(\mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H)\bar{\mathbf{T}}_n \tag{18}$$

which leads to

$$(\mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H)^{-1}\bar{\mathbf{K}}\bar{\mathbf{K}}^H(\mathbf{I}_{n_{\mathrm{E}}} - \bar{\mathbf{K}}\bar{\mathbf{K}}^H)^{-1} = \bar{\mathbf{T}}_n\bar{\mathbf{T}}_n^H. \tag{19}$$

Let $\bar{\mathbf{K}}\bar{\mathbf{K}}^H = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^H$ be the EVD of $\bar{\mathbf{K}}\bar{\mathbf{K}}^H$, then (19) can be rewritten as

$$\mathbf{Z}(\mathbf{I}_{n_{\mathrm{E}}} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{I}_{n_{\mathrm{E}}} - \boldsymbol{\Gamma})^{-1}\mathbf{Z}^H = \mathbf{U}\mathrm{diag}(\rho_1, \ldots, \rho_{n_{\mathrm{E}}})\mathbf{U} \tag{20}$$

where the EVD of $\bar{\mathbf{T}}_n\bar{\mathbf{T}}_n^H = \mathbf{U}\mathrm{diag}(\rho_1, \ldots, \rho_{n_{\mathrm{E}}})\mathbf{U}^H$. Thus we can conclude $\mathbf{Z} = \mathbf{W}$ and $(\mathbf{I}_{n_{\mathrm{E}}} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{I}_{n_{\mathrm{E}}} - \boldsymbol{\Gamma})^{-1} = \mathrm{diag}(\rho_1, \ldots, \rho_{n_{\mathrm{E}}})$, resulting in

$$\boldsymbol{\Gamma} = 4\mathrm{diag}\left(\frac{\rho_1}{\left(1+\sqrt{1+4\rho_1}\right)^2}, \ldots, \frac{\rho_{n_{\mathrm{E}}}}{\left(1+\sqrt{1+4\rho_{n_{\mathrm{E}}}}\right)^2}\right) \prec \mathbf{I}_{n_{\mathrm{E}}} \tag{21}$$

which completes the proof. ■

Appendix B
Proof of Lemma 3

The steps of the proof follow those in [8]. Since the term $\log|\mathbf{K}+\bar{\mathbf{H}}\mathbf{W}\bar{\mathbf{H}}^H|$ is jointly concave with $\mathbf{K}$ and $\mathbf{W}$, it holds that

$$\begin{aligned}
f_t(\mathbf{K}_{n+1}, \mathbf{W}_{n+1}) &= \log|\mathbf{K}_{n+1} + \bar{\mathbf{H}}\mathbf{W}_{n+1}\bar{\mathbf{H}}^H| \\
&\quad - \log|\mathbf{K}_{n+1}| - \mathrm{Tr}(\mathbf{Q}_t\mathbf{W}_{n+1}) + c \\
&\le \log|\underbrace{\mathbf{K}_n + \bar{\mathbf{H}}\mathbf{W}_n\bar{\mathbf{H}}^H}_{\mathbf{T}_n}| + \mathrm{Tr}\big(\bar{\mathbf{H}}^H\mathbf{T}_n^{-1}\bar{\mathbf{H}}(\mathbf{W}_{n+1} - \mathbf{W}_n)\big) \\
&\quad + \mathrm{Tr}\big(\mathbf{T}_n^{-1}(\mathbf{K}_{n+1} - \mathbf{K}_n)\big) - \log|\mathbf{K}_{n+1}| \\
&\quad - \mathrm{Tr}(\mathbf{Q}_t\mathbf{W}_{n+1}) + c \qquad (22) \\
&\le \log|\mathbf{T}_n| - \mathrm{Tr}(\mathbf{Q}_t\mathbf{W}_n) \\
&\quad + \mathrm{Tr}\big(\mathbf{T}_n^{-1}(\mathbf{K}_{n+1} - \mathbf{K}_n)\big) - \log|\mathbf{K}_{n+1}| + c \qquad (23) \\
&\le \log|\boldsymbol{\Psi}_n| - \mathrm{Tr}(\mathbf{Q}_t\mathbf{W}_n) - \log|\mathbf{K}_n| + c \qquad (24) \\
&= f_t(\mathbf{K}_n, \mathbf{W}_n) \qquad (25)
\end{aligned}$$

where $c = \mathrm{Tr}(\mathbf{Q}_t\mathbf{S}_t) - \log|\mathbf{I} + \mathbf{G}\mathbf{S}_t\mathbf{G}^H|$, (22) is due to the first order approximation of a concave function, (23) is due to the fact that $\mathbf{W}_{n+1}$ maximizes $\log|\mathbf{K}_n + \bar{\mathbf{H}}\mathbf{W}\bar{\mathbf{H}}^H| - \mathrm{Tr}(\mathbf{Q}_t\mathbf{W})$, yielding $\mathrm{Tr}\left(\bar{\mathbf{H}}^H\boldsymbol{\Psi}_n^{-1}\bar{\mathbf{H}}\left(\mathbf{W}_{n+1} - \mathbf{W}_n\right)\right) - \mathrm{Tr}\left(\mathbf{Q}_t\left(\mathbf{W}_{n+1} - \mathbf{W}_n\right)\right) \le 0$, and (24) holds since $\mathbf{K}_{n+1}$ minimizes $\mathrm{Tr}\left(\boldsymbol{\Psi}_n^{-1}\mathbf{K}\right) - \log|\mathbf{K}|$. Obviously, $f_t(\mathbf{K}_n, \mathbf{W}_n)$ is bounded below and thus $f_t(\mathbf{K}_n, \mathbf{W}_n)$ is convergent.

Next, it follows from Bolzano-Weierstrass theorem [11] that Algorithm 2 produces at least a convergent subsequence $\left\{\mathbf{K}_{[n]}, \mathbf{W}_{[n]}\right\}_{n=0}^{\infty}$ of the sequence $\{\mathbf{K}_n, \mathbf{W}_n\}_{n=0}^{\infty}$. Let $(\mathbf{K}_t^\star, \mathbf{W}_t^\star)$ be the limit point of this subsequence, we shall show that $(\mathbf{K}_t^\star, \mathbf{W}_t^\star)$ is a saddle point of (7). To do so, we first prove that $\mathbf{K}_t^\star \succ \mathbf{0}$ for the case of

$\bar{\mathbf{H}}\bar{\mathbf{H}}^H \succ \mathbf{0}$ by contrary.[1] Specifically, suppose that $\mathbf{K}_t^\star$ is singular. Let $\mathbf{K}_{[n]} = \mathbf{U}_{[n]}\mathbf{\Delta}_{[n]}\mathbf{U}_{[n]}^H$ be the EVD of $\mathbf{K}_{[n]}$ where $\mathbf{\Delta}_{[n]} = \operatorname{diag}\left(\delta_{[n],1}, \ldots, \delta_{[n],n_{\mathrm{R}}+n_{\mathrm{E}}}\right)$ and $\mathbf{U}_{[n]} = \left[\mathbf{u}_{[n],1} \ldots \mathbf{u}_{[n],n_{\mathrm{R}}+n_{\mathrm{E}}}\right]$, then without loss of generality, we can assume that $\left\{\delta_{[n],1}\right\} \to 0$, leading to

$$\begin{aligned}
&f_t(\mathbf{K}_{[n]}, \mathbf{W}_{[n]}) \\
&\geq \log|\mathbf{K}_{[n]} + \tilde{P}\bar{\mathbf{H}}\bar{\mathbf{H}}^H| - \log|\mathbf{K}_{[n]}| - \tilde{P}\operatorname{Tr}(\mathbf{Q}_k) + c \\
&= \log|\mathbf{I} + \tilde{P}\bar{\mathbf{H}}^H\mathbf{K}_{[n]}^{-1}\bar{\mathbf{H}}| - \tilde{P}\operatorname{Tr}(\mathbf{Q}_k) + c \\
&= \log|\mathbf{I} + \tilde{P}\dot{\mathbf{H}}^H\mathbf{\Delta}_{[n]}^{-1}\dot{\mathbf{H}}| - \tilde{P}\operatorname{Tr}(\mathbf{Q}_k) + c \\
&= \log|\mathbf{B}_{[n]}| + \log(1 + \tilde{P}\delta_{[n],1}^{-1}\dot{\mathbf{h}}_1^H\mathbf{B}_{[n]}^{-1}\dot{\mathbf{h}}_1) - \tilde{P}\operatorname{Tr}(\mathbf{Q}_k) + c
\end{aligned} \tag{26}$$

where $\tilde{P} = \frac{P}{n_{\mathrm{T}}}$, $\dot{\mathbf{H}} = \bar{\mathbf{H}}^H\mathbf{U}_{[n]} = \left[\dot{\mathbf{h}}_1 \ldots \dot{\mathbf{h}}_{n_{\mathrm{R}}+n_{\mathrm{E}}}\right]$ and $\mathbf{B}_{[n]} = \mathbf{I} + \tilde{P}\sum_{l=2}^{n_{\mathrm{T}}}\delta_{[n],1}^{-1}\dot{\mathbf{h}}_l^H\mathbf{B}_{[n]}^{-1}\dot{\mathbf{h}}_l$ with the maximum eigenvalue $\lambda_{\max}\left(\mathbf{B}_{[n]}\right) \geq 1$. Therefore, we have

$$\begin{aligned}
&f_t(\mathbf{K}_{[n]}, \mathbf{W}_{[n]}) \\
&\quad\geq \log\left(1 + \lambda_{\max}\left(\mathbf{B}_{[n]}\right)\right) - \tilde{P}\operatorname{Tr}(\mathbf{Q}_k) \\
&\quad\quad + \log\left(1 + \frac{\tilde{P}\delta_{[n],1}^{-1}}{\lambda_{\max}\left(\mathbf{B}_{[n]}\right)}\mathbf{u}_{[n],1}^H\bar{\mathbf{H}}\bar{\mathbf{H}}^H\mathbf{u}_{[n],1}\right) + c \\
&\quad\geq \log\left(1 + \lambda_{\max}\left(\mathbf{B}_{[n]}\right)\right) - \tilde{P}\operatorname{Tr}(\mathbf{Q}_k) \\
&\quad\quad + \log\left(1 + \tilde{P}\frac{\delta_{[n],1}^{-1}}{\lambda_{\max}\left(\mathbf{B}_{[n]}\right)}\lambda_{\min}\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H\right)\right) + c
\end{aligned} \tag{27}$$

where $\lambda_{\min}\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H\right) > 0$ is the minimum eigenvalue of $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$. It follows from (27) that $f_t(\mathbf{K}_{[n]}, \mathbf{W}_{[n]}) \to \infty$ as $\delta_{[n],1} \to 0$ in both cases of bounded and unbounded of $\lambda_{\max}\left(\mathbf{B}_{[n]}\right)$, leading to the contradiction with the first statement of the lemma. Therefore, we have $\mathbf{K}_t^\star \succ \mathbf{0}$ and $\lim_{n\to\infty} f_t(\mathbf{K}_n, \mathbf{W}_n) = f_t(\mathbf{K}_t^\star, \mathbf{W}_t^\star)$. With these results and by using the same arguments as in [8, Appendix B], we can show that $(\mathbf{K}_t^\star, \mathbf{W}_t^\star)$ is a saddle point of (7).

Appendix C
Proof of Lemma 4

Proof: For each t, we have

$$\begin{aligned}
&f\left(\mathbf{\Omega}_{t+1}^\star, \mathbf{S}_{t+1}^\star\right) \\
&\geq f_t\left(\mathbf{\Omega}_{t+1}^\star, \mathbf{S}_{t+1}^\star\right) && (28)\\
&\geq f_t\left(\mathbf{\Omega}_{t+1}^\star, \mathbf{S}_t^\star\right) && (29)\\
&= \log|\mathbf{\Omega}_{t+1}^\star + \bar{\mathbf{H}}\mathbf{S}_t^\star\bar{\mathbf{H}}^H| - \log|\mathbf{\Omega}_{t+1}^\star| - \log|\mathbf{I} + \mathbf{G}\mathbf{S}_t^\star\mathbf{G}^H| && (30)\\
&\geq f\left(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star\right) && (31)
\end{aligned}$$

where (28) is explained below (6), (29) is due to the fact that $\mathbf{S}_{t+1}^\star$ maximizes $f_t\left(\mathbf{\Omega}_{t+1}^\star, \mathbf{S}\right)$, and (31) is from the fact that $\mathbf{\Omega}_t^\star$ minimizes $\log|\mathbf{\Omega} + \bar{\mathbf{H}}\mathbf{S}_t^\star\bar{\mathbf{H}}^H| - \log|\mathbf{\Omega}|$. Hence, $\left\{f(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star)\right\}_t$ is an increasing sequence. Furthermore, $f\left(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star\right)$ is upper-bounded, and thus converges.

Again, according to Bolzano-Weierstrass theorem, we can extract a convergent subsequence $\left\{\left(\mathbf{\Omega}_{[t]}^\star, \mathbf{S}_{[t]}^\star\right)\right\}_{t=0}^\infty$ from the sequence $\left\{\left(\mathbf{\Omega}_t^\star, \mathbf{S}_t^\star\right)\right\}_{t=0}^\infty$. Let $\left(\mathbf{\Omega}^\star, \mathbf{S}^\star\right)$ be the limit point of this subsequence. The proof that $\left(\mathbf{\Omega}^\star, \mathbf{S}^\star\right)$ is a saddle point of (3) is straightforward and skipped for the sake of brevity. ■

[1] If $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ is singular, we can add a small regularization term into $\bar{\mathbf{H}}\mathbf{S}\bar{\mathbf{H}}^H$ to form $\bar{\mathbf{H}}\mathbf{S}\bar{\mathbf{H}}^H + \epsilon\mathbf{I}$ and find the min-max solution for new problem. Then, by the continuity property of $f_t(\mathbf{K}, \mathbf{W})$ with respect to ϵ for $\epsilon \geq 0$, we can reach the solution for the original problem.

Acknowledgment

This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 17/CDA/4786. This work was supported in part by the Academy of Finland under projects FURMESFuN (Grant31089), and 6Genesis Flagship (Grant 318927).

References

[1] Y. Shiu, S. Y. Chang, H. Wu, S. C. . Huang, and H. Chen, "Physical layer security in wireless networks: a tutorial," *IEEE Wireless Commun. Mag.*, vol. 18, no. 2, pp. 66–74, April 2011.

[2] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Wireless Commun. Mag.*, vol. 53, no. 4, pp. 20–27, April 2015.

[3] 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.

[4] 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.

[5] S. Loyka and C. D. Charalambous, "An algorithm for global maximization of secrecy rates in Gaussian MIMO wiretap channels," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2288–2299, June 2015.

[6] A. Ghosh and S. Boyd, "Minimax and convex-concave games," https://web.stanford.edu/class/ee392o/cvxccv.pdf, 2004.

[7] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, June 2007.

[8] T. M. Pham, R. Farrell, and L. Tran, "Revisiting the MIMO capacity with per-antenna power constraint: Fixed-point iteration and alternating optimization," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 388–401, Jan 2019.

[9] Q. Li, M. Hong, H. Wai, Y. Liu, W. Ma, and Z. Luo, "Transmit solutions for MIMO wiretap channels using alternating optimization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1714–1727, Sep. 2013.

[10] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st ed., S. Boyd, Ed. Cambridge, 2004.

[11] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*, 4th ed. Wiley, 2011.