On the Optimality of the Stationary Solution of Secrecy Rate Maximization for MIMO Wiretap Channel

Anshu Mukherjee, *Graduate Student Member, IEEE*, Vaibhav Kumar, *Member, IEEE*, Eduard Jorswieck, *Fellow, IEEE*, Björn Ottersten, *Fellow, IEEE*, and Le-Nam Tran, *Senior Member, IEEE*

***Abstract*—To achieve perfect secrecy in a multiple-input multiple-output (MIMO) Gaussian wiretap channel (WTC), we need to find its secrecy capacity and optimal signaling, which involves solving a difference of convex functions program known to be non-convex for the non-degraded case. To deal with this, a class of existing solutions have been developed but only local optimality is guaranteed by standard convergence analysis. Interestingly, our extensive numerical experiments have shown that these local optimization methods *indeed achieve global optimality*. In this letter, we provide an analytical proof for this observation. To achieve this, we show that the Karush-Kuhn-Tucker (KKT) conditions of the secrecy rate maximization problem admit a unique solution for *both degraded and non-degraded cases*. Motivated by this, we also propose a low-complexity algorithm to find a stationary point. Numerical results are presented to verify the theoretical analysis.**

***Index Terms*—MIMO, wiretap channel, secrecy capacity, sum power constraint, KKT conditions.**

I. INTRODUCTION

WITH the advent of new wireless communication applications including social networking, financial transactions, and military-related communications, there has been an ever increasing demand for privacy-preserving communication services. Traditional data security schemes such as cryptography are implemented at a higher layer of the communication network. However, Wyner in his seminal work introduced an information-theoretic paradigm of physical layer security (PLS) for discrete memoryless wiretap channel (WTC) [1]. Building on [1], the authors in [2] derived the secrecy capacity of a Gaussian WTC, and established that the achievable rate region of a Gaussian WTC can be completely characterized by the corresponding secrecy capacity.

It was established in [3] that the secrecy rate maximization (SRM) problem for the general multiple-input multiple-output (MIMO) Gaussian WTC is non-convex and thus is difficult to solve. The authors in [3] therefore considered a special case of multiple-input single-output (MISO) Gaussian WTC with perfect instantaneous channel state information at the transmitter (CSIT), and derived an analytical solution for the (reformulated) optimization problem. In [4], a beamforming strategy was applied and was shown to be secrecy capacity-achieving for the case of a multi-antenna transmitter, a single-antenna receiver and a multi-antenna eavesdropper (MISOME) system. The problem of computing the secrecy capacity for the case of a multi-antenna transmitter, a multi-antenna receiver and a multi-antenna eavesdropper (MIMOME) was studied in [5], where the authors showed that Gaussian signaling indeed achieves the secrecy capacity and also derived the optimal covariance structure. Independently, the secrecy capacity of the MIMO WTC was also studied in [6]. There is indeed a rich literature concerning analytical and numerical solutions to find the secrecy capacity of the MIMO WTC channel [7]–[14].

It is well known that the secrecy capacity problem for the non-degraded MIMO WTC is non-convex in the original form.[1] Consequently, there has been a concern that algorithms based on convex optimization techniques applied to the original form can be trapped in a locally optimal solution far from the secrecy capacity. To avoid this, the equivalent convex-concave reformulation of the secrecy capacity problem has been used to find the optimal signaling for the non-degraded MIMO WTC [6], [10]. Against this background, the main contributions in this letter are as follows:

- We give a rigorous analytical proof that there exists *a unique Karush–Kuhn–Tucker (KKT) point* for the secrecy capacity problem for both degraded and non-degraded Gaussian MIMO WTC. This interesting result in fact establishes that existing local optimization methods such as [8] for the secrecy problem indeed yield the globally-optimal solution.
- Motivated by this result, we propose an accelerated gradient projection algorithm with adaptive momentum parameters that solves the secrecy capacity problem directly, rather than the equivalent convex-concave form.

Manuscript received August 27, 2021; accepted November 8, 2021. Date of publication November 16, 2021; date of current version February 17, 2022. This work was supported by a Grant from Science Foundation Ireland under Grant 17/CDA/4786. The work of Eduard Jorswieck was supported in part by the German Research Foundation (DFG) under Grant JO 801/25-1. The work of Björn Ottersten was supported in part by ERC ADG AGNOSTIC under Grant 742648. The associate editor coordinating the review of this article and approving it for publication was X. Li. *(Corresponding author: Anshu Mukherjee.)*

Anshu Mukherjee, Vaibhav Kumar, and Le-Nam Tran are with the School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, D04 V1W8 Ireland (e-mail: anshu.mukherjee@ucdconnect.ie; vaibhav.kumar@ieee.org; nam.tran@ucd.ie).

Eduard Jorswieck is with the Institute of Communications Technology, Technische Universität Braunschweig, 38106 Braunschweig, Germany (e-mail: jorswieck@ifn.ing.tu-bs.de).

Björn Ottersten is with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg (e-mail: bjorn.ottersten@uni.lu).

Digital Object Identifier 10.1109/LWC.2021.3128331

[1]The MIMO WTC is said to be non-degraded if the difference between the receiver's channel and the eavesdropper's channel is indefinite.

The proposed algorithm is provably convergent to a KKT point, and thus solves the MIMOME precoding problem globally and efficiently.

Notation: We use bold uppercase and lowercase letters to denote matrices and vectors, respectively. $\mathrm{tr}(\mathbf{X}), |\mathbf{X}|$ and $\|\mathbf{X}\|$ denote the trace, determinant and Frobenius norm of $\mathbf{X}$, respectively. By $\mathbf{X}_{i,j}$ we denote the *j*-th element of the *i*-th row of matrix $\mathbf{X}$. $(\cdot)^{\dagger}$ and $(\cdot)^{\mathsf{T}}$ denote the Hermitian transpose and (ordinary) transpose, respectively. $\mathbb{C}^{M\times N}$ denotes the space of complex matrices of size $M \times N$, and $\mathbb{E}\{\cdot\}$ is the expectation operator. $\mathrm{diag}(\mathbf{x})$ denotes the square diagonal matrix which has the elements of $\mathbf{x}$ on the main diagonal, and $[x]_{+} \triangleq \max\{x, 0\}$. $\mathbf{I}$ and $\mathbf{0}$ represent identity and zero matrices, respectively. By $\mathbf{A} \succeq (\succ)\mathbf{B}$ we mean that $\mathbf{A}-\mathbf{B}$ is positive semidefinite (definite). The maximum eigenvalue of $\mathbf{X}$ is denoted by $\sigma_{\max}(\mathbf{X})$.

II. System Model and Secrecy Capacity Problem

We consider a communication system that includes Alice as the transmitter, Bob as the legitimate receiver, and Eve as the eavesdropper. In this MIMO system, Alice wants to transmit information to Bob in the presence of Eve, where Alice is equipped with N_t number of transmitting antenna, Bob and Eve are equipped with N_r and N_e number of antennas respectively. Let us denote, $\mathbf{H} \in \mathbb{C}^{N_r\times N_t}$ and $\mathbf{G} \in \mathbb{C}^{N_e\times N_t}$ as the channel matrices corresponding to Bob and Eve. The received signal at legitimate receiver and at eavesdropper can be expressed as

$$\mathbf{y}_b = \mathbf{H}\mathbf{q} + \mathbf{z}_b;\ \mathbf{y}_e = \mathbf{G}\mathbf{q} + \mathbf{z}_e \tag{1}$$

where $\mathbf{q}$ is the transmitted signal $\mathbf{q} \in \mathbb{C}^{N_t\times 1}$. The additive white Gaussian noise at the legitimate receiver and at the eavesdropper represented as $\mathbf{z}_b \in \mathbb{C}^{N_r\times 1} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ and $\mathbf{z}_e \in \mathbb{C}^{N_e\times 1} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$. In this letter, we assume that $\mathbf{H}$ and $\mathbf{G}$ are perfectly known at Alice and Bob. Let $\mathbf{Q} = \mathbb{E}\{\mathbf{q}\mathbf{q}^{\dagger}\} \succeq \mathbf{0}$ be the input covariance matrix. Then the secrecy capacity under a sum power constraint (SPC) has been expressed as [6]

$$C_s = \max\big\{ C_s(\mathbf{Q}) \big|\ \mathbf{Q} \in \mathcal{Q}\big\} \tag{2}$$

where $\mathcal{Q} = \{\mathbf{Q} \mid \mathrm{tr}(\mathbf{Q}) \le P_T; \mathbf{Q} \succeq \mathbf{0}\}$, $C_s(\mathbf{Q}) = \ln|\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}| - \ln|\mathbf{I} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\dagger}|$, $P_T > 0$ is the total transmit power. Problem (2) is non-convex in general.

III. Uniqueness of KKT Point of (2)

We remark that if $\mathbf{H}^{\dagger}\mathbf{H} - \mathbf{G}^{\dagger}\mathbf{G}$ is negative semi-definite, C_s is zero. Thus, the next theorem provides a complete characterization of the uniqueness of the stationary point of (2).

Theorem 1: Assume $\mathbf{H}^{\dagger}\mathbf{H} - \mathbf{G}^{\dagger}\mathbf{G}$ is not negative semi-definite. Then problem (2) has a unique KKT point. Therefore the KKT conditions are necessary and sufficient for the optimality of problem (2).

Proof: The Lagrangian function of problem (2) is

$$\mathcal{L}(\mathbf{Q},\mathbf{Z},\lambda) = C_s(\mathbf{Q}) - \lambda(\mathrm{tr}(\mathbf{Q}) - P_T) + \mathrm{tr}(\mathbf{Q}\mathbf{Z}) \tag{3}$$

where $\lambda \ge 0$ and $\mathbf{Z} \succeq \mathbf{0}$ are the Lagrangian multiplier for the constraints $\mathrm{tr}(\mathbf{Q}) \le P_T$ and $\mathbf{Q} \succeq \mathbf{0}$, respectively. Note that the gradient of $C_s(\mathbf{Q})$ is

$$\nabla C_s(\mathbf{Q}) = \mathbf{H}^{\dagger}\Big(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\Big)^{-1}\mathbf{H} - \mathbf{G}^{\dagger}\Big(\mathbf{I} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\dagger}\Big)^{-1}\mathbf{G}. \tag{4}$$

Thus, the KKT conditions of (2) are given by

$$\mathbf{H}^{\dagger}\big(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\big)^{-1}\mathbf{H} - \mathbf{G}^{\dagger}\big(\mathbf{I} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\dagger}\big)^{-1}\mathbf{G} - \lambda\mathbf{I} + \mathbf{Z} = \mathbf{0} \tag{5a}$$

$$\mathrm{tr}(\mathbf{Q}) \le P_T; \lambda \ge 0; \lambda\big(\mathrm{tr}(\mathbf{Q}) - P_T\big) = 0 \tag{5b}$$

$$\mathbf{Z} \succeq \mathbf{0}; \mathbf{Q} \succeq \mathbf{0}; \mathbf{Q}\mathbf{Z} = \mathbf{0}. \tag{5c}$$

Throughout the proof we use the following equality which is a special case of the matrix inversion lemma [15, eq. (2.8.20)]

$$\mathbf{Y}(\mathbf{I} + \mathbf{X}\mathbf{Y})^{-1} = (\mathbf{I} + \mathbf{Y}\mathbf{X})^{-1}\mathbf{Y}. \tag{6}$$

In the first part of the proof we assert that $\lambda > 0$ and thus $\mathrm{tr}(\mathbf{Q}) = P_T$ if $\mathbf{Q}$ is a KKT point. To proceed we note that applying (6) to (5a) yields

$$(\mathbf{I} + \mathbf{H}^{\dagger}\mathbf{H}\mathbf{Q})^{-1}\mathbf{H}^{\dagger}\mathbf{H} - \mathbf{G}^{\dagger}\mathbf{G}(\mathbf{I} + \mathbf{Q}\mathbf{G}^{\dagger}\mathbf{G})^{-\mathbf{1}} = \lambda\mathbf{I} - \mathbf{Z}. \tag{7}$$

Suppose to the contrary that $\lambda = 0$. Then (7) reduces to

$$(\mathbf{I} + \mathbf{H}^{\dagger}\mathbf{H}\mathbf{Q})^{-1}\mathbf{H}^{\dagger}\mathbf{H} - \mathbf{G}^{\dagger}\mathbf{G}(\mathbf{I} + \mathbf{Q}\mathbf{G}^{\dagger}\mathbf{G})^{-1} = -\mathbf{Z} \tag{8}$$

which is equivalent to

$$\mathbf{H}^{\dagger}\mathbf{H} - \mathbf{G}^{\dagger}\mathbf{G} = -\Big(\mathbf{I} + \mathbf{H}^{\dagger}\mathbf{H}\mathbf{Q}\Big)\mathbf{Z}\Big(\mathbf{I} + \mathbf{Q}\mathbf{G}^{\dagger}\mathbf{G}\Big). \tag{9}$$

It is clear that the right hand side of (9) is negative semidefinite, which contradicts the assumption that $\mathbf{H}^{\dagger}\mathbf{H} - \mathbf{G}^{\dagger}\mathbf{G}$ is positive semidefinite or indefinite. Thus we can conclude that $\lambda > 0$ and thus $\mathrm{tr}(\mathbf{Q}) = P_T$.

Next, adopting proof by contradiction, we show that there is a unique solution $\mathbf{Q}$ to the KKT conditions. Suppose to the contrary that $(\mathbf{Q}_1, \mathbf{Z}_1, \lambda_1)$ and $(\mathbf{Q}_2, \mathbf{Z}_2, \lambda_2)$ are two different KKT points of (2). Let us define

$$\mathbf{\Phi}_i = \mathbf{H}^{\dagger}\big(\mathbf{I} + \mathbf{H}\mathbf{Q}_i\mathbf{H}^{\dagger}\big)^{-1}\mathbf{H}; \mathbf{\Psi}_i = \mathbf{G}^{\dagger}\big(\mathbf{I} + \mathbf{G}\mathbf{Q}_i\mathbf{G}^{\dagger}\big)^{-1}\mathbf{G} \tag{10}$$

for $i = \{1, 2\}$. Then (5a) for those two KKT points is

$$\mathbf{Z}_1 + \mathbf{\Phi}_1 - \mathbf{\Psi}_1 = \lambda_1\mathbf{I} \tag{11a}$$

$$\mathbf{Z}_2 + \mathbf{\Phi}_2 - \mathbf{\Psi}_2 = \lambda_2\mathbf{I}. \tag{11b}$$

Since, $\mathrm{tr}(\mathbf{Q}_1) = \mathrm{tr}(\mathbf{Q}_2) = P_T$, it is easy to check that

$$\mathrm{tr}\big(\big(\mathbf{\Phi}_1 - \mathbf{\Psi}_1 + \mathbf{Z}_1\big)\big(\mathbf{Q}_1 - \mathbf{Q}_2\big)\big) = \lambda_1\,\mathrm{tr}\big(\mathbf{Q}_1 - \mathbf{Q}_2\big) = 0, \tag{12}$$

$$\mathrm{tr}\big(\big(\mathbf{\Phi}_2 - \mathbf{\Psi}_2 + \mathbf{Z}_2\big)\big(\mathbf{Q}_1 - \mathbf{Q}_2\big)\big) = \lambda_2\,\mathrm{tr}\big(\mathbf{Q}_1 - \mathbf{Q}_2\big) = 0. \tag{13}$$

Now, subtracting (13) from (12), we obtain

$$\mathrm{tr}\big((\mathbf{\Phi}_1 - \mathbf{\Phi}_2 - (\mathbf{\Psi}_1 - \mathbf{\Psi}_2))(\mathbf{Q}_1 - \mathbf{Q}_2)\big) - \mathrm{tr}(\mathbf{Z}_1\mathbf{Q}_2 + \mathbf{Z}_2\mathbf{Q}_1) = 0. \tag{14}$$

Our purpose in the sequel is to show that (14) is impossible if $\mathbf{Q}_1 \ne \mathbf{Q}_2$. To this end we first express $\mathbf{\Phi}_1 - \mathbf{\Phi}_2$ as in (15) shown at the bottom of the next page. Note that in (15b) we have used the fact that $\mathbf{X}^{-1} - \mathbf{Y}^{-1} = \mathbf{X}^{-1}(\mathbf{Y} - \mathbf{X})\mathbf{Y}^{-1}$ for invertible matrices $\mathbf{X}$ and $\mathbf{Y}$, and that (15c) is due to (11). Similarly we can write

$$\mathbf{\Psi}_1 - \mathbf{\Psi}_2 = \mathbf{\Psi}_1(\mathbf{Q}_2 - \mathbf{Q}_1)\mathbf{\Psi}_2. \tag{16}$$

Combining (16) with (15c) produces (17) shown at the bottom of the next page. Substituting (17) into (14) we can rewrite the first term in the left-hand side of (14) as in (18) shown at the bottom of the page.

Comparing (18) and (14), our next step is to bound the first two terms of the right-hand side (RHS) of (18) properly. To this end we multiply both sides of (11a) with $\mathbf{Q}_1$, and note that $\mathbf{Q}_1\mathbf{Z}_1 = 0$, which yields

$$\lambda_1\mathbf{Q}_1 + \boldsymbol{\Psi}_1\mathbf{Q}_1 = \boldsymbol{\Phi}_1\mathbf{Q}_1 \tag{19}$$

and thus we have

$$\begin{aligned}&\operatorname{tr}((\lambda_1\mathbf{Q}_1 + \boldsymbol{\Psi}_1\mathbf{Q}_1)\mathbf{Z}_2\mathbf{Q}_1) - \operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1) = \operatorname{tr}(\boldsymbol{\Phi}_1\mathbf{Q}_1\mathbf{Z}_2\mathbf{Q}_1)\\ &- \operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1) = \operatorname{tr}((\boldsymbol{\Phi}_1\mathbf{Q}_1 - \mathbf{I})\mathbf{Z}_2\mathbf{Q}_1) = \operatorname{tr}\Big((\mathbf{I}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma} - \mathbf{I}\Big)\mathbf{Z}_2\mathbf{Q}_1\end{aligned}$$

where $\boldsymbol{\Gamma} = \mathbf{H}^\dagger\mathbf{H}\mathbf{Q}_1$. It is easy to see that $(\mathbf{I}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma} \preceq \mathbf{I}$ and thus $((\mathbf{I}+\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma} - \mathbf{I})\mathbf{Z}_2\mathbf{Q}_1 \preceq \mathbf{0}$, which leads to

$$\operatorname{tr}((\lambda_1\mathbf{Q}_1 + \boldsymbol{\Psi}_1\mathbf{Q}_1)\mathbf{Z}_2\mathbf{Q}_1) - \operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1) \leq 0. \tag{20}$$

Similarly, we have

$$\operatorname{tr}((\lambda_2\mathbf{Q}_2 + \boldsymbol{\Psi}_2\mathbf{Q}_2)\mathbf{Z}_1\mathbf{Q}_2) - \operatorname{tr}(\mathbf{Z}_1\mathbf{Q}_2) \leq 0. \tag{21}$$

Adding (20) and (21) gives

$$\begin{aligned}&\operatorname{tr}((\lambda_1\mathbf{Q}_1 + \boldsymbol{\Psi}_1\mathbf{Q}_1)\mathbf{Z}_2\mathbf{Q}_1) + \operatorname{tr}((\lambda_2\mathbf{Q}_2 + \boldsymbol{\Psi}_2\mathbf{Q}_2)\mathbf{Z}_1\mathbf{Q}_2)\\ &\quad - \operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1 + \mathbf{Z}_1\mathbf{Q}_2) \leq 0.\end{aligned} \tag{22}$$

which results in (23) shown at the bottom of the page. It is not difficult to check that (23) holds due to (22) and also the fact that the last four terms of the RHS of (23) are non-positive. Now subtracting both sides of (18) by $\operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1 + \mathbf{Z}_1\mathbf{Q}_2)$ and using (23) produces

$$\operatorname{tr}\Big(\big(\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2)\big)(\mathbf{Q}_1 - \mathbf{Q}_2)\Big) - \operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1 + \mathbf{Z}_1\mathbf{Q}_2) \leq 0. \tag{24}$$

We remark that if $\mathbf{Q}_1 \neq \mathbf{Q}_2$, then $\lambda_1\lambda_2\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)(\mathbf{Q}_2 - \mathbf{Q}_1)) > 0$. Consequently, the inequality in (23) is strict, and so is (24), which contradicts (14) and completes the proof. ■

An immediate consequence of Theorem 1 is that any local optimization method is also a global optimization method for the secrecy problem. In the next we exploit this property to derive an efficient numerical method to solve (2).

Algorithm 1: Iterative Algorithm for Solving (2)

Input: $\mathbf{Y}_1 = \mathbf{Q}_0 \in \mathcal{Q}$, $\beta_0 = L_0 > 0$, $\alpha = 1$, $\xi \in (0,1)$, $\gamma_u > 1$, $\epsilon > 0$, $k \leftarrow 1$

1 **repeat**
2   **repeat**   `/* line search */`
3         $\mathbf{Q}_k = \Pi_{\mathcal{Q}}\big(\mathbf{Y}_k + \frac{1}{\beta_{k-1}}\nabla C_S(\mathbf{Y}_k)\big)$
4   **if** $C_s(\mathbf{Q}_k) < \mu_{\beta_{k-1}}(\mathbf{Y}_k;\mathbf{Q}_k)$ **then**
5             $\beta_{k-1} = \gamma_u\beta_{k-1}$
6   **end**
7   **until** $C_s(\mathbf{Q}_k) \geq \mu_{\beta_{k-1}}(\mathbf{Y}_k;\mathbf{Q}_k)$
8     $\beta_k = \max\{L_0, \beta_{k-1}/\gamma_u\}$
9     $\mathbf{Z}_k = \mathbf{Q}_k + \alpha(\mathbf{Q}_k - \mathbf{Q}_{k-1})$
10   **if** $C_s(\mathbf{Z}_k) \geq C_s(\mathbf{Q}_k)$ *and* $\mathbf{Z}_k$ *is feasible* **then**
11   Update $\alpha = \min\Big\{\frac{\alpha}{\xi}, 1\Big\}$ and $\mathbf{Y}_{k+1} = \mathbf{Z}_k$
12   **else**
13   Update $\alpha = \xi\alpha$ and $\mathbf{Y}_{k+1} = \mathbf{Q}_k$
14   **end**
15     $k \leftarrow k+1$
16 **until** $|C_s(\mathbf{Q}_k)| - |C_s(\mathbf{Q}_{k-1})| \leq \epsilon$

Output: $\mathbf{Q}_k$

IV. A Low-Complexity Method

A. Algorithm Description

The proposed method is based on the accelerated projected gradient method for non-convex programming with adaptive momentum presented in [16]. The pseudo-code of the proposed method is provided in Algorithm 1, which is explained in detail as follows. Let $\mathbf{Y}_k$ be the current operating point. Then we take a projected gradient step to obtain the current iterate $\mathbf{Q}_k$ (see Line 3). Note that the notation $\Pi_{\mathcal{Q}}(\mathbf{X})$ in Line 3 denotes the projection of a given point $\mathbf{X}$ onto the feasible set $\mathcal{Q}$, i.e., $\Pi_{\mathcal{Q}}(\mathbf{X}) = \operatorname{argmin}\{||\mathbf{U} - \mathbf{X}|| \mid \mathbf{U} \in \mathcal{Q}\}$. In contrast to [16] where a constant stepsize is used, we implement a backtracking line search as done in Lines 2-7 to find a proper step size, which is adopted from [17]. For this purpose we define a quadratic model of $C_s(\mathbf{Q})$ as

$$\mu_\beta(\mathbf{Q};\bar{\mathbf{Q}}) = C_s(\mathbf{Q}) + \operatorname{tr}\big(\nabla C_s(\mathbf{Q})\big(\bar{\mathbf{Q}} - \mathbf{Q}\big)\big) - \frac{\beta}{2}\big\|\bar{\mathbf{Q}} - \mathbf{Q}\big\|^2. \tag{25}$$

Recall that if $\beta \geq L$, where $L > 0$ is a Lipschitz constant of $\nabla C_s(\mathbf{Q})$ on $\mathcal{Q}$, then the inequality $C_s(\bar{\mathbf{Q}}) \geq \mu_\beta(\mathbf{Q};\bar{\mathbf{Q}})$ holds [17]. In the Appendix we show that $L = \sigma^2_{\max}(\mathbf{H}^\dagger\mathbf{H}) + \sigma^2_{\max}(\mathbf{G}^\dagger\mathbf{G})$ is a Lipschitz constant for of $\nabla C_s(\mathbf{Q})$. To find a proper β in each iteration, we start from the value of β

$$\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 = \mathbf{H}^\dagger\Big(\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^\dagger\Big)^{-1}\mathbf{H} - \mathbf{H}^\dagger\Big(\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^\dagger\Big)^{-1}\mathbf{H} = \mathbf{H}^\dagger\bigg(\Big(\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^\dagger\Big)^{-1} - \Big(\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^\dagger\Big)^{-1}\bigg)\mathbf{H} \tag{15a}$$

$$= \mathbf{H}^\dagger\Big(\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^\dagger\Big)^{-1}\mathbf{H}(\mathbf{Q}_2 - \mathbf{Q}_1)\mathbf{H}^\dagger\Big(\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^\dagger\Big)^{-1}\mathbf{H} = \boldsymbol{\Phi}_1(\mathbf{Q}_2 - \mathbf{Q}_1)\boldsymbol{\Phi}_2 \tag{15b}$$

$$= (\boldsymbol{\Psi}_1 + \lambda_1\mathbf{I} - \mathbf{Z}_1)(\mathbf{Q}_2 - \mathbf{Q}_1)(\boldsymbol{\Psi}_2 + \lambda_2\mathbf{I} - \mathbf{Z}_2) \tag{15c}$$

$$(\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2) - (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2) = \boldsymbol{\Psi}_1\big(\mathbf{Q}_2 - \mathbf{Q}_1\big)\big(\lambda_2\mathbf{I} - \mathbf{Z}_2\big) + \big(\lambda_1\mathbf{I} - \mathbf{Z}_1\big)\big(\mathbf{Q}_2 - \mathbf{Q}_1\big)\boldsymbol{\Psi}_2 + \big(\lambda_1\mathbf{I} - \mathbf{Z}_1\big)\big(\mathbf{Q}_2 - \mathbf{Q}_1\big)\big(\lambda_2\mathbf{I} - \mathbf{Z}_2\big) \tag{17}$$

$$\begin{aligned}&\operatorname{tr}((\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2))(\mathbf{Q}_1 - \mathbf{Q}_2)) = \operatorname{tr}((\lambda_1\mathbf{Q}_1 + \boldsymbol{\Psi}_1\mathbf{Q}_1)\mathbf{Z}_2\mathbf{Q}_1) + \operatorname{tr}((\lambda_2\mathbf{Q}_2 + \boldsymbol{\Psi}_2\mathbf{Q}_2)\mathbf{Z}_1\mathbf{Q}_2)\\ &\quad -\lambda_1\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)\boldsymbol{\Psi}_2(\mathbf{Q}_2 - \mathbf{Q}_1)) - \lambda_2\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)\boldsymbol{\Psi}_1(\mathbf{Q}_2 - \mathbf{Q}_1)) - \lambda_1\lambda_2\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)(\mathbf{Q}_2 - \mathbf{Q}_1))\end{aligned} \tag{18}$$

$$\begin{aligned}&\operatorname{tr}((\lambda_1\mathbf{Q}_1 + \boldsymbol{\Psi}_1\mathbf{Q}_1)\mathbf{Z}_2\mathbf{Q}_1) + \operatorname{tr}((\lambda_2\mathbf{Q}_2 + \boldsymbol{\Psi}_2\mathbf{Q}_2)\mathbf{Z}_1\mathbf{Q}_2) - \operatorname{tr}(\mathbf{Z}_2\mathbf{Q}_1 + \mathbf{Z}_1\mathbf{Q}_2)\\ &\quad -\lambda_1\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)\boldsymbol{\Psi}_2(\mathbf{Q}_2 - \mathbf{Q}_1)) - \lambda_2\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)\boldsymbol{\Psi}_1(\mathbf{Q}_2 - \mathbf{Q}_1)) - \lambda_1\lambda_2\operatorname{tr}((\mathbf{Q}_2 - \mathbf{Q}_1)(\mathbf{Q}_2 - \mathbf{Q}_1)) \leq 0\end{aligned} \tag{23}$$

in the previous iteration and increase it by $\gamma_u > 1$ until $\mu_\beta(\mathbf{Y}_k; \mathbf{Q}_k)$ becomes a lower bound of $C_s(\mathbf{Q}_k)$. In this way, the projected gradient step always produces an improved iterate. Next, for acceleration, we compute the extrapolated point $\mathbf{Z}_k = \mathbf{Q}_k + \alpha(\mathbf{Q}_k - \mathbf{Q}_{k-1})$, where α is called the momentum parameter (see Line 9). For convex optimization, the momentum parameter is fixed. However, since the objective in (2) is non-convex, the extrapolation can be bad and thus α needs to be adapted in accordance with the extrapolated point. To this end a monitor process needs to be considered [16]. Specifically, if the extrapolation reduces the current objective (i.e., bad extrapolation), then the current iteration is taken for the next iteration and α is reduced with a rate ξ (see Line 13). Otherwise, the extrapolated point is taken to the next iteration and α is increased by $1/\xi$ (see Line 11). The stopping criterion for Algorithm 1 is when the increase in the last iteration is less than a small pre-determined parameter ϵ.

Algorithm 1 is very simple to implement because $\nabla C_s(\mathbf{Q})$ is given in closed-form in (4) and the projection of a given point $\mathbf{X}$ onto $\mathcal{Q}$, $\Pi_{\mathcal{Q}}(\mathbf{X})$, admits a water-filling like algorithm as[18]

$$\Pi_{\mathcal{Q}}(\mathbf{X}) = \mathbf{U}\,\mathrm{diag}\big([\mathbf{x} - c]_+\big)\mathbf{U} \tag{26}$$

where $\mathbf{X} = \mathbf{U}\,\mathrm{diag}(\mathbf{x})\mathbf{U}^\dagger$ is the eigenvalue decomposition of $\mathbf{X}$ and c is the root of the following equation

$$\sum\nolimits_{i=1}^{N_t} [x_i - c]_+ = P_T. \tag{27}$$

B. Convergence Analysis

We now show that the iterate sequence $\{\mathbf{Q}_k\}$ returned by Algorithm 1 converges to the stationary solution of (2), which, by Theorem 1, is also the only optimal solution. First, since $\nabla C_s(\mathbf{Q})$ is Lipschitz continuous, the backtracking line search terminates in an finite number of steps. More specifically, the number of line search steps at iteration $k \geq 1$ is bounded by $2 + \log_{\gamma_u}(\beta_k/\beta_{k-1})$. Now suppose that $C_s(\mathbf{Z}_k) \geq C_s(\mathbf{Q}_k)$ and $\mathbf{Z}_k$ is feasible. Then we have $\mathbf{Y}_{k+1} = \mathbf{Z}_k$ (see Line 11). The projection at iteration $k + 1$ can be explicitly written as

$$\mathbf{Q}_{k+1} = \Pi_{\mathcal{Q}}\big(\mathbf{Z}_k + \tfrac{1}{\beta_k}\nabla C_s(\mathbf{Z}_k)\big) \tag{28a}$$

$$= \arg\min_{\mathbf{Q}\in\mathcal{Q}} \big\|\mathbf{Q} - \mathbf{Z}_k - \tfrac{1}{\beta_k}\nabla C_s(\mathbf{Z}_k)\big\|^2 \tag{28b}$$

$$= \arg\min_{\mathbf{Q}\in\mathcal{Q}} \tfrac{\beta_k}{2}\big\|\mathbf{Q} - \mathbf{Z}_k\big\|^2 - \mathrm{tr}\big(\nabla C_s(\mathbf{Z}_k)(\mathbf{Q} - \mathbf{Z}_k)\big). \tag{28c}$$

Note that (28c) implies

$$\tfrac{\beta_k}{2}\big\|\mathbf{Q}_{k+1} - \mathbf{Z}_k\big\|^2 - \mathrm{tr}(\nabla C_s(\mathbf{Z}_k)(\mathbf{Q}_{k+1} - \mathbf{Z}_k)) \leq 0. \tag{29}$$

It follows from the condition in Line 7 that

$$\begin{aligned} C_s(\mathbf{Q}_{k+1}) &\geq \mu_{\beta_k}(\mathbf{Z}_k; \mathbf{Q}_{k+1}) = C_s(\mathbf{Z}_k) \\ &+ \mathrm{tr}(\nabla C_s(\mathbf{Z}_k)(\mathbf{Q}_{k+1} - \mathbf{Z}_k) - \tfrac{\beta}{2}\big\|\mathbf{Q}_{k+1} - \mathbf{Z}_k\big\|^2 \end{aligned} \tag{30}$$

which, by using (29), yields

$$C_s(\mathbf{Q}_{k+1}) \geq C_s(\mathbf{Z}_k) \geq C_s(\mathbf{Q}_k). \tag{31}$$

Similarly, if $\mathbf{Y}_{k+1} = \mathbf{Q}_k$, then we can also prove that $C_s(\mathbf{Q}_{k+1}) \geq C_s(\mathbf{Q}_k)$ following the same procedure. Note that the inequality is strict if $\mathbf{Q}_{k+1} \neq \mathbf{Q}_k$. By noting that the feasible set is compact convex, we can conclude the objective sequence $\{C_s(\mathbf{Q}_k)\}$ is convergent and there exists a subsequence $\{\mathbf{Q}_k\}$ converging to a limit point $\mathbf{Q}^*$. The proof that $\mathbf{Q}^*$ is a stationary point of (2) is standard and thus omitted here for the sake of brevity [16].

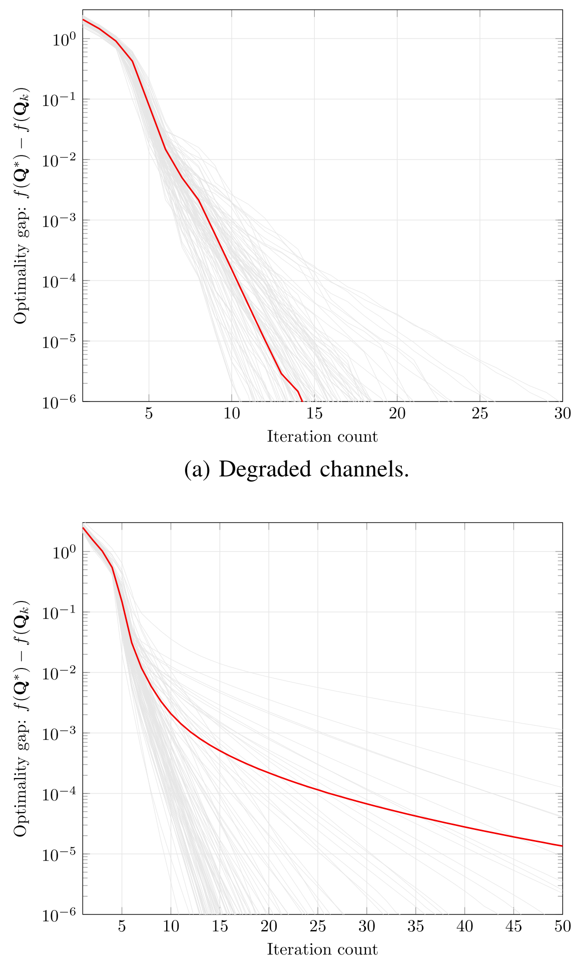

Fig. 1. Convergence results of Algorithm 1. $N_t = N_r = N_e = 4$ at $P_t = 15$ dB.

V. Numerical Results

To illustrate Theorem 1 and also the convergence rate of Algorithm 1, we plot the residual error (i.e., $C_s - C_s(\mathbf{Q}_k)$) for both degraded and non-degraded channels in Figs. 1(a) and 1(b), respectively. The channels $\mathbf{H}$ and $\mathbf{G}$ are generated as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. The secrecy capacity C_S is found using existing optimal algorithms. More specifically, for the degraded MIMO WTC, problem (2) can be reformulated as a standard semidefinite program and thus can be optimally solved by off-the-shelf solvers such as MOSEK [19]. For the non-degraded case, we implement the barrier method [10, Algorithm 3] and [13, Algorithm 3], both of which can find the secrecy capacity but are based on the equivalent convex-concave reformulation. Note that Algorithm 1 is applied to problem (2) directly. As can be seen clearly in Fig. 1, Algorithm 1 achieves monotonic convergence as proved in (31). Also, the optimality gap is reduced quickly to zero as the iteration process continues for both degraded and non-degraded cases, meaning that Algorithm 1 can achieve the same secrecy

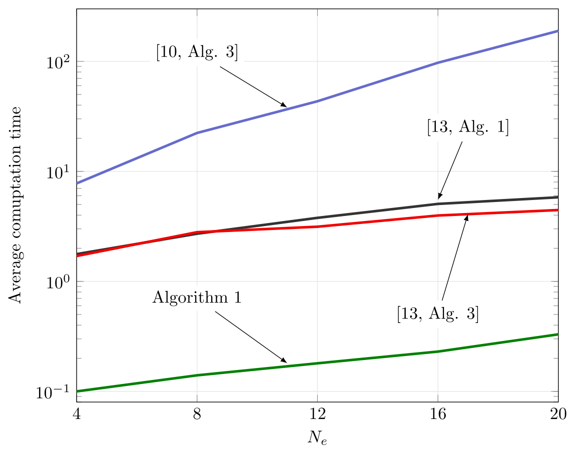

Fig. 2. Time comparison of various algorithms for $N_t = 4$ and $N_r = 3$.

capacity performance compared to other known optimal methods. These results indeed confirm that Algorithm 1, even when applied to the nonconvex form of the secrecy capacity problem, can still compute the optimal solution, which is explained by Theorem 1.

To further demonstrate the benefit of Theorem 1 and Algorithm 1, in Fig. 2, we compare the run time of Algorithm 1 as a function of the number of the antennas at Eve. The simulation codes are built on MATLAB and executed in a 64-bit Windows PC system with 16 GB RAM and Intel Core-i7, 3.20 GHz processor. We plotted the average actual run-time for 200 different channel realizations. The stopping criteria for all the algorithms is when the increase in the resulting objective is less than 10^{-5} during the last 5 iterations. We can see our proposed algorithm outperforms other known methods such as [13, Algorithms 1 and 3] and [10, Algorithm 3] in terms of time complexity.

VI. Conclusion

We have proved that the secrecy rate maximization problem of the general MIMOME WTC (i.e., no assumption is made on whether the channel is degraded) under a sum power constraint, despite its non-convexity, has a unique KKT solution. The proof basically implies that any local optimization method that aims to find a stationary solution can indeed solve secrecy capacity problem for non-degraded MIMO wiretap channels which are known to be non-convex. Motivated by this interesting result, we have also presented an accelerated projected gradient method with adaptive momentum to solve the secrecy problem. Simulation results have demonstrated that the proposed algorithm can find the optimal solution very fast.

Appendix
Lipschitz Constant of $\nabla C_s(\cdot)$

Recall that L > 0 is a Lipschitz constant of $\nabla C_s(\mathbf{Q})$ on $\mathcal{Q}$ if the following inequality holds

$$\big\|\nabla C_s(\mathbf{X}) - \nabla C_s(\mathbf{Y})\big\| \leq L\big\|\mathbf{X}-\mathbf{Y}\big\|, \forall \mathbf{X},\mathbf{Y}\in\mathcal{Q}. \quad (32)$$

Using (4) and the norm inequality, it is easy to see that

$$\begin{aligned}&\big\|\nabla C_s(\mathbf{X}) - \nabla C_s(\mathbf{Y})\big\|\\ &\leq \big\|\mathbf{H}^\dagger\big(\mathbf{I}+\mathbf{HXH}^\dagger\big)^{-1}\mathbf{H}(\mathbf{Y}-\mathbf{X})\mathbf{H}^\dagger\big(\mathbf{I}+\mathbf{HYH}^\dagger\big)^{-1}\mathbf{H}\big\|\\ &\quad+\big\|\mathbf{G}^\dagger\big(\mathbf{I}+\mathbf{GXG}^\dagger\big)^{-1}\mathbf{G}(\mathbf{Y}-\mathbf{X})\mathbf{G}^\dagger\big(\mathbf{I}+\mathbf{GYG}^\dagger\big)^{-1}\mathbf{G}\big\|. \quad (33)\end{aligned}$$

We now recall the following well known inequality:

$$||\mathbf{AB}|| \leq \lambda_{\max}(\mathbf{A})||\mathbf{B}|| \quad (34)$$

where $\lambda_{\max}(\cdot)$ denotes the maximum singular value of the matrix in the argument. Applying the above inequality and by noting that $\lambda_{\max}((\mathbf{I}+\mathbf{HXH}^\dagger)^{-1}) \leq 1$ we have

$$\big\|\nabla C_s(\mathbf{X}) - \nabla C_s(\mathbf{Y})\big\| \leq \big(\sigma\big(\mathbf{H}^\dagger\mathbf{H}\big)+\sigma^2_{\max}\big(\mathbf{G}^\dagger\mathbf{G}\big)\big)\big\|\mathbf{X}-\mathbf{Y}\big\|$$

which means that $L = \sigma^2_{\max}(\mathbf{H}^\dagger\mathbf{H}) + \sigma^2_{\max}(\mathbf{G}^\dagger\mathbf{G})$ is a Lipschitz constant of $\nabla C_s(\cdot)$.

References

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.
[2] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.
[3] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Proc. 41st Annu. Conf. Inf. Sci. Syst.*, Mar. 2007, pp. 905–910.
[4] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2010.
[5] 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.
[6] 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.
[7] S. A. A. Fakoorian and A. L. Swindlehurst, "Full rank solutions for the MIMO Gaussian wiretap channel with an average power constraint," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2620–2631, May 2013.
[8] Q. Li, M. Hong, H.-T. Wai, Y.-F. Liu, W.-K. Ma, and Z.-Q. Luo, "Transmit solutions for MIMO wiretap channels using alternating optimization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1714–1727, Sep. 2013.
[9] 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 2014, pp. 5686–5690.
[10] 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, Jun. 2015.
[11] S. Loyka and C. D. Charalambous, "Optimal signaling for secure communications over Gaussian MIMO wiretap channels," in *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7207–7215, Dec. 2016.
[12] T. V. Nguyen, Q.-D. Vu, M. Juntti, and L.-N. Tran, "A low-complexity algorithm for achieving secrecy capacity in MIMO wiretap channels," in *Proc. IEEE ICC*, Jun. 2020, pp. 1–6.
[13] A. Mukherjee, B. Ottersten, and L.-N. Tran, "Efficient numerical methods for secrecy capacity of Gaussian MIMO wiretap channel," in *Proc. IEEE VTC*, May 2021, pp. 1–5.
[14] A. Mukherjee, B. Ottersten, and L.-N. Tran, "On the secrecy capacity of MIMO wiretap channels: Convex reformulation and efficient numerical methods," *IEEE Trans. Wireless Commun.*, vol. 69, no. 10, pp. 6865–6878, Oct. 2021.
[15] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[16] Q. Li, Y. Zhou, Y. Liang, and P. K. Varshney, "Convergence analysis of proximal gradient with momentum for nonconvex optimization," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 2111–2119.
[17] Y. Nesterov, "Gradient methods for minimizing composite functions," *Math. Program.*, vol. 140, pp. 125–161, Aug. 2013.
[18] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.
[19] M. ApS, *The MOSEK Optimization Toolbox for MATLAB Manual. Version 9.2*, 2020. [Online]. Available: https://docs.mosek.com/9.2/toolbox/index.html