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

Anshu Mukherjee, 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 paper, 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. 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. 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. $\text{tr}(X)$, $\|X\|$ denote the trace, determinant and Frobenius norm of $X$, respectively. By $X_{i,j}$ we denote the $j$-th element of the $i$-th row of matrix $X$. $(\cdot)^{\dagger}$ and $(\cdot)^{\top}$ 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. $\text{diag}(x)$ denotes the square diagonal matrix which has the elements of $x$ on the main diagonal, and $\lfloor x \rfloor_+ \triangleq \max\{x, 0\}$. $\mathbf{I}$ and $\mathbf{0}$ represent identity and zero

\footnote{The MIMO WTC is said to be non-degraded if the difference between the receiver’s channel and the eavesdropper’s channel is indefinite.}
matrices, respectively. By \( A \succeq (\succ) B \) we mean that \( A - B \) is positive semidefinite (definite). The maximum eigenvalue of \( X \) is denoted by \( \sigma_{\text{max}}(X) \).

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, \( H \in \mathbb{C}^{N_r \times N_t} \) and \( G \in \mathbb{C}^{N_e \times N_t} \), as the channel matrices corresponding to Bob and Eve. The received signal at legitimate receiver and at eavesdropper can be expressed as

\[
y_b = Hq + z_b; \quad y_e = Gq + z_e
\]

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

\[
C_s = \max \left\{ C_s(Q) \mid Q \in \mathcal{Q} \right\}
\]

where \( Q = \left\{ Q \mid \text{tr}(Q) \leq P_T; Q \succeq 0 \right\} \), \( C_s(Q) = \ln |I + HQH^\dagger| - \ln |I + GGQG^\dagger| \), \( P_T > 0 \) is the total transmit power. Problem (2) is non-convex in general.

III. UNIQUENESS OF KKT POINT OF (2)

We remark that if \( H^\dagger H - G^\dagger G \) 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 \( H^\dagger H - G^\dagger G \) 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

\[
\mathcal{L}(Q, Z, \lambda) = C_s(Q) - \lambda (\text{tr}(Q) - P_T) + \text{tr}(QZ)
\]

where \( \lambda \geq 0 \) and \( Z \geq 0 \) are the Lagrangian multiplier for the constraints \( \text{tr}(Q) \leq P_T \) and \( Q \succeq 0 \), respectively. Note that the gradient of \( C_s(Q) \) is

\[
\nabla C_s(Q) = H^\dagger (I + HQH^\dagger)^{-1} H - G^\dagger (I + GGQG^\dagger)^{-1} G.
\]

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

\[
H^\dagger (I + HQH^\dagger)^{-1} H - G^\dagger (I + GGQG^\dagger)^{-1} G - \lambda I + Z = 0 \quad \text{(5a)}
\]

\[
\text{tr}(Q) \leq P_T; \lambda \geq 0; \lambda (\text{tr}(Q) - P_T) = 0 \quad \text{(5b)}
\]

\[
Z \succeq 0; Q \succeq 0; QZ = 0. \quad \text{(5c)}
\]

Throughout the proof we use the following equality which is a special case of the matrix inversion lemma \([15\text{ Eqn. (2.8.20)]}\)

\[
Y (I + XY)^{-1} = (I + YX)^{-1} Y.
\]

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

\[
(I + H^\dagger HQ)^{-1} H^\dagger H - G^\dagger G(I + QGQ^\dagger)^{-1} = \lambda I - Z. \quad (7)
\]

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

\[
(I + H^\dagger HQ)^{-1} H^\dagger H - G^\dagger G(I + QGQ^\dagger)^{-1} = -Z \quad (8)
\]

which is equivalent to

\[
H^\dagger H - G^\dagger G = -(I + H^\dagger HQ)Z(I + QGQ^\dagger). \quad (9)
\]

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

Next, we show that there is a unique solution \( Q \) to the KKT conditions. Suppose to the contrary that \( (Q_1, Z_1, \lambda_1) \) and \( (Q_2, Z_2, \lambda_2) \) are two different KKT points of (2). Let us define

\[
\Phi_i = H^\dagger (I + HQ_i H^\dagger)^{-1} H; \quad \Psi_i = G^\dagger (I + GQ_i Q^\dagger)^{-1} G \quad (10)
\]

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

\[
Z_1 + \Phi_1 - \Psi_1 = \lambda_1 I \quad \text{(11a)}
\]

\[
Z_2 + \Phi_2 - \Psi_2 = \lambda_2 I \quad \text{(11b)}
\]

Since, \( \text{tr}(Q_1) = \text{tr}(Q_2) = P_T \), it is easy to check that

\[
\text{tr}((\Phi_1 - \Psi_1 + Z_1)(Q_1 - Q_2)) = \lambda_1 \text{tr}(Q_1 - Q_2) = 0 \quad \text{(12)}
\]

\[
\text{tr}((\Phi_2 - \Psi_2 + Z_2)(Q_1 - Q_2)) = \lambda_2 \text{tr}(Q_1 - Q_2) = 0. \quad \text{(13)}
\]

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

\[
\text{tr}((\Phi_1 - \Phi_2 - (\Psi_1 - \Psi_2))(Q_1 - Q_2)) - \text{tr}(Z_1 Z_2 + Z_2 Z_1) = 0. \quad \text{(14)}
\]

Our purpose in the sequel is to show that (14) is impossible if \( Q_1 \neq Q_2 \). To this end we first express \( \Phi_1 - \Phi_2 \) as in (15) shown at the top of the following page. Note that in (15) we have used the fact that \( X^{-1} - Y^{-1} = X^{-1}(Y - X)Y^{-1} \) for invertible matrices \( X \) and \( Y \), and that (15) is due to (11). Similarly we can write

\[
\Psi_1 - \Psi_2 = \Phi_1 (Q_2 - Q_1) \Psi_2. \quad (16)
\]

Combining (16) with (15c) produces (17) shown at the top of the following page. Substituting (17) into (14) we can rewrite the first term in the left-hand side of (14) as in (18) shown at the top of the following 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 \( Q_1 \), and note that \( Q_1 Z_1 = 0 \), which yields

\[
\lambda_1 Q_1 + \Psi_1 Q_1 = \Phi_1 Q_1 \quad \text{(19)}
\]

and thus we have

\[
\text{tr}(\lambda_1 Q_1 + \Psi_1 Q_1) Z_2 Q_1 - \text{tr}(Z_2 Q_1) = \text{tr}(\Phi_1 Q_1 Z_2 Q_1)
\]

\[
- \text{tr}(Z_2 Q_1) = \text{tr}(\Phi_1 Q_1 - I) Z_2 Q_1 = \text{tr}((I + \Gamma)^{-1} \Gamma - I) Z_2 Q_1
\]
\[ \Phi_1 - \Phi_2 = H(I + HQ_2H^\dagger)^{-1}H - H(I + HQ_1H^\dagger)^{-1}H = H((I + HQ_1H^\dagger)^{-1} - (I + HQ_2H^\dagger)^{-1})H \]  
(15a)

\[ = (\Psi_1 + \lambda_1I - Z_1)(Q_2 - Q_1)(\Psi_2 + \lambda_2I - Z_2) \]  
(15b)

\[ \Theta = (1 - \Theta)(Q_2 - Q_1)(\Psi_2 + \lambda_2I - Z_2) \]  
(15c)

\[ (\Phi_1 - \Phi_2) - (\Psi_1 - \Psi_2) = \Psi_1(Q_2 - Q_1)(\lambda_2I - Z_2) + (\lambda_1I - Z_1)(Q_2 - Q_1)\Psi_2 + (\lambda_1I - Z_1)(Q_2 - Q_1)(\lambda_2I - Z_2). \]  
(17)

\[\text{tr}((\Phi_1 - \Phi_2 - (\Psi_1 - \Psi_2))(Q_1 - Q_2)) = \text{tr}((\lambda_1Q_1 + \Psi_1Q_1)Z_2Q_1) + \text{tr}((\lambda_2Q_2 + \Psi_2Q_2)Z_2Q_2) \]

\[- \lambda_1\text{tr}((Q_2 - Q_1)\Psi_2(Q_2 - Q_1)) - \lambda_2\text{tr}((Q_2 - Q_1)\Psi_1(Q_2 - Q_1)) - \lambda_1\lambda_2\text{tr}((Q_2 - Q_1)(Q_2 - Q_1)) \]  
(18)

\[\text{tr}((\lambda_1Q_1 + \Psi_1Q_1)Z_2Q_1) + \text{tr}((\lambda_2Q_2 + \Psi_2Q_2)Z_2Q_2) - \text{tr}(Z_2Q_1 + Z_1Q_2) \]

\[- \lambda_1\text{tr}((Q_2 - Q_1)\Psi_2(Q_2 - Q_1)) - \lambda_2\text{tr}((Q_2 - Q_1)\Psi_1(Q_2 - Q_1)) - \lambda_1\lambda_2\text{tr}((Q_2 - Q_1)(Q_2 - Q_1)) \leq 0 \]  
(23)

where \( \Gamma = H^\dagger HQ_1 \). It is easy to see that \((I + \Gamma)^{-1} \Gamma \preceq I \) and thus \((I + \Gamma)^{-1} (I - \Gamma)Z_2 Q_1 \preceq 0 \), which leads to

\[\text{tr}((\lambda_1Q_1 + \Psi_1Q_1)Z_2Q_1) - \text{tr}(Z_2Q_1) \leq 0. \]  
(20)

Similarly, we have

\[\text{tr}((\lambda_2Q_2 + \Psi_2Q_2)Z_2Q_2) - \text{tr}(Z_2Q_1) \leq 0. \]  
(21)

Adding (20) and (21) gives

\[\text{tr}((\lambda_1Q_1 + \Psi_1Q_1)Z_2Q_1) + \text{tr}((\lambda_2Q_2 + \Psi_2Q_2)Z_2Q_2) - \text{tr}(Z_2Q_1 + Z_1Q_2) \leq 0. \]  
(22)

which results in (23) shown at the top of this page. It is not difficult to check that (22) 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 \(\text{tr}(Z_2Q_1 + Z_1Q_2)\) and using (23) produces

\[\text{tr}((\Phi_1 - \Phi_2 - (\Psi_1 - \Psi_2))(Q_1 - Q_2)) - \text{tr}(Z_2Q_1 + Z_1Q_2) \leq 0. \]  
(24)

We remark that if \(Q_1 \neq Q_2\), then \(\lambda_1\lambda_2 \text{tr}((Q_2 - Q_1)(Q_2 - Q_1)) > 0\). Consequently, the inequality in (23) is strict, and so is (24), which contradicts (14) and completes the proof. \(\square\)

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).

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 \(Y_k\) be the current operating point. Then we take a projected gradient step to obtain the current iterate \(Q_k\) (cf. Line 3). Note that the notation \(\Pi_Q(X)\) in Line 3 denotes the projection of a given point \(X\) onto the feasible set \(Q\). i.e., \(\Pi_Q(X) = \text{argmin}||U - X|| \mid U \in Q\). In contrast to [16] where a constant stepsize is used, we implement a backtracking line search as done in Lines 2 [7].

| Algorithm 1: Iterative algorithm for solving (2) |
|-----------------------------------------------|
| **Input:** \(Y_1 = Q_0 \in Q\), \(\beta_0 = \lambda_0 > 0\), \(\alpha = 1\), \(\xi \in (0,1)\), \(\gamma_u > 1\), \(\epsilon > 0\), \(k \leftarrow 1\) |
| repeat |
| \(Q_k = \Pi_Q(Y_k + \frac{1}{\beta_{k-1}} \nabla C_s(Y_k))\) |
| if \(C_s(Q_k) < \mu_{\beta_{k-1}}(Y_k; Q_k)\) then |
| \(\beta_{k-1} = \gamma_u \beta_{k-1}\) |
| until \(C_s(Q_k) \geq \mu_{\beta_{k-1}}(Y_k; Q_k)\) |
| Update \(\alpha = \min \left\{ \frac{\alpha}{\xi}, 1 \right\} \) and \(Y_{k+1} = Z_k\) |
| end |
| \(k \leftarrow k + 1\) |
| until \(|C_s(Q_k)| - |C_s(Q_{k-1})| \leq \epsilon\) |
| **Output:** \(Q_k\) |

| Algorithm 1: Iterative algorithm for solving (2) |
|-----------------------------------------------|
| **Input:** \(Y_1 = Q_0 \in Q\), \(\beta_0 = \lambda_0 > 0\), \(\alpha = 1\), \(\xi \in (0,1)\), \(\gamma_u > 1\), \(\epsilon > 0\), \(k \leftarrow 1\) |
| repeat |
| \(Q_k = \Pi_Q(Y_k + \frac{1}{\beta_{k-1}} \nabla C_s(Y_k))\) |
| if \(C_s(Q_k) < \mu_{\beta_{k-1}}(Y_k; Q_k)\) then |
| \(\beta_{k-1} = \gamma_u \beta_{k-1}\) |
| until \(C_s(Q_k) \geq \mu_{\beta_{k-1}}(Y_k; Q_k)\) |
| Update \(\alpha = \min \left\{ \frac{\alpha}{\xi}, 1 \right\} \) and \(Y_{k+1} = Z_k\) |
| end |
| \(k \leftarrow k + 1\) |
| until \(|C_s(Q_k)| - |C_s(Q_{k-1})| \leq \epsilon\) |
| **Output:** \(Q_k\) |

To find a proper step size, which is adopted from [17]. For this purpose we define a quadratic model of \(C_s(Q)\) as

\[\mu_{\beta}(Q, Q) = C_s(Q) + \text{tr}(\nabla C_s(Q)(Q - Q)) - \frac{\beta}{2} ||Q - Q||^2.\]  
(25)

Recall that if \(\beta \geq L\), where \(L > 0\) is a Lipschitz constant of \(\nabla C_s(Q)\) on \(Q\), then the inequality \(C_s(Q) \geq \mu_{\beta}(Q, Q)\) holds [17]. In the Appendix we show that \(L = c_s^2\max\{H^\dagger H, G^\dagger G\}\) is a Lipschitz constant for of \(\nabla C_s(Q)\). To find a proper \(\beta\) in each iteration, we start from the value of \(\beta\) in the previous iteration and increase it by \(\gamma_u > 1\) until \(\mu_{\beta}(Y_k; Q_k)\) becomes a lower bound of \(C_s(Q_k)\). In this way, the projected gradient step always produces an improved iterate. Next, for acceleration, we compute the extrapolation point \(Z_k = Q_k + \alpha(Q_k - Q_{k-1})\), where \(\alpha\) is called the momentum parameter (cf. Line 3). For convex optimization, the momentum parameter is fixed. However, since the objective in (2) is non-convex, the extrapolation can be bad and thus \(\alpha\) needs to be adapted in accordance with
Fig. 1. Convergence results of Algorithm 1. $N_t = N_r = N_c = 4$ at $P_t = 15$ dB.

Algorithm 1 is very simple to implement because $\nabla C_s(Q)$ is given in closed-form in (4) and the projection of a given point $Q$ onto $Q$. $\Pi_Q(X)$, admits a water-filling like algorithm as

$$
\Pi_Q(X) = U \text{diag}([x - c]^+) U
$$

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

$$
\sum_{i=1}^{N_t} [x_i - c]^+ = P_t.
$$
Using (4) and the norm inequality, it is easy to see that if the following inequality holds
\[
\| H F_{\text{I}} \| \leq \sigma_{\max}(H) \leq 1
\]
Figs. [(a)] and [(b)] respectively. The channels $H$ and $G$ are generated as $CN(0, I)$. 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, Alg. 3] and [13, Alg. 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. (a), Algorithm 1 achieves monotonic convergence as proved in (31). Also, the residual error is reduced quickly to zero as the iteration process continues for both degraded and non-degraded cases. 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 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, Algorithm 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 $\nabla C_S(Q)$ on $Q$ if the following inequality holds
\[
\| \nabla C_S(X) - \nabla C_S(Y) \| \leq L \| X - Y \|, \forall X, Y \in Q. \tag{32}
\]
Using (4) and the norm inequality, it is easy to see that
\[
\| \nabla C_S(X) - \nabla C_S(Y) \| \leq
\| H^T (I + HXH^T)^{-1} H (Y - X) H^T (I + HYH^T)^{-1} H \| +
\| G^T (I + GXG^T)^{-1} G (Y - X) G^T (I + GYG^T)^{-1} G \|. \tag{33}
\]
We now recall the following well known inequality:
\[
\| AB \| \leq \lambda_{\max}(A) \| B \|	ag{34}
\]
where $\lambda_{\max}(\cdot)$ denotes the maximum singular value of the matrix in the argument. Applying the above inequality and by noting that $\lambda_{\max}((I + HXH^T)^{-1}) \leq 1$ we have
\[
\| \nabla C_S(X) - \nabla C_S(Y) \| \leq \sigma_{\max}(H^T H) + \sigma_{\max}(G^T G) \| X - Y \|
\]
which means that $L = \sigma_{\max}(H^T H) + \sigma_{\max}(G^T G)$ is a Lipschitz constant of $\nabla C_S(\cdot)$.

REFERENCES

[1] A. D. Wyner, “The wire-tap channel,” Bell System Technical Journal, 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 41st Annual Conference on Information Sciences and Systems 2007, Mar. 2007, pp. 905–910.
[4] A. Khisti and G. W. Wornell, “Secure transmission with multiple antennas I: The MISO-M wiretap channel,” IEEE Trans. Inf. Theory, vol. 56, no. 7, pp. 3088–3104, Jun. 2010.
[5] ——, “Secure transmission with multiple antennas part II: The MIMO-M wiretap channel,” IEEE Inf. Theory, vol. 56, no. 11, pp. 5515–5532, Oct. 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. 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, Mar. 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 2014, 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, Dec. 2016, pp. 7207–7215.
[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 2020, Jun. 2020.
[13] A. Mukherjee, B. Ottersten, and L.-N. Tran, “Efficient numerical methods for secrecy capacity of Gaussian MIMO wiretap channel,” in Proc. IEEE VTC 2021, May 2021. [Online]. Available: https://arxiv.org/abs/2102.10396
[14] ——, “On the MIMO secrecy capacity of MIMO wiretap channels: Convex reformulation and efficient numerical methods,” in Proc. IEEE Trans. Commun., 2020. [Online]. Available: https://arxiv.org/abs/2012.05667
[15] D. S. Bernstein, Matrix Mathematics: Theory, Facts, and Formulas (Second Edition). Princeton University 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. the 34th International Conference on Machine Learning, vol. 70. PMLR, Aug. 2017, pp. 2111–2119. [Online]. Available: http://proceedings.mlr.press/v70/lin17b.html
[17] Y. Nesterov, “Gradient methods for minimizing composite functions,” Math. Program., vol. 140, pp. 125–161, 2013.
[18] S. Ye and R. 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