Secret-Key Agreement with Public Discussion over Multi-Antenna Transmitters with Amplitude Constraints

Zouheir Rezki  
ECE Department, University of Idaho  
Moscow, ID, USA  
zrezki@uidaho.edu

Mohamed-Slim Alouini  
EE Program, CEMSE Division, KAUST  
Thuwal, Mekkah Province, Saudi Arabia  
slim.alouini@kaust.edu.sa

Abstract—We consider secret-key agreement with public discussion over a multiple-input single output (MISO) Gaussian channel with an amplitude constraint. We prove that the capacity is achieved by a discrete input, i.e., an input whose support is sparse. The proof follows from the concavity of the conditional mutual information in terms of the input distribution and hence the Karush-Kuhn-Tucker (KKT) condition provides a necessary and sufficient condition for optimality. Then, a contradiction argument that rules out the non-sparsity of any optimal input’s support is utilized. The latter approach is essential to apply the identity theorem in a multidimensional setting as $\mathbb{R}^n$ is not an open subset of $\mathbb{C}^n$.

Index Terms—Secret-key agreement, information-theoretic security, Karush-Kuhn-Tucker (KKT) conditions, discrete input.

I. INTRODUCTION

Information theoretic secret-key agreement provides provably secure mechanisms for generating secret-keys between two or more legitimate terminals. In such protocols, the legitimate terminals need to have access to a source of correlated randomness e.g., communication channels or correlated sources [1], [2]. Furthermore a discussion channel of unlimited capacity is also available for communication, but is public to the wiretapper. The legitimate terminals distill a common secret-key that satisfies an equivocation constraint with respect to the eavesdropper.

This paper addresses capacity limits of secret-key agreement when the sender has multiple antennas, i.e., a multiple-input single-output (MISO) channel. We further assume that the transmitter inputs are subject to an amplitude constraint. Such a constraint appears in several important setting, e.g., [3].

Note that in our proposed channel model, the outputs at the legitimate receiver and the eavesdropper are conditionally independent given the channel input. A class of discrete memoryless channel models with this property was studied in [1], [2] and a single-letter capacity expression was characterized. While their result can be extended using standard techniques to the (continuous-valued) channels studied in this work, finding the optimal input distribution is difficult in general. When the transmitter has a single antenna, the problem has been solved in [4] and the optimal input has been shown to be discrete with a finite support. A key observation for solving the single antenna case in [4] was the ability to convert the channel into an equivalent scalar Gaussian wiretap channel with an amplitude constraint and apply the results in [5]. Unfortunately, such an approach cannot be applied for the MISO channel, let alone that the amplitude-constrained MISO wiretap secrecy capacity is still not known.

In the present work, we show that for the MISO case, the capacity achieving distribution is discrete. We adopt a direct approach similar to the ones in e.g., [3], [5]–[11], to cite only a few. However, differently from previous work, our objective function includes a conditional mutual information that requires several non-straightforward modifications of the previous proofs. In [12], the secret-key agreement capacity of the non-coherent channel model has been solved and the optimal input has been shown to be discrete. A key observation made in [12] is that information is only conveyed through the norm of the input which, once again, allowed converting the MISO channel into a scalar channel thus simplifying the proof enormously. Adopting a similar approach to our setting is not straightforward neither. As a matter of fact and differently from previous work, we claim nothing about the number of mass points (finite or not) of an optimal input distribution. This is essentially due to the fact that a bounded subset of real numbers is sparse\textsuperscript{1} if and only if it is finite. A property that does not hold true for a high-dimensional space like in the MISO case [13]. However, owing to a necessary and sufficient condition for optimality, the number of mass points may be found numerically.

II. THE CHANNEL MODEL

Consider a MISO Gaussian Channel consisting of a transmitter with $n_T$ antennas, a legitimate receiver and an eavesdropper, with a single antenna, each. The outputs at both the legitimate destination and the eavesdropper are given by:

\[
\begin{aligned}
  y(i) &= h^\top x(i) + n_i(i), \\
  z(i) &= g^\top x(i) + n_{z}(i),
\end{aligned}
\]

for $i = 1, \ldots, n$, where $x(i) \in \mathbb{R}^{N_t}$ is the transmitted signal, and $h \in \mathbb{R}^{N_t}$, $g \in \mathbb{R}^{N_r}$ represent the main channel and the eavesdropper channel gains, respectively; and $n_i(i) \in \mathbb{R}$, $n_z(i) \in \mathbb{R}$ are zero-mean white Gaussian noises with variances $\sigma^2_p$ and $\sigma^2_n$.

\textsuperscript{1}A formal definition of a sparse set is given in Section IV.
\( \sigma_{E_i}^2 \), respectively. The source is subject to a peak-amplitude constraint
\[ |x(i,j)| \leq A, \]  
for all \( j = 1, \ldots, N_i \), where \( x(i,j) \) designates the \( j \)-th component of \( x(i) \). For convenience, the amplitude constraint (2) is also referred to as \( x(i) \leq A \).

III. Secret-Key Capacity

In [1, Theorem 2], a single-letter expression of key-capacity has been established and is given by:
\[ C = \sup_{F_X(\cdot) \in \mathcal{T}} I(X;Y), \]  
where \( \mathcal{T} \) is the set of all possible input distribution functions \( F_X(\cdot) \) that satisfy the peak amplitude constraints (2), i.e.,
\[ \mathcal{T} = \left\{ F_X(\cdot) : F_X(A) \triangleq \int_A dF_X(x) = 1 \right\}, \]  
where \( A = [-A, A]^N \). Note that \( Y \) and \( Z \) are conditionally independent given \( X \) and their distributions are \( N(h(x), \sigma_E^2) \) and \( N(g(x), \sigma_E^2) \), respectively. Since the latter marginals fully describe the channel of interest, it is then clear that the secret-key capacity depends on the input via \( X \) rather than via \( ||X|| \) as in [12].

Before stating our main result, we simplify the target in (3) as follows:
\[ I(X;Y) = h(Y) - h(Y|X), \]  
\[ = \int_A \left[ i(x;F_X)dF_X(x) \right], \]  
where \( i(x;F_X) \) is the conditional mutual information density defined by:
\[ i(x;F_X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{YZ}(y,z|x)p_{X|Y}(x|y) \log \left( \frac{p_{Y|X}(y|x)}{p_{Z|Y}(z|y;F_X)} \right) dy \, dz. \]  
To obtain (5), we have used the fact that \( Y \rightarrow X \rightarrow Z \) forms a Markov chain. In (7), \( p_{Y|X}(y|x) \) represents the conditional probability distribution of \( Y \) given \( Z \) induced by the input distribution \( F_X(\cdot) \). We now state our main result.

Theorem 1: The supremum in (3) is achievable by at least one input distribution \( F_X^*(\cdot) \). Furthermore, any solution \( F_X^*(\cdot) \) of (3) is discrete.

Proof: For convenience, the proof is presented in Section IV. 

With the help of Theorem 1, the concavity of \( I(X;Y) \) in \( F_X(\cdot) \) and the necessary and sufficient KKT condition (see Corollary 2 in Section IV), the secret-key capacity can be efficiently computed numerically via a maximization of the right hand side (RHS) of (52) over all inputs of the form
\[ F_X(x) = \sum_{i=1}^{N} p_i u(x-x_i), \]  
where \( u(\cdot) \) is the multidimensional unit step, \( x_i \leq A \) is the \( i \)-th mass point location and \( p_i \) is its probability of occurrence, \( i = 1, \ldots, N \). Another useful property of a capacity-achieving input is its symmetry formalized as follows.

Corollary 1: For any amplitude constraint \( A > 0 \), the secret-key capacity is achieved by a symmetric input of the form:
\[ F_X^*(x) = \sum_{i=1}^{N} p_i u(x \pm x_i). \]  
That is, if \( x_i \in S_{F_X^*} \), then \( -x_i \in S_{F_X^*} \) and they both have the same probability of occurrence \( p_i \), where \( S_{F_X^*} \) designates the support of \( F_X^*(\cdot) \).

Proof: For convenience, the proof is presented in the appendix.

IV. Proof of the Main Result

First, for convenience, let us designate by \( I(F_X) \) the conditional mutual information \( I(X;Y,Z) \) induced by the input distribution \( F_X(\cdot) \). The proof is then streamlined into a few lemmas which we state below.

Lemma 1: \( \mathcal{T} \) is Convex and Sequentially Compact.

Proof: Let \( (F_{X_k},F_{X_k}) \in \mathcal{T}^2 \) and \( t \in [0,1] \). One can easily verify that:
\[ \int_{\mathcal{A}} d \left[ (1-t)F_{X_k} + t F_{X_k} \right] (x) = 1, \]  
which confirms that \( \mathcal{T} \) is a convex set. To prove that \( \mathcal{T} \) is sequentially compact, it suffices to show that it is tight and closed. Since \( \mathcal{A} \) is the cartesian product of compact (closed and bounded) sets, then \( \mathcal{A} \) is also compact, and since \( \forall \geq 0 \), \( F_X(\mathcal{A}) = \int_{\mathcal{A}} dF_X(x) = 1 > 1 - \epsilon \), for any \( F_X(\cdot) \in \mathcal{T} \), then \( \mathcal{T} \) is tight [14, Chap. 3].

To prove that \( \mathcal{T} \) is closed, let \( \{F_{X_k}(\cdot)\}_{k=1}^{\infty} \) be a convergent sequence in \( \mathcal{T} \) and let \( F_{X_k}(\cdot) \) be its limit. Then since \( \mathcal{A} \) is closed, we have \( 1 = \lim_{k \rightarrow \infty} F_{X_k}(\mathcal{A}) = F_{X_k}(\mathcal{A}) \), implying that \( F_{X_k}(\mathcal{A}) = 1 \), i.e., \( F_X(\mathcal{A}) \in \mathcal{T} \). Hence, \( \mathcal{T} \) is also closed. Therefore, being a tight and closed set, \( \mathcal{T} \) is sequentially compact [14, Theorem 3.19].

Lemma 2: \( I(F_X) \) is continuous over \( \mathcal{T} \).

Proof: First, we expand \( I(X;Y,Z) \) as \( I(X;Y,Z) = I(X;Y,Z) - I(X;Z) \). Then, it suffices to show that \( I(X;Y,Z) \) is continuous on every \( F_X \in \mathcal{T} \) as the continuity of \( I(X;Z) \) follows along similar lines. Then, noting that \( I(X;Y,Z) = h(Y,Z;F_X) - \frac{1}{2} \log \left( 2\pi e \sigma_Y^2 \sigma_Z^2 \right) \), we only need to show that \( h(Y,Z;F_X) \) is continuous over \( \mathcal{T} \). Recall
Hence, for any convergent sequence \( p_{YZ}(y, z; F_X) \) is continuous over \( F \) and so is \( p_{YZ}(y, z; F_X) \). Hence, for any convergent sequence \( \{F_X(\cdot)\}_{i=1}^{\infty} \) in \( F \) whose limit is \( F_X(\cdot) \), we have:
\[
\lim_{n \to \infty} p_{YZ}(y, z; F_X) = p_{YZ}(y, z; F_X).
\]

Now, we have:
\[
\lim_{n \to \infty} h(Y, Z; F_X) = \lim_{n \to \infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{YZ}(y, z; F_X) \log(p_{YZ}(y, z; F_X)) \, dydz
\]
\[
= -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lim_{n \to \infty} [p_{YZ}(y, z; F_X) \log(p_{YZ}(y, z; F_X))] \, dydz
\]
\[
= -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{YZ}(y, z; F_X) \log(p_{YZ}(y, z; F_X)) \, dydz
\]
\[
h(Y, Z; F_X).
\]

where to obtain (15), we used the dominated convergence theorem since for any \( F_X \in F \),
\[
\left| p_{YZ}(y, z; F_X) \log(p_{YZ}(y, z; F_X)) \right| 
\leq \alpha e^{-\beta_1 y^2 - \beta_2 z^2} + \beta e^{-\beta_1 y^2 - \beta_2 z^2},
\]
for some positive constants \( \alpha, \beta, i = 1, \ldots, 4 \), whose integral exists and is finite [3, Lemma 3]; and (16) is due to (12). Therefore, \( h(Y, Z; F_X) \) is continuous over \( F \) and so is \( I(X; Y, Z) \). We conclude that \( I(F_X) \) is continuous over \( F \) and Lemma 2 is thus proved.

a) Proof of the first statement in Theorem 1: From Lemma 1 and Lemma 2, \( I(F_X) \) is continuous in \( F_X(\cdot) \) over \( F \) which itself is a compact set, then by the Extreme Value Theorem, \( I(F_X) \) is bounded above and attains its supremum. That is, the supremum in (3) is actually a maximum which is achievable by at least one input distribution \( F_X(\cdot) \).

Lemma 3: \( I(F_X) \) is concave and weak-differentiable over \( F \). Furthermore, the weak derivative of \( I(F_X) \) at any \( F_X \in F \) is given by:
\[
I_{F_{X_0}}(F_X) = \int_{\Omega} i(x; F_X) \, dF_X(x) - I(F_X).
\]

Proof: The concavity of \( I(F_X) \) in \( F_X(\cdot) \) has been proved in [15, Fact 2]. The weak-differentiability follows along similar lines as [12], with the difference that the integration on the right-hand side (RHS) of (7) is multidimensional.

b) The Karush-Kuhn-Tucker (KKT) Condition: From Lemma 1 and Lemma 3, we have a concave and weak-differentiable function \( I(F_X) \) over \( F \) which is a convex set, then a sufficient and necessary condition for an input distribution \( F_X(\cdot) \) to be optimal is
\[
I_{F_{X_0}}(F_X) \leq 0,
\]
for any \( F_X(\cdot) \in F \). Condition (20), in turn, translates into the following optimality condition.

Corollary 2: An input distribution \( F_X(\cdot) \) is secret-key capacity-achieving if and only if:
\[
i(x; F_X) \leq C,
\]
for any \( x \in \Omega \). Furthermore, if \( x \) is in the support (or a point of increase) of \( F_X \), then (21) is satisfied with equality.

Proof: Assume that \( F_X \) is secret-key capacity-achieving, i.e., \( I(F_X) = C \). Let \( x \in \Omega \) and let \( F_X(\cdot) \) be an input distribution with a unique mass point at \( x \). Since \( F_X(\cdot) \in F \), then applying (20) to \( F_X(\cdot) \) along with (7) give:
\[
i(x; F_X) \leq C.
\]

Assume now that \( x_0 \) is a point of increase of \( F_X \) and that \( i(x_0; F_X) < C \). Then, by definition of point of increase, there exists an open \( O \subset \Omega \) containing \( x_0 \) such that \( \int_O dF_X = \delta > 0 \). In addition:
\[
C = I(F_X) = \int_{\Omega} i(x; F_X) \, dF_X(x) \leq \int_O i(x; F_X) \, dF_X(x) + \int_{\Omega \setminus O} i(x; F_X) \, dF_X(x)
\]
\[
< C\delta + C(1 - \delta) = C,
\]
which is a contradiction. Thus, if \( x_0 \) is a point of increase of \( F_X \), then it must hold that \( i(x_0; F_X) = C \) which completes the proof of the direct part of Corollary 2. The converse is immediate and follows by averaging both sides of (21) with respect to \( F_X(\cdot) \) over \( \Omega \) and verifying (20).

c) Proof that \( S_{F_X} \) is Sparse: Recall that a probability distribution is said to be discrete if its set of points of increase (or support) is sparse. The notion of sparsity can be formally defined as follows [3].

Definition 1: A subset \( F \) of \( \mathbb{R}^N \) is sparse if there exists a nonzero holomorphic function \( f(\cdot) \) defined on a connected open subset \( U \subset \mathbb{C}^N \) containing the closure of \( F \) such that \( f(\cdot) = 0 \) for all \( \omega \in F \).

We proceed by contradiction. Assume that \( S_{F_X} \) is not sparse. Let us first extend the KKT condition in Corollary 2 to \( \mathbb{C}^N \) and define \( a(\omega) \) as
\[
a(\omega) = C - i \left( \omega; F_X \right).
\]
Note that \( a(\cdot) \) is holomorphic on \( \mathbb{C}^N \) since it is holomorphic in each variable \( w(i) \) of \( \omega \) separately, for \( i = 1, \ldots, N \) [15, Theorem B6]. Also, \( a(\cdot) \) is zero on \( S_{F_X} \) due to Corollary 2. Then, we have a holomorphic function defined on \( \mathbb{C}^N \), a connected open set, which contains the closure of \( S_{F_X} \), and such a function is zero over a non-sparse set \( S_{F_X} \). By the identity theorem, we conclude that \( a(\cdot) \) is zero on \( \mathbb{C}^N \), and in particular over \( \mathbb{R}^N \) [17]. Let us examine the implication of
this statement. First, one can verify easily (see Appendix B) that
\[ a(x) = \left[ p_{Y_G, Z_G}(y, z) \ast g(y, z) \right] (h^T x, g^T x), \]  
(26)
where \( Y_G \sim \mathcal{N}(0, \sigma_G^2) \), \( Z_G \sim \mathcal{N}(0, \sigma_Z^2) \), the operator \( \ast \) designates the convolution and \( g(y, z) \doteq \log \left( p_{YZ} \left( y|z; F_X^* \right) \right) + h \left( y|z; F_X^* \right) \). Note that \( g(y, z) \) is continuous over \( R^2 \). Furthermore, \( g(y, z) \) is polynomially-bounded as formalized below.

**Lemma 4:** There exist \( \alpha > 0 \) and \( \beta > 0 \) such that
\[ |g(y, z)| \leq (\alpha + \beta \left( y^2 + z^2 \right)). \]
(27)

**Proof:** For convenience, the proof is presented in Appendix C.

Combining (26), Lemma 4 and [3, Corollary 9], we establish that \( g(y, z) \) is the zero function.

To reach a contradiction, it suffices to compute the limit of \( g(y, z) \) as \( (y, z) \to (\infty, 0) \), for instance. For this purpose, and in regard of (50), we have
\[
\lim_{(y, z) \to (\infty, 0)} \frac{p_{YZ}(y, z; F_X^*)}{p_Z(z; F_X^*)} = \lim_{(y, z) \to (\infty, 0)} \frac{E_{F_X}(p_{YZ}(y, z|X))}{p_Z(z; F_X^*)} \\
\leq \lim_{(y, z) \to (\infty, 0)} e^{(\alpha^* + \beta^* z^2)}E_{F_X}(p_{YZ}(y, z|X)) \\
= e^{\alpha^*}E_{F_X} \left[ \lim_{(y, z) \to (\infty, 0)} p_{YZ}(y, z|X) \right] \\
= 0,
\]
(28)
(29)
(30)
(31)
where (29) follows because \( |p_Z(z; F_X^*)| \leq e^{-\left( \alpha^* + \beta^* z^2 \right)} \), for some positive constant \( \alpha^* \) and \( \beta^* \), and (30) holds true since \( |p_{YZ}(y, z|X)| \leq \frac{1}{2\pi \sqrt{\sigma_Y^2 \sigma_X^2}} \) and \( E_{F_X} \left[ \frac{1}{2\pi \sqrt{\sigma_Y^2 \sigma_X^2}} \right] = \frac{1}{2\pi \sqrt{\sigma_Y^2 \sigma_X^2}} < \infty \), then by the dominated convergence theorem, it is sound to insert the limit inside the expectation. Thus,
\[
\lim_{(y, z) \to (\infty, 0)} g(y, z) = \lim_{(y, z) \to (\infty, 0)} \left[ \log \left( p_{YZ} \left( y|z; F_X^* \right) \right) + h \left( y|z; F_X^* \right) \right] \\
= -\infty,
\]
(32)
(33)
which contradicts the fact that \( g(y, z) \) is the zero function. Hence, \( S_{F_X} \) must be sparse, i.e., \( F_X^*(\cdot) \) is discrete as claimed.

**V. Numerical Results**

We present selected numerical results for secret-key agreement Gaussian channels with \( N_t = 1 \) and \( N_r = 2 \), respectively. In all cases, and for each value of the amplitude constraint \( A \), we maximize the RHS of (52) numerically using inputs of the form of (9). The outputs of this algorithm are then checked against the necessary and sufficient KKT condition in Corollary 2. If the outputs fulfill the KKT condition (21), then we assert that an optimal input is found along with the secret-key capacity for that particular \( A \). Otherwise, the number of mass points \( N \) is increased, the maximization algorithm is run again and its outputs verified against the KKT condition anew. We repeat this process until an optimal input is found together with the secret-key capacity.

For the SISO case, it can be seen in Fig. 1 that as \( A \) increases, the number of mass points of an optimal input increases gradually from 2 to 3 and then 4. An evidence of optimality of an input is the KKT condition as shown in Fig. 2.

For the MISO case with \( N_t = 2 \) transmit antennas, it can be seen in Fig. 3 that as \( A \) increases, the number of mass points increases discontinuously from 2 to 4 and then 6. This suggests that there is apparently no mass point at 0 for these particular channel gains. While verifying the KKT condition in the SISO case is relatively feasible, it becomes readily tedious in the MISO case, especially as the number of mass points increases. In Fig. 4, the KKT condition is illustrated for \( A^2 = 0.1 \), where two-mass-point input achieves the capacity.

**VI. Conclusion**

We addressed the secret-key agreement with public discussion over a multi-antenna transmitter subject to amplitude constraints. We showed that the capacity is achieved by a discrete input. While we report no results regarding the finiteness of the optimal input’s support and the unicity of
for instance, and thus there is no loss of optimality by making

\[ F_X^*(x) = \sum_{i \in \mathcal{B}} p_i^{(B)} u(x + x_i^{(B)}) + \sum_{D} p_i^{(D)} u(x - x_i^{(D)}), \]  

(36)

where \(\mathcal{B} = \{ x_i^{(B)} \in S_{F_X} \mid -x_i^{(B)} \in S_{F_X} \}\) and \(D = S_{F_X} \setminus \mathcal{B}\). Let us now consider a new input defined by

\[ F_X^{**}(x) = \sum_{i \in \mathcal{B}} p_i^{(B)} u(x + x_i^{(B)}) + \sum_{D} p_i^{(D)} u(x + x_i^{(D)}). \]  

(37)

Clearly, \(S_{F_X}^{**} = (\mathcal{D} \cup \mathcal{B})\) and it can be easily verified (via a simple change of variables) that

\[ i(x_i^{(B)}; F_X^{**}) = i(-x_i^{(B)}; F_X) \quad \forall x_i^{(B)} \in \mathcal{B} \]  

(38)

\[ i(-x_i^{(D)}; F_X^{**}) = i(x_i^{(D)}; F_X) \quad \forall x_i^{(D)} \in \mathcal{D} \]  

(39)

Hence, \(I(F_X^{**}) = I(F_X) = C\). Finally, using Jensen’s inequality, we obtain:

\[ C = \frac{1}{2} I(F_X) + \frac{1}{2} I(F_X) \leq \frac{1}{2} I(F_X^{**}) + \frac{1}{2} I(F_X^{**}). \]  

(40)

(41)

The argument on the RHS of (41) is a legitimate input and coincides with the one given by (9). This completes the proof of Corollary 1.

APPENDIX B

VERIFICATION OF (26)

Departing from the definition (25), we write:

\[ a(x) = -i(x; F_X) + C \]  

(42)

\[ = -i(x; F_X) + h(Y; F_X) - \frac{1}{2} \log(2\pi e\sigma_D^2) \]  

(43)

\[ = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{Y|x}(y|x)p_{Z|x}(z|x) \log \left( \frac{p_{Y|x}(y|x)}{p_{YZ}(y|z; F_X)} \right) dy dz \]  

(44)

\[ + h(Y; F_X) - \frac{1}{2} \log(2\pi e\sigma_D^2) \]  

\[ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{Y|x}(y|x)p_{Z|x}(z|x) \log \left( \frac{p_{Y|x}(y|x)}{p_{YZ}(y|z; F_X)} \right) dy dz \]  

(45)

\[ + h(Y; F_X) \]  

\[ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{Y|x}(y|x)p_{Z|x}(z|x) \log \left( \frac{p_{Y|x}(y|x)}{p_{YZ}(y|z; F_X)} \right) dy dz \]  

(46)

\[ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{2\pi \sqrt{\sigma^2 + \sigma^2 E}} \]  

\[ \left[ -\frac{y^2 + x^2 - 2x'y}{\sigma^2} \right] \]  

\[ \log \left( \frac{p_{Y|x}(y|x)}{p_{YZ}(y|z; F_X)} \right) \]  

(47)

\[ \log \left( \frac{p_{Y|x}(y|x)}{p_{YZ}(y|z; F_X)} \right) \]  

(48)

\[ = \log \left( \frac{p_{Y|x}(y|x)}{p_{YZ}(y|z; F_X)} \right) + h(Y; F_X) \]  

we claimed; where (48) follows because \(p_{YZ}(y|z; F_X)\) is a function of \((y, z)\) and \(h(Y; F_X)\) is a constant independent of \((y, z)\).
To prove the lemma, it is equivalent to prove that there exist \( \alpha > 0 \) and \( \beta > 0 \) such that

\[
e^{-a'-\beta(y^2+z^2)} \leq p_{Y|Z}(y|Z; F_x^*) \leq e^{a''+\beta(y^2+z^2)},
\]

where \( a' = \alpha + h(Y|Z; F_x^*) \) and \( a'' = \alpha - h(Y|Z; F_x^*) \). For this purpose, let us first write

\[
p_{Y|Z}(y|Z; F_x^*) = \frac{p_{YZ}(y,z; F_x^*)}{p_{Z}(z; F_x^*)}.
\]

Then, by [3, Lemma 3], we have:

\[
e^{-a_1-\beta(y^2+z^2)} \leq p_{YZ}(y,z; F_x^*) \leq e^{a_1+\beta(y^2+z^2)}
\]

\[
e^{-a_2-\beta(y^2+z^2)} \leq p_{Z}(z; F_x^*) \leq e^{a_2+\beta(y^2+z^2)},
\]

for some positive constants \( a_i \), \( i = 1, 2 \) and \( \beta_i \), \( i = 1,\ldots,4 \). Hence, departing from (51) and (52), we upper-bound

\[
p_{Y|Z}(y|Z; F_x^*) \geq e^{a_1-a_2-\beta(y^2+z^2)} (a_{12}-\beta(y^2+z^2)),
\]

where (54) follows from the fact that \( e^{-ax} \leq e^{a_2} \), for any \( a \geq 0 \) any real number \( x \). Similarly, using once again (51) and (52), we establish the following lower bound:

\[
p_{Y|Z}(y|Z; F_x^*) \geq e^{-a_1-a_2-\beta(y^2+z^2)} (a_{12}+\beta(y^2+z^2)).
\]

Let \( \alpha = \alpha_1 + \alpha_2 + h(Y|Z; F_x^*) \). Then, it can be readily verified that

\[
e^{a_1+a_2} \geq e^{a_1+a_2+\beta(Y^2+F_x^*)} = e^{a_1-a_2(Y|Z; F_x^*)} \Delta = e^{\Delta}
\]

\[
= e^{-a_1-a_2} \Delta = e^{-a_1-a_2} \geq e^{-a_1} e^{-h(Y|Z; F_x^*)} \Delta = e^{-a}.\n\]

Similarly, let \( \beta = \max \{ \beta_1 + \beta_2, \beta + \beta_2 \} \). Then, it can be immediately verified that

\[
e^{\beta_1 y^2+\beta_2 y^2+z^2} \leq e^{\beta(y^2+z^2)}
\]

\[
e^{-\beta_1(y^2+z^2)} \geq e^{-\beta(y^2+z^2)}.
\]

Combining (54) and (55) along with (56)–(59), we establish (49).