The Convex–Concave Characteristics of Gaussian Channel Capacity Functions

Han Wu Chen and Kenjiro Yanagi, *Member, IEEE*

***Abstract*—In this correspondence, we give several inherent properties of the capacity function of a Gaussian channel with and without feedback by using some operator inequalities and matrix analysis. We give a new proof method which is different from the method appearing in: K. Yanagi and H. W. Chen, "Operator inequality and its application to information theory," *Taiwanese J. Math.*, vol. 4, no. 3, pp. 407–416, Sep. 2000. We obtain the following results: $C_{n,Z}(P)$ and $C_{n,FB,Z}(P)$ are both concave functions of P, $C_{n,Z}(P)$ is a convex function of the noise covariance matrix and $C_{n,FB,Z}(P)$ is a convex-like function of the noise covariance matrix. This new proof method is very elementary and the results shall help study the capacity of Gaussian channel. Finally, we state a conjecture concerning the convexity of $C_{n,FB,\cdot}(P)$.**

***Index Terms*—Capacity, feedback, Gaussian channel, Shannon theory.**

I. Introduction

The following model for the discrete time Gaussian channel with feedback is considered:

$$Y_n = S_n + Z_n, \quad n = 1, 2, \ldots$$

where $Z = \{Z_n; n = 1, 2, \ldots\}$ is a nondegenerate, zero mean Gaussian process representing the noise and $S = \{S_n; n = 1, 2, \ldots\}$ and $Y = \{Y_n; n = 1, 2, \ldots\}$ are stochastic processes representing input signals and output signals, respectively. The channel is used with noiseless feedback, so S_n is a function of a message W to be transmitted and the output signals $Y_1, \ldots, Y_{n-1}$. For code rate R, the message $W \in \{1, 2, \ldots, 2^{nR}\}$ is uniformly distributed and independent of Z^n. The codewords are denoted as $x^n(W, Y^{n-1})$, and the channel output is given by $Y^n = x^n(W, Y^{n-1}) + Z^n$. If $g_n : \mathbb{R}^n \to \{1, \ldots, 2^{nR}\}$ denotes the decoding function, then the probability of decoding error can be written as $Pe^{(n)} = Pr\{g_n(Y^n) \neq W\}$. The signal is subject to an expected power constraint

$$\frac{1}{n}\sum_{i=1}^{n} E\left[S_i^2\right] \leq P$$

and the feedback is causal, i.e., S_i depends on $Z_1, \ldots, Z_{i-1}$ for $i = 1, 2, \ldots, n$. Similarly, when there is no feedback, S_i is independent of Z^n. We denote by $R_S^{(n)}, R_Z^{(n)}, R_{S+Z}^{(n)}$ the covariance matrices of S, Z, $S+Z$, respectively, and we denote the determinant of a matrix A by $|A|$. It is well-known that a finite block length capacity without feedback is given by [7]

$$C_{n,Z}(P) = \max_{\mathrm{Tr}\left[R_S^{(n)}\right] \leq nP} \frac{1}{2n} \log \frac{\left|R_S^{(n)} + R_Z^{(n)}\right|}{\left|R_Z^{(n)}\right|}$$

Manuscript received January 28, 2005; revised February 5, 2006. This work was supported in part by National Natural Science Foundation of China under Grants 90412014 and 60572071.

H. W. Chen is with the Department of Computer Science and Engineering, Southeast University, Nanjing 210096, People's Republic of China (e-mail: hw_chen@seu.edu.cn).

K. Yanagi is with the Department of Applied Science, Faculty of Engineering, Yamaguchi University, Ube 755-8611, Japan (e-mail: yanagi@yamaguchi-u.ac.jp).

Communicated by R. W. Yeung, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2006.872851

and with feedback is given by [7]

$$C_{n,FB,Z}(P) = \max_{Tr\left[R_S^{(n)}\right] \leq nP} \frac{1}{2n} \log \frac{\left|R_{S+Z}^{(n)}\right|}{\left|R_Z^{(n)}\right|}.$$

We can also write $C_{n,FB,Z}(P)$ using the following formula:

$$C_{n,FB,Z}(P) = \max \frac{1}{2n} \log \frac{\left|R_X^{(n)} + R_Z^{(n)}\right|}{\left|R_Z^{(n)}\right|}$$

where $X = S - TY$ and T is strictly lower triangular, and the maximum is taken under the constraint

$$\mathrm{Tr}\left[(I+B)R_X^{(n)}(I+B^t) + BR_Z^{(n)}B^t\right] \leq nP$$

where $R_X^{(n)}$ is symmetric, nonnegative definite, and B is strictly lower triangular.

Proposition 1 (Cover and Pombra [6]): For every $\epsilon > 0$ there exist codes, with block length n and $2^{n(C_{n,FB,Z}(P)-\epsilon)}$ codewords, $n = 1, 2, \ldots$, such that $Pe^{(n)} \to 0$, as $n \to \infty$. Conversely, for every $\epsilon > 0$ and any sequence of codes with $2^{n(C_{n,FB,Z}(P)+\epsilon)}$ codewords and block length n, $Pe^{(n)}$ is bounded away from zero for all n. (The same theorem holds in the special case without feedback upon replacing $C_{n,FB,Z}(P)$ by $C_{n,Z}(P)$.)

When the block length n is fixed, $C_{n,Z}(P)$ is given in the following.

Proposition 2 (Gallager [11], Theorem 7.5.1):

$$C_{n,Z}(P) = \frac{1}{2n}\sum_{i=1}^{k} \log \frac{nP + r_1 + \cdots + r_k}{kr_i}$$

where $0 < r_1 \leq r_2 \leq \cdots \leq r_n$ are eigenvalues of $R_Z^{(n)}$ and $k(\leq n)$ is the largest integer satisfying $nP + r_1 + \cdots + r_k > kr_k$.

II. Concavity of $C_{n,Z}(P)$ and $C_{n,FB,Z}(P)$ Relative to P

Before proving the concavity of $C_{n,Z}(P)$ and $C_{n,FB,Z}(P)$ with respect to P, we first give some known results. We denote the range of A and the kernel of A by $\mathrm{ran}A$ and $\ker A$, respectively.

Proposition 3 (Cover and Pombra [6]): Let A and B be nonnegative definite matrices. For any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$, we have

$$|\alpha A + \beta B| \geq |A|^{\alpha}|B|^{\beta}.$$

Proposition 4 (Douglas [8]): Let $\mathcal{H}$ be a real Hilbert space and let $\mathcal{B}(\mathcal{H})$ be the set of all bounded linear operators on $\mathcal{H}$. And let $A, B \in \mathcal{B}(\mathcal{H})$. Then the following assertions are equivalent:

1) $\mathrm{ran}A \subset \mathrm{ran}B$;
2) there exists $\alpha \geq 0$ such that $AA^* \leq \alpha BB^*$, where A^* denotes the conjugate of A;
3) there exists $C \in \mathcal{B}(\mathcal{H})$ such that $A = BC$.

Furthermore when the above condition 3) holds, C is uniquely determined and the following three conditions are satisfied:

1) $\|C\|^2 = \inf\{\alpha : AA^* \leq \alpha BB^*\}$, where $\|\cdot\|$ denotes the matrix norm;
2) $\ker A = \ker C$;
3) $\overline{\mathrm{ran}C} \subset (\ker B)^{\perp}$, where $\overline{\mathrm{ran}C}$ denotes the closure of $\mathrm{ran}C$, and $(\ker B)^{\perp}$ denotes the orthogonal complement of $\ker B$.

Proposition 5 (Baker [1]): Let $\mathcal{H}_1$ (resp. $\mathcal{H}_2$) be a real and separable Hilbert space with Borel σ-field Γ_1 (resp. Γ_2). Let μ_X (resp. μ_Y) be a probability measure on $(\mathcal{H}_1, \Gamma_1)$ (resp. $(\mathcal{H}_2, \Gamma_2)$) satisfying

$$\int_{\mathcal{H}_1} \|x\|_1^2 d\mu_X(x) < \infty \quad \left(resp. \int_{\mathcal{H}_2} \|y\|_2^2 d\mu_Y(y) < \infty\right).$$

Let R_X and m_X (resp. R_Y and m_Y) denote the covariance operator and mean element of μ_X (resp. μ_Y). Let $(\mathcal{H}_1 \times \mathcal{H}_2, \Gamma_1 \times \Gamma_2)$ be the product measurable space generated by the measurable rectangles. Let μ_{XY}, having $\mathcal{R}$ as covariance and m as mean element, be a joint measure on $(\mathcal{H}_1 \times \mathcal{H}_2, \Gamma_1 \times \Gamma_2)$ with projections μ_X and μ_Y. Then the cross-covariance operator R_{XY} of the μ_{XY} has a decomposition

$$R_{XY} = R_X^{\frac{1}{2}} V R_Y^{\frac{1}{2}}$$

where V is a unique bounded linear operator such that $V: \mathcal{H}_2 \to \mathcal{H}_1$, $\|V\| \leq 1$, $\ker R_Y \subset \ker V$ and $\overline{\mathrm{ran} V} \subset \overline{\mathrm{ran} R_X}$.

Lemma 1: Let R_{S_1} and R_{S_2} be the covariance matrices of S_1 and S_2, respectively. For any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$, the following formulas hold:

1) $\alpha R_{S_1} + \beta R_{S_2} = R_{\alpha S_1 + \beta S_2} + \alpha\beta R_{S_1 - S_2}$;
2) $\alpha R_{S_1} + \beta R_{S_2} \geq R_{\alpha S_1 + \beta S_2}$, where the equality holds if and only if $S_1 = S_2$ (for $0 < \alpha < 1$);
3) $\alpha R_{S_1 + Z} + \beta R_{S_2 + Z} = R_{\alpha S_1 + \beta S_2 + Z} + \alpha\beta R_{S_1 - S_2}$;
4) $R^{\frac{1}{2}}_{\alpha S_1 + \beta S_2} = (\alpha R_{S_1} + \beta R_{S_2})^{\frac{1}{2}} W$, where $\|W\| \leq 1$.

Proof of Lemma 1:

1) It is easy to obtain the following relations by the properties of nonnegative definite matrices:

$$\begin{aligned} &R_{\alpha S_1 + \beta S_2} + \alpha\beta R_{S_1 - S_2} \\ &\quad = \alpha^2 R_{S_1} + \alpha\beta R_{S_1 S_2} + \alpha\beta R_{S_2 S_1} + \beta^2 R_{S_2} \\ &\qquad + \alpha\beta R_{S_1} - \alpha\beta R_{S_1 S_2} - \alpha\beta R_{S_2 S_1} + \alpha\beta R_{S_2} \\ &\quad = \alpha(\alpha + \beta) R_{S_1} + \beta(\alpha + \beta) R_{S_2} \\ &\quad = \alpha R_{S_1} + \beta R_{S_2}. \end{aligned}$$

2) We can directly get the result 2) from 1), because $R_{S_1 - S_2}$ is a nonnegative definite matrix.
3) It is easy to see from 1). Let $S_1 = \hat{S}_1 + Z$ and $S_2 = \hat{S}_2 + Z$, then

$$\begin{aligned} \alpha S_1 + \beta S_2 &= \alpha(\hat{S}_1 + Z) + \beta(\hat{S}_2 + Z) = \alpha\hat{S}_1 + \beta\hat{S}_2 + Z \\ S_1 - S_2 &= \hat{S}_1 + Z - \hat{S}_2 - Z = \hat{S}_1 - \hat{S}_2. \end{aligned}$$

Therefore

$$\begin{aligned} \alpha R_{\hat{S}_1 + Z} + \beta R_{\hat{S}_2 + Z} &= \alpha R_{S_1} + \beta R_{S_2} \\ &= R_{\alpha S_1 + \beta S_2} + \alpha\beta R_{S_1 - S_2} \\ &= R_{\alpha\hat{S}_1 + \beta\hat{S}_2 + Z} + \alpha\beta R_{\hat{S}_1 - \hat{S}_2}. \end{aligned}$$

Then we have the result 3).

4) We can directly get the result 4) from 2) of Lemma 1 and 2), 3) of Proposition 4.

By 2) of Lemma 1, we have

$$R_{\alpha S_1 + \beta S_2} \leq \alpha R_{S_1} + \beta R_{S_2}$$

and linear operators $R_{\alpha S_1 + \beta S_2}$ and $\alpha R_{S_1} + \beta R_{S_2}$ satisfy the conditions of Proposition 4. Therefore by Proposition 4, there exists W such that $\|W\| \leq 1$ and

$$R^{\frac{1}{2}}_{\alpha S_1 + \beta S_2} = (\alpha R_{S_1} + \beta R_{S_2})^{\frac{1}{2}} W. \qquad \text{Q.E.D.}$$

Theorem 1: Let S_1 and S_2 be two statistically independent, zero-mean random vectors, and let Z be the zero-mean random vector. For any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$, the following formula holds:

$$\left| R_{\sqrt{\alpha} S_1 + \sqrt{\beta} S_2} + R_Z \right| \geq |R_{S_1} + R_Z|^{\alpha} |R_{S_2} + R_Z|^{\beta}.$$

Proof of Theorem 1: Since

$$\begin{aligned} R_{\sqrt{\alpha} S_1 + \sqrt{\beta} S_2} &= E\left(\sqrt{\alpha} S_1 + \sqrt{\beta} S_2\right)^2 \\ &= E\left(\alpha S_1^2 + \beta S_2^2\right) \\ &= \alpha E S_1^2 + \beta E S_2^2 = \alpha R_{S_1} + \beta R_{S_2} \end{aligned}$$

then

$$\begin{aligned} R_{\sqrt{\alpha} S_1 + \sqrt{\beta} S_2} + R_Z &= \alpha R_{S_1} + \beta R_{S_2} + R_Z \\ &= \alpha(R_{S_1} + R_Z) + \beta(R_{S_2} + R_Z). \end{aligned}$$

By taking determinants on both sides of the above equality, we have

$$\begin{aligned} \left| R_{\sqrt{\alpha} S_1 + \sqrt{\beta} S_2} + R_Z \right| &= |\alpha(R_{S_1} + R_Z) + \beta(R_{S_2} + R_Z)| \\ &\overset{(a)}{\geq} |R_{S_1} + R_Z|^{\alpha} |R_{S_2} + R_Z|^{\beta}. \end{aligned} \qquad (1)$$

Here, (a) follows from Proposition 3. Q.E.D.

Corollary 1: $C_{n,Z}(P)$ is a concave function with respect to P. That is, for any $P_1, P_2 \geq 0$ and for any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$

$$C_{n,Z}(\alpha P_1 + \beta P_2) \geq \alpha C_{n,Z}(P_1) + \beta C_{n,Z}(P_2).$$

Proof of Corollary 1: We can write $C_{n,Z}(P)$ as the follows:

$$C_{n,Z}(P) = \max_{S \in \Gamma(P)} \frac{1}{2n} \log \frac{\left| R_S^{(n)} + R_Z^{(n)} \right|}{\left| R_Z^{(n)} \right|}$$

where $\Gamma(P) = \{S; Tr[R_S] \leq nP\}$. By Theorem 1, dividing by the determinant of $R_Z^{(n)}$ and taking the logarithm on both sides of (1), we have

$$\begin{aligned} \frac{1}{2n} \log \frac{\left| R^{(n)}_{\sqrt{\alpha} S_1 + \sqrt{\beta} S_2} + R_Z^{(n)} \right|}{\left| R_Z^{(n)} \right|} &\geq \alpha \frac{1}{2n} \log \frac{\left| R_{S_1}^{(n)} + R_Z^{(n)} \right|}{\left| R_Z^{(n)} \right|} \\ &\quad + \beta \frac{1}{2n} \log \frac{\left| R_{S_2}^{(n)} + R_Z^{(n)} \right|}{\left| R_Z^{(n)} \right|}. \end{aligned} \qquad (2)$$

Let S_1 attain $C_{n,Z}(P_1)$ with $S_1 \in \Gamma(P_1)$ and let S_2 attain $C_{n,Z}(P_2)$ with $S_2 \in \Gamma(P_2)$. Then the right-hand side (RHS) of (2) equals

$$\text{RHS} = \alpha C_{n,Z}(P_1) + \beta C_{n,Z}(P_2).$$

Since $\sqrt{\alpha} S_1 + \sqrt{\beta} S_2 \in \Gamma(\alpha P_1 + \beta P_2)$, we maximize the left-hand side (LHS) of (2) over $\Gamma(\alpha P_1 + \beta P_2)$ and get

$$C_{n,Z}(\alpha P_1 + \beta P_2) = \text{LHS}.$$

Thus we have

$$C_{n,Z}(\alpha P_1 + \beta P_2) \geq \alpha C_{n,Z}(P_1) + \beta C_{n,Z}(P_2). \qquad \text{Q.E.D.}$$

Theorem 2: Let R_{S_i} be the covariance matrix of a zero-mean random vector S_i, where $i \in \{1,2\}$. For any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$, the following formula holds:

$$|\alpha R_{S_1+Z} + \beta R_{S_2+Z}| = |R_{\tilde{S}} + U + U^t + R_Z| \\ \geq |R_{S_1+Z}|^\alpha |R_{S_2+Z}|^\beta$$

where

$$R_{\tilde{S}} = \alpha R_{S_1} + \beta R_{S_2}$$

and

$$U = (R_{\tilde{S}})^{\frac{1}{2}} WVR_Z^{\frac{1}{2}}, \quad \|W\| < 1, \quad \|V\| < 1.$$

Proof of Theorem 2: By Lemma 1 1), we have

$$\begin{aligned} &\alpha R_{S_1+Z} + \beta R_{S_2+Z} \\ &= R_{\alpha S_1+\beta S_2+Z} + \alpha\beta R_{S_1-S_2} \\ &= R_{\alpha S_1+\beta S_2} + R_{\alpha S_1+\beta S_2, Z} \\ &\quad + R_{Z,\alpha S_1+\beta S_2} + R_Z + \alpha\beta R_{S_1-S_2} \\ &\overset{(b)}{=} \alpha R_{S_1} + \beta R_{S_2} + R_{\alpha S_1+\beta S_2, Z} + R_{Z,\alpha S_1+\beta S_2} + R_Z \\ &\overset{(c)}{=} \alpha R_{S_1} + \beta R_{S_2} + R^{\frac{1}{2}}_{\alpha S_1+\beta S_2} V R_Z^{\frac{1}{2}} \\ &\quad + R_Z^{\frac{1}{2}} V^t R^{\frac{1}{2}}_{\alpha S_1+\beta S_2} + R_Z \\ &\overset{(d)}{=} \alpha R_{S_1} + \beta R_{S_2} + (\alpha R_{S_1} + \beta R_{S_2})^{\frac{1}{2}} WVR_Z^{\frac{1}{2}} \\ &\quad + R_Z^{\frac{1}{2}} (WV)^t (\alpha R_{S_1} + \beta R_{S_2})^{\frac{1}{2}} + R_Z \\ &= R_{\tilde{S}} + U + U^t + R_Z. \end{aligned}$$

Here (b) follows from the Lemma 1 1), and (c) follows from Proposition 5, where $\|V\| \leq 1$, and (d) follows from the fact that we can obtain $R_{\alpha S_1+\beta S_2} \leq \alpha R_{S_1} + \beta R_{S_2}$ by Lemma 1 (ii) and $(R_{\alpha S_1+\beta S_2})^{\frac{1}{2}} = (\alpha R_{S_1} + \beta R_{S_2})^{\frac{1}{2}} W$ by by Lemma 1 (iv), where $\|W\| \leq 1$. By taking determinants on both sides of the equality above, we have

$$\begin{aligned} \left|R_{\tilde{S}} + U + U^t + R_Z\right| &= |\alpha R_{S_1+Z} + \beta R_{S_2+Z}| \\ &\overset{(e)}{\geq} |R_{S_1+Z}|^\alpha |R_{S_2+Z}|^\beta. \end{aligned} \tag{3}$$

Here (e) follows from Proposition 3. Q.E.D.

Corollary 2: $C_{n,FB,Z}(P)$ is a concave function with respect to P. That is, for any $P_1, P_2 \geq 0$ and for any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$,

$$C_{n,FB,Z}(\alpha P_1 + \beta P_2) \geq \alpha C_{n,FB,Z}(P_1) + \beta C_{n,FB,Z}(P_2).$$

Proof of Corollary 2: We can write $C_{n,FB,Z}(P)$ as follows:

$$C_{n,FB,Z}(P) = \max_{S \in \Gamma(P)} \frac{1}{2n} \log \frac{\left|R^{(n)}_{S+Z}\right|}{\left|R^{(n)}_Z\right|},$$

where $\Gamma(P) = \{S; Tr[R_S] \leq nP\}$. By Theorem 2, dividing by the determinant of $R_Z^{(n)}$ and taking the logarithm on both sides of inequality (3), we have

$$\frac{1}{2n} \log \frac{\left|R^{(n)}_{\tilde{S}} + U + U^t + R^{(n)}_Z\right|}{\left|R^{(n)}_Z\right|} \geq \frac{1}{2n} \log \frac{\left|R^{(n)}_{S_1+Z}\right|^\alpha \left|R^{(n)}_{S_2+Z}\right|^\beta}{\left|R^{(n)}_Z\right|}. \tag{4}$$

Let S_1 attain $C_{n,FB,Z}(P_1)$ with $S_1 \in \Gamma(P_1)$ and let S_2 attain $C_{n,FB,Z}(P_2)$ with $S_2 \in \Gamma(P_2)$, then the RHS of (4) is

$$\mathbf{RHS} = \alpha C_{n,FB,Z}(P_1) + \beta C_{n,FB,Z}(P_2).$$

Since

$$\begin{aligned} Tr\left[\alpha R^{(n)}_{S_1} + \beta R^{(n)}_{S_2}\right] &= \alpha \mathrm{Tr}\left[R^{(n)}_{S_1}\right] + \beta \mathrm{Tr}\left[R^{(n)}_{S_2}\right] \\ &\leq \alpha n P_1 + \beta n P_2 = n(\alpha P_1 + \beta P_2) \end{aligned}$$

and $\|WV\| \leq \|W\|\|V\| \leq 1$, we maximize the LHS of (4) over $\Gamma(\alpha P_1 + \beta P_2)$ and we get

$$C_{n,FB,Z}(\alpha P_1 + \beta P_2) \geq \mathbf{LHS}.$$

Thus, we have

$$C_{n,FB,Z}(\alpha P_1 + \beta P_2) \geq \alpha C_{n,FB,Z}(P_1) + \beta C_{n,FB,Z}(P_2). \text{ Q.E.D.}$$

III. Operator Inequality

Before proving that $C_{n,Z}(P)$ and $C_{n,FB,Z}(P)$ are convex functions of the covariance matrix of additive Gaussian noise Z, we need to introduce some operator inequalities of the real Hilbert space.

Let $\mathcal{H}$ be a Hilbert space. Let $\mathcal{B}(\mathcal{H})$ be the set of all bounded linear operators on $\mathcal{H}$ and $\mathcal{B}(\mathcal{H})_+ = \{A \in \mathcal{B}(\mathcal{H}); A \geq 0\}$. Let J be any interval of $\mathbb{R}$ and $S(A)$ be spectrum of $A \in \mathcal{B}(\mathcal{H})$.

Definition 1: Let $f : J \to \mathbb{R}$ be continuous.

1) f is called operator monotone if for any self-adjoint $A, B \in \mathcal{B}(\mathcal{H})$ satisfying $S(A), S(B) \subset J$,

$$A \leq B \text{ implies } f(A) \leq f(B).$$

2) f is called operator convex if for any self-adjoint $A, B \in \mathcal{B}(\mathcal{H})$ satisfying $S(A), S(B) \subset J$

$$f\left(\frac{A+B}{2}\right) \leq \frac{f(A) + f(B)}{2}.$$

By the continuity of f, it is equivalent to

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B)$$

for any $0 \leq \lambda \leq 1$.

3) f is called operator concave if $-f$ is operator convex.

Proposition 6 ([12]): Let f be nonnegative continuous function on $[0,\infty)$. Then f is operator monotone if and only if f is operator concave.

Proposition 7 ([12]): $f(t) = t^{-1}$ is operator convex on $[0,\infty)$.

Definition 2 (Kubo and Ando [15]): σ is called operator connection if σ is binary operation on $\mathcal{B}(\mathcal{H})_+$ satisfying the following axioms.

1) (Monotonicity)

$$A \leq C \text{ and } B \leq D \text{ implies } A\sigma B \leq C\sigma D.$$

2) (Transform Inequality)

$$C(A\sigma B)C \leq (CAC)\sigma(CBC).$$

3) (Upper Continuity)

$$A_n \downarrow A \text{ and } B_n \downarrow B \text{ implies } A_n \sigma B_n \downarrow A\sigma B$$

where $A_n \downarrow A$ represents

$$A_1 \geq A_2 \geq \cdots$$

and

$$A_n \to A(\text{strong operator topology}).$$

σ is called operator mean if σ is operator connection satisfying $I\sigma I = I$.

Proposition 8 ([15]): For any operator connection σ, there exsits a unique nonnegative operator monotone function f on $[0, \infty)$ such that

$$f(t)I = I\sigma(tI), t \geq 0.$$

Then we have the followings:

1) $\sigma \to f$ is an affine order isomorphism between the class of connections and the class of nonnegative operator monotone functions on $[0, \infty)$.
2) For invertible $A \in \mathcal{B}(\mathcal{H})_+$

$$A\sigma B = A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2}.$$

3) σ is operator mean if and only if $f(1) = 1$.

Proposition 9 ([15]): Let σ be operator connection and $A, B, C \in \mathcal{B}(\mathcal{H})_+$.

1) For any invertible C

$$C(A\sigma B)C = (CAC)\sigma(CBC).$$

2) For any $\alpha \geq 0$

$$\alpha(A\sigma B) = (\alpha A)\sigma(\alpha B).$$

Definition 3 (Kubo and Ando [15]): For invertible $A, B \in \mathcal{B}(\mathcal{H})_+$, parallel sum is difined by

$$A : B = (A^{-1} + B^{-1})^{-1}.$$

In general for $A, B \in \mathcal{B}(\mathcal{H})_+$, it is defined by

$$A : B = s - \lim_{\epsilon \downarrow 0} (A + \epsilon I) : (B + \epsilon I)$$

where $s - \lim A_n$ represents the limit of A_n relative to strong operator topology. Harmonic mean is defined by

$$A!B = 2(A : B).$$

Proposition 10 ([15]): Let σ be operator connection and $A, B, C, D \in \mathcal{B}(\mathcal{H})_+$. Then

$$(A\sigma B) : (C\sigma D) \geq (A : C)\sigma(B : D).$$

Lemma 2: Let f be nonnegative continuous function on $[0, \infty)$. If f is operator monotone, then for any $A, B \in \mathcal{B}(\mathcal{H})_+$

$$f(A!B) \leq f(A)!f(B).$$

Proof of Lemma 2: By Proposition 10, let $U, V, X, Y \in \mathcal{B}(\mathcal{H})_+$ then

$$(U\sigma V) : (X\sigma Y) \geq (U : V)\sigma(X : Y).$$

Let $U = I, V = A, X = I, Y = B$ then

$$\begin{aligned}(I\sigma A) : (I\sigma B) &\geq (I : I)\sigma(A : B) \\ &= (I^{-1} : I^{-1})\sigma(A : B) \\ &= (2I)^{-1}\sigma(A : B) \\ &= \left(\frac{1}{2}I\right)\sigma(A : B) \\ &= \left(\frac{1}{2}I\right)\sigma\left(\frac{1}{2}(2(A : B))\right) \\ &= \frac{1}{2}(I\sigma(2(A : B))) \\ &= \frac{1}{2}(I\sigma(A!B)).\end{aligned}$$

Then

$$2((I\sigma A) : (I\sigma B)) \geq I\sigma(A!B).$$

Hence

$$(I\sigma A)!(I\sigma B) \geq I\sigma(A!B).$$

By Proposition 8, for this operator connection σ, there exists a unique operator monotone function $f \geq 0$ let $f(A)I = I\sigma(AI)$, therefore

$$f(A)!f(B) \geq f(A!B). \qquad \text{Q.E.D.}$$

Lemma 3: Let f be positive continuous function on $[0, \infty)$. If $f(t)$ is operator monotone, then $f(t^{-1})$ is operator convex.

Proof of Lemma 3: For any invertible $A, B \in \mathcal{B}(\mathcal{H})_+$, we have

$$\begin{aligned}f\left(\left(\frac{A+B}{2}\right)^{-1}\right) &= f(A^{-1}!B^{-1}) \\ &\overset{(g)}{\geq} f(A^{-1})!f(B^{-1}) \\ &= \left\{\frac{(f(A^{-1}))^{-1} + (f(B^{-1}))^{-1}}{2}\right\}^{-1} \\ &\overset{(h)}{\leq} \frac{1}{2}f(A^{-1}) + \frac{1}{2}f(B^{-1}).\end{aligned}$$

Here (g) following from the Lemma 2 and (h) following from the Proposition 7. Q.E.D.

Remark 1: We remark that it is shown that $f(x) = \log(1 + \frac{1}{x})$ is operator convex in [21].

IV. Convexity of $C_{n,Z}(P)$ and $C_{n,FB,Z}(P)$ With Respect to the Noise Covariance

Theorem 3: Let R_{Z_1} and R_{Z_2} denote covariance matrices of zero-mean random vectors Z_1 and Z_2, respectively. For any $\alpha, \beta \geq 0$ satisfying $\alpha + \beta = 1$, we set $R_{\tilde{Z}} = \alpha R_{Z_1} + \beta R_{Z_2}$, then the following formula holds:

$$\log \frac{|R_S + R_{\tilde{Z}}|}{|R_{\tilde{Z}}|} \leq \alpha \log \frac{|R_S + R_{Z_1}|}{|R_{Z_1}|} + \beta \log \frac{|R_S + R_{Z_2}|}{|R_{Z_2}|}.$$

Proof of Theorem 3: Let R_S and R_Z denote covariance matrices of random vectors S and Z with mean zero. Thus we have

$$\frac{|R_S+R_Z|}{|R_Z|}=|R_S+R_Z||R_Z|^{-1}=\left|R_S^{1/2}R_Z^{-1}R_S^{1/2}+I\right|. \quad (5)$$

Let $A=R_S^{-1/2}R_{Z_1}R_S^{-1/2}$ and $B=R_S^{-1/2}R_{Z_2}R_S^{-1/2}$. Then $\alpha A+\beta B=R_S^{-1/2}R_{\tilde{Z}}R_S^{-1/2}$. Let $f(x)=\log(1+x), x\in[0,\infty)$. Then $f(x)$ is a positive continuous function on $[0,\infty)$. It is well known that $f(x)$ is operator monotone. By Lemma 3, $f(x^{-1})$ is operator convex. Then we have

$$f((\alpha A+\beta B)^{-1})\leq \alpha f(A^{-1})+\beta f(B^{-1}).$$

That is

$$\begin{aligned}\log&\left(I+R_S^{1/2}R_{\tilde{Z}}^{-1}R_S^{1/2}\right)\\&=\log\left(I+\left(R_S^{-1/2}R_{\tilde{Z}}R_S^{-1/2}\right)^{-1}\right)\\&\leq\alpha\log\left(I+\left(R_S^{-1/2}R_{Z_1}R_S^{-1/2}\right)^{-1}\right)\\&\quad+\beta\log\left(I+\left(R_S^{-1/2}R_{Z_2}R_S^{-1/2}\right)^{-1}\right)\\&=\alpha\log\left(I+R_S^{1/2}R_{Z_1}^{-1}R_S^{1/2}\right)\\&\quad+\beta\log\left(I+R_S^{1/2}R_{Z_2}^{-1}R_S^{1/2}\right).\end{aligned}$$

By taking the trace on both sides

$$\log\left|I+R_S^{1/2}R_{\tilde{Z}}^{-1}R_S^{1/2}\right|\leq\alpha\log\left|I+R_S^{1/2}R_{Z_1}^{-1}R_S^{1/2}\right|+\beta\log\left|I+R_S^{1/2}R_{Z_2}^{-1}R_S^{1/2}\right|.$$

It follows from (5) that

$$\log\frac{|R_S+R_{\tilde{Z}}|}{|R_{\tilde{Z}}|}\leq\alpha\log\frac{|R_S+R_{Z_1}|}{|R_{Z_1}|}+\beta\log\frac{|R_S+R_{Z_2}|}{|R_{Z_2}|}.$$ Q.E.D.

Corollary 3: $C_{n,Z}(P)$ is a convex function of the noise covariance matrix. That is, for any Z_1, Z_2, for any $P\geq 0$ and for any $\alpha,\beta\geq 0$ satisfying $\alpha+\beta=1$, let $R_{\tilde{Z}}^{(n)}=\alpha R_{Z_1}^{(n)}+\beta R_{Z_2}^{(n)}$, where $R_{Z_1}^{(n)}$ and $R_{Z_2}^{(n)}$ denote the covariance matrices of Z_1 and Z_2, respectively, then the following inequality holds:

$$C_{n,\tilde{Z}}(P)\leq\alpha C_{n,Z_1}(P)+\beta C_{n,Z_2}(P).$$

Proof of Corollary 3: We define $C_{n,Z}(P)$ as the following:

$$C_{n,Z}(P)=\max_{\mathrm{Tr}\left[R_S^{(n)}\right]\leq nP}\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_Z^{(n)}\right|}{\left|R_Z^{(n)}\right|}.$$

By Theorem 3, then

$$\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_{\tilde{Z}}^{(n)}\right|}{\left|R_{\tilde{Z}}^{(n)}\right|}\leq\alpha\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_{Z_1}^{(n)}\right|}{\left|R_{Z_1}^{(n)}\right|}+\beta\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_{Z_2}^{(n)}\right|}{\left|R_{Z_2}^{(n)}\right|}. \quad (6)$$

Let $S\in\Gamma(P)$ attain $C_{n,\tilde{Z}}(P)$, where $\Gamma(P)=\{S;Tr[R_S]\leq nP\}$. By taking the maximization of the RHS of (6), we get

$$\begin{aligned}\max_{\mathrm{Tr}\left[R_S^{(n)}\right]\leq nP}&\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_{\tilde{Z}}^{(n)}\right|}{\left|R_{\tilde{Z}}^{(n)}\right|}\\&\leq\max_{\mathrm{Tr}\left[R_S^{(n)}\right]\leq nP}\alpha\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_{Z_1}^{(n)}\right|}{\left|R_{Z_1}^{(n)}\right|}\\&\quad+\max_{\mathrm{Tr}\left[R_S^{(n)}\right]\leq nP}\beta\frac{1}{2n}\log\frac{\left|R_S^{(n)}+R_{Z_2}^{(n)}\right|}{\left|R_{Z_2}^{(n)}\right|}.\end{aligned}$$

We obtain the proof. Q.E.D.

Now we have the following convex-like property of $C_{n,FB,\cdot}(P)$.

Corollary 4: For any Z_1, Z_2, for any $P\geq 0$ and for any $\alpha,\beta\geq 0$ satisfying $\alpha+\beta=1$, there exist $P_1,P_2\geq 0$ satisfying $P=\alpha P_1+\beta P_2$ such that

$$C_{n,FB,\tilde{Z}}(P)\leq\alpha C_{n,FB,Z_1}(P_1)+\beta C_{n,FB,Z_2}(P_2).$$

Proof of Corollary 4: We can write $C_{n,FB,Z}(P)$ as follows:

$$C_{n,FB,Z}(P)=\max\frac{1}{2n}\log\frac{\left|R_X^{(n)}+R_Z^{(n)}\right|}{\left|R_Z^{(n)}\right|}$$

where $X=S-TY$ and T is a strictly lower triangular, and the maximum is taken subject to the constraint

$$\mathrm{Tr}\left[(I+B)R_X^{(n)}(I+B^t)+BR_Z^{(n)}B^t\right]\leq nP$$

where $R_X^{(n)}$ is symmetric, nonnegative definite, and B is strictly lower triangular. By Theorem 3

$$\frac{1}{2n}\log\frac{\left|R_X^{(n)}+R_{\tilde{Z}}^{(n)}\right|}{\left|R_{\tilde{Z}}^{(n)}\right|}\leq\alpha\frac{1}{2n}\log\frac{\left|R_X^{(n)}+R_{Z_1}^{(n)}\right|}{\left|R_{Z_1}^{(n)}\right|}+\beta\frac{1}{2n}\log\frac{\left|R_X^{(n)}+R_{Z_2}^{(n)}\right|}{\left|R_{Z_2}^{(n)}\right|}. \quad (7)$$

Let $(\hat{X},\hat{B})\in\Delta(P)$ attain $C_{n,FB,\tilde{Z}}(P)$, where

$$\Delta(P)=\left\{(X,B);\mathrm{Tr}\left[(I+B)R_X^{(n)}(I+B^t)+BR_{\tilde{Z}}^{(n)}B^t\right]\leq nP\right\}.$$

Since

$$\begin{aligned}\mathrm{Tr}&\left[(I+\hat{B})R_{\hat{X}}^{(n)}(I+(\hat{B})^t)+\hat{B}R_{\tilde{Z}}^{(n)}(\hat{B})^t\right]\\&=\alpha\mathrm{Tr}\left[(I+\hat{B})R_{\hat{X}}^{(n)}(I+(\hat{B})^t)+\hat{B}R_{Z_1}^{(n)}(\hat{B})^t\right]\\&\quad+\beta\mathrm{Tr}\left[(I+\hat{B})R_{\hat{X}}^{(n)}(I+(\hat{B})^t)+\hat{B}R_{Z_2}^{(n)}(\hat{B})^t\right]\end{aligned}$$

we have $\alpha P_1+\beta P_2=P$, where

$$\mathrm{Tr}\left[(I+\hat{B})R_{\hat{X}}^{(n)}(I+(\hat{B})^t)+\hat{B}R_{Z_1}^{(n)}(\hat{B})^t\right]=nP_1$$

and

$$\mathrm{Tr}\left[(I+\hat{B})R_{\hat{X}}^{(n)}(I+(\hat{B})^t)+\hat{B}R_{Z_2}^{(n)}(\hat{B})^t\right]=nP_2.$$

By taking the maximization of the right hand side of (7), we have the result. Q.E.D.

Finally we state the following conjecture.

Conjecture: For any Z_1, Z_2, for any $P \geq 0$ and for any $\alpha, \beta \geq 0 (\alpha + \beta = 1)$

$$C_{n,FB,\tilde{Z}}(P) \leq \alpha C_{n,FB,Z_1}(P) + \beta C_{n,FB,Z_2}(P).$$

V. Conclusion

We gave several inherent prperties of the capacity function of Gaussian channel with and without feedback by using operator inequalities and matrix analysis. By using the operator concavity of $\log x$ we showed that $C_{n,FB,Z}(P)$ is a concave function of P. And also by using the operator convexity of $\log\left(1+\frac{1}{t}\right)$ we showed that $C_{n,FB,Z}(P)$ is a convex-like function of the noise covariance R_Z. The operator convexity of $\log\left(1+\frac{1}{t}\right)$ is generalized to the operator convexity of $f(t^{-1})$ as a function of t, where $f(t)$ is operator monotone. Though the nonfeedback capacity $C_{n,Z}(P)$ is a convex function of R_Z, the feedback capacity $C_{n,FB,Z}(P)$ is a convex-like function of R_Z. Strict convexity of $C_{n,FB,Z}(P)$ as a function of R_Z remains an open problem.

References

[1] C. R. Baker, "Joint measures and cross covariance operators," *Trans. Amer. Math. Soc.*, vol. 186, pp. 273–289, 1973.

[2] H. W. Chen and K. Yanagi, "On the cover's conjecture on capacity of Gaussian channel with feedback," *IEICE Trans. Fund.*, vol. E80-A, no. 11, pp. 2272–2275, Nov. 1997.

[3] ——, "Refinements of the half-bit and factor-of-two bounds for capacity in Gaussian channels with feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 319–325, Jan. 1999.

[4] ——, "Upper bounds on the capacity of discrete time blockwise white Gaussian channels with feedback," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 1125–1131, May 2000.

[5] T. M. Cover, "Conjecture: Feedback does not help much," in *Open Problems in Communication and Computation*, T. Covert and B. Gopinath, Eds. New York: Springer-Verlag, 1987, pp. 70–71.

[6] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 37–43, Jan. 1989.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[8] R. G. Douglas, "On majorization, factorization, and range inclusion of operators on Hilbert space," *Proc. Amer. Math. Soc.*, vol. 17, pp. 413–415, 1966.

[9] A. Dembo, "On Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. IT-35, no. 5, pp. 1072–1089, Sep. 1989.

[10] P. Ebert, "The capacity of the Gaussian channel with feedback," *Bell. Syst. Tech. J.*, vol. 49, pp. 1705–1712, 1970.

[11] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[12] F. Hansen and G. K. Pedersen, "Jensen's inequality for operators and Löwer's theorem," *Math. Ann.*, vol. 258, pp. 229–241, 1982.

[13] F. Hiai and K. Yanagi, *Hilbert Spaces and Linear Operators* (in Japanese): Makino-Shoten, 1995.

[14] S. Ihara and K. Yanagi, "Capacity of discrete time Gaussian channel with and without feedback, II," *Japan J. Appl. Math.*, vol. 6, pp. 245–258, 1989.

[15] F. Kubo and T. Ando, "Means of positive linear operators," *Math. Ann.*, vol. 240, pp. 205–224, 1980.

[16] M. Pinsker, *Talk Delivered at the Soviet Information Theory Meeting*, 1969.

[17] K. Yanagi, "An upper bound to the capacity of discrete time Gaussian channel with feedback," *Lecture Notes Math.*, vol. 1299, pp. 565–570, 1988.

[18] ——, "Necessary and sufficient condition for capacity of the discrete time Gaussian channel to be increased by feedback," *IEEE Trans. Inf. Theory*, vol. IT-38, no. 6, pp. 1788–1791, Nov. 1992.

[19] ——, "An upper bound to the capacity of discrete time Gaussian channel with feedback, II," *IEEE Trans. Information Theory*, vol. 40, no. 2, pp. 588–593, Mar. 1994.

[20] ——, "An upper bound to the capacity of discrete time Gaussian channel with feedback, III," *Bull. Kyushu Inst. Tech., Pure Appl. Math.*, vol. 45, pp. 1–8, 1998.

[21] K. Yanagi, H. W. Chen, and J. W. Yu, "Operator inequality and its application to information theory," *Taiwanese J. Math.*, vol. 4, no. 3, pp. 407–416, Sep. 2000.