Linear Sum Capacity for Gaussian Multiple Access Channel with Feedback

Ehsan Ardestanizadeh, Michèle A. Wigger, Young-Han Kim, and Tara Javidi

***Abstract*—This paper studies the class of *generalized linear feedback codes* for additive white Gaussian noise multiple access channel. This class includes (nonlinear) nonfeedback codes at one extreme and linear feedback codes by Schalkwijk and Kailath, Ozarow, and Kramer at the other extreme. The *linear sum capacity* $C_{\mathsf{L}}(P)$, the maximum sum-rate achieved by the generalized linear feedback codes, is characterized under symmetric block power constraints P for all the senders. In particular, it is shown that the Kramer linear code achieves $C_{\mathsf{L}}(P)$. Based on the properties of the *conditional maximal correlation*, an extension of the Hirschfeld–Gebelein–Renyi maximal correlation, it is conjectured that Kramer's linear code achieves not only the *linear* sum capacity, but also the general sum capacity, i.e., the maximum sum-rate achieved by *arbitrary* feedback codes.**

I. Introduction

The capacity region for the N-sender additive white Gaussian noise (AWGN) multiple access channel (MAC) with feedback is not known except for the case of $N = 2$, which was found by Ozarow [1]. The capacity achieving feedback code for $N = 2$ is an extension of the linear feedback code by Schalkwijk and Kailath [2] for the single-user AWGN channel. Two decades later, Kramer [3] further generalized Ozarow's code to the case $N \geq 3$ and established the sum capacity for the case of equal and high enough power constraints (cf. (29)). However, it is not known whether the Kramer code achieves the sum-capacity in general.

In this paper, we focus on the class of *generalized linear feedback codes* (or *linear codes* in short), where linearity refers to how the feedback signals are incorporated into the transmitted signals. This class of generalized linear feedback codes includes the linear feedback codes by Schalkwijk and Kailath [2], Ozarow [1], and Kramer [3] as well as any nonfeedback (nonlinear) code.

We characterize the *linear sum capacity* $C_{\mathsf{L}}(P)$, which is the maximum sum-rate achieved by generalized linear feedback codes under symmetric block power constraints P for all the senders (see Theorem 1). The main contribution is the proof of the converse. We first prove an upper bound on $C_{\mathsf{L}}(P)$, which is a multi-letter optimization problem over Gaussian distributions (cf. Cover and Pombra [4]). Next, we derive an equivalent optimization problem over the set of positive semidefinite (covariance) matrices by considering a dependence balance condition, introduced by Hekstra and Willems [5] and refined by Kramer and Gastpar [6]. Lastly, we carefully analyze this (nonconvex) optimization problem via Lagrange dual formulation.

Achievability is proved by the Kramer linear code. Hence, this rather simple code which refines receiver's knowledge iteratively is sum-rate optimal among the class of generalized linear feedback codes.

Complete characterization of $C(P)$, the maximum sum-rate among all feedback codes, still remains open. We conjecture that $C(P) = C_{\mathsf{L}}(P)$, based on the properties of *conditional maximal correlation*, which is an extension of the Hirschfeld–Gebelein–Renyi maximal correlation [7] to the case where an extra common random variable is shared (see Section IV).

The rest of the paper is organized as follows. In Sections II and III, we state the main result and provide the proof of the converse, respectively. Section IV concludes the paper by discussion on the aforementioned conjecture.

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, 92093-0407, USA, email: ehsan@ucsd.edu, michele.wigger@telecom-paristech.fr, yhk@ucsd.edu, tjavidi@ucsd.edu.

II. Main Result

Consider the communication problem over AWGN-MAC where each sender $j \in \{1, \dots, N\}$ wishes to transmit a message $M_j \in \mathcal{M}_j := \{1, \dots, 2^{nR_j}\}$ reliably to the common receiver. At each time $i = 1, \dots, n$, the output of the channel is

$$Y_i = \sum_{j=1}^{N} X_{ji} + Z_i \tag{1}$$

where $\{Z_i\}$ is a discrete-time zero-mean white Gaussian noise process with unit average power, i.e., $\mathsf{E}(Z_i^2) = 1$, and is independent of $M_1, \dots, M_N$. We assume that the output symbols are causally fed back to each sender and the transmitted symbol X_{ji} from sender j at time i depends on both the previous channel output sequence $Y^{i-1} := \{Y_1, Y_2, \dots, Y_{i-1}\}$ and the message M_j.

We define a $(2^{nR_1}, \dots, 2^{nR_N}, n)$ code with power constraints $P_1, \dots, P_N$ as

1) N message sets $\mathcal{M}_1, \dots, \mathcal{M}_N$, where $\mathcal{M}_j = \{1, 2, \dots, 2^{nR_j}\}$,
2) a set of N encoders, where encoder j at each time i maps the pair (m_j, y^{i-1}) to a symbol x_{ji} such that X_{ji} satisfy the *block power constraint*
$$\sum_{i=1}^{n} \mathsf{E}(X_{ji}^2(m_j, Y^{i-1})) \leq nP_j, \quad m_j \in \mathcal{M}_j,$$
and
3) a decoder map which assigns message estimates $\hat{m}_j \in \mathcal{M}_j$, $j \in \{1, \dots, n\}$, to each received sequence y^n.

978-1-4244-7892-7/10/$26.00 ©2010 IEEE

We assume throughout that $M(\mathcal{S}) := (M_1, \ldots, M_N)$ is a random vector uniformly distributed over $\mathcal{M}_1 \times \cdots \times \mathcal{M}_N$. The probability of error is defined as $P_e^{(n)} := \mathsf{P}\{\hat{M}(\mathcal{S}) \neq M(\mathcal{S})\}$. A rate-tuple $(R_1, \ldots, R_N)$ is called achievable if there exists a sequence of $(2^{nR_1}, \ldots, 2^{nR_N}, n)$ codes such that $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity region $\mathscr{C}$ is defined as the closure of the set of achievable rate-tuples and the sum capacity C is defined as

$$C := \sup \left\{ \sum_{j=1}^{N} R_j : \ (R_1, \ldots, R_N) \in \mathscr{C} \right\}.$$

We refer to $R = \sum_{j=1}^{N} R_j$ as the sum-rate of a given code.

Definition 1: A $(2^{nR_1}, \ldots, 2^{nR_N}, n)$ code is called a *generalized linear feedback code*, if the encoding maps can be decomposed as follows.

1) Nonfeedback (nonlinear) mappings: The message m_j is mapped to a vector $\mathbf{\Theta}_j \in \mathbb{R}^k, k \in \{1, \ldots, n\}$, which we refer to as the message point.
2) Linear feedback mappings: At each time i, the pair $(\mathbf{\Theta}_j, Y^{i-1})$ is mapped to a symbol X_{ji} such that $X_{ji} = \mathsf{L}_{ji}(\mathbf{\Theta}_j, Y^{i-1})$ is linear in $(\mathbf{\Theta}_j, Y^{i-1})$.

As mentioned earlier, any nonfeedback code is a generalized linear feedback code by picking $k = n$ and $\mathbf{\Theta}_j \in \mathbb{R}^n$ to be the codeword of the j-th user.

Let $\mathscr{C}_{\mathsf{L}}$ be the closure of the set of rate-tuples achievable by linear codes and the linear sum capacity C_{L} be

$$C_{\mathsf{L}} := \sup \left\{ \sum_{j=1}^{N} R_j : (R_1, \ldots, R_N) \in \mathscr{C}_{\mathsf{L}} \right\}.$$

The following theorem characterizes $C_{\mathsf{L}}(P)$, the linear sum capacity under symmetric block power constraints P.

Theorem 1: For the N-sender AWGN-MAC with symmetric block power constraints $P_j = P$, $j \in \{1, \ldots, N\}$, we have

$$C_{\mathsf{L}}(P) = \frac{1}{2} \log(1 + NP\phi(P)) \tag{2}$$

where $\phi(P) \in \mathbb{R}$ is the unique solution in interval $[1, N]$ for

$$(1 + NP\phi)^{N-1} = (1 + P\phi(N - \phi))^N. \tag{3}$$

The proof of the upper bound is provided in Section III. This establishes the theorem since it is already known that Kramer linear code [3] achieves sum-rates arbitrarily close to (2).

Note that $\phi(P) \in [1, N]$ captures the amount of cooperation among the senders such that $\phi = 1$ corresponds to no cooperation whereas $\phi = N$ corresponds to full cooperation.

III. Proof of the Converse

In this section we show that sum rate R achievable by a linear code under symmetric block power constraints P satisfies

$$R \leq C_1(P, \phi(P)) := \frac{1}{2} \log(1 + NP\phi(P)) \tag{4}$$

where $\phi(P) \in \mathbb{R}$ is the unique solution in interval $[1, N]$ for

$$(1 + NP\phi)^{N-1} = (1 + P\phi(N - \phi))^N.$$

The proof can be summarized in four steps. Consider an upper bound based on Fano's inequality for the sum-rate, which is a multi-letter optimization problem over all causal distributions. First, based on linearity of the code, we prove that by limiting the distributions to be Gaussian we can still upper bound the sum-rate (see Lemma 1). Second, using a dependence balance condition [5], [6] and the fact that for Gaussian distributions mutual information terms can be substituted by covariance matrices, we obtain an equivalent optimization problem (see (10)) over positive semi-definite matrices which is nonconvex due to the added dependence balance condition. Third, based on the Lagrange dual formulation and the symmetry of the functions in the problem, we derive an upper bound as a function of Lagrange multipliers which involves an optimization over only two variables (see Lemma 3). Finally, using a few technical tricks and strong duality, we show that with appropriate choices of Lagrange multipliers this upper bound is equal to $C_1(P, \phi(P))$ in (4) (see Lemma 4).

Details are as follows.

Step 1: For linear codes, we prove that the sum-rate can be upper bounded by considering only Gaussian distributions, in the multi-letter optimization problem based on Fano's inequality.

Lemma 1: A sum-rate R achievable by a linear code under symmetric block power constraints P is bounded as $R \leq \lim_{n\to\infty} C_n(P)$, where

$$C_n(P) := \max \frac{1}{n} \sum_{i=1}^{n} I(X_{1i}, \ldots, X_{Ni}; Y_i | Y^{i-1}). \tag{5}$$

Here the maximization is over X_{ji} of the form

$$\begin{aligned} X_{ji} &= \mathsf{L}_{ji}(\mathbf{V}_j, Y^{i-1}),\ i = 1, \ldots, n \\ \sum_{i=1}^{n} \mathsf{E}(X_{ji}^2) &\leq nP,\ j = 1, \ldots, N \end{aligned} \tag{6}$$

where L_{ji} is some linear function and $\mathbf{V}_j \in \mathbb{R}^n \sim \mathrm{N}(0, K_{\mathbf{V}_j})$ is Gaussian and independent of Z^n and $\{\mathbf{V}_{j'} : j' \neq j\}$.

Proof: From Fano's inequality [8] and memoryless property of the channel, it can be shown that if $P_e^{(n)} \to 0$ as $n \to \infty$ then

$$nR = n \sum_{k=1}^{N} R_j \leq \sum_{i=1}^{n} I(X_i(\mathcal{S}); Y_i | Y^{i-1}) + n\epsilon_n \tag{7}$$

where $\{\epsilon_n\}$ denotes a sequence such that $\epsilon_n \to 0$ as $n \to \infty$. Hence,

$$R \leq \lim_{n\to\infty} \max \frac{1}{n} \sum_{i=1}^{n} I(X_i(\mathcal{S}); Y_i | Y^{i-1}) \tag{8}$$

where the maximization is over all linear codes such that $X_{ji} = \mathsf{L}_{ji}(\mathbf{\Theta}_j, Y^{i-1})$.

Given a linear code with message points $\mathbf{\Theta}(\mathcal{S})$, let $\mathbf{V}(\mathcal{S}) \sim \mathrm{N}(0, K_{\mathbf{\Theta}(\mathcal{S})})$. We use $\mathbf{V}(\mathcal{S})$ with same linear functions as in the given code to generate

$$\tilde{X}_{ji} = \mathsf{L}_{ji}(\mathbf{V}_j, \tilde{Y}^{i-1})$$

where $\tilde{Y}_i$ is the output of the AWGN-MAC corresponding to $\tilde{X}_i(\mathcal{S})$. It is not hard to see that

$$(\tilde{X}_i(\mathcal{S}), \tilde{Y}^i) \sim \mathrm{N}(0, K_{X_i(\mathcal{S}),Y^i}).$$

By the conditional maximum entropy theorem [9, Lemma 1] we have

$$I(X_i(\mathcal{S}); Y_i|Y^{i-1}) \leq I(\tilde{X}_i(\mathcal{S}); \tilde{Y}_i|\tilde{Y}^{i-1}). \tag{9}$$

Combining (8) and (9) we have

$$R \leq \lim_{n\to\infty} \max \frac{1}{n}\sum_{i=1}^{n} I(\tilde{X}_i(\mathcal{S}); \tilde{Y}_i|\tilde{Y}^{i-1})$$

where $\tilde{X}_{ji} = \mathsf{L}_{ji}(\mathbf{V}_j, \tilde{Y}^{i-1})$. ■

Step 2: We show that optimization problem defining $C_n(P)$ in (5) is equivalent to the following optimization problem

$$\begin{aligned} \text{maximize} \quad & \tfrac{1}{n}\textstyle\sum_{i=1}^n f_1(K_i) \\ \text{subject to} \quad & K_i \succeq 0, \quad i = 1, \ldots, n \\ & \textstyle\sum_{i=1}^n (K_i)_{jj} \leq nP, \quad j = 1, \ldots, N \\ & \textstyle\sum_{i=1}^n f_1(K_i) - f_2(K_i) \leq 0. \end{aligned} \tag{10}$$

where

$$f_1(K_i) := \frac{1}{2}\log\Big(1 + \sum_{j,j'}(K_i)_{jj'}\Big) \tag{11}$$

and

$$f_2(K_i) := \frac{1}{2(N-1)}\sum_{j=1}^{N}\log\Bigg[1 + \sum_{j',j''}(K_i)_{j'j''} - \frac{\Big(\sum_{j'}(K_i)_{jj'}\Big)^2}{(K_i)_{jj}}\Bigg]. \tag{12}$$

The following lemma provides a necessary condition for any (causal) functional relationship of the form (6).

Lemma 2 ([6], Theorem 1): Consider independent random vectors $\mathbf{V}_j \in \mathbb{R}^n$ and let $X_{ji}, i = 1, \ldots, n, j = 1, \ldots, N$ be defined as in (6). Then,

$$\sum_{i=1}^{n}\Big(I(X_i(\mathcal{S}); Y_i|Y^{i-1}) \leq \frac{1}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{N} I(X_i(\mathcal{S}\backslash\{j\}); Y_i|Y^{i-1}, X_{ji})\Big). \tag{13}$$

Remark 1: The proof of Lemma 2 is valid also in the more general case where the inputs $X_{ji} = f_{ji}(\mathbf{V}_j, Y^{i-1})$ are obtained using arbitrary functions $\{f_{ji}\}$.

Since random vectors $\mathbf{V}_j$ in (6) are jointly Gaussian and the functions L_{ji} are linear, the random variables $(X^n(\mathcal{S}), Y^n)$ generated according to (6) are also jointly Gaussian and we can replace the mutual information terms in condition (13) with functions of the covariance matrices. Specifically, let $\mathbf{X}_i = (X_{1i}, \ldots, X_{Ni})^T \sim \mathrm{N}(0, K_i)$ where $K_i := K_{\mathbf{X}_i} \succeq 0$. Then

$$I(X_{1i}, \ldots, X_{Ni}; Y_i|Y^{i-1}) = f_1(K_i)$$

$$\frac{1}{N-1}\sum_{j=1}^{N} I(X_i(\mathcal{S}\backslash\{j\}); Y_i|Y^{i-1}, X_{ji}) = f_2(K_i)$$

Hence, the condition (13) reduces to

$$\sum_{i=1}^{n} f_1(K_i) - f_2(K_i) \leq 0. \tag{14}$$

Recall that the condition (14) follows from the functional realtionship (6). Hence, by representing the objective function and the power constraints in terms of K_i, the optimization problem in (5) becomes equivalent to (10).

Notice that even though both functions $f_1(K)$ and $f_2(K)$ are concave (see [10]), their difference $f_1(K) - f_2(K)$ is neither concave nor convex. Hence, the optimizatin problem (10) is nonconvex [11] due to the constraint (14).

Step 3: Using Lagrange multipliers $\lambda, \gamma \geq 0$, we provide a general upper bound $U(\lambda, \gamma)$ for the solution of the optimization problem given in (10). We further simplify this upper bound exploiting symmetry.

Consider the dual problem of (10) with equal Lagrange multipliers $\lambda_j = \lambda \geq 0$, $j \in \{1, \ldots, N\}$ for the power constraints $\sum_{i=1}^n P - (K_i)_{jj} \geq 0, \quad j \in \{1, \ldots, N\}$, and the Lagrange multiplier $\gamma \geq 0$ for the constraint $\sum_{i=1}^n f_2(K_i) - f_1(K_i) \geq 0$. Since the dual problem is an average of a function of K_i, it is not hard to see that the maximum can be upper bounded by the maximum of this function, which is given by

$$U(\lambda, \gamma) := \max_{K\succeq 0}(1-\gamma)f_1(K) + \gamma f_2(K) + \lambda\sum_{j=1}^{N}(P - K_{jj}). \tag{15}$$

It is easy to see that, for any $\lambda, \gamma \geq 0$, the solution of (10) is upper bounded by the solution of the dual problem. Hence, $U(\lambda, \gamma)$ is an upper bound for the solution of (10).

Lemma 3: Let $\lambda, \gamma \geq 0$. Then, the upper bound $U(\lambda, \gamma)$ can be simplified as follows.

$$U(\lambda, \gamma) = \max_{x\geq 0}\max_{0\leq\phi\leq N} g(\gamma, x, \phi) + \lambda N(P - x). \tag{16}$$

where

$$g(\gamma, x, \phi) := (1-\gamma)C_1(x, \phi) + \gamma C_2(x, \phi). \tag{17}$$

and

$$\begin{aligned} C_1(x, \phi) &:= \frac{1}{2}\log(1 + Nx\phi) \\ C_2(x, \phi) &:= \frac{N}{2(N-1)}\log(1 + (N-\phi)x\phi). \end{aligned} \tag{18}$$

Proof:

It can be shown [10] that there exists a matrix K of the following form

$$K = x \cdot \begin{pmatrix} 1 & \rho & \rho & \ldots & \rho \\ \rho & 1 & \rho & \ldots & \rho \\ \rho & \rho & 1 & \ldots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \ldots & 1 \end{pmatrix} \tag{19}$$

which achieves the maximum in (15). Thus, we continue our analysis with matrices of the form in (19) and by letting

$$\phi = 1 + (N-1)\rho$$

we have

$$f_1(K) = C_1(x,\phi)$$
$$f_2(K) = C_2(x,\phi). \tag{20}$$

Since K is positive semidefinite, $x \geq 0$ and $-1/(N-1) \leq \rho \leq 1$, where the lower bound on ρ comes from the fact that $\sum_{i,j=1}^{N} K_{ij}$ is nonnegative for $K \succeq 0$. Hence, $0 \leq \phi \leq N$ and (15) reduces to (16). ■

The form of K in (19) was also considered in [9], [3]. However, in those cases the objective function was concave. In our case if $\gamma > 1$ the objective function is not necessarily concave and proving this claim needs further treatment (see [10]).

Step 4: We show that there exists Lagrange multipliers (λ^*, γ^*) such that $U(\lambda^*, \gamma^*) = C_1(P, \phi(P))$, which completes the proof of the converse.

Lemma 4: There exists $\lambda^*, \gamma^* \geq 0$ such that

$$U(\lambda^*, \gamma^*) = C_1(P, \phi(P)).$$

where $C_1(P, \phi(P))$ is given in (4).

Proof: Consider the optimization problem over (x,ϕ) which gives $U(\lambda,\gamma)$ in (16). Note that $g(\gamma, x, \phi)$ given by (17) is neither concave or convex in (x,ϕ) for $\gamma > 1$. Let

$$U(\gamma) := U(\lambda^*(\gamma), \gamma) = \min_{\lambda \geq 0} U(\lambda, \gamma). \tag{21}$$

where $\lambda^*(\gamma)$ is the minimizer corresponding to γ. we use the following lemma to find $U(\gamma)$.

Lemma 5: $g(\gamma, x, \phi)$ is concave in ϕ for fixed $x, \gamma \geq 0$.

Proof: See [10].

By concavity of $g(\gamma, x, \phi)$ in ϕ for a fixed x, the inner maximum in (16) happens at $0 < \phi^*(\gamma, x) < N$ such that

$$\frac{\partial g(\gamma, x, \phi)}{\partial \phi} = 0$$
$$\Leftrightarrow \quad \frac{(1-\gamma)(N-1)}{1 + \ Nx\phi^*} = \frac{\gamma(2\phi^* - N)}{1 + x\phi^*(N - \phi^*)} \tag{22}$$

or at the boundaries $\phi^*(\gamma, x) \in \{0, N\}$. Therefore,

$$U(\gamma) = \min_{\lambda \geq 0} \max_{x \geq 0} \max_{0 \leq \phi \leq N} g(\gamma, x, \phi) + \lambda N(P - x)$$
$$= \min_{\lambda \geq 0} \max_{x \geq 0} g(\gamma, x, \phi^*(\gamma, x)) + \lambda N(P - x). \tag{23}$$

for any $\gamma \geq 0$. To evaluate the last expression we use the following lemma.

Lemma 6: Let $\gamma, x \geq 0$ and $\phi^*(\gamma, x) > 0$ be the positive solution to (22). Then, $g(\gamma, x, \phi^*(\gamma, x))$ is increasing and concave in x.

Proof: See [10].

Remark 2: As pointed out earlier, for $\gamma > 1$, $g(\gamma, x, \phi)$ is not concave in both x, ϕ in general. However, this lemma shows that $g(\gamma, x, \phi^*(\gamma, x))$ is concave in x for all $\gamma > 1$ and this is sufficient for the rest of the proof.

By concavity of $g(\gamma, x, \phi^*(\gamma, x))$ and Slater's condition [11] we have strong duality as follows.

$$\min_{\lambda \geq 0} \max_{x} g(\gamma, x, \phi^*(\gamma, x)) + \lambda N(P - x)$$
$$= \max_{x \leq P} g(\gamma, x, \phi^*(\gamma, x)) = g(\gamma, P, \phi^*(\gamma, P)) \tag{24}$$

where the last equality follows from the fact that $g(\gamma, x, \phi^*(\gamma, x))$ is increasing in x (see Lemma 6). Combining (23) and (24) we have

$$U(\gamma) = g(\gamma, P, \phi^*(\gamma, P)). \tag{25}$$

Lastly, we find γ^* such that $U(\gamma^*) = C_1(P, \phi(P))$.

Lemma 7: For a fixed $x \geq 0$, the equation $C_1(x,\phi) - C_2(x,\phi) = 0$ has a unique solution $1 \leq \phi(x) \leq N$. Moreover,

$$1 + \frac{(2\phi(x) - N)(1 + Nx\phi(x))}{(N-1)(1 + x\phi(x)(N - \phi(x)))} > 0. \tag{26}$$

Proof: See [10].

Let $\phi(P) \in [1, N]$ be the unique solution for $C_1(P,\phi) = C_2(P,\phi)$. Given P and $\phi(P)$, we pick $\gamma^*(P, \phi(P))$ such that it satisfies (22) for $x = P$ and $\phi^* = \phi(P)$. It is easy to check that $\gamma^* := \gamma^*(P, \phi(P)) > 0$ is greater than zero by plugging $x = P$ and $\phi(x) = \phi(P)$ in (26). Since γ^*, P and $\phi(P)$ satisfy (22) we conclude that $\phi(P)$ is equal to $\phi^*(\gamma^*, P)$, the positive solution of (22). Plugging $\gamma^* > 0$ and $\phi^*(\gamma^*, P)$ into (25) we have

$$U(\gamma^*) = g(\gamma^*, P, \phi^*(\gamma^*, P))$$
$$= (1 - \gamma^*) C_1(P, \phi^*(\gamma^*, P)) + \gamma^* C_2(P, \phi^*(\gamma^*, P))$$
$$= (1 - \gamma^*) C_1(P, \phi(P)) + \gamma^* C_2(P, \phi(P)) \tag{27}$$
$$= C_1(P, \phi(P)) \tag{28}$$

where (27) and (28) follow from $\phi^*(\gamma^*, P) = \phi(P)$ and $C_1(P,\phi) = C_2(P,\phi)$, respectively. Hence, $U(\lambda^*(\gamma^*), \gamma^*) = U(\gamma^*) = C_1(P, \phi(P))$. ■

Combining the four steps we have $R \leq C_1(P, \phi(P))$, and the proof of the converse is complete.

IV. Discussion

It is still unknown whether the linear sum capacity $C_{\mathsf{L}}(P)$ is in general equal to the sum capacity $C(P)$. However, there exists [3] a threshold P_c, which depends on the number of users N, such that for $P \geq P_c$, a linear code can achieve the sum capacity $C(P)$. More specifically,

$$C_{\mathsf{L}}(P) = C(P), \quad P \geq P_c$$

where $P_c \geq 0$ is the unique solution to

$$(1 + N^2 P/2)^{N-1} = (1 + N^2 P/4)^N. \tag{29}$$

The condition (29) corresponds to the case for which the well-known cut-set upper bound [9] on the sum capacity,

$$C(P) \leq \max_{\phi} \min \Big\{ C_1(P,\phi), C_2(P,\phi) \Big\}. \tag{30}$$

meets with linear sum capacity $C_{\mathsf{L}}(P)$, where functions $C_1(P,\phi), C_2(P,\phi)$ are same as in (18).

For $P < P_c$, we conjecture that we still have $C(P) = C_{\mathsf{L}}(P)$ based on the properties of Hirschfeld–Gebelein–Rényi maximal correlation [7]. In the following we provide some insights. Let $\rho^*(\Theta_1, \Theta_2)$ denote the maximal correlation between two random variables Θ_1 and Θ_2, that is,

$$\rho^*(\Theta_1, \Theta_2) = \sup_{g_1, g_2} \mathsf{E}\left(g_1(\Theta_1) g_2(\Theta_2)\right) \tag{31}$$

where the supremum is over all g_1, g_2 such that $\mathsf{E}(g_1) = \mathsf{E}(g_2) = 0, \mathsf{E}(g_1^2) = \mathsf{E}(g_2^2) = 1$. We extend this notion of maximal correlation to *conditional maximal correlation* as follows. Let

$$\rho^*(\Theta_1, \Theta_2|Y) = \sup_{g_1,g_2} \mathsf{E}\left(g_1(\Theta_1, Y) g_2(\Theta_2, Y)\right) \quad (32)$$

where g_1, g_2 satisfy $\mathsf{E}(g_1|Y) = \mathsf{E}(g_2|Y) = 0, \mathsf{E}(g_1^2) = \mathsf{E}(g_2^2) = 1$ be the conditional maximal correlation between Θ_1 and Θ_2 given a common random variable Y. The assumption $\mathsf{E}(g_1|Y) = \mathsf{E}(g_2|Y) = 0$ is crucial; otherwise, g_1 and g_2 can be picked as Y and $\rho^*(\Theta_1, \Theta_2|Y) = 1$ trivially.

For simplicity, consider $N = 2$ and equal per-symbol power constraint $\mathsf{E}(X_{ji}^2) \leq P, j = 1, 2$. Then,

$$\begin{aligned} R &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i|Y^{i-1}) \\ &= \frac{1}{n}\sum_{i=1}^{n} I(\tilde{X}_{1i}, \tilde{X}_{2i}; \tilde{Y}_i|Y^{i-1}) \quad (33) \end{aligned}$$

where $\tilde{X}_{ji} = X_{ji} - \mathsf{E}(X_{ji}|Y^{i-1})$, $j = 1, 2$, and $\tilde{Y}_i = \tilde{X}_{1i} + \tilde{X}_{2i} + Z_i$. The equality (33) holds since $\mathsf{E}(X_{ji}|Y^{i-1})$ is a funcion of Y^{i-1}. On the other hand, it can be easily seen that $\mathsf{E}(\tilde{X}_{ji}^2) \leq \mathsf{E}(X_{ji}^2) \leq P$. Therefore, there is no loss of optimality in considering X_{ji} such that $\mathsf{E}(X_{ji}|Y^{i-1}) = 0$. Under this assumption, consider

$$\begin{aligned} R &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i|Y^{i-1}) \\ &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) \\ &\leq \frac{1}{2n}\sum_{i=1}^{n} \log\Big(1 + 2P + 2P\,\mathsf{E}\left(\rho(X_{1i}X_{2i})\right)\Big) \quad (34) \\ &\leq \frac{1}{2n}\sum_{i=1}^{n} \log\Big(1 + 2P\Big(1 + \rho^*\big(\Theta_{1i}, \Theta_{2i}\big|Y^{i-1}\big)\Big)\Big) \quad (35) \end{aligned}$$

where $\rho(X_{1i}, X_{2i}) = \mathsf{E}\left(\frac{X_{1i}}{\sqrt{\mathsf{E}(X_{1i}^2)}} \cdot \frac{X_{2i}}{\sqrt{\mathsf{E}(X_{2i}^2)}}\right)$ is the correlation coefficient between X_{1i} and X_{2i} (Recall by assumption $\mathsf{E}(X_{ji}|Y^{i-1}) = 0$). Inequality (34) follows from the maximum entropy theorem [8] and equal per-symbol power constraint $\mathsf{E}(X_{ji}^2) \leq P$. Inequality (35) follows from the definition of conditional maximal correlation (32). The following lemma provides a useful property of conditional maximal correlation.

Lemma 8: If (Θ_1, Θ_2, Y) are jointly Gaussian, then

$$\rho^*(\Theta_1, \Theta_2|Y) = \rho(\Theta_1, \Theta_2|Y)$$

and linear functions $g_1^{\mathsf{L}}, g_2^{\mathsf{L}}$ of the form

$$\begin{aligned} g_1^{\mathsf{L}}(\Theta_1, Y) &= \frac{\Theta_1 - \mathsf{E}(\Theta_1|Y)}{\sqrt{\mathsf{E}\left((\Theta_1 - \mathsf{E}(\Theta_1|Y))^2\right)}} \\ g_2^{\mathsf{L}}(\Theta_2, Y) &= \frac{\Theta_2 - \mathsf{E}(\Theta_2|Y)}{\sqrt{\mathsf{E}\left((\Theta_2 - \mathsf{E}(\Theta_2|Y))^2\right)}} \quad (36) \end{aligned}$$

attain the supremum in $\rho^*(\Theta_1, \Theta_2|Y)$.

Proof: The proof follows from the fact [12] that $\rho^*(U, V) = \rho(U, V)$ for jointly Gaussian random variables (U, V) and that given $Y = y$, $(\Theta_1, \Theta_2)\big|_{Y=y}$ is Gaussian with some correlation $\rho(\Theta_1, \Theta_2\big|Y = y) = \rho$ independent of y. For details see [10]. ■

Let (Θ_1, Θ_2) be Gaussian and $\rho^*(\Theta_1, \Theta_2|Y^{i-1})$ be defined similar to (32) with a common collection of random variables Y^{i-1}. By Lemma 8 we know that if $(\Theta_1, \Theta_2, Y^{i-1})$ are jointly Gaussian, $\rho^*(\Theta_1, \Theta_2|Y^{i-1}) = \rho(\Theta_1, \Theta_2|Y^{i-1})$ and $X_{ji} = \mathsf{L}_{ji}(\Theta_j, Y^{i-1})$, where L_{ji} is of the form (36), achieve the upper bound on $I(X_{1i}, X_{2i}; Y_i|Y^{i-1})$ (see (35)). This implies that linear functions are greedy optimal for maximizing the right hand side of (33) since using linear functions up to time $i-1$, results in Gaussian $(\Theta_1, \Theta_2, Y^{i-1})$.

A similar argument holds for any number of senders N, where we have the the following upper bound,

$$\begin{aligned} R &\leq \frac{1}{n}\sum_{i=1}^{n} I(X(\mathcal{S}); Y_i|Y^{i-1}) \\ &\leq \frac{1}{2n}\sum_{i=1}^{n} \log\Big(1 + NP + P\sum_{j\neq k}\rho^*\big(\Theta_{ji}, \Theta_{ki}\big|Y^{i-1}\big)\Big). \end{aligned}$$

However, it is not clear whether linear functions are in general optimal. This is because the distribution of $(\Theta_1, \Theta_2, Y^{i-1})$ depends on all the previous functions up to time $i-1$. For instance using nonlinear functions at time 1, although hurting the current mutual information term $I(X(\mathcal{S}); Y_1)$, might be advantageous for the future terms $I(X(\mathcal{S}); Y_i|Y^{i-1}), i > 1$.

References

[1] L. H. Ozarow, "The capacity of the white gaussian multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, pp. 623–629, 1984.

[2] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–i: No bandlimited constraint," *IEEE Trans. Inf. Theory*, vol. 12, pp. 172–182, 1966.

[3] G. Kramer, "Feedback strategies for white gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1423–1438, 2002.

[4] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 35, pp. 37–43, 1989.

[5] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, pp. 44–53, 1989.

[6] G. Kramer and M. Gastpar, "Dependece balance and the gaussian multiaccess channel with feedback," in *IEEE Information Theory Workshop*, 2006, pp. 198–202.

[7] A. Réyni, "On measures of dependence," *Acta Mathematica Hungarica*, vol. 10, pp. 441–451, 1959.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York: Wiley, 2006.

[9] J. A. Thomas, "Feedback can at most double gaussain multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, pp. 711–716, 1987.

[10] E. Ardestanizadeh, M. A. Wigger, Y.-H. Kim, and T. Javidi, "Linear sum capacity for Gaussian multiple access channel with feedback," Feb 2010, arXiv:1002.1781.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[12] H. O. Lancaster, "Some properties of the bivariate normal distribution considered in the form of a contingency table," *Biometrika*, vol. 44, pp. 289–292, 1957.