Sum Capacity of One-Sided Parallel Gaussian Interference Channels

Chi Wan Sung, Kenneth W. K. Lui, Kenneth W. Shum, and H. C. So

***Abstract*—The sum capacity of the one-sided parallel Gaussian interference channel is shown to be a concave function of user powers. Exploiting the inherent structure of the problem, we construct a numerical algorithm to compute it. Two suboptimal schemes are compared with the capacity-achieving scheme. One of the suboptimal schemes, namely iterative waterfilling, yields close-to-capacity performance when the cross link gain is small.**

***Index Terms*—Gaussian interference channels, iterative waterfilling, sum capacity.**

I. Introduction

The capacity region of the interference channel (IC) has long been an open problem. The largest achievable rate region known today was given by Han and Kobayashi in 1981 [1]. Some special classes of IC have been defined in the literature. One of them is called *one-sided IC* in which one user interferes with the other but *not* vice versa. In terms of capacity region, it was proven to be equivalent to the so-called degraded IC by Costa [2], and its sum capacity was derived by Sason [3]. Nevertheless, its capacity region is still unknown. Some recent results can be found in [4]. In this correspondence, we focus on the parallel one-sided Gaussian IC. It can be used to model a frequency-selective environment, where each channel represents one frequency carrier.

Our contribution in this work is twofold. First, we devise an algorithm to compute the sum capacity of this channel based on the idea of alternating optimization. Second, we evaluate the performances of two sub-optimal transmission schemes. One is a power allocation method called *iterative waterfilling* (IW), together with the suboptimal coding scheme by treating the signal of the other user as Gaussian noise [5]. The other one is simply allocating power uniformly across all parallel channels, assuming that optimal coding is used.

II. Channel Model and Problem Formulation

Consider L independent one-sided parallel Gaussian IC's between two users in standard form. For $l = 1, 2, \ldots, L$, the channel inputs and outputs are related by the following linear relations:

$$Y^{(l)}{}_1 = X^{(l)}{}_1 + Z^{(l)}{}_1 \quad (1)$$
$$Y^{(l)}{}_2 = \sqrt{a^{(l)}} X^{(l)}{}_1 + X^{(l)}{}_2 + Z^{(l)}{}_2 \quad (2)$$

where $Z^{(l)}{}_1, Z^{(l)}{}_2$ are Gaussian noises with zero mean and unit variance. The power constraints $(i = 1, 2)$ are

$$\sum_{l=1}^{L} w^{(l)}{}_i p^{(l)}{}_i \le P_i \quad (3)$$

Manuscript received April 26, 2007; revised September 30, 2007. This work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 119806).

The authors are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: albert.sung@cityu.edu.hk; 50469990@student.cityu.edu.hk; kshum2008@gmail.com; hcso@ee.cityu.edu.hk).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2007.911178

where $p^{(l)}{}_i$ is the average power constraint for $X^{(l)}{}_i$. For $i = 1, 2$, denote the feasible region of $\boldsymbol{p}_i \triangleq (p_i^{(1)}, \ldots, p_i^{(L)})$ by

$$\mathcal{P}_i \triangleq \left\{ \boldsymbol{p}_i : \sum_{l=1}^{L} w^{(l)}{}_i p^{(l)}{}_i \le P_i,\ p^{(l)}{}_i \ge 0\ \forall l \right\}. \quad (4)$$

The Cartesian product of $\mathcal{P}_1$ and $\mathcal{P}_2$ is denoted by $\mathcal{P}$, and the *system power allocation* $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ by $\boldsymbol{p}$.

The sum capacity of a one-sided Gaussian IC was given in [3]. Based on that result, we will show in the next section that the sum capacity of a parallel one-sided Gaussian IC can be found by solving the following optimization problem:

$$\max_{\boldsymbol{p} \in \mathcal{P}} \left(f(\boldsymbol{p}_1, \boldsymbol{p}_2) \triangleq \kappa \sum_{l=1}^{L} C_{a^{(l)}} \left(p^{(l)}{}_1, p^{(l)}{}_2 \right) \right). \quad (5)$$

A normalization constant $\kappa \triangleq 2 \ln 2$ is introduced for the sake of simplifying notations. The function $C_a(p_1, p_2)$ is the sum capacity of a single one-sided Gaussian IC given by ([3, Th. 2]

$$\begin{cases} \gamma(p_1) + \gamma\left(\frac{p_2}{1+ap_1}\right), & \text{if } 0 \le a \le 1 \\ \gamma(ap_1 + p_2), & \text{if } 1 \le a \le 1 + p_2 \\ \gamma(p_1) + \gamma(p_2), & \text{if } a \ge 1 + p_2 \end{cases} \quad (6)$$

where $\gamma(x)$ is the function $\frac{1}{2}\log_2(1 + x)$. Note that C_a is a smooth function of p_1 and p_2 when $a \le 1$. It is not differentiable at $p_2 = a - 1$ when $a > 1$.

We say that channel l is under *weak* interference if $a^{(l)}$ is less than or equal to one. Otherwise, it is under *strong* interference. When under strong interference, a channel may operate in two modes. We say that channel l operates in *interfering* mode if $p^{(l)}{}_1 > 0$ and $p^{(l)}{}_2 > a^{(l)} - 1$, or in *noninterfering* mode otherwise.

III. Characterization of the Sum Capacity

In this section, we verify that the sum capacity can be obtained by solving a concave optimization problem, and present some properties of the solution.

Lemma 1: The function $C_a(p_1, p_2)$ is strictly concave on $\mathbb{R}_+^2$ for any $a \ge 0$.

Proof: For $0 \le a \le 1$, let $H_a(p_1, p_2)$ be $-\kappa C_a(p_1, p_2)$. We are going to prove that H_a is strictly convex for $0 \le a \le 1$. Let $\boldsymbol{H}$ be its Hessian matrix whose entries are given by $h_{ij} \triangleq \partial^2 H_a / \partial p_i \partial p_j$. We have

$$h_{11} = \frac{1}{(1+p_1)^2} - \frac{a^2}{(1+ap_1)^2} + \frac{a^2}{(1+ap_1+p_2)^2} \quad (7)$$
$$h_{22} = \frac{1}{(1+ap_1+p_2)^2} = \frac{h_{12}}{a}. \quad (8)$$

The strict convexity of H_a follows from $h_{11}, h_{22} > 0$ and $h_{11}h_{22} > h_{12}^2$, which can be verified as follows:

$$\begin{aligned} h_{11}h_{22} - h_{12}^2 &= h_{22}(h_{11} - ah_{12}) \\ &= h_{22}\left[\frac{1}{(1+p_1)^2} - \frac{a^2}{(1+ap_1)^2}\right] \end{aligned} \quad (9)$$

Since $0 \le a \le 1$ and the function $\beta(x) = \frac{x}{1+xp_1}$ is monotonic increasing for $p_1 > 0$, then $h_{11}h_{22} - h_{12}^2 > 0$. Finally, the positivity of h_{11} follows again since the function β is positive and $0 \le a \le 1$.

Next we consider the second case where $a \geq 1$. It is easy to check that $C_a(p_1, p_2)$ is equal to the minimum of $\gamma(ap_1 + p_2)$ and $\gamma(p_1) + \gamma(p_2)$. Since both functions within the minimum operator are strictly concave, their pointwise minimum is also strictly concave. □

Theorem 2: The function f in (5) is strictly concave on $\mathbb{R}_+^2$, and there exists a unique $\boldsymbol{p}^* \in \mathcal{P}$ that maximizes f.

Proof: Lemma 1 implies immediately that f is also strictly concave. Since the domain $\mathcal{P}$ is compact, there exists a unique $\boldsymbol{p}^* \in \mathcal{P}$ that maximizes f. □

Theorem 3: The sum capacity of a parallel one-sided IC is given by

$$C_{\text{sum}} \triangleq \max_{\boldsymbol{p} \in \mathcal{P}} \sum_{l=1}^{L} C_{a^{(l)}} \left({p^{(l)}}_1, {p^{(l)}}_2 \right). \tag{10}$$

The maximum in (10) is well defined by Theorem 2. The proof of Theorem 3 will be given in the Appendix. The theorem says that the capacity of a parallel one-sided Gaussian IC can be achieved by allocating powers over the subchannels and coding independently over each subchannel. In other words, coding jointly across the subchannels is not needed and the sum capacity can be achieved by solving a power allocation problem.

Consider the Lagrangian

$$\begin{aligned} \mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2, \lambda_1, \lambda_2) \triangleq f(\boldsymbol{p}_1, \boldsymbol{p}_2) + \lambda_1 \left(P_1 - \sum_{l=1}^{L} w_1^{(l)} {p^{(l)}}_1 \right) \\ + \lambda_2 \left(P_2 - \sum_{l=1}^{L} w_2^{(l)} {p^{(l)}}_2 \right). \end{aligned} \tag{11}$$

A point $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*, \lambda_1^*, \lambda_2^*)$ is said to be its *saddle point* if

$$\mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2, \lambda_1^*, \lambda_2^*) \leq \mathfrak{L}(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*, \lambda_1^*, \lambda_2^*) \leq \mathfrak{L}(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*, \lambda_1, \lambda_2) \tag{12}$$

for all $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, and for all $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ with nonnegative components. The following proposition will be useful ([6, Th. 1.B.5].

Proposition 4 (Saddle-Point Condition): Provided that there exists a power allocation $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ satisfying

$$\sum_{l=1}^{L} w_i^{(l)} {p^{(l)}}_i < P_i \tag{13}$$

for $i = 1, 2$, then $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ is the optimal solution to the maximization problem in (10) if and only if we can find λ_1^* and λ_2^* such that $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*, \lambda_1^*, \lambda_2^*)$ is a saddle point of $\mathfrak{L}$.

It is easy to see that the condition in (13), known as the Slater condition, holds by picking ${p^{(l)}}_i = P_i/(2{w^{(l)}}_i L)$ for $i = 1, 2$ and $l = 1, \ldots, L$.

Theorem 5: At the optimal solution to the maximization problem in (10), at most one of the channels under strong interference operates in interfering mode, provided that $a^{(l)} {w^{(l)}}_2 / {w^{(l)}}_1$'s are distinct for all those channels.

Proof: Let $\boldsymbol{p}^* = (\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ denote the optimal solution. Denote the set of strong-interference channels operating in interfering mode at $\boldsymbol{p}^*$ by $\mathcal{S}$. By definition, ${p^{(l)}}_1 > 0$ and ${p^{(l)}}_2 > a^{(l)} - 1 > 0$ for all $l \in \mathcal{S}$. The partial derivatives of the Lagrangian with respect to ${p^{(l)}}_1$ and ${p^{(l)}}_2$ exist. Suppose that λ_1^* and λ_2^* are the Lagrange multipliers that satisfy the saddle point condition in Prop. 4. Since $\mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2, \lambda_1^*, \lambda_2^*)$ is maximized at $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$, these two partial derivatives equal zero for $l \in \mathcal{S}$. Thus

$$\frac{a^{(l)}}{a^{(l)} p_1^{(l)*} + p_2^{(l)*} + 1} = {w^{(l)}}_1 \lambda_1^* \tag{14}$$

$$\frac{1}{a^{(l)} p_1^{(l)*} + p_2^{(l)*} + 1} = {w^{(l)}}_2 \lambda_2^*. \tag{15}$$

Therefore, both λ_1^* and λ_2^* are nonzero, and $\frac{\lambda_1^*}{\lambda_2^*} = \frac{a^{(l)} w_2^{(l)}}{{w^{(l)}}_1}$, for all $l \in \mathcal{S}$. Hence, $|\mathcal{S}|$ must be strictly less than two if $a^{(l)} {w^{(l)}}_2 / {w^{(l)}}_1$ are distinct for all $l = 1, 2, \ldots, L$. □

IV. Computation of the Sum Capacity

The idea of our proposed method is to solve two subproblems alternately. We will prove that this method yields the optimal solution after convergence.

A. First Subproblem

In the first subproblem, the power allocation of sender 2 is fixed at $\hat{\mathbf{p}}_2$, and we want to maximize $f(\boldsymbol{p}_1, \hat{\mathbf{p}}_2)$ over all feasible $\boldsymbol{p}_1$. The Lagrangian $\mathfrak{L}(\boldsymbol{p}_1, \hat{\mathbf{p}}_2, \lambda_1, \hat{\lambda}_2)$ can be written as

$$\sum_{l=1}^{L} \left[\kappa C_{a^{(l)}} \left({p^{(l)}}_1, \hat{p}_2^{(l)} \right) - \lambda_1 {w^{(l)}}_1 {p^{(l)}}_1 \right] + K_1 \tag{16}$$

where the constant K_1 does not involve $\boldsymbol{p}_1$.

For fixed $\hat{\mathbf{p}}_2, \lambda_1$ and $\hat{\lambda}_2$, we can find the $\boldsymbol{p}_1$ that maximizes $\mathfrak{L}(\boldsymbol{p}_1 \hat{\mathbf{p}}_2, \lambda_1, \hat{\lambda}_2)$ by dual decomposition. Each summand in (16) is a strictly concave function in ${p^{(l)}}_1$ with domain $\mathbb{R}_+$. We set, for $l = 1, 2, \ldots, L$

$$p_1^{(l)} = \arg\max_{p \geq 0} \left(\kappa C_{a^{(l)}} \left(p, \hat{p}_2^{(l)} \right) - \lambda_1 w_1 p \right). \tag{17}$$

Let ${\beta^{(l)}}_1(\lambda_1; \hat{\mathbf{p}}_2)$ denote the zero to the equation $\partial \mathfrak{L}_1 / \partial {p^{(l)}}_1 = 0$. We can compute ${\beta^{(l)}}_1(\lambda_1; \hat{\mathbf{p}}_2)$ by solving for ${p^{(l)}}_1$ in

$$\begin{aligned} 0 = \frac{\partial \mathfrak{L}}{\partial {p^{(l)}}_1} = \frac{1 - a^{(l)}}{(1 + {p^{(l)}}_1)(1 + a^{(l)} {p^{(l)}}_1)} \\ + \frac{a^{(l)}}{1 + a^{(l)} {p^{(l)}}_1 + \hat{p}_2^{(l)}} - \lambda_1 {w^{(l)}}_1 \end{aligned} \tag{18}$$

when $0 \leq a^{(l)} \leq 1$, or

$$0 = \frac{\partial \mathfrak{L}}{\partial {p^{(l)}}_1} = \frac{a^{(l)}}{a^{(l)} {p^{(l)}}_1 + \max\{a^{(l)}, \hat{p}_2^{(l)} + 1\}} - \lambda_1 {w^{(l)}}_1 \tag{19}$$

when $a^{(l)} \geq 1$. Note the ${\beta^{(l)}}_1(\lambda_1; \hat{\mathbf{p}}_2)$ may be negative. Define x^+ as $\max(x, 0)$. The power allocation ${p^{(l)}}_1 = {\beta^{(l)}}_1(\lambda_1; \hat{\mathbf{p}}_2)^+$ for $l = 1, 2, \ldots, L$, satisfies the Karush-Kuhn-Tucker condition for the optimality

$$\frac{\partial \mathfrak{L}}{\partial {p^{(l)}}_1} \begin{cases} = 0, & \text{if } {p^{(l)}}_1 > 0 \\ \leq 0, & \text{if } {p^{(l)}}_1 = 0. \end{cases} \tag{20}$$

It can be seen from (18) and (19) that ${\beta^{(l)}}_1(\lambda_1; \hat{\mathbf{p}}_2)$ is a monotonically decreasing function of λ_1. We can adjust λ_1 such that the power constraint is met. The optimal λ_1 can be found by a simple binary search.

B. Second Subproblem

Given $\hat{\mathbf{p}}_1$, we want to find the optimal $\boldsymbol{p}_2$. The Lagrangian $\mathfrak{L}(\hat{\mathbf{p}}_1, \boldsymbol{p}_2, \hat{\lambda}_1, \lambda_2)$ becomes

$$\sum_{l=1}^{L} \left[\kappa C_{a^{(l)}} (\hat{\mathbf{p}}_1, \boldsymbol{p}_2) - \lambda_1 {w^{(l)}}_2 {p^{(l)}}_2 \right] + K_2 \tag{21}$$

where the constant K_2 is independent of $\boldsymbol{p}_2$. Each summand in the above summation is a strictly concave function of ${p^{(l)}}_2$. We can thus find the optimal

$$p_2^{(l)} = \arg\max_{p\geq 0}\left(\kappa C_{a^{(l)}}\left(\hat{p}_1^{(l)}, p\right) - \lambda_2 w_2 p\right) \tag{22}$$

for $l = 1, 2, \ldots, L$. The computation is slightly more complicated than the first subproblem, because the function to be maximized now is not differentiable at ${p^{(l)}}_2 = a^{(l)} - 1$, when $a^{(l)} > 1$. When $0 \leq a^{(l)} \leq 1 + {p^{(l)}}_2$,

$$\frac{\partial C_{a^{(l)}}}{\partial {p^{(l)}}_2} = \frac{1}{1 + a^{(l)}{p^{(l)}}_1 + {p^{(l)}}_2} \tag{23}$$

and when $a^{(l)} > 1 + {p^{(l)}}_2$

$$\frac{\partial C_{a^{(l)}}}{\partial {p^{(l)}}_2} = \frac{1}{1 + {p^{(l)}}_2}. \tag{24}$$

When ${p^{(l)}}_2 = a^{(l)} - 1, C_{a^{(l)}}$ is not differentiable with respect to ${p^{(l)}}_2$. Let $\partial C_{a^{(l)}}$ be the subdifferential of $C_{a^{(l)}}$ at ${p^{(l)}}_2$. When ${p^{(l)}}_2 = a^{(l)} - 1$

$$\partial C_{a^{(l)}} = \left[\frac{1}{1 + a^{(l)}{p^{(l)}}_1 + {p^{(l)}}_2}, \frac{1}{1 + {p^{(l)}}_2}\right]. \tag{25}$$

In summary, when $a^{(l)} \neq 1 + {p^{(l)}}_2$, the subdifferential $\partial C_{a^{(l)}}$ is a singleton consisting of the number in either (23) or (24); when $a^{(l)} = 1 + {p^{(l)}}_2$, $\partial C_{a^{(l)}}$ is the interval in (25).

Given λ_2, the optimal ${p^{(l)}}_2$ should satisfy $\lambda_2 {w^{(l)}}_2 \in \partial C_{a^{(l)}}$ for ${p^{(l)}}_2 > 0$, and

$$\lambda_2 {w^{(l)}}_2 \geq \begin{cases} \left(1 + a^{(l)}{p^{(l)}}_1 + {p^{(l)}}_2\right)^{-1}, & \text{if } a^{(l)} \leq 1 \\ \left(1 + {p^{(l)}}_2\right)^{-1}, & \text{if } a^{(l)} > 1 \end{cases} \tag{26}$$

for ${p^{(l)}}_2 = 0$, and $l = 1, 2, \ldots, L$. (see ([7, Sec. 28])

Consider the case where $a^{(l)} \leq 1$. The optimal ${p^{(l)}}_2$ is simply given by

$${p^{(l)}}_2 = \left[\left(\lambda_2 {w^{(l)}}_2\right)^{-1} - 1 - a^{(l)}{p^{(l)}}_1\right]^+. \tag{27}$$

Next consider the case where $a^{(l)} > 1$. Define

$$\alpha \triangleq \left[(\lambda_2 {w^{(l)}}_2)^{-1} - 1\right]^+ \tag{28}$$

and

$$\beta \triangleq \left(\lambda_2 {w^{(l)}}_2\right)^{-1} - 1 - a^{(l)}{p^{(l)}}_1. \tag{29}$$

Then the optimal ${p^{(l)}}_2$ is given by

$${p^{(l)}}_2 = \begin{cases} \alpha, & \text{if } \alpha \leq a^{(l)} - 1 \\ \beta, & \text{if } \beta > a^{(l)} - 1 \\ a^{(l)} - 1, & \text{otherwise.} \end{cases} \tag{30}$$

Notice that given any $a^{(l)}$, the optimal ${p^{(l)}}_2$ is a continuous, decreasing function of λ_2. Hence, we can use binary search to find the optimal λ_2 such that the power constraint is met.

C. *Alternating Optimization*

Solving the above two subproblems alternatively, we have the following algorithm.

1) Initially, let n be zero and $\boldsymbol{p}^{(0)}$ be the zero vector.
2) Update the power allocation as follows:

$$\boldsymbol{p}_1^{(n+1)} \triangleq \arg\max_{\boldsymbol{p}_1\in\mathcal{P}_1} f\left(\boldsymbol{p}_1, \boldsymbol{p}_2^{(n)}\right) \tag{31}$$

$$\boldsymbol{p}_2^{(n+1)} \triangleq \arg\max_{\boldsymbol{p}_2\in\mathcal{P}_2} f\left(\boldsymbol{p}_1^{(n+1)}, \boldsymbol{p}_2\right). \tag{32}$$

3) Increase n by one and repeat step 2).

Since f is a strictly concave function, given the power allocation of a user, there is a unique power allocation for the other user to maximize f. Hence, the algorithm is well defined. Note that $f(\boldsymbol{p}^{(n+1)}) \geq f(\boldsymbol{p}^{(n)})$ with equality holds only if $\boldsymbol{p}^{(n+1)} = \boldsymbol{p}^{(n)}$.

We denote the iterative function in (31) and (32) by I_1 and I_2 respectively. The overall iteration $(I_1(\boldsymbol{p}_2), I_2(I_1(\boldsymbol{p}_1))$ is denoted by I. A vector $\boldsymbol{p}$ is said to be a fixed point of I if $\boldsymbol{p} = I(\boldsymbol{p})$. The proof of the following result is in the Appendix.

Theorem 6: Given the iterative function I and any initial point $\boldsymbol{p}^{(0)}$, the sequence $\{\boldsymbol{p}^{(n)}\}$ converges to the unique point that maximizes f in $\mathcal{P}$.

V. Comparison With Suboptimal Schemes

We consider two suboptimal schemes. In the first scheme, each user employs optimal code for transmission. The sum rate is also given by $f(\boldsymbol{p}_1, \boldsymbol{p}_2)$. However, both users allocate equal power to each of their L channels. In other words ${p^{(l)}}_i = P_i/L$ for $i = 1, 2$ and $l = 1, 2, \ldots, L$. In the second scheme, suboptimal code is used. Interference from the other user is treated as additive Gaussian noise, and the codebook appropriate for the Gaussian channel is used. Under this scheme, the sum rate is given by

$$\kappa \sum_{l=1}^{L}\left[\gamma(p_1) + \gamma\left(\frac{p_2}{1 + a^{(l)}p_1}\right)\right]. \tag{33}$$

Note that the two terms in the above expression correspond to the rate of user 1 and user 2, respectively. Under the iterative waterfilling method with sequential update, each user takes turns to maximize his own rate, treating the other user's signal as Gaussian noise. This method, when applied to the one-sided parallel Gaussian IC, will always stop after one round, since user 2's transmission does not induce any interference to user 1. More precisely, user 1 first allocates his power by waterfilling method. Then user 2 will waterfill his power, treating user 1's signal as noise. The waterfilling method can be found, for example, in [5].

Fig. 1 shows the performance of the optimal scheme and the two suboptimal schemes. In this example, there are 32 channels in the system. The cross link gains $a^{(l)}$'s are equal to a value of g_{21}, for all l. The weights, ${w^{(l)}}_i$'s, are independent random variables uniformly distributed between zero and one. The maximum powers of both users are equal to 100. We plot the sum rate against the cross link gain, g_{21}. Each data point is obtained by averaging over 1000 independent runs. It can be seen that all curves decrease monotonically when g_{21} is less than one. The reason is that in this regime, user 1's signal is treated as Gaussian noise even under optimal coding scheme. Therefore, when g_{21} increases, more interference is experienced by receiver 2, thus deteriorating system performance. When g_{21} is greater than one, interference may be canceled if optimal code is used. Therefore, only the curve for IW continues to decrease when g_{21} increases beyond one. The curves for the other two schemes have exactly the same

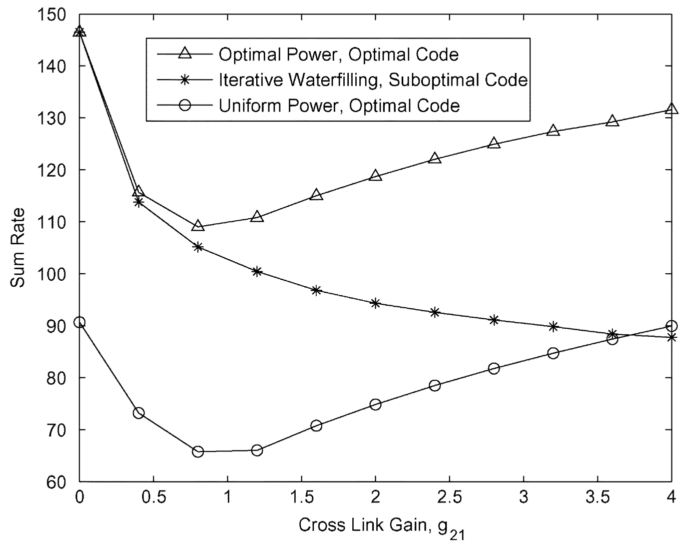

Fig. 1. Comparison of three different transmission schemes.

shape. From the figure, we can see that IWs performance is close to capacity when g_{21} is small. On the other hand, when g_{21} is much larger than one, IW performs poorly, since interference cancellation is not employed.

VI. Conclusion

We have shown that the computation of sum capacity can be reduced to allocating powers over subchannels. The power allocation problem turns out to be convex. The solution can be efficiently found by alternately optimizing the power allocation of each user. The convergence of this method is guaranteed. Our result provides a fundamental limit on the capacity of the channel, which is informative for evaluation of practical, suboptimal schemes.

Appendix
Proof of Theorem 3

It is clear that $C_{\rm sum}$ can be achieved by allocating power according to the solution to the optimization problem.

Conversely, suppose that there exists a sequence of codes $\{\mathcal{C}_N : N = 1, 2, \ldots\}$, which asymptotically achieves vanishing block error probability as we let N tend to infinity. The block length of $\mathcal{C}_N$ is NL. The message of user i, W_i, is uniformly distributed over the set $\{1, 2, \ldots, 2^{NR_i}\}$, where R_i is the code rate of user i, measured in number of bits per channel use. Note that L symbols are transmitted in parallel for each channel use. An encoder maps W_i into codeword $\boldsymbol{X}_i$, which is partitioned into L equal parts, each for one subchannel: $\boldsymbol{X}_i = (\boldsymbol{X}_i^{(1)}, \boldsymbol{X}_i^{(2)}, \ldots, \boldsymbol{X}_i^{(L)})$, where $\boldsymbol{X}_i^{(l)} \triangleq (X_i^{(l)}(1), X_i^{(l)}(2), \ldots, X_i^{(l)}(N))$. Any codeword must satisfy the following power constraint:

$$\sum_{l=1}^{L} w_i^{(l)} \left[\frac{1}{N} \sum_{t=1}^{N} X_i^{(l)}(t) \right] \le P_i, \quad \text{for } i = 1, 2. \tag{34}$$

Let $\boldsymbol{Y}_i^{(l)}$ and $\boldsymbol{Z}_i^{(l)}$ be the received vector and noise vector at receiver i and channel l. Define $\boldsymbol{Y}_i$ and $\boldsymbol{Z}_i$ as the concatenation of the respective L vectors. The sum rate is bounded

$$\begin{aligned} N(R_1 + R_2) &= H(W_1, W_2) \\ &\le I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_1, \boldsymbol{Y}_2) + H(W_1, W_2 \,|\, \boldsymbol{Y}_1, \boldsymbol{Y}_2) \end{aligned} \tag{35}$$

where the inequality follows from the data processing theorem.

By Fano's inequality, we have $H(W_1, W_2|\boldsymbol{Y}_1, \boldsymbol{Y}_2) \le N\epsilon_N$, where $\epsilon_N \to 0$ as $P_e^{(N)} \to 0$.

As $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are independent with $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, we obtain

$$\begin{aligned} I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_1, \boldsymbol{Y}_2) &= h(\boldsymbol{Y}_1, \boldsymbol{Y}_2) - h(\boldsymbol{Z}_1, \boldsymbol{Z}_2) \\ &\le \sum_{l=1}^{L} \sum_{t=1}^{N} I(X_1^{(l)}(t), X_2^{(l)}(t); Y_1^{(l)}(t), Y_2^{(l)}(t)). \end{aligned} \tag{36}$$

Denote the second moment of $X_i^{(l)}(t)$ and the average by

$$\rho_i^{(l)}(t) \triangleq \mathbb{E}_{W_i} \left[\left(X_i^{(l)}(t) \right)^2 \right], \quad \bar{\rho}_i^{(l)} \triangleq \frac{1}{N} \sum_{t=1}^{N} \rho_i^{(l)}(t). \tag{37}$$

It follows that

$$\begin{aligned} &\frac{1}{N} \sum_{l=1}^{L} \sum_{t=1}^{N} I\left(X_1^{(l)}(t), X_2^{(l)}(t); Y_1^{(l)}(t), Y_2^{(l)}(t) \right) \\ &\le \frac{1}{N} \sum_{l=1}^{L} \sum_{t=1}^{N} C_{a^{(l)}} \left(\rho_1^{(l)}(t), \rho_2^{(l)}(t) \right) \\ &\le \sum_{l=1}^{L} C_{a^{(l)}} \left(\bar{\rho}_1^{(l)}, \bar{\rho}_2^{(l)} \right) \end{aligned} \tag{38}$$

where the last inequality follows from the concavity of $C_{a^{(l)}}$.

Since each codeword satisfies (34), so does their average. Therefore, $\sum_{l=1}^{L} C_{a^{(l)}}(\bar{\rho}_1^{(l)}, \bar{\rho}_2^{(l)}) \le C_{\rm sum}$. Hence, $R_1 + R_2 \le C_{\rm sum} + \epsilon_N$. The converse is completed by letting $P_e^{(N)} \to 0$.

Appendix
Proof of Theorem 6

Lemma 7: The iterative function I is a continuous mapping from $\mathcal{P}$ to $\mathcal{P}$.

Proof: The following "Maximum Theorem" ([8, p. 116]) implies that both I_1 and I_2, and hence their composition I, are continuous functions.

(*Maximum Theorem*) Let $\phi(x, y)$ be a real-valued continuous function with domain $X \times Y$, where $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ are compact sets. Suppose $\phi(x, y)$ is strictly concave in x for each y. The functions $M(y) = \max\{\phi(x, y) : x \in X\}$ and $\Phi(y) = \arg\max\{\phi(x, y) : x \in X\}$ are well defined for all $y \in Y$, and are both continuous. □

Lemma 8: Given the iterative function I and any initial point $\boldsymbol{p}^{(0)}$, there exists a convergent subsequence of $\{\boldsymbol{p}^{(n)}\}$, whose limit is the unique point $\boldsymbol{p}^*$ that maximizes f.

Proof: The proof is divided into three parts: 1) The sequence $\{\boldsymbol{p}^{(n)}\}$ contains a convergent subsequence. It follows immediately from the fact that the domain of $\boldsymbol{p}$ is compact.

2) The limit of any convergent subsequence is a fixed point of I. Suppose that $\{\boldsymbol{p}^{(\alpha_n)}\}$ is a subsequence of $\{\boldsymbol{p}^{(n)}\}$ that converges to the limit point $\tilde{\boldsymbol{p}}$. Suppose $\tilde{\boldsymbol{p}}$ is not a fixed point, that is, $I(\tilde{\boldsymbol{p}}) \ne \tilde{\boldsymbol{p}}$. Then $f(\tilde{\boldsymbol{p}}) \le f(I(\tilde{\boldsymbol{p}}))$. By the continuity of f, we can find radii $\tilde{r}$ and r' such that the spherical neighborhoods, $\mathcal{N}(\tilde{\boldsymbol{p}}; \tilde{r})$ and $\mathcal{N}(I(\tilde{\boldsymbol{p}}); r')$, are disjoint and $f(\boldsymbol{x}) \le f(\boldsymbol{y})$, for any $\boldsymbol{x} \in \mathcal{N}(\tilde{\boldsymbol{p}}; \tilde{r})$ and $\boldsymbol{y} \in \mathcal{N}(I(\tilde{\boldsymbol{p}}); r')$.

Since I is continuous by Lemma 7, $I(\boldsymbol{p}^{(\alpha_n)})$ converges to $I(\tilde{\boldsymbol{p}})$. So we can find a sufficiently large N such that $I(\boldsymbol{p}^{(\alpha_n)}) \in \mathcal{N}(I(\tilde{\boldsymbol{p}}); r')$ for all $n \ge N$. However, the sequence $\{f(\boldsymbol{p}^{(n)})\}$ is nondecreasing, thereby $f(I(\boldsymbol{p}^{(\alpha_N)})) \le f(\boldsymbol{p}^{(\alpha_{N+m})})$, for all $m \ge 1$. This yields a contradiction because $I(\boldsymbol{p}^{(\alpha_N)}) \in \mathcal{N}(I(\tilde{\boldsymbol{p}}); r')$ but there exists a sufficiently large $m_0 \geq 1$ such that $\boldsymbol{p}^{(\alpha_N + m_0)} \in \mathcal{N}(\tilde{\boldsymbol{p}}; \tilde{r})$.

3) Any fixed point of I maximizes f over $\mathcal{P}$. Let $\tilde{\boldsymbol{p}} = (\tilde{\boldsymbol{p}}_1, \tilde{\boldsymbol{p}}_2)$ be a fixed point of I. By this we mean that, $f(\tilde{\boldsymbol{p}}_1, \tilde{\boldsymbol{p}}_2)$ is the maximum of $f(\boldsymbol{p}_1, \tilde{\boldsymbol{p}}_2)$ over all feasible $\boldsymbol{p}_1$, and also the maximum of $f(\tilde{\boldsymbol{p}}_1, \boldsymbol{p}_2)$ over all feasible $\boldsymbol{p}_2$. Here, we extend the Lagrangian so that nonnegativity constraints are included. We denote the new Lagrangian also by $\mathfrak{L}$.

$$\begin{aligned}\mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2, \lambda_1, \lambda_2, \boldsymbol{\mu}) &\triangleq f(\boldsymbol{p}_1, \boldsymbol{p}_2) + \lambda_1 \left(P_1 - \sum_{l=1}^{L} w_1 {p^{(l)}}_1 \right) \\ &+ \lambda_2 \left(P_2 - \sum_{l=1}^{L} w_2 {p^{(l)}}_2 \right) + \sum_{i=1}^{2} \sum_{l=1}^{L} \mu_i^{(l)} p_i^{(l)}. \end{aligned} \tag{39}$$

where $\boldsymbol{\mu}$ is the vector $(\mu_1^{(1)}, \ldots \mu_1^{(L)}, \mu_2^{(1)}, \ldots \mu_2^{(L)})$.

The condition of being a fixed point is rephrased as follows: we can find nonnegative $\tilde{\lambda}_1$, $\tilde{\lambda}_2$ and $\tilde{\boldsymbol{\mu}}$ such that

$$\begin{aligned}\mathfrak{L}(\tilde{\boldsymbol{p}}_1, \tilde{\boldsymbol{p}}_2, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\boldsymbol{\mu}}) &= \max_{\boldsymbol{p}_1} \mathfrak{L}(\boldsymbol{p}_1, \tilde{\boldsymbol{p}}_2, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\boldsymbol{\mu}}) \\ &= \max_{\boldsymbol{p}_2} \mathfrak{L}(\tilde{\boldsymbol{p}}_1, \boldsymbol{p}_2, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\boldsymbol{\mu}}) \end{aligned} \tag{40}$$

where both maximization problems are unconstrained, i.e., the powers are not necessarily restricted to be nonnegative.

We fix the multipliers $\tilde{\lambda}_1$, $\tilde{\lambda}_2$ and $\tilde{\boldsymbol{\mu}}$. The Lagrangian $\mathfrak{L}$ is considered as a function of $\boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2)$, and written as $\mathfrak{L}(\boldsymbol{p}) \triangleq \mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\boldsymbol{\mu}})$. We want to show that the zero vector is included in the subdifferential, i.e., $\mathbf{0} \in \partial \mathfrak{L}(\tilde{\boldsymbol{p}})$. Without loss of generality, suppose that $\tilde{p}_2^{(l)} = a^{(l)} - 1$ for $l = 1, 2, \ldots, L_0$.

The subdifferential has the following property ([7, p. 222–223]): for convex function $f_1, \ldots, f_n$ defined on the same domain, and constant $c_1, \ldots, c_n, \partial(c_1 f_1 + \ldots + c_n f_n) = c_1 \partial f_1 + \ldots + c_n \partial f_n$ where the sum of two sets is defined as $A + B \triangleq \{a + b : a \in A, b \in B\}$. We have

$$\begin{aligned}\partial \mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2) = \kappa &\sum_{l=1}^{L} \partial C_{a^{(l)}} \left({p^{(l)}}_1, {p^{(l)}}_2 \right) \\ &- \sum_{l=1}^{L} \left(\tilde{\lambda}_1 w_1 \partial {p^{(l)}}_1 + \tilde{\lambda}_2 w_2 \partial {p^{(l)}}_2 \right) \\ &+ \sum_{i=1}^{2} \sum_{l=1}^{L} \tilde{\mu}_i^{(l)} \partial p_i^{(l)}. \end{aligned} \tag{41}$$

The subdifferential $\partial p_i^{(l)}$ consists of a single element for all i and l. The subdifferential $\partial C_{a^{(l)}}({p^{(l)}}_1, {p^{(l)}}_2)$ also consists of a single element for $l > L_0$, because it is a smooth function. For $l = 1, \ldots, L_0$, the subgradient in $\partial C_{a^{(l)}}({p^{(l)}}_1, {p^{(l)}}_2)$ has the form $(0, \ldots, 0, \theta_l, 0, \ldots, 0, \phi_l, 0, \ldots, 0)$ where θ_l, the lth component, is a positive constant, and ϕ_l, the $(L + l)$th component, is in some interval $[r_l, s_l]$ for the explicit form of r_l and s_l. After summation, $\partial \mathfrak{L}(\boldsymbol{p}_1, \boldsymbol{p}_2)$ is the Cartesian product

$$\begin{aligned}&\left\{ v_1^{(1)} \right\} \times \cdots \times \left\{ v_1^{(L)} \right\} \times [\rho_1, \sigma_1] \times \cdots \times [\rho_{L_0}, \sigma_{L_0}] \\ &\qquad \times \left\{ v_2^{(L_0+1)} \right\} \times \left\{ v_2^{(L_0+2)} \right\} \times \cdots \left\{ v_2^{(L)} \right\}\end{aligned}$$

Since $\tilde{\boldsymbol{p}}$ is a fixed point, we have $v_1^{(l)} = 0$ for all l, $v_2^{(l)} = 0$ for $l > L_0$ and $0 \in [\rho_l, \sigma_l]$ for $l \leq L_0$. Consequently, $\mathbf{0} \in \partial \mathfrak{L}(\tilde{\boldsymbol{p}})$, which implies the saddle-point condition in Proposition 4. □

Lemma 9: For all open set $\mathcal{A} \subset \mathcal{P}$ that contains the optimal point $\boldsymbol{p}^*$, there exists $c \leq f(\boldsymbol{p}^*)$ such that $f^{-1}([c, f(\boldsymbol{p}^*)])$ is a subset of $\mathcal{A}$.

Proof: For any $c \leq f(\boldsymbol{p}^*)$, let $\mathcal{S}_c$ be the open set $\{\boldsymbol{p} \in \mathcal{P} : f(p) \leq c\}$. Then $\mathcal{P} = \mathcal{A} \cup (\bigcup_{c \leq M} \mathcal{S}_c)$. Since $\mathcal{P}$ is compact, it has a finite subcover, say, $\mathcal{A} \cup (\bigcup_{k=1}^{K} \mathcal{S}_{c_k})$. It follows that $\mathcal{A} \cup \mathcal{S}_{c_0} = \mathcal{P}$, where $c_0 = \max\{c_1, \ldots, c_K\}$. Therefore, the complement of $\mathcal{S}_{c_0}$ in $\mathcal{P}$ is contained in $\mathcal{A}$, i.e., $\{\boldsymbol{p} \in \mathcal{P} : c_0 \leq f(\boldsymbol{p}) \leq f(\boldsymbol{p}^*)\}$ must be a subset of $\mathcal{A}$. □

Proof of Theorem 6: Consider an ϵ-neighborhood of the optimal point $\boldsymbol{p}^*$, denoted by $\mathcal{N}(\boldsymbol{p}^*; \epsilon)$. By Lemma 9, we can find $c \leq f(\boldsymbol{p}^*)$ such that $\boldsymbol{p} \in \mathcal{N}(\boldsymbol{p}^*; \epsilon)$ if $f(\boldsymbol{p}) > c$. By Lemma 8, there exists a subsequence $\{\boldsymbol{p}^{(\alpha_n)}\}$ that converges to $\boldsymbol{p}^*$. The continuity of f implies that $\{f(\boldsymbol{p}^{(\alpha_n)})\}$ converges to $f(\boldsymbol{p}^*)$, and thus we can find N such that $f(\boldsymbol{p}^{(\alpha_N)}) > c$. Since $f(\boldsymbol{p}^{(n)})$ is increasing in n, $f(\boldsymbol{p}^{(n)}) > c$ for all $n \geq \alpha_N$, and hence $\boldsymbol{p}^{(n)}$ is in $\mathcal{N}(\boldsymbol{p}^*; \epsilon)$ for all $n \geq \alpha_N$. As ϵ is an arbitrarily small positive constant, $\{\boldsymbol{p}^{(n)}\}$ converges to $\boldsymbol{p}^*$. □

REFERENCES

[1] T. S. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 49–60, Jan. 1981.

[2] M. H. M. Costa, “On the Gaussian interference channel,” *IEEE Trans. Inf. Theory*, vol. IT-31, no. 5, pp. 607–615, Sep. 1985.

[3] I. Sason, “On achievable rate regions for the Gaussian interference channel,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, Jun. 2004.

[4] H. F. Chong, M. Montani, and H. K. Garg, “Capacity theorems for the “Z” channel,” *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1348–1365, Apr. 2007.

[5] W. Yu and J. M. Cioffi, “Competitive equilibrium in the Gaussian interference channel,” in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, Jun. 2000, p. 431.

[6] A. Takayama, *Mathematical Economics*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 1985.

[7] R. T. Rockafellar, *Convex Analysis*, ser. Princeton Landmarks in Mathematics and Physics. Princeton, NJ: Princeton University Press, 1997.

[8] C. Berge, *Topological Spaces*. New York: Dover, 1997.