Um Algoritmo para Escolha de Parâmetros da Pré-codificação de Forçagem a Inteiros para Canais *Downlink* MU–MIMO

Ricardo Bohaczuk Venturelli e Danilo Silva

***Resumo*— A técnica de pré-codificação de forçagem a inteiros (IF) é uma alternativa às técnicas tradicionais de pré-codificação linear em canais *downlink* MIMO com múltiplos usuários (MU-MIMO). A pré-codificação IF aproxima o canal efetivo, observado por cada usuário, em uma combinação linear inteira dos vetores transmitidos, generalizando métodos tradicionais de pré-codificação linear. Recentemente, Silva *et al.* propuseram um método para encontrar parâmetros ótimos para pré-codificação IF em alta SNR, porém restrito ao caso de $K = 2$ usuários. Nesse artigo, propomos um método para encontrar bons parâmetros para $K \geq 2$ usuários com complexidade $\mathcal{O}(K^3)$. Resultados de simulação mostram que o método proposto atinge um desempenho muito superior aos métodos tradicionais de pré-codificação linear.**

***Palavras-Chave*— MIMO com múltiplos usuários, canal downlink, forçagem a inteiros, pré-codificação linear.**

***Abstract*— Integer-Forcing (IF) precoding is an alternative to traditional methods of linear precoding in multiuser MIMO (MU-MIMO) downlink channels. The IF precoding tries to convert the effective channel, as it is seen by each receiver, into an integer linear combination of the transmitted vectors, generalizing traditional linear precoding. Recently, Silva *et al.* proposed a method to find optimal parameters for IF precoding in high SNR, restricted, however, to $K = 2$ users. In this paper, we propose a method to find good parameters for $K \geq 2$ users with complexity $\mathcal{O}(K^3)$. Simulations results show that the proposed method can outperform traditional methods of linear precoding.**

***Keywords*— Multiuser MIMO, downlink channel, integer-forcing, linear precoding.**

I. Introdução

Canais de múltiplas entradas e múltiplas saídas (MIMO) com múltiplos usuários (MU-MIMO) vem sendo largamente usados em sistemas de telecomunicação, principalmente devido ao fato de que a taxa-soma do sistema cresce linearmente com o número de antenas utilizadas [1]. Em particular, estamos interessados no canal *downlink* (também chamado de MIMO *broadcast*), no qual uma estação rádio-base deseja transmitir para vários usuários, os quais não cooperam entre si.

Dentre os métodos mais tradicionais para canais *downlink* MU-MIMO destaca-se o uso de pré-codificação linear (também conhecida como pré-equalização ou *beamforming*), a qual faz uso de uma matriz de pré-codificação [2]. Uma técnica bem conhecida de pré-codificação linear é a chamada forçagem a zero (ZF), onde a matriz de pré-codificação tem o intuito de pré-inverter o canal para que o usuário receba apenas o sinal de interesse (acrescido do ruído) [2]. Outra técnicaxa também bem conhecida é a forçagem a zero regularizada (RZF), que ao invés de uma simples pré-inversão da matriz do canal, tenta minimizar o efeito da interferência acrescida do ruído do canal [2]. Entretanto, o desempenho desses métodos fica bastante aquém da capacidade-soma do canal, principalmente quando o número de antenas do transmissor é próximo ao número de antenas dos usuários combinados, mesmo em alta SNR [3].

Uma alternativa a essas técnicas tradicionais é a pré-codificação de forçagem a inteiros (IF) [4]–[6]. A estrutura da técnica de pré-codificação IF é semelhante aos métodos de pré-codificação linear, no sentido que também é utilizada uma matriz de pré-codificação. Entretanto, na técnica IF, essa matriz é utilizada para transformar o canal efetivo em uma matriz de coeficientes inteiros $\mathbf{A}$, dessa forma generalizando a pré-codificação linear. Além disso, na pré-codificação IF os vetores transmitidos são pontos de um reticulado, o que significa que qualquer combinação linear inteira de pontos do reticulado também é um ponto de reticulado [7]. Dessa maneira, cada usuário consegue recuperar essa combinação linear usando uma decodificação de reticulado. Em conjunto com a *pré-codificação de mensagem*, na qual o transmissor pré-multiplica as mensagens pela matriz inversa de $\mathbf{A}$ antes de codificá-las, cada usuário consegue recuperar sua mensagem de interesse livre de interferência [6].

Em [6] são propostos dois esquemas de pré-codificação IF, o primeiro chamado DIF (*diagonally-scaled exact IF*), que é análogo ao método ZF, e o segundo chamado de RDIF (DIF regularizado), que por sua vez, é análogo ao método RZF. Esses métodos fornecem uma forma de encontrar os parâmetros ótimos da pré-codificação IF de forma analítica para $K = 2$ usuários e alta SNR. Além disso, nesse caso particular, também é mostrado que esses métodos alcançam uma taxa-soma muito próxima a capacidade do canal [6].

Nesse artigo, propomos um método para encontrar bons parâmetros da pré-codificação IF para qualquer $K \geq 2$ usuários. O método proposto faz uso de um relaxamento do problema original para encontrar esses parâmetros com complexidade $\mathcal{O}(K^3)$. Resultados de simulação mostram que o método proposto tem um desempenho muito superior ao ZF e RZF e, em alguns casos, próximo da capacidade-soma do canal *downlink* MIMO [8]–[11].

Ricardo Bohaczuk Venturelli e Danilo Silva¸ Departamento de Engenharia Elétrica e Eletrônica, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil, E-mails: ricardo.bventurelli@gmail.com, danilo.silva@ufsc.br.

II. PRELIMINARES

A. Modelo do Canal

Considere um canal *downlink* MIMO com um transmissor com M antenas e $K \leq M$ usuários receptores com uma antena cada. Seja $\mathbf{w}_i \in \mathcal{W}_i$ a mensagem destinada ao i-ésimo usuário, $i = 1, \ldots, K$. O transmissor codifica as mensagens $\mathbf{w}_1, \ldots, \mathbf{w}_K$ em uma matriz $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1^{\mathrm{T}} & \cdots & \mathbf{x}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{K\times n}$, onde cada vetor $\mathbf{x}_i \in \mathbb{R}^n$ satisfaz uma restrição de potência $\mathbb{E}[\|\mathbf{x}_i\|^2] \leq n\mathrm{SNR}$, e em seguida aplica uma pré-codificação definida por uma matriz $\mathbf{T} \in \mathbb{R}^{M\times K}$ que satisfaz

$$\mathrm{Tr}(\mathbf{T}\mathbf{T}^{\mathrm{T}}) \leq 1 \tag{1}$$

a qual é chamada matriz de pré-codificação ou matriz de *beamforming*.

Seja $\mathbf{Y} \in \mathbb{R}^{K\times n}$ a matriz cuja i-ésima linha, denotada por $\mathbf{y}_i$, é o sinal recebido pelo i-ésimo usuário. O canal pode ser modelado por

$$\mathbf{Y} = \mathbf{H}\mathbf{T}\mathbf{X} + \mathbf{Z} \tag{2}$$

em que $\mathbf{H} = \begin{bmatrix}\mathbf{h}_1^{\mathrm{T}} & \cdots & \mathbf{h}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{K\times M}$ e $\mathbf{h}_i \in \mathbb{R}^M$ são os coeficientes do canal associados ao i-ésimo usuário e $\mathbf{Z} \in \mathbb{R}^{K\times n}$ é a matriz de ruído gaussiano, cujas entradas são i.i.d. de média nula e variância unitária.

O i-ésimo receptor irá inferir, a partir de $\mathbf{y}_i$, a mensagem $\hat{\mathbf{w}}_i \in \mathcal{W}_i$. Dizemos que um erro ocorre se $\hat{\mathbf{w}}_i \neq \mathbf{w}_i$, $i = 1, \ldots, K$. A taxa-soma do sistema é dada por $R_{\mathrm{sum}} = R_1 + \cdots + R_K$, em que $R_i = \frac{1}{n}\log_2|\mathcal{W}_i|$. Uma taxa-soma R é dita ser alcançável se, para qualquer $\epsilon > 0$ e para algum n suficientemente grande, existe um esquema de codificação e decodificação com probabilidade de erro menor que ϵ.

A capacidade-soma do canal é o supremo das taxas alcançáveis, a qual é dada por [9]–[11]

$$C = \sup_{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})\leq 1} \log_2\det\left(\mathbf{I} + \mathrm{SNR}\mathbf{H}^{\mathrm{T}}\mathbf{Q}\mathbf{H}\right) \tag{3}$$

em que $\mathbf{Q} \in \mathbb{R}^{K\times K}$ é uma matriz diagonal com entradas não negativas.

B. Pré-codificação de Forçagem a Inteiros

A abordagem de forçagem a inteiros faz uso de reticulados. Um reticulado $\Lambda \subseteq \mathbb{R}^n$ é sub-grupo discreto de $\mathbb{R}^n$ [7]. Pode-se expressar um reticulado como $\Lambda = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{u}\mathbf{G}, \mathbf{u} \in \mathbb{Z}^n\}$, em que $\mathbf{G} \in \mathbb{R}^{n\times n}$ é a matriz geradora. Se $\lambda_i \in \Lambda$, $i = 1, \ldots, K$, então $a_1\lambda_1 + \cdots + a_K\lambda_K \in \Lambda$, em que $a_1, \ldots, a_K \in \mathbb{Z}$.

Seja p um número primo e considere que $\mathcal{W}_i \subseteq \mathcal{W}$, em que $\mathcal{W} = \mathbb{Z}_p^n$. Seja $\Lambda \subseteq \mathbb{R}^n$ um reticulado e sejam $\Lambda_i \subseteq \Lambda$, $i = 1, \ldots, K$. Além disso, seja $\varphi : \Lambda \to \mathcal{W}$ um mapeamento linear tal que $\varphi(\Lambda_i) = \mathcal{W}_i$ e seja $\tilde{\varphi} : \mathcal{W} \to \Lambda$ uma função bijetiva tal que $\varphi(\tilde{\varphi}(\mathcal{W}_i)) = \mathcal{W}_i$.

Seja $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1^{\mathrm{T}} & \cdots & \mathbf{a}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{Z}^{K\times K}$ uma matriz de posto completo, $\mathbf{A}' \in \mathbb{Z}^{K\times K}$ uma matriz tal que $\mathbf{A}\mathbf{A}' = \mathbf{I} \mod p$ e seja $\mathbf{W}$ a matriz cuja i-ésima linha é dada por $\mathbf{w}_i \in \mathcal{W}_i$. O transmissor calcula [6]

$$\mathbf{W}' = \mathbf{A}'\mathbf{W} \tag{4}$$

e então aplica a função $\tilde{\varphi}$ nas linhas de $\mathbf{W}'$ gerando a matriz $\mathbf{X}$, em que, agora, as linhas são pontos do reticulado.

O i-ésimo usuário receptor aplica o coeficiente de equalização $\alpha_i \in \mathbb{R}$ no vetor $\mathbf{y}_i$ gerando [6]

$$\mathbf{y}_{\mathrm{eff},i} = \alpha_i\mathbf{y}_i = \lambda_i + \mathbf{z}_{\mathrm{eff},i} \tag{5}$$

em que $\mathbf{z}_{\mathrm{eff},i}$ é o ruído efetivo e $\lambda_i = \mathbf{a}_i\mathbf{X} \in \Lambda$ é um ponto de reticulado, tal que $\varphi(\lambda_i) = \mathbf{w}_i$.

Teorema 1: [6], [12], [13] Para p e n suficientemente grandes existe um esquema de pré-codificação IF que alcança a seguinte taxa-soma

$$R_{\mathrm{IF}}(\mathbf{H}, \mathbf{A}, \mathbf{T}) = \sum_{i=1}^{K} R(\mathbf{h}_i, \mathbf{a}_i, \mathbf{T}) \tag{6}$$

em que

$$R(\mathbf{h}_i, \mathbf{a}_i, \mathbf{T}) = \log_2^+\left(\frac{\mathrm{SNR}}{\mathbf{a}_i\left(\mathbf{I} - \frac{\mathrm{SNR}}{\mathrm{SNR}\|\mathbf{h}_i\mathbf{T}\|^2+1}\mathbf{T}^{\mathrm{T}}\mathbf{h}_i^{\mathrm{T}}\mathbf{h}_i\mathbf{T}\right)\mathbf{a}_i^{\mathrm{T}}}\right) \tag{7}$$

são as taxas de cada usuário [14], [15].
Nosso objetivo é maximizar (6) através das escolhas de $\mathbf{A}$ e $\mathbf{T}$.

Note que a pré-codificação IF pode ser vista como uma generalização de métodos mais tradicionais. Por exemplo, se escolheremos $\mathbf{A} = \mathbf{I}$ e $\mathbf{T} = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{D}$, em que $\mathbf{D}$ é uma matriz diagonal, então temos a pré-codificação com forçagem a zero (ZF, do inglês *zero-forcing*), já se fizermos $\mathbf{A} = \mathbf{I}$ e $\mathbf{T} = \mathbf{H}^{\mathrm{T}}(\frac{K}{\mathrm{SNR}}\mathbf{I} + \mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{D}$ então recaímos na pré-codificação com forçagem a zero regularizada (RZF, do inglês *regularized zero-forcing*) [2], [16].

C. Métodos DIF e RDIF

Em [6] são propostos dois métodos para escolher $\mathbf{T}$ e $\mathbf{A}$. O primeiro método, chamado de pré-codificação DIF (*diagonally-scaled exact integer-forcing*), escolhe

$$\mathbf{T} = c\mathbf{T}_0 \tag{8}$$

$$\mathbf{T}_0 = \mathbf{H}^{\mathrm{T}}\mathbf{M}\mathbf{D}_0\mathbf{A} \tag{9}$$

onde $\mathbf{M} = \left(\mathbf{H}\mathbf{H}^{\mathrm{T}}\right)^{-1}$, $\mathbf{D}_0$ é uma matriz diagonal tal que $|\det\mathbf{D}_0| = 1$ e c é escolhido para satisfazer (1). A taxa-soma desse método é limitada inferiormente por

$$R_{\mathrm{DIF}}^{\mathrm{HI}}(\mathbf{H}, \mathbf{A}, \mathbf{D}_0) \triangleq K\log_2\left(\frac{\mathrm{SNR}}{\mathrm{Tr}(\mathbf{T}_0^{\mathrm{T}}\mathbf{T}_0)}\right). \tag{10}$$

Embora (10) seja um limitante inferior, mostra-se em [6] que, em alta SNR, esta taxa é a máxima alcançável por qualquer esquema de pré-codificação IF. Em [6] é proposto um método analítico para encontrar os valores de $\mathbf{A}$ e $\mathbf{D}_0$ que maximizam (10) no caso especial $K = 2$.

Embora ótima para alta SNR, a pré-codificação DIF sofre uma deterioração para valores moderados de SNR. Assim, um segundo método é proposto em [6], chamado de pré-codificação RDIF (*regularized DIF*), o qual apresenta um melhor desempenho que o DIF para qualquer SNR finita. A única diferença entre os métodos está na substituição da matriz $\mathbf{M}$ por $\mathbf{M} = \left(\frac{K}{\mathrm{SNR}}\mathbf{I} + \mathbf{H}\mathbf{H}^{\mathrm{T}}\right)^{-1}$. Em particular, no

caso especial $K = 2$, o mesmo método proposto em [6] permite determinar analiticamente os parâmetros $\mathbf{A}$ e $\mathbf{D}_0$ que maximizam o limitante inferior (10).

III. Método RDIF para $K \geq 2$

Seguindo a abordagem de [6], nosso objetivo é maximizar o limitante inferior (10), para qualquer $K \geq 2$. Isso equivale a minimizar $\mathrm{Tr}(\mathbf{T}_0^{\mathrm{T}}\mathbf{T}_0)$ em que $\mathbf{T}_0$ é dado por (9), com as restrições de que $\mathbf{A} \in \mathbb{Z}^{K\times K}$ tenha posto completo e que $\mathbf{D}_0 \in \mathbb{R}^{K\times K}$ seja uma matriz diagonal tal que $|\det \mathbf{D}_0| = 1$.

Seja $\mathbf{D}_0 = \mathrm{diag}(d_1, d_2, \ldots, d_K)$, $\left[\mathbf{M}\right]_{ij} = M_{ij}$ e $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1^{\mathrm{T}} & \cdots & \mathbf{a}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. Podemos escrever

$$\mathrm{Tr}(\mathbf{T}_0^{\mathrm{T}}\mathbf{T}_0) = \mathrm{Tr}(\mathbf{A}^{\mathrm{T}}\mathbf{D}_0^{\mathrm{T}}\mathbf{M}\mathbf{D}_0\mathbf{A}) \tag{11}$$

$$= \sum_{i=1}^{K} M_{ii}\left\|\mathbf{a}_i\right\|^2 d_i^2 + \sum_{i=1}^{K}\sum_{j=i+1}^{K} 2M_{ji}\mathbf{a}_i\mathbf{a}_j^{\mathrm{T}} d_i d_j. \tag{12}$$

A. Estrutura de $\mathbf{A}$

Note que em (12) cada termo do primeiro somatório é sempre positivo, enquanto cada termo do segundo somatório pode conter valores positivos ou negativos. No melhor caso, em que esses termos são negativos, queremos escolher $\mathbf{A}$ de modo que a norma euclidiana de cada linha seja minimizada, enquanto maximizamos o valor absoluto dos produtos interno entre as linhas.

Se pensarmos em, exclusivamente, minimizar $\|\mathbf{a}_i\|$, então é fácil perceber que uma escolha ótima seria $\mathbf{A} = \mathbf{I}$. Entretanto, queremos ter alguns graus de liberdade para maximizar os produtos internos. Assim, propomos escolher $\mathbf{A}$ como uma matriz triangular inferior, isto é,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & 1 \end{bmatrix}. \tag{13}$$

Note que, esta escolha é uma extensão direta do caso ótimo para $K = 2$ [6].

Diferentemente do caso $K = 2$, uma permutação das linhas de $\mathbf{A}$ pode fazer com que a taxa-soma seja diferente, isto é,

$$R_{\mathrm{IF}}(\mathbf{H}, \mathbf{PA}, \mathbf{T}) \neq R_{\mathrm{IF}}(\mathbf{H}, \mathbf{A}, \mathbf{T})$$

em que $\mathbf{P}$ é uma matriz de permutação. No entanto, é fácil ver que

$$R_{\mathrm{IF}}(\mathbf{H}, \mathbf{PA}, \mathbf{T}) = R_{\mathrm{IF}}(\mathbf{P}^{\mathrm{T}}\mathbf{H}, \mathbf{A}, \mathbf{T}) \tag{14}$$

isto é, uma permutação de $\mathbf{A}$ pode ser avaliada equivalentemente aplicando a permutação inversa em $\mathbf{H}$. Assim, para o restante do método, consideraremos somente a estrutura em (13). Note que, com essa estrutura, a restrição de posto completo já é satisfeita, não sendo mais necessária considerá-la.

B. Problema Relaxado

A otimização de (11) é complexa devido à restrição $\mathbf{A} \in \mathbb{Z}^{K\times K}$. Uma maneira de simplificar o problema é considerar um relaxamento em que os coeficientes de $\mathbf{A}$ possam assumir qualquer valor em $\mathbb{R}$. Mais precisamente, seja $\tilde{\mathbf{A}} \in \mathbb{R}^{K\times K}$ com estrutura dada em (13), seja $\tilde{\mathbf{D}}_0 \in \mathbb{R}^{K\times K}$ uma matriz diagonal e considere a seguinte função objetivo $f(\tilde{\mathbf{A}}, \tilde{\mathbf{D}}_0) \triangleq \mathrm{Tr}(\tilde{\mathbf{A}}^{\mathrm{T}}\tilde{\mathbf{D}}_0^{\mathrm{T}}\mathbf{M}\tilde{\mathbf{D}}_0\tilde{\mathbf{A}})$. Deseja-se encontrar $\tilde{\mathbf{D}}_0^{\mathrm{opt}}$ e $\tilde{\mathbf{A}}^{\mathrm{opt}}(\tilde{\mathbf{D}}_0^{\mathrm{opt}})$ tais que

$$\tilde{\mathbf{A}}^{\mathrm{opt}}(\tilde{\mathbf{D}}_0) = \arg\min_{\tilde{\mathbf{A}}} f(\tilde{\mathbf{A}}, \tilde{\mathbf{D}}_0) \tag{15}$$

$$\tilde{\mathbf{D}}_0^{\mathrm{opt}} = \arg\min_{\tilde{\mathbf{D}}_0 : \left|\det \tilde{\mathbf{D}}_0\right| = 1} f(\tilde{\mathbf{A}}^{\mathrm{opt}}(\tilde{\mathbf{D}}_0), \tilde{\mathbf{D}}_0) \tag{16}$$

isto é, podemos dividir o problema em, primeiramente, encontrar $\tilde{\mathbf{A}}^{\mathrm{opt}}$ dado por (15) (em função de $\tilde{\mathbf{D}}_0$), e então encontrar $\tilde{\mathbf{D}}_0^{\mathrm{opt}}$ dado por (16) propriamente [17].

Teorema 2: O valor $\tilde{\mathbf{A}}$ que soluciona (15) é dado por

$$\tilde{\mathbf{A}}^{\mathrm{opt}} = \tilde{\mathbf{D}}_0^{-1}\mathbf{V}\tilde{\mathbf{D}}_0. \tag{17}$$

em que $\mathbf{V}$ é uma matriz diagonal inferior com diagonal unitária tal que $\mathbf{M}^{-1} = \mathbf{V}\mathbf{C}^{-1}\mathbf{V}^{\mathrm{T}}$ e $\mathbf{C}^{-1}$ é uma matriz diagonal.

Demonstração: A derivada da função f em relação a $\tilde{\mathbf{A}}$ é dada por $\nabla f = \frac{\partial f}{\partial \tilde{\mathbf{A}}} = 2\tilde{\mathbf{D}}_0^{\mathrm{T}}\mathbf{M}\tilde{\mathbf{D}}_0\tilde{\mathbf{A}}$. Note que, como apenas os coeficientes abaixo da diagonal principal são variáveis, é necessário que

$$(\nabla f)_{ij} = (2\tilde{\mathbf{D}}_0^{\mathrm{T}}\mathbf{M}\tilde{\mathbf{D}}_0\tilde{\mathbf{A}})_{ij} = 0 \tag{18}$$

$i = j+1, \ldots, K$ e $j = 1, \ldots, K-1$. Seja $\mathbf{V} = \tilde{\mathbf{D}}_0\tilde{\mathbf{A}}\tilde{\mathbf{D}}_0^{-1}$, podemos reescrever o problema (18) como $(\mathbf{MV})_{ij} = 0$, $i = j+1, \ldots, K$, $j = 1, \ldots, K-1$.

Para encontrar uma solução, considere a decomposição LDL [18] da matriz $\mathbf{M}^{-1}$, isto é

$$\mathbf{M}^{-1} = \mathbf{L}\mathbf{C}^{-1}\mathbf{L}^{\mathrm{T}} \tag{19}$$

em que $\mathbf{L}$ é uma matriz triangular inferior com diagonal unitária, e $\mathbf{C}^{-1}$ é uma matriz diagonal. Note que isso significa que $\mathbf{M} = \mathbf{L}^{-\mathrm{T}}\mathbf{C}\mathbf{L}^{-1}$ e portanto $\mathbf{ML} = \mathbf{L}^{-\mathrm{T}}\mathbf{C}$. Como $\mathbf{L}^{-\mathrm{T}}$ é uma matriz diagonal superior, temos que $(\mathbf{ML})_{ij} = 0$, para $i = j+1, \ldots, K$, $j = 1, \ldots, K-1$. Portanto escolher $\mathbf{V} = \mathbf{L}$ é uma solução para $(\mathbf{MV})_{ij} = 0$.

Por fim, temos que $\tilde{\mathbf{A}} = \tilde{\mathbf{D}}_0^{-1}\mathbf{V}\tilde{\mathbf{D}}_0$ completando a prova. ■

Note que as matrizes $\mathbf{V}$ e $\mathbf{C}^{-1}$ podem ser calculadas apenas com o conhecimento dos coeficientes do canal. Além disso, como $\tilde{\mathbf{A}}^{\mathrm{opt}}$ está em função de $\tilde{\mathbf{D}}_0$, podemos resolver o problema (16).

Teorema 3: Considerando a matriz $\tilde{\mathbf{A}}^{\mathrm{opt}}$ encontrada no Teorema 2, a solução para (16) é dada por

$$\tilde{\mathbf{D}}_0\tilde{\mathbf{D}}_0^{\mathrm{T}} = (\det \mathbf{M})^{1/K}\mathbf{C}^{-1} \tag{20}$$

em que $\mathbf{C}^{-1}$ é uma matriz diagonal obtida da decomposição LDL de $\mathbf{M}^{-1}$.

Demonstração: Substituindo o valor de $\tilde{\mathbf{A}}$ encontrado no Teorema 2, temos que

$$\begin{aligned} f(\tilde{\mathbf{D}}_0) &= f(\tilde{\mathbf{A}}^{\text{opt}}, \tilde{\mathbf{D}}_0) \\ &= \text{Tr}(\tilde{\mathbf{D}}^{\text{T}}\mathbf{V}^{\text{T}}\mathbf{M}\mathbf{V}\tilde{\mathbf{D}}_0) \\ &= \text{Tr}(\tilde{\mathbf{D}}_0^{\text{T}}\mathbf{C}\tilde{\mathbf{D}}_0) \\ &= \sum_{i=1}^{K} c_i \tilde{d}_i^2 \end{aligned}$$

em que c_i e $\tilde{d}_i$ são o i-ésimo elemento da diagonal de $\mathbf{C}$ e $\tilde{\mathbf{D}}_0$, respectivamente.

Note que a restrição $\left|\det \tilde{\mathbf{D}}_0\right| = \left|\prod_{i=1}^{K} \tilde{d}_i\right| = 1$ implica em $\prod_{i=1}^{K} \tilde{d}_i^2 = 1$ que, por sua vez, implica em $\prod_{i=1}^{K} c_i \tilde{d}_i^2 = \prod_{i=1}^{K} c_i = \det \mathbf{C}$. Além disso, note que $\det \mathbf{C} = \det(\mathbf{V}^{\text{T}}\mathbf{M}\mathbf{V}) = \det \mathbf{M}$ uma vez que $\mathbf{V}$ é uma matriz triangular inferior com diagonal unitária. Por fim, pela desigualdade entre as médias aritmética e geométrica, temos que $1/K \sum_{i=1}^{K} c_i \tilde{d}_i^2 \geq \left(\prod_{i=1}^{K} c_i \tilde{d}_i^2\right)^{1/K} = (\det \mathbf{M})^{1/K}$, com igualdade se e somente se $c_i \tilde{d}_i^2 = (\det \mathbf{M})^{1/K}$. Assim temos que $\tilde{d}_i^2 = (\det \mathbf{M})^{1/K}/c_i$ para $i = 1, \ldots, K$ ou, em notação matricial, $\tilde{\mathbf{D}}_0 \tilde{\mathbf{D}}_0^{\text{T}} = (\det \mathbf{M})^{1/K} \mathbf{C}^{-1}$.

■

C. Otimização de $\mathbf{A}$

Para a pré-codificação IF, é necessário que $\mathbf{A}$ seja uma matriz com coeficientes inteiros. Portanto, é necessário realizar uma quantização em $\tilde{\mathbf{A}}$. Note que, dado $\mathbf{D}_0$, a solução para $\tilde{\mathbf{A}}$ encontrada no teorema 2 apresenta um único ponto crítico. Além disso, como se trata de uma otimização sem restrição, podemos afirmar que os valores encontrados são, de fato, um ótimo global. Como a função cresce monotonicamente a partir do ponto crítico, os valores ótimos dos coeficientes de $\mathbf{A}$ estão ao redor dos valores ótimos dos coeficientes de $\tilde{\mathbf{A}}$.

Note que existem $(K^2-K)/2$ coeficientes em $\tilde{\mathbf{A}}$, o que nos daria uma complexidade $\mathcal{O}(2^{K^2})$ se quisermos testar todas as quantizações possíveis[1]. Uma abordagem mais simples, porém sub-ótima, seria quantizar cada coeficiente para o inteiro mais próximo, isto é, sejam $\left[\tilde{\mathbf{A}}\right]_{ij} = \tilde{a}_{ij}$ e $\left[\mathbf{A}\right]_{ij} = a_{ij}$ então

$$a_{ij} = \lceil \tilde{a}_{ij} \rfloor \tag{21}$$

$i = j+1, \ldots, K$, $j = 1, \ldots, K-1$.

D. Otimização de $\mathbf{D}_0$

Note que, ao quantizarmos $\tilde{\mathbf{A}}$ em $\mathbf{A}$, o teorema 3 perde a validade. Além disso, não é mais possível expressar os coeficientes de $\mathbf{A}$ em função de $\mathbf{D}_0$. Seja $\mathbf{d} = \begin{bmatrix} d_1 & d_2 & \cdots & d_K \end{bmatrix}$. Nesse caso, o problema de otimização (11) pode ser descrito como

$$\begin{aligned} \text{minimize} \quad & \mathbf{d}\left(\mathbf{A}\mathbf{A}^{\text{H}} \circ \mathbf{M}^{\text{T}}\right)\mathbf{d}^{\text{T}} \\ \text{s.t.} \quad & \left|\prod_i d_i\right| = 1 \end{aligned} \tag{22}$$

[1]Mais precisamente, as quantizações possíveis são $a_{ij} = \lceil \tilde{a}_{ij} \rceil$ ou $a_{ij} = \lfloor \tilde{a}_{ij} \rfloor$, para $i = j+1, \ldots, K$, $j = 1, \ldots, K-1$.

em que $\circ$ representa o produto de Hadamard. Como os coeficientes de $\left(\mathbf{A}\mathbf{A}^{\text{H}} \circ \mathbf{M}^{\text{T}}\right)$ podem ser negativos, não existe garantia que o problema de otimização seja geométrico, e portanto, possível de converter em um problema de otimização convexo.

Esse problema de otimização pode ser convertido em um problema sem restrições o que permite que algoritmos mais simples de otimização, como o método do gradiente descendente [17], possam ser utilizados para encontrar ótimos locais. Para isso, é importante indicar um ponto inicial, como por exemplo, usar os valores de $\tilde{\mathbf{D}}_0$ calculados na seção anterior.

E. Sumário do Algoritmo

Nessa seção descrevemos um sumário do algoritmo para encontrar $\mathbf{A}$ e $\mathbf{D}_0$ com uma complexidade controlada.

Entrada: $\mathbf{H}$ e N_p (em que N_p é o número de permutações testadas)

1) Crie uma matriz de permutação aleatória $\mathbf{P}$.
2) Calcule $\mathbf{M} = \mathbf{P}^{\text{T}}\left(\frac{K}{\text{SNR}}\mathbf{I} + \mathbf{H}\mathbf{H}^{\text{T}}\right)^{-1}\mathbf{P}$
3) Calcule a decomposição LDL (19), e então, calcule a matriz $\tilde{\mathbf{D}}_0$ com (20).
4) Calcule (17) para encontrar $\tilde{\mathbf{A}}$.
5) Utilize (21) em $\tilde{\mathbf{A}}$ para encontrar $\mathbf{A}$.
6) Utilize algum algoritmo de otimização local em (22) para encontrar $\mathbf{D}_0$, tendo como ponto inicial $\tilde{\mathbf{D}}_0$.
7) Calcule $\mathbf{T}_0$ como (9) e então $\mathbf{T}$ como (8).
8) Calcule a taxa-soma (6).
9) Se esta for a maior taxa-soma calculada até então, faça $\mathbf{A}^* = \mathbf{A}$, $\mathbf{T}^* = \mathbf{T}$ e $\mathbf{P}^* = \mathbf{P}$.
10) Repita N_p vezes os passos 1-9.

Saída: $\mathbf{A} = \mathbf{P}^*\mathbf{A}^*$ e $\mathbf{T} = \mathbf{T}^*$.

F. Análise de Complexidade

Note que a complexidade do algoritmo é dominada pela decomposição LDL no passo (3), que pode ser resolvido com complexidade $\mathcal{O}(K^3)$.

Note ainda que é interessante escolher um número fixo de permutações a serem testadas, uma vez que testar todas as permutações torna-se proibitivo a medida que K aumenta.

IV. Resultados

Nessa seção iremos mostrar alguns resultados da pré-codificação RDIF em comparação com os métodos tradicionais de pré-codificação como o ZF e o RZF. Em nossas simulações, as taxas-somas são obtidas através da média de 1000 realizações de um canal com coeficientes reais. Em cada realização, os coeficientes do canal são obtidos aleatoriamente de uma distribuição gaussiana de média nula e variância unitária. Além disso, em todos os casos consideramos $M = K$.

A Fig. 1 mostra o desempenho em um canal $K = 8$ para valores moderados de SNR. Note que o desempenho do método RDIF é muito superior ao método RZF, mesmo com $N_p = 1$. Por exemplo, para uma $\text{SNR} = 5$ dB, a taxa-soma do método RZF é quase 1 bit/uso de canal menor do que a taxa-soma do RDIF com $N_p = 1$ e essa diferença cresce ainda mais conforme a SNR aumenta. Além disso, note que o

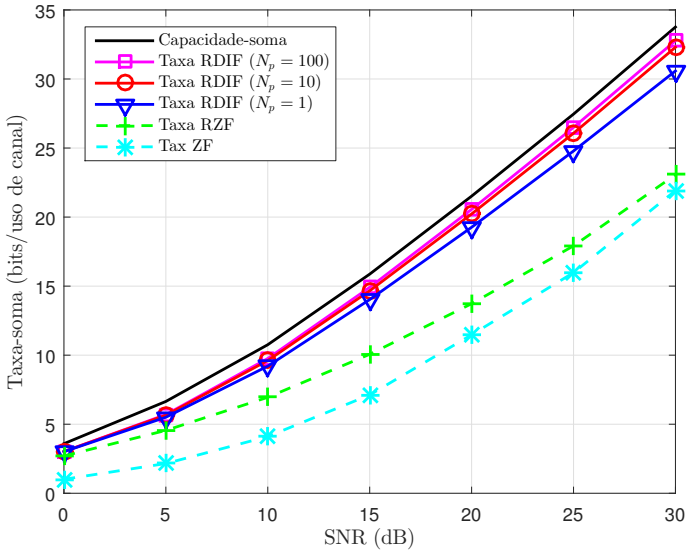

Fig. 1. Taxa-soma para valores moderados de SNR em um canal $K = 8$.

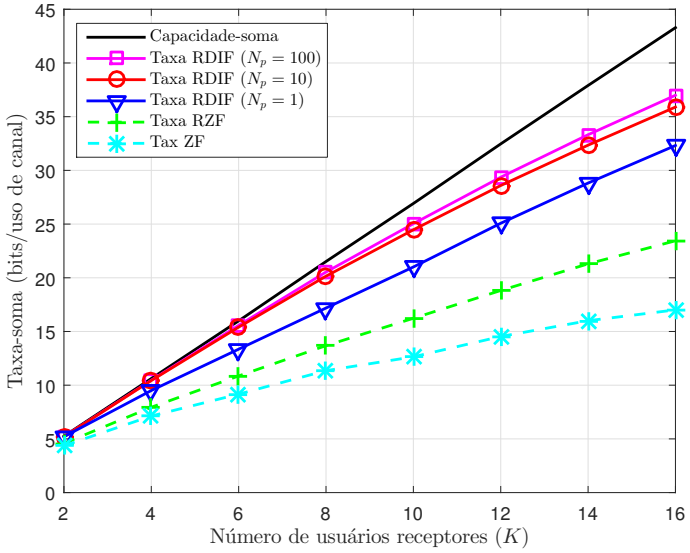

Fig. 2. Taxa-soma variando o número de usuários receptores (assim como o número de antenas no transmissor) considerando SNR $= 20$ dB.

método RDIF com $N_p = 10$ se aproxima do desempenho da capacidade-soma. Por exemplo, para SNR $= 5$ dB, o método RDIF consegue alcançar 85% da capacidade-soma, e para uma SNR $= 25$ dB, consegue alcançar quase 95% da capacidade.

Já a Fig. 2 mostra a taxa-soma para SNR $= 20$ dB variando o número de usuários. Note que, a medida que K aumenta há uma degradação do RDIF em relação a capacidade-soma do canal. Entretanto, mesmo considerando $N_p = 1$, o desempenho do RDIF é muito superior em relação ao método RZF. Além disso, note que não há uma melhora significativa quando $N_p = 100$ em relação a $N_p = 10$.

V. Conclusões

Nesse artigo, propomos uma maneira para encontrar bons parâmetros para o método RDIF para $K \geq 2$. Para isso, utilizamos um relaxamento do problema de otimização original, o qual pode ser resolvido analiticamente com complexidade $\mathcal{O}(K^3)$. Resultados de simulação mostraram que o método proposto é muito superior aos métodos RZF e ZF nos cenários simulados. Em particular, para K não muito grande ($K \leq 8$), a taxa do RDIF se mostra significativamente próxima da capacidade-soma do canal.

Uma limitação do método proposto está no fato que a matriz de permutação para $\mathbf{A}$ é escolhida de forma aleatória, e portanto, pode haver uma degradação para uma única realização de canal. Uma melhor heurística para escolher a matriz de permutação sub-ótima poderia solucionar essa limitação.

Referências

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge University Press, Jul. 2005.

[2] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal Multiuser Transmit Beamforming: A Difficult Problem with a Simple Solution Structure [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[3] J. Lee and N. Jindal, "High SNR Analysis for MIMO Broadcast Channels: Dirty Paper Coding Versus Linear Precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4787–4792, Dec. 2007.

[4] W. He, B. Nazer, and S. S. Shitz, "Uplink-Downlink Duality for Integer-Forcing," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1992–2011, Mar. 2018.

[5] S. N. Hong and G. Caire, "Reverse compute and forward: A low-complexity architecture for downlink distributed antenna systems," in *2012 IEEE International Symposium on Information Theory Proceedings*, Jul. 2012, pp. 1147–1151.

[6] D. Silva, G. Pivaro, G. Fraidenraich, and B. Aazhang, "On Integer-Forcing Precoding for the Gaussian MIMO Broadcast Channel," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 7, pp. 4476–4488, Jul. 2017.

[7] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.

[8] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[9] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[10] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[11] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[12] U. Erez and R. Zamir, "Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[13] B. Nazer and M. Gastpar, "Compute-and-Forward: Harnessing Interference Through Structured Codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[14] C. Feng, D. Silva, and F. R. Kschischang, "An Algebraic Approach to Physical-Layer Network Coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7576–7596, Nov. 2013.

[15] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-Forcing Linear Receivers," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[16] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-Forcing Precoding and Generalized Inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[17] S. Boyd, L. Vandenberghe, and B. Stephen, *Convex Optimization*. Cambridge University Press, Mar. 2004.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Oct. 2012.