Conformação de Feixe e Controle de Potência em Arranjo de Antenas utilizando Aprendizagem por Reforço

Náthalee Cavalcanti de Almeida
UFERSA
Pau dos Ferros, Brasil
nathalee.almeida@ufersa.edu.br

Marcelo A. C. Fernandes e Adrião D. D. Neto
DCA – CT - UFRN
Natal, Brasil
mfernandes@dca.ufrn.br e adriao@dca.ufrn.br

***Resumo*—O emprego isolado ou em conjunto de conformação de feixe e controle de potência apresenta vantagens e desvantagens, dependendo da aplicação. A utilização em conjunto destas técnicas tem demonstrado ser eficaz em aplicações que envolvem a supressão de sinais interferentes provenientes de diferentes fontes. No entanto, é necessário identificar metodologias eficientes para aplicação entre as duas técnicas. Este trabalho propõe um algoritmo utilizando Aprendizagem por Reforço (AR), mais especificamente o Q-*learning*, para determinar a técnica mais adequada entre conformação de feixe e controle de potência em arranjo de antenas. Simulações mostram que o AR é eficaz para a aplicação de uma política de comutação envolvendo conformação de feixe e controle de potência, buscando aproveitar características positivas de cada umas delas na recepção do sinais.**

Palavras-chave—conformação de feixe; controle de potência; Arranjo de antenas; Q-learning.

I. INTRODUÇÃO

O desenvolvimento de arranjos adaptativos de antenas utilizando conformação de feixe em conjunto com controle de potência pode ser utilizado para obter um melhor desempenho do sistema, com menor consumo de energia para transmissão [1-5]. Técnicas de Inteligência Artificial (IA) tem sido bastante utilizadas em problemas que utilizam conformação de feixe e controle de potência [6-10], entre outros.

Em diversos trabalhos, a atualização da potência de transmissão ocorre apenas após a convergência do processo de conformação de feixe, ocasionando testes sucessivos de convergência e perda de desempenho em ambientes não-estacionários. Diante disso, a proposta deste artigo é desenvolver um agente (algoritmo) inteligente, utilizando aprendizagem por reforço (AR), para estabelecer uma política ótima de acionamento entre duas técnicas: conformação de feixe (CF) e controle de potência (CP) em arranjo de antenas, buscando aproveitar características individuais de cada uma das técnicas de acordo com um limiar de Relação Sinal-Interferência mais Ruído (SINR). O grande desafio da AR, neste caso, é escolher a melhor ação (CF ou CP) que, baseado no estado do aprendiz (SINR), possa fornecer a técnica mais adequada ao sistema. O algoritmo de AR a ser utilizado é o *Q-learning*, com uma política ε-gulosa, treinado através do método *off-line*. Neste caso, a aprendizagem das técnicas mais adequadas, que indicam a ação a escolher para cada estado do aprendiz, dar-se-á por reforço, através de uma estrutura de parâmetros adaptativos sobre os quais o algoritmo opera.

Como contribuições deste artigo estão: o algoritmo AR proposto apresenta comutação inteligente entre CF e CP acarretando independência de execução entre eles; método com complexidade reduzida, com menor número de operações necessárias, visto que uma técnica não é seguida da outra; a metodologia AR proposta é independente da técnica de ajuste de parâmetros da CF e do algoritmo de controle de potência.

Este artigo está organizado como segue. A seção II contempla o funcionamento de arranjo adaptativo de antenas, com uma descrição básica de arranjo de antenas, seguida pelo modelo dos sinais de entrada e saída do arranjo, além da forma como é resolvido o problema de conformação de feixe, utilizando o algoritmo LMS para ajuste de pesos e o controle de potência; Na seção III é apresentado o algoritmo *Q-learning*, baseado na AR; na seção IV é mostrado a proposta do agente inteligente utilizando AR; na Seção V são mostrados os resultados obtidos mediante aplicação do agente ao arranjo de antenas. As considerações finais são apresentadas na Seção VI.

II. ARRANJO ADAPTATIVO DE ANTENAS

A Figura 1 apresenta um diagrama funcional de um arranjo linear de antenas adaptativas com *K* elementos. Cada *k*-ésimo elemento do arranjo de antenas possui espaçamento *d* que geralmente deve ser igual a *λ/2* (λ é o comprimento de onda). O sinal $x_{i,k}(n)$ recebido pelo *k*-ésimo elemento da antenna é dado por (1):

$$x_{i,k}(n) = \sum_{i=1}^{M} \rho_i v_i(n) e^{-j(k-1)\left(\frac{2\pi}{\lambda}\right)d\cos(\theta_i)} + r_k(n) \quad (1)$$

onde $v_i(n)$, ρ_i, *e* θ_i são o sinal, a atenuação e o ângulo de chegada para a *i*-ésima fonte, respectivamente. Todas as fontes tem o mesmo comprimento de onda e $r_k(n)$ é o ruído

associado com cada elemento da antenna. O sinal $v_i(n)$ é modelado como um sinal de banda estreita.

O desempenho da recepção, em geral, é medido de forma eficaz utilizando a SINR na saída do arranjo. A SINR estima a razão entre o sinal de interesse com vista ao ruído mais interferência, fornecendo uma medida da qualidade da comunicação.

Em arranjo de antenas, o processo adaptativo é realizado por meio do ajuste dos coeficientes, associados com cada um dos elementos do arranjo. Este ajuste é realizado utilizando um processador de sinal e considera um critério estabelecido para o desempenho do sistema, o que poderia ser o SINR, o erro quadrático médio, a taxa de erro de bit (BER), ou qualquer outro parâmetro [11-12].

No sistema ilustrado na Fig. 1, o sinal na saída do arranjo para o *i*-ésimo usuário é dado por (2):

$$y_i(n) = \boldsymbol{w}_{i,K}^H(n)\boldsymbol{x}(n) \tag{2}$$

onde $\boldsymbol{w}_{i,K}^H(n)$ é o vetor de pesos para o *i*-ésimo usuário do sistema, e o índice H representa o transposto conjugado hermitiano. O sinal de saída do arranjo $y_i(n)$ é comparado com a resposta desejada $d_i(n)$, a diferença entre eles é denominado erro de estimação, conforme ilustrado na Fig. 1. Como apresentado em [13], o sinal de referência (ou sinal desejado) é uma sequência de treinamento compreendida pelo arranjo de antenas, enviado periodicamente pelas fontes.

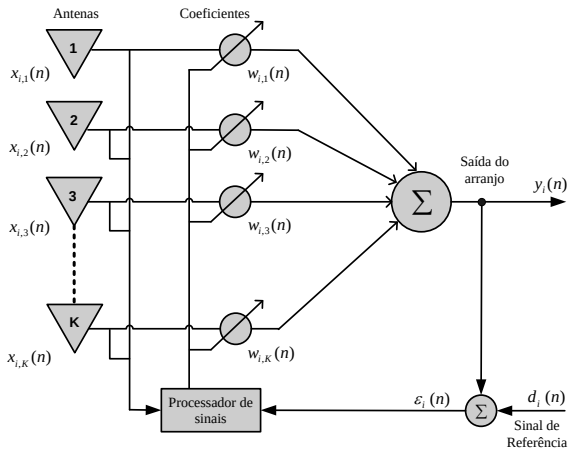


Fig. 1. Diagrama Funcional de um Arranjo Adaptativo

A. *Conformação de Feixe*

O objetivo da conformação de feixe é ajustar o vetor de pesos de tal forma que se consiga obter a máxima SINR na saída do arranjo. Isto pode ser alcançado pela minimização da interferência total na saída do arranjo, mantendo-se constante o ganho para o sinal de interesse [12-14]. Tomando-se a saída do arranjo dada em (2), a potência média total de saída (em *W*) para o *i*-ésimo usuário pode ser escrita como:

$$P_i = E[y_i(n)^2] = E[\boldsymbol{w}_{i,K}^H(n)\boldsymbol{x}(n)\boldsymbol{x}^H(n)\boldsymbol{w}_{i,K}(n)] \tag{3}$$

onde *E* é operador da média. Definindo

$$R = E[\boldsymbol{x}(n)\boldsymbol{x}(n)^H] \tag{4}$$

como a matriz de autocorrelação do sinal de entrada, a potência média total de saída do arranjo é expressa como:

$$P_i = \boldsymbol{w}_{i,K}^H(n)R\boldsymbol{w}_{i,K}(n) \tag{5}$$

O vetor de pesos que maximiza a SINR de saída do arranjo pode ser então encontrado pelo seguinte problema de minimização:

$$\min_{(\boldsymbol{w}_{i,1},..,\boldsymbol{w}_{i,K},P_1,...,P_K)} \left(\sum_{i=1}^{K} P_i\right) \quad \text{subject to } SINR_i \geq \delta_i \tag{6}$$

onde δ_i, P_i e $\mathbf{w}_{i,K}$ denotam, respectivamente, o menor valor admissível de SINR em dB (nível limiar preestabelecido), a potência de transmissão e o vetor de conformação de feixe para o *i*-ésimo usuário. O objetivo é obter um par ótimo formado pelos vetores de peso e potência de transmissão, ou seja, procura-se minimizar a potência total de transmissão, mantendo-se a razão sinal-interferência mais ruído (SINR) acima de um limiar preestabelecido δ_i.

B. *Algoritmo LMS*

O algoritmo LMS é um método baseado em técnicas de procura do gradiente aplicado a funções de erro quadrático médio, utilizado para a aproximação da solução ótima da equação de Wiener-Hopf. O algoritmo é baseado no método da descida mais íngreme [15], em que mudanças no vetor de pesos são feitas ao longo da direção contrária do vetor gradiente estimado. Desta forma:

$$\boldsymbol{w}_{i,K}(n+1) = \boldsymbol{w}_{i,K}(n) - \mu\hat{\nabla}(n) \tag{7}$$

onde μ é uma constante escalar que controla a taxa de convergência e estabilidade do algoritmo (passo de adaptação), e $\hat{\nabla}(n)$ é o vetor gradiente estimado do erro quadrático em relação a $\boldsymbol{w}_{i,K}^H$. O erro é tomado entre a saída do filtro ($y_i(n) = \boldsymbol{w}_{i,K}^H(n)\boldsymbol{x}(n)$) e o sinal de referência $d_i(n)$, ou seja,

$$\varepsilon_i(n) = d_i(n) - \boldsymbol{w}_{i,K}^H(n)x(n). \tag{8}$$

O critério de parada ou critério de convergência dos pesos utilizando o algoritmo LMS é a variação do erro médio quadrático em cada iteração, isto é:

$$|\varepsilon_i{}^2(n) - \varepsilon_i{}^2(n-1)| < \xi \tag{9}$$

onde o limiar ξ é definido pelo projetista com base na sua aplicação que limita a quantidade de iterações.

C. *Controle de Potência*

Um importante benefício resultante da conformação de feixe e do consequente aumento da SINR, é a possibilidade de reduzir-se a potência de transmissão dos sinais. Essa redução possibilita uma melhora da eficiência energética dos sistemas e diminui a interferência entre os usuários, ou seja, estabelece

uma garantia que cada sinal transmita com a menor potência necessária para manter um enlace de boa qualidade.

O controle de potência em arranjo de antenas baseia-se em algum critério de qualidade, neste trabalho, o critério considerado é a SINR. Durante esse processo, as potências são reduzidas de forma que as restrições existentes em (6) sejam satisfeitas. Baseado nos artigos em [1] e [2], a atualização da potência é dada por:

$$P_i(n+1) = P_i(n)\frac{\delta_i}{SINR_i} \tag{10}$$

onde pode ser notado que quando $SINR_i = \delta_i$, a convergência é atingida. No cálculo de (10) é necessário a medição da $SINR_i$ no terminal móvel, porém, esta pode ser estimada através do mínimo erro quadrático médio da etapa de conformação de feixe (LMS) conforme descrito em (11):

$$SINR_i = \frac{1 - min\, E[{\varepsilon_i}^2(n)]}{min\, E[{\varepsilon_i}^2(n)]} \tag{11}$$

em que $SINR_i$ representa a estimativa da relação sinal-interferência mais ruído para cada usuário:

Reesecrevendo (11) em função do mínimo erro quadrático médio, chega-se a expressão final para o cálculo da potência:

$$P_i(n+1) = \delta_i P_i(n)\frac{min\, E[{\varepsilon_i}^2(n)]}{1 - min\, E[{\varepsilon_i}^2(n)]} \tag{12}$$

III. Aprendizagem por Reforço

Aprendizagem por Reforço é um paradigma de aprendizagem em que um agente aprendiz busca maximizar uma medida de desempenho baseada nos reforços que recebe ao interagir com um ambiente desconhecido. Sua utilização é recomendada quando não se dispõe de modelos a priori, ou quando não se consegue obter exemplos apropriados das situações as quais o agente aprendiz irá enfrentar. O agente sem conhecimentos prévios tem como objetivo aprender através da interação com o ambiente, recebendo recompensas por suas ações e assim descobrindo a política ótima [16].

Em um sistema de aprendizagem por reforço, o estado do ambiente é representado por um conjunto de variáveis, denominados de espaço de estados, que são percebidas pelos sensores do agente. Uma ação escolhida pelo agente muda o estado do ambiente e o valor desta transição de estados é passado ao ambiente através de um sinal escalar de reforço, denominado recompensa. O objetivo do método é levar o agente a escolher a sequência de ações que tendem a aumentar a soma dos valores de recompensa.

O agente se move autonomamente no espaço de estados interagindo com o ambiente e aprendendo sobre ele através de ações por experimentação. Cada vez que o agente realiza uma ação, uma entidade externa treinadora (crítico) ou mesmo o ambiente pode lhe dar uma recompensa ou uma penalidade, indicando o quão desejável é alcançar o estado resultante [16]. Desta forma, o reforço nem sempre significa avanço, como pode também inibir o agente em relação à ação executada. A Fig.2 representa um esquema genérico da ideia de aprendizagem por reforço.

A AR busca orientar o agente a tomar ações que venham a maximizar (ou minimizar) o somatório dos sinais de reforço (recompensa ou punição numérica) recebidos ao longo do tempo, denominado retorno esperado, o que nem sempre significa maximizar o reforço imediato a receber [16].

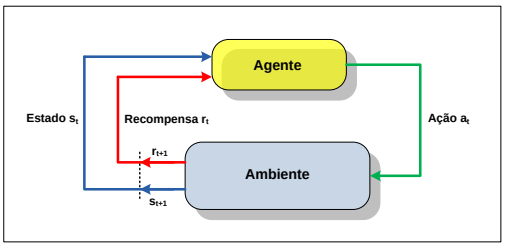


Fig. 2. Esquema de interação entre o agente e o ambiente.

De acordo com [16], a expressão para o somatório dos sinais de reforço em um horizonte infinito é dada por :

$$R_t = r_{t+1} + \gamma . r_{t+2} + \cdots = r_{t+1} + \sum_{k=1}^{\infty} \gamma^k . r_{t+k+1} \tag{13}$$

onde R_t representa o retorno (somatório dos reforços) recebido ao longo do tempo, r_{t+k} o sinal de reforço imediato, e γ é o fator de desconto, definido no interval $0 \leq \gamma \leq 1$ para garantir que R_t seja finito. Se $\gamma = 0$, o agente tem uma visão míope dos reforços, maximizando apenas os reforços imediatos. Se γ =1, a visão do reforço abrange todos os estados futuros dando a mesma importância para ganhos neste momento e qualquer ganho futuro.

O comportamento que o agente deve seguir para alcançar a maximização (ou minimização) do retorno é chamado de política e pode ser expresso por π. Segundo [16], uma política $\pi(s,a)$ é um mapeamento de estados s em ações a, tomadas naquele estado, e representa a probabilidade de seleção de cada uma das ações possíveis, de tal forma que as melhores ações correspondem às maiores probabilidades de escolha. Quando este mapeamento maximiza a soma das recompensas, tem-se o que se chama política ótima.

Para avaliar a qualidade das ações tomadas pelo agente, aplica-se o conceito de "função de valor estado-ação" $Q(s,a)$, a qual é um valor que representa uma estimativa do quão bom é para o agente estar em um dado estado s e tomar uma dada ação a, quando se está seguindo uma política π qualquer.

Nesse trabalho optou-se pelo uso do algoritmo *Q-learning*, desenvolvido por [17], que dentre suas vantagens está o fato dele aproximar diretamente o valor ótimo de $Q(s,a)$, independente da política utilizada. Os valores de $Q(s,a)$ são atualizados de acordo com (14):

$$Q(s,a) = Q(s,a) + \alpha[r_{t+1} + \gamma . max_a Q(s_{t+1}, a) - Q(s,a)] \tag{14}$$

onde α é a taxa de aprendizagem $0 \leq \alpha \leq 1$ e γ é a taxa de desconto $0 \leq \gamma \leq 1$.

O Algoritmo 1 apresenta o algoritmo *Q-learning*. O episódio mencionado neste algoritmo é caracterizado pela sequência de estados terminando em um estado final [16].

Algoritmo 1**.** Algoritmo *Q-Learning*

1: Inicialize *Q (s, a)* de forma arbitrária;
2: **Repita** (para cada episódio)
3: Inicializar *s;*
4: **Repita**
5: Escolher *a* para *s* usando a política π;
6: Dado a ação *a*, observe *r, s'*;
7: $Q(s,a) = Q(s,a) + \alpha[r_{t+1} + \gamma . max_a Q(s_{t+1}, a) - Q(s,a)]$
8: $s \rightarrow s'$;
9: **até** encontrar um estado final;
10: **até** atingir número de episódios

IV. TÉCNICA PROPOSTA

O objetivo a ser alcançado pela modelagem da Aprendizagem por Reforço deste problema é encontrar a política ótima que indique quais são as técnicas mais adequadas que o agente deve escolher dentre as ações disponíveis: Conformação de Feixe e Controle de Potência.

O Algoritmo 2 apresenta o processo de obtenção da matriz *Q* (14) que indicará o funcionamento do agente.

Algoritmo 2. Algoritmo *Q-learning* em Arranjo de Antenas

1: **Requer:** $(\alpha, \varepsilon, \gamma, \xi, M)$
2: Inicialize *Q (s, a)*
3: **Enquanto** episódios máximos não atingidos **Faça**
4: Inicialize *s* *% estado inicial*
5: **Enquanto** não encontrar um estado final **Faça**
6: Selecione *a* de acordo com a regra *ε*-gulosa; *% a =1(CF) ou a=2 (CP)*
7: **Se** a==1 *% Conformação de feixe*
8: **Enquanto** $|{\varepsilon_i}^2(n) - {\varepsilon_i}^2(n-1)| < \xi$ **Faça**
9: **Para** i=1, 2, …, K
10: $w_i(n+1) = w_i(n) + \mu \boldsymbol{x}(n)\varepsilon_i(n)$
11: **Fim_Para**
12: **Fim- Enquanto**
13: **Fim - Se**
14: **Se** a==2 *% Controle de Potência*
15: **Para** i=1, 2,…, M
16: $P_i(n+1) = \delta P_i(n) \frac{minE[\varepsilon_i^2(n)]}{1 - minE[\varepsilon_i^2(n)]}$
17: **Fim-Para**
18: **Fim - Se**
19: Observe *r* *% Recompensa*
20: $s' = (1 - min\_mse)/min\_mse$ *% Novo Estado*
21: $Q(s,a) = Q(s,a) + \alpha[r_{t+1} + \gamma . max_a Q(s_{t+1}, a) - Q(s,a)]$
22: s → s'
23: **Fim-Enquanto**
24: **Fim-Enquanto**
25: **retorno** *Q(s,a)* *%Matriz dos Q-valores*

Do ponto de vista da Aprendizagem por Reforço, o problema pode ser modelado como segue: o estado do ambiente é representado por um conjunto discreto de valores de SINR, ou seja, $S = \{SINR_1, SINR_2, \ldots, SINRm\}$, onde *m* representa o número de estados. Em cada estado $s \in S$, o agente decisor deve escolher uma ação *a* de um conjunto de ações disponíveis no estado *s*, denotada por *A(s)*. As possíveis ações disponíveis para cada estado são: Conformação de Feixe (CF) e Controle de Potência (CP). O algoritmo Q-*learning* pode gerenciar a decisão de explorar ou tirar proveito, utilizando uma política denominada *ε*-gulosa. Esta política é definida no algoritmo por escolher a ação que possui o maior valor de utilidade do estado (critério guloso), com probabilidade (1- *ε*) e de ação aleatória com probabilidade *ε*.

Como consequência da escolha de uma ação $a \in A(s)$ a partir do estado s no instante de decisão *t*, o agente decisor recebe uma recompensa $r_t(s,a)$. A escolha da ação *a* altera a percepção do agente em relação ao ambiente, levando-o a um novo estado s_{t+1} que o conduzirá ao novo instante de decisão *t+1*. Quando a recompensa é positiva é visto como lucro ou prêmio, quando negativa é vista como custo ou punição. Na definição da função de retorno (recompensa), deve-se indicar o objetivo a ser alcançado pelo algoritmo. Portanto, a recompensa foi definida como positiva quando o SIR se aproxima do limiar e negativo, caso contrário.

O Algoritmo 3 mostra o funcionamento do agente onde *j* é chamado de ciclo de processamento e corresponde a escolher um estado inicial, executar uma ação e alcançar um novo estado.

Algoritmo 3**.** Algoritmo de Funcionamento do agente

1: **Se** (altera ambiente)
2: treinamento *% Algoritmo 2*
3: **Senão**
4: Selecione um estado inicial *s*
5: **Para** j **de** 1 **até** valor_max **Faça**
6: Escolha a_{max} para o estado *s* e execute
7: $s \rightarrow s'$
8: ***Fim-Para***

V. SIMULAÇÕES E RESULTADOS

O funcionamento do algoritmo proposto foi demonstrado para duas fontes (*M=2*) com ângulos de chegada de 90° e 30° para sinais de interesse e interferente, respectivamente. O SINR alvo é de 2 dB, e as potências iniciais foram fixados em 1*W* para ambas as fontes. Todas as fontes foram modeladas com sinais binários com distribuição uniforme e aleatória. O ruído em cada elemento da antena foi modelado como um ruído gaussiano aditivo branco (AWGN) com variância ($\boldsymbol{\sigma^2}$). Os parâmetros utilizados na simulação são listados na Tabela 1.

Tabela 1**.** Parâmetros de Simulação

Paramêtros	**Valores**
Número de elementos do arranjo (K)	8
Número de sinais envolvidos (M)	2
Potência inicial de transmissão (P_0)	1
Restrição da SINR (δ)	2
Passo de adaptação (μ)	0.001
Variância de ruído (σ^2)	0.1
Distância entre cada elemento do arranjo (d)	0.5
Coeficiente de Aprendizagem (α)	0.1
Taxa de Desconto (γ)	0.9
Parâmetro que regula o critério guloso (ε)	0.2

Neste artigo, um arranjo linear de antenas com 8 elementos e duas fontes de sinal foi utilizada: uma fonte de interesse e uma fonte interferente. Um arranjo linear de *K* elementos pode criar até *K-1* nulos na direção da fonte interferente. Quando o número de fontes indesejáveis (interferentes) está próximo de tal limite, a atenuação de sinais indesejáveis é reduzido e há ganhos em excesso (maiores que o ganho atribuído ao sinal desejado) na proximidade dos ângulos desejados e interferentes. Assim, a utilização de duas fontes está dentro dos limites estabelecidos para o desempenho do sistema.

A distância entre os elementos do arranjo (d) é limitado pelo valor $\boldsymbol{\lambda/2}$. Esta limitação evita a produção e sobreposição de lóbulos laterais. As potências de transmissão unitárias foram usadas para inicializar o algoritmo de controle de potência e facilitar os cálculos. A escolha do $\boldsymbol{\mu}$ no passo de adaptação foi determinada experimentalmente de modo a proporcionar estabilidade do algoritmo. Quanto maior o passo de adaptação, maior a velocidade de convergência. Entretanto, o excesso de erro também se torna maior, o que é indesejável. Os valores de $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, e $\boldsymbol{\varepsilon}$ foram obtidas depois de várias simulações.

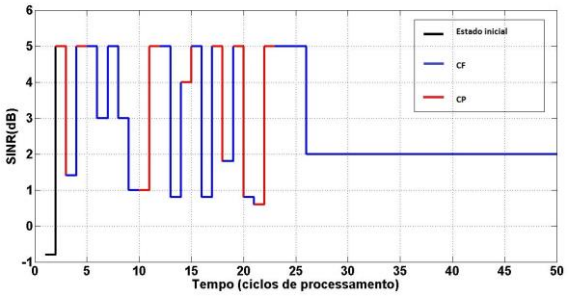
Os estados do ambiente são representados pelos valores de SINR discretos entre -0.8 a 5 referenciados por índices (1 a 18). Cada ação é referenciada por um índice (1 ou 2), em que a ação 1 representa Conformação de Feixe e a ação 2 representa o Controle de Potência. Foi adotada uma política ε-gulosa, com 80% de chance de escolha da ação com maior estimativa.

Os resultados da simulação são apresentados na Tabela 2 e indicam as políticas obtidas ao final de cada um dos episódios destacados, onde cada linha corresponde a um valor de SINR e cada coluna corresponde ao processo de conformação de feixe (CF) ou controle de potência (CP). Os valores referenciados na Tabela 2 correspondem à probabilidade de escolha de cada técnica para as faixas de valor do SINR discretizado. A política obtida após o 250° foi destacada e diz respeito à política ótima que foi adotada para o teste do agente.

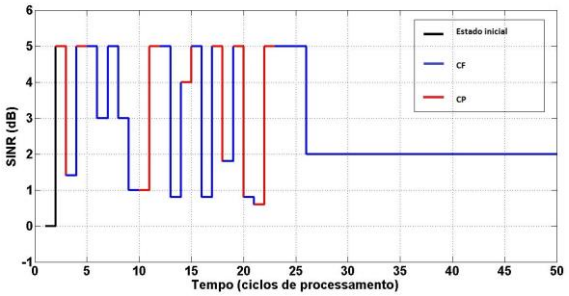
Tabela 2. Evolução da política do agente

		Política 10		**Política 50**		**Política 250**	
ÍNDICE	**SINR (dB)**	**CF**	**CP**	**CF**	**CP**	**CF**	**CP**
1	-0.8	0.5	0.5	1	0	0	1
2	-0.6	0.5	0.5	1	0	1	0
3	-0.4	0.5	0.5	1	0	0	1
4	-0.2	0	1	1	0	1	0
5	0	0	1	0	1	0	1
6	0.2	0	1	0	1	0	1
7	0.4	0	1	0	1	0	1
8	0.6	0	1	0	1	0	1
9	0.8	1	0	1	0	0	1
10	1	0	1	0	1	0	1
11	1.2	0	1	1	0	0	1
12	1.4	0	1	0	1	0	1
13	1.6	0	1	0	1	0	1
14	1.8	0	1	0	1	0	1
15	2	**Destino**		**Destino**		**Destino**	
16	3	1	0	0	1	1	0
17	4	1	0	1	0	1	0
18	5	1	0	1	0	1	0

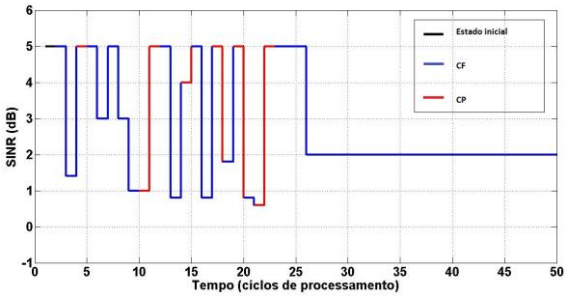
As Figuras 3a, 3b e 3c apresentam no eixo das ordenadas os estados (SINR) e a curva indica a evolução do SINR até alcançar a SINR alvo (δ=2).



(a)



(b)



(c)

Fig. 3. (a) Resposta do sistema. Agente partindo do SINR=-0.8 dB; (b) Resposta do sistema. Agente partindo do SINR= 0 dB; (c) Resposta do sistema. Agente partindo do SINR=5 dB.

Analisando a Tabela 2 que representa a matriz Q resultante da execução do algoritmo 2, percebe-se que o Controle de Potência foi escolhido em muitos estados em detrimento da Conformação de Feixe, isso demonstra a independência do algoritmo em escolher a política ótima, diferentes dos algoritmos apresentados em [1] e [2] que somente utiliza a potência quando a convergência da conformação de feixe é alcançada.

A sequência de chaveamento das ações para um SINR inicial de -0.8 dB está ilustrada na Fig. 4, com o eixo das ordenadas representando as ações (Conformação de Feixe e Controle de Potência) executadas e a curva indica a ordem de execução em cada ciclo de processamento. O índice 1 e 2

representam a Conformação de Feixe e Controle de Potência, respectivamente. Esta curva também comprova que a política de comutação ótima entre as duas técnicas (CF e CP) não segue o modelo tradicional apresentados nos artigos citados no qual, espera-se a convergência da conformação de feixe para comutar para o controle de potência.

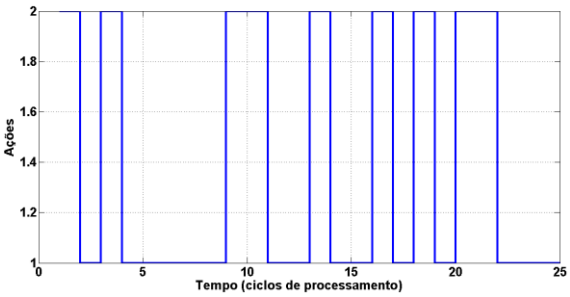


Fig. 4. Sequência de Chaveamento entre as duas técnicas (CF e CP)

VI. CONCLUSÕES

A utilização de conformação de feixe e controle de potência em arranjo de antenas de forma individual tem suas vantagens, porém ao utilizá-los em conjunto ocasiona um melhor desempenho do sistema. Diversos estudos tem utilizado conformação de feixe e controle de potência de forma conjunta, porém este artigo apresenta um novo método de controle empregando uma combinação dessas técnicas. O algoritmo apresentado utiliza aprendizagem por reforço para obter a política ótima de seleção entre conformação de feixe e controle de potência.

Pode-se observar que a técnica proposta reduz o custo computacional, visto que as técnicas são selecionadas independentemente. Por exemplo, de acordo com a Tabela 2, utilizando a política ótima obtida após 250 episódios, verifica-se que em muitos casos não foi necessário a execução do algoritmo LMS. Isto resultou em menor custo computacional e redução da complexidade do método proposto.

Mediante as simulações realizadas, pode-se concluir que a aprendizagem por reforço oferece uma maneira eficaz de implementar uma política ótima de alternância entre conformação de feixe e contole de potência em arranjo de antenas, beneficiando-se das vantagens de ambas as técnicas e fazendo com que seja escolhida a técnica mais adequada ao sistema de acordo com a situação de SINR.

REFERÊNCIAS

[1] C. A. Pitz, M.G. Vanti, O. J. Tobias e R. Seara, "Adaptive Beamforming for Antenna Arrays in Cellular Systems Based on a Duality between Uplink and Downlink Channels," In Proceedings of 7th International Telecommunications Symposium (ITS), Manaus, September 2010.

[2] Y. Huang, C.W. Tan e B.D. Rao,"Joint beamforming and power control in coordinated multicell: Max-min duality, effective network and large system transition," IEEE Trans. Wirel. Commun., vol. 12, pp. 2730-2742, 2013.

[3] M. R. A. Khandaker e Y. Rong, "Joint power control and beamforming for peer-to-peer MIMO relay systems," In Proceedings of the 2011 International Conference on Wireless Communicantions Signal Processing (WCSP), Nanjing, China, November 2011.

[4] X. Lu, W. Li, A. Tolli, M. Juntti, E. Kunnari e O. Piirainen, "Joint power control, receiver beamforming and adaptive multi base station coordination for uplink wireless communications," In Proceedings of the 21st IEEE International Symposium on Personal, Indoor and Mobile radio Communications Workshops, Instanbul, Turkey, pp. 446–450, September 2010.

[5] Z. Li, C. Yin e G. Yue, "A novel approach to joint beamforming and power control for the coordinated multicell multi-antenna system," In Proceedings of the 8th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), Shanghai,China, pp. 1–4, September 2012.

[6] N. Noori, S. M. Razavizadeh e A. Attar, "Joint beamforming and power control in MIMO cognitive radio networks," IEICE Electonics Express,vol *7*, pp. 203–208, 2010.

[7] J. Yoshida e A. Hirose, "Beamforming for impulse-radio UWB communication systems based on complex-valued spatio-temporal neural networks," In Proceedings of the International Symposium on Electromagnetic Theory, Hiroshima, Japan, pp. 848–851, May 2013.

[8] Z. D. Zaharis, C. Skeberies, T. D. Xenos, P. I. Lazaridis e J. Cosmas,"Design of a novel antenna array beamformer using neural networks trained by modified adaptive dispersion invasive weed optimization based data", *IEEE Trans. Broadcast.*, *59*, 455–460, 2013..

[9] K. Terabayashi e A. Hirose, "Ultra-short-pulse acoustic imaging using complex-valued spatio-temporal neural-network for null-steering: Experimental results," In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Beijing, China, pp. 3410–3413, July 2014.

[10] Y. Liu, P. Zhang e T. Hain, "Using neural network front-ends on far field multiple microphones based speech recognition," In Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 5542–5546, May 2014.

[11] C. A. Balanis, P. I. Ioannides, "Introduction to Smart Antennas," Morgan & Claypool Publishers, Synthesis Lecture on Antennas, California, EUA, 2007.

[12] R. A. Monzingo e T. W. Miller, "Introduction to Adaptive Arrays,"Scitech Publishing Inc.: Raleigh, NC, USA, 2004.

[13] S. Haykin, "Adaptive Filter Theory," 3rd ed., Prentice Hall: Upper Saddle River, NJ, USA, 1996.

[14] C. A. Balanis, "Antenna Theory: Analysis and Design," 3rd ed.; Wiley-Interscience: New York, NY, USA, 2005.

[15] B. Widrow, P. E. Mantey, L. J. Griffiths, B. B Goode, "AdaptiveAntenna Systems," In Proceedings of the IEEE, California, EUA, Vol. 55, pp. 2143-2159, December 1967.

[16] R. Sutton e A. Barto, "Reinforcement Learning: An Introduction," MIT Press:Cambridge,MA, USA, 1998.

[17] C. J. C. H Watkins e P. Dayan, "Q-Learning: Machine Learning," Kluwer Academic Publishers: Dordrecht, The Netherlands, pp. 279–292, 1992.