

Busca do Equilíbrio de Nash em Jogos Não-cooperativos com Atrasos^{*}

Tiago Roux Oliveira^{*} Victor Hugo Pereira Rodrigues^{**}
Miroslav Krstić^{***} Tamer Başar^{****}

^{*} *Universidade do Estado do Rio de Janeiro (UERJ),
Rio de Janeiro, RJ, Brasil
(e-mail: tiagoroux@uerj.br)*

^{**} *Universidade Federal do Rio de Janeiro (UFRJ),
Rio de Janeiro, RJ, Brasil
(e-mail: rodrigues.vhp@gmail.com)*

^{***} *University of California at San Diego (UCSD),
San Diego – CA, USA
(e-mail: krstic@ucsd.edu)*

^{****} *University of Illinois at Urbana-Champaign,
Urbana – IL, USA
(e-mail: basarl@illinois.edu)*

Abstract: In this paper we propose a non-model based strategy for locally stable convergence to Nash equilibrium in a quadratic noncooperative (duopoly) game with arbitrarily delayed player actions. In our noncooperative scenario the players have access only to their own payoff values. In order to compensate distinct delays in the inputs of the two players, we have employed an extremum seeking boundary control via predictor feedback with averaging-based estimates. We apply a small-gain analysis for the resulting Input-to-State Stable hyperbolic PDE-ODE loop as well as averaging theory in infinite dimensions, due to the infinite-dimensional state of the time delays, in order to obtain local convergence results to a small neighborhood of the Nash equilibrium. We quantify the size of these residual sets and corroborate the theoretical results numerically on an example of a two-player game with delays.

Resumo: Neste artigo, propomos uma estratégia de Busca pelo Equilíbrio de Nash baseada no não-conhecimento do modelo que garanta ao sistema de malha fechada convergência local com propriedades de estabilidade em jogos quadrático não-cooperativos, do tipo duopólio, com ações arbitrariamente atrasadas. Em nosso cenário não-cooperativo, os jogadores têm acesso apenas aos seus próprios lucros. Para compensar atrasos distintos nas entradas dos dois jogadores, emprega-se a busca extremal por controle de fronteira via realimentação por predição com estimativas baseadas na média. Aplica-se uma análise de pequenos ganhos para a malha formada por equações diferenciais ordinárias (EDO) - equações diferenciais parciais (EDP), bem como a teoria da média em dimensões infinitas, devido ao estado dimensional infinito dos atrasos temporais, a fim de obter resultados de convergência local a uma pequena vizinhança do Equilíbrio de Nash. Quantifica-se o tamanho desse conjunto residual e os resultados teóricos validados através de um exemplo de simulação.

Keywords: Extremum seeking; Nash equilibrium; (Non)cooperative games; Time delays; Predictor feedback; Averaging in infinite dimensions.

Palavras-chaves: Busca extremal; Equilíbrio de Nash; Jogos Não-cooperativos; Atraso Temporal; Realimentação por Preditor; Média em dimensões infinitas.

1. INTRODUÇÃO

A teoria dos jogos fornece uma estrutura teórica para representar situações sociais entre concorrentes com a utilização de modelos matemáticos de interação estratégica entre esses tomadores de decisões racionais (Fudenberg and Tirole, 1991). O emprego da teoria dos jogos para projetar, modelar e otimizar

sistemas de engenharia, comportamentais e financeiros tornam este tópico de pesquisa uma ferramenta extremamente importante com uma ampla gama de aplicações (Han et al., 2019; Başar and Zaccour, 2018). De fato, podemos encontrar numerosos resultados para teoria e prática na literatura correspondente a jogos diferenciais (Wang et al., 2018; Cotrina and Zúñiga, 2018; Aussel and Svensson, 2019; Alasseur et al., 2020).

Os jogos podem ser classificados em duas categorias: *cooperativos* e *não-cooperativos* (Başar and Olsder, 1999). Um jogo é cooperativo se os jogadores forem capazes de formar compromissos vinculativos resultando em um lucro coletivo. Um jogo é não-cooperativo se os jogadores não puderem formar alianças.

^{*} O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Os autores também agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Em jogos não-cooperativos, uma boa estratégia para tomada de decisão é a análise do *Equilíbrio de Nash*. O Equilíbrio de Nash pode ser interpretado como um perfil específico de ações (entradas) que otimiza o lucro (*payoff*) de todos os jogadores. Dessa forma, o Equilíbrio de Nash assegura a impossibilidade de que qualquer jogador possa aumentar seu *payoff* tomando decisões unilateralmente (Başar and Olsder, 1999).

O desenvolvimento de algoritmos para alcançar a convergência ao equilíbrio de Nash tem sido foco de pesquisadores por várias décadas. Em relação à busca extremal, os autores Frihauf et al. (2012) estudam o problema da busca do equilíbrio de Nash, em tempo real, em jogos estáticos não-cooperativos com N jogadores. Ao utilizar a busca extremal com perturbações sinusoidais (Krstić and Wang, 2000), os jogadores alcançam de maneira local e estável suas estratégias de Nash sem a necessidade de qualquer informação do modelo.

Por outro lado, atrasos são uns dos fenômenos mais comuns que surgem na prática da engenharia e indústria, envolvendo problemas em áreas como: virtualização de redes, redes definidas por *software*, computação em nuvem, internet das coisas, redes verdes e segurança (Han et al., 2019; Alpcan and Başar, 2011). Portanto, a motivação para empregar a busca extremal para otimização desses processos com atrasos, comumente modelados com a estrutura teórica dos jogos, é clara e justificada. Podemos até encontrar publicações sobre jogos diferenciais com atrasos (Pamen, 2015; Carmona et al., 2018), mas a literatura não abordou isso no contexto de busca extremal.

Nesse sentido, este artigo apresenta uma proposta para a Busca do Equilíbrio de Nash em jogos não-cooperativos, do tipo duopólio, quando são considerados atrasos nas ações dos tomadores de decisões (jogadores). Desenvolve-se um resultado para jogos com atrasos, em que os jogadores estimam apenas as partes diagonais da matriz Hessiana de forma que os termos fora da diagonal principal são suficientemente pequenos permitido a utilização da teoria de pequenos ganhos para o sistema médio formado por uma cascata EDP-EDO (Karafyllis and Krstić, 2018).

2. NOTAÇÃO E TERMINOLOGIA

Denota-se a derivada parcial de uma função $u(x, t)$ como $\partial_x u(x, t) = \partial u(x, t) / \partial x$, $\partial_t u(x, t) = \partial u(x, t) / \partial t$, ou, convenientemente, $u_x(x, t)$ e $u_t(x, t)$, respectivamente. A norma-2 (Euclidina) de um vetor de estado de dimensão finita $\vartheta(t)$ é denotada por barras simples, $|\vartheta(t)|$. Denota-se a norma espacial $\mathcal{L}_2[0, D]$ de $u(x, t)$ como $\|u(t)\|_{\mathcal{L}_2([0, D])}^2 := \int_0^D u^2(x, t) dx$. Para simplificar a notação, omite-se o índice $\mathcal{L}_2([0, D])$ e, portanto, $\|\cdot\| = \|\cdot\|_{\mathcal{L}_2([0, D])}$ (Krstić, 2009). Como definido em Khalil (2002), uma função vetorial $f(t, \epsilon) \in \mathbb{R}^n$ é dita ser de ordem $\mathcal{O}(\epsilon)$ em um intervalo $[t_1, t_2]$, se $\exists k, \bar{\epsilon} : |f(t, \epsilon)| \leq k\epsilon, \forall \epsilon \in [0, \bar{\epsilon}]$ and $\forall t \in [t_1, t_2]$. Na maioria dos casos, não se fornece estimativas precisas para as constantes k e $\bar{\epsilon}$, e interpreta-se $\mathcal{O}(\epsilon)$ como um resíduo de ordem de magnitude ϵ suficientemente pequena. A definição de *Input-to-State Stability* (ISS) para sistemas baseados em EDO, bem como para sistemas baseados em EDP, é considerada como em Khalil (2002) e Karafyllis and Krstić (2018), respectivamente.

3. DUOPÓLIO COM LUCROS QUADRÁTICOS E ATRASOS NAS DECISÕES

Em um duopólio, a otimalidade dos lucros (saídas) dos jogadores J_1 e J_2 , respectivamente $y_1(t) \in \mathbb{R}$ e $y_2(t) \in \mathbb{R}$, não

dependem exclusivamente de suas próprias ações/estratégias (sinais de entrada) $\theta_1(t) \in \mathbb{R}$ e $\theta_2(t) \in \mathbb{R}$. Existe um perfil de sinais de entrada $\theta^* = [\theta_1^*, \theta_2^*]^T \in \mathbb{R}^2$, de modo que a estratégia de cada jogador seja uma entrada ideal para a função de lucro do outro jogador J_1 e J_2 . Quando cada jogador encontra sua entrada ideal, $\theta(t) = \theta^*$, diz-se que o equilíbrio de Nash é alcançado (Fudenberg and Tirole, 1991). Apesar do grande número de publicações sobre a Busca pelo Equilíbrio de Nash (Frihauf et al., 2012), o estudo do problema na presença de atrasos ainda está aberto.

Portanto, consideramos jogos em que lucro de cada jogador é dado por uma função quadrática, expressa como uma combinação estritamente côncava de suas ações atrasadas

$$\begin{aligned} J_1(\theta(t-D)) &= \frac{H_{11}^1}{2} \theta_1^2(t-D_1) + \frac{H_{22}^1}{2} \theta_2^2(t-D_2) + \\ &+ \epsilon H_{12}^1 \theta_1(t-D_1) \theta_2(t-D_2) + \\ &+ h_1^1 \theta_1(t-D_1) + h_2^1 \theta_2(t-D_2) + c_1, \quad (1) \\ J_2(\theta(t-D)) &= \frac{H_{11}^2}{2} \theta_1^2(t-D_1) + \frac{H_{22}^2}{2} \theta_2^2(t-D_2) + \\ &+ \epsilon H_{21}^2 \theta_1(t-D_1) \theta_2(t-D_2) + \\ &+ h_1^2 \theta_1(t-D_1) + h_2^2 \theta_2(t-D_2) + c_2, \quad (2) \end{aligned}$$

onde $J_1(\theta), J_2(\theta) : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\theta_1(t-D_1), \theta_2(t-D_2) \in \mathbb{R}$ são variáveis que representam as ações dos Jogadores atrasadas de $D_1, D_2 \in \mathbb{R}^+$ unidades de tempo, $H_{jk}^i, h_j^i, c_i \in \mathbb{R}$ são constantes, $H_{ii}^i < 0$, e $\epsilon_{jk}^i = \epsilon_{kj}^i > 0, \forall i, j, k \in \{1, 2\}$.

Assume-se que as entradas possuem atrasos conhecidos, constantes, distintos e que são ordenados para que

$$D = \text{diag}\{D_1, D_2\}, \quad 0 \leq D_1 \leq D_2. \quad (3)$$

Além disso, um sinal em $f \in \mathbb{R}^N$, a notação f^D indica

$$f^D(t) := f(t-D) = [f_1(t-D_1), \dots, f_N(t-D_N)]^T. \quad (4)$$

As funções de *payoff* quadráticas são de particular interesse na teoria dos jogos, primeiro porque constituem aproximações de segunda ordem a outros tipos de funções não quadráticas e, segundo, porque são analiticamente tratáveis, levando, em geral, a soluções que fornecem *insights* sobre as propriedades e características conceituais do equilíbrio de Nash (Başar and Olsder, 1999).

Fornecemos aqui, em termos matemáticos, a definição de um Equilíbrio de Nash, $\theta^* = [\theta_1^*, \theta_2^*]^T$, em um Jogo de dois jogadores:

$$J_1(\theta_1^*, \theta_2^*) \geq J_1(\theta_1, \theta_2^*) \text{ e } J_2(\theta_1^*, \theta_2^*) \geq J_2(\theta_1^*, \theta_2) \quad (5)$$

Portanto, nenhum jogador é motivado a alterar unilateralmente sua ação de θ^* . No exemplo de um duopólio, $\Theta_1, \Theta_2 \in \mathbb{R}_+$, onde \mathbb{R}_+ indica o conjunto de números reais positivos.

Para determinar a solução de equilíbrio de Nash em jogos quadráticos estritamente côncavos com dois jogadores, deve-se diferenciar J_1 e J_2 , respectivamente, com relação a $\theta_1(t-D_1)$ e $\theta_2(t-D_2)$, igualando-se as expressões resultantes a zero e resolvendo o conjunto de equações assim obtidas, *i.e.*,

$$\begin{cases} H_{11}^1 \theta_1^* + \epsilon H_{12}^1 \theta_2^* + h_1^1 = 0 \\ \epsilon H_{21}^2 \theta_1^* + H_{22}^2 \theta_2^* + h_2^2 = 0 \end{cases}, \quad (6)$$

que pode ser escrito de forma matricial como

$$\begin{bmatrix} H_{11}^1 & \epsilon H_{12}^1 \\ \epsilon H_{21}^2 & H_{22}^2 \end{bmatrix} \begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix} = - \begin{bmatrix} h_1^1 \\ h_2^2 \end{bmatrix}. \quad (7)$$

Definindo a matriz Hessiana H e vetores θ^* e h por

$$H := \begin{bmatrix} H_{11}^1 & \epsilon H_{12}^1 \\ \epsilon H_{21}^2 & H_{22}^2 \end{bmatrix}, \quad \theta^* := \begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix}, \quad h := \begin{bmatrix} h_1^1 \\ h_2^2 \end{bmatrix}, \quad (8)$$

existe apenas um Equilíbrio de Nash em $\theta^* = -H^{-1}h$, se H for inversível:

$$\begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix} = - \begin{bmatrix} H_{11}^1 & \epsilon H_{12}^1 \\ \epsilon H_{21}^2 & H_{22}^2 \end{bmatrix}^{-1} \begin{bmatrix} h_1^1 \\ h_2^2 \end{bmatrix}. \quad (9)$$

Para mais detalhes, consulte (Başar and Olsder, 1999, Capítulo 4).

O objetivo de controle é projetar uma nova estratégia baseada na busca extremal para alcançar o Equilíbrio de Nash em jogos não-cooperativos sujeitos a atrasos distintos nas variáveis de decisão dos jogadores (sinais de entrada).

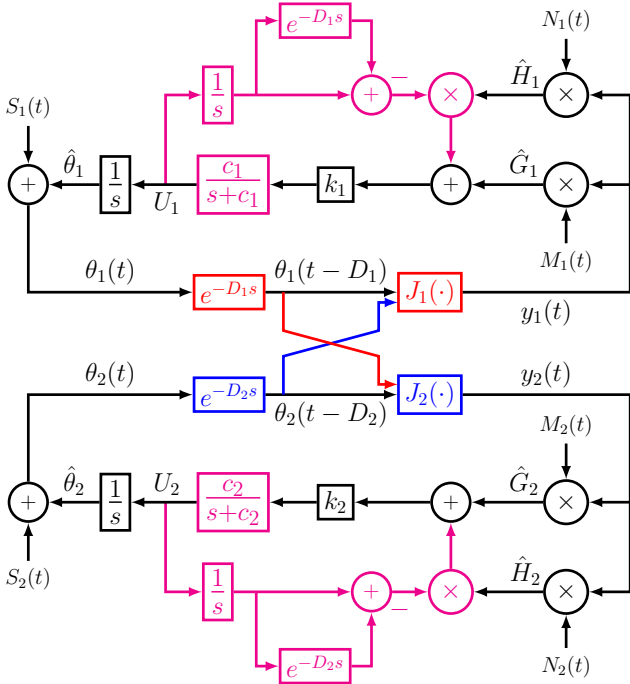


Figura 1. Diagrama de blocos que ilustra a estratégia de Busca pelo Equilíbrio de Nash realizada por cada jogador. Na cor magenta, a realimentação por preditor usada para compensar os atrasos D_1 e D_2 no jogo não-cooperativo.

A Fig. 1 contém um diagrama esquemático que resume a estratégia proposta de Busca pelo Equilíbrio de Nash de cada jogador, no qual as saídas são dada por

$$\begin{cases} y_1(t) = J_1(\theta(t-D)) \\ y_2(t) = J_2(\theta(t-D)) \end{cases}. \quad (10)$$

Os sinais aditivos são

$$\begin{cases} S_1(t) = a_1 \sin(\omega_1 t + \omega_1 D_1) \\ S_2(t) = a_2 \sin(\omega_2 t + \omega_2 D_2) \end{cases}, \quad (11)$$

e os multiplicativos são dados por

$$\begin{cases} M_1(t) = \frac{2}{a_1} \sin(\omega_1 t) \\ M_2(t) = \frac{2}{a_2} \sin(\omega_2 t) \end{cases}, \quad (12)$$

com amplitudes constantes $a_1, a_2 > 0$ e frequências $\omega_1 \neq \omega_2$ tal que

$$\omega_i = \omega'_i \omega = \mathcal{O}(\omega), \quad i = 1 \text{ ou } 2, \quad (13)$$

na qual ω é uma constante positiva e ω'_i é um número racional - uma opção possível é dada em Ghaffari et al. (2012).

Seguindo o paradigma de não-compartilhamento de informações, com base na teoria da média, apenas os elementos da diagonal de H podem ser individualmente recuperados pelos jogadores. Neste sentido, os sinais $N_1(t)$ e $N_2(t)$ são definidos como:

$$\begin{cases} N_1(t) = \frac{16}{a_1^2} \left(\sin^2(\omega_1 t) - \frac{1}{2} \right) \\ N_2(t) = \frac{16}{a_2^2} \left(\sin^2(\omega_2 t) - \frac{1}{2} \right) \end{cases}. \quad (14)$$

Então, a versão média de

$$\begin{cases} \hat{H}_1(t) = N_1(t)y_1(t) \\ \hat{H}_2(t) = N_2(t)y_2(t) \end{cases}, \quad (15)$$

é dada por

$$\begin{cases} \hat{H}_1^{\text{av}}(t) = [N_1(t)y_1(t)]_{\text{av}} = H_{11}^1 \\ \hat{H}_2^{\text{av}}(t) = [N_2(t)y_2(t)]_{\text{av}} = H_{22}^2 \end{cases}. \quad (16)$$

Considerando $\hat{\theta}_1(t)$ e $\hat{\theta}_2(t)$ como estimativas de θ_1^* e θ_2^* , pode-se definir o erro de estimação como:

$$\begin{cases} \tilde{\theta}_1(t) = \hat{\theta}_1(t) - \theta_1^* \\ \tilde{\theta}_2(t) = \hat{\theta}_2(t) - \theta_2^* \end{cases}. \quad (17)$$

A estimativa do gradiente das funções de payoff é dada por

$$\begin{cases} \hat{G}_1(t) = M_1(t)y_1(t) \\ \hat{G}_2(t) = M_2(t)y_2(t) \end{cases}, \quad (18)$$

cujas versões média é

$$\begin{cases} \hat{G}_1^{\text{av}}(t) = H_{11}^1 \tilde{\theta}_1^{\text{av}}(t-D) + \epsilon H_{12}^1 \tilde{\theta}_2^{\text{av}}(t-D) \\ \hat{G}_2^{\text{av}}(t) = \epsilon H_{21}^2 \tilde{\theta}_1^{\text{av}}(t-D) + H_{22}^2 \tilde{\theta}_2^{\text{av}}(t-D) \end{cases}. \quad (19)$$

Nesse ponto, negligencia-se a malha de predição e o filtro passa-baixas (ambos indicados na cor magenta na Fig. 1), as leis de controle $U_1(t) = k_1 \hat{G}_1(t)$ e $U_2(t) = k_2 \hat{G}_2(t)$ poderiam ser obtidas como na abordagem clássica da busca extremal. Neste caso, das equações (17) e (19), tem-se

$$\dot{\tilde{\theta}}^{\text{av}}(t) = KH \tilde{\theta}^{\text{av}}(t-D), \quad (20)$$

onde $\tilde{\theta}^{\text{av}}(t) := [\tilde{\theta}_1^{\text{av}}(t), \tilde{\theta}_2^{\text{av}}(t)]^T \in \mathbb{R}^2$, a matriz $K := \text{diag}\{k_1, k_2\}$ e H dados por (8). A equação (20) significa que, mesmo que KH fosse uma matriz de Hurwitz, o equilíbrio $\tilde{\theta}_e^{\text{av}} = 0$ do sistema médio de malha fechada não seria necessariamente estável para valores arbitrários do atraso D . Isso reforça a necessidade de se empregar a realimentação por predição $U(t) = K\hat{G}(t+D)$ onde $\hat{G}(t) := [\hat{G}_1(t), \hat{G}_2(t)]^T \in \mathbb{R}^2$ - ou sua versão filtrada - conforme ilustrado na Fig. 1.

Por outro lado, é possível encontrar uma forma compacta para o gradiente médio como

$$\hat{G}^{\text{av}}(t) = H \tilde{\theta}^{\text{av}}(t-D), \quad (21)$$

$$\dot{\hat{G}}^{\text{av}}(t) = H U^{\text{av}}(t-D), \quad (22)$$

onde $U(t) := [U_1^{\text{av}}(t), U_2^{\text{av}}(t)]^T \in \mathbb{R}^2$.

Ao longo do artigo, a ideia principal é o projeto de leis de controle (políticas) para cada jogador, a fim de que se alcance uma pequena vizinhança do Equilíbrio de Nash. Para isso, usamos uma estratégia de busca extremal baseada na realimentação por

predição para compensar atrasos múltiplos e distintos nas ações dos jogadores. Basicamente, as leis de controle são capazes de garantir a estabilização exponencial de $\hat{G}^{\text{av}}(t)$ e, consequentemente, de $\hat{\theta}^{\text{av}}(t)$. De (21), fica claro que, se H for invertível, $\hat{\theta}^{\text{av}}(t) \rightarrow 0$ como $\hat{G}^{\text{av}}(t) \rightarrow 0$. Portanto, a convergência de $\hat{\theta}^{\text{av}}(t)$ a uma vizinhança da origem resulta na convergência de $\theta(t)$ a uma vizinhança de (5) via teoria da média Hale and Lunel (1990).

4. DUOPÓLIO NÃO-COOPERATIVO COM ATRASOS

Em nosso jogo não-cooperativo sujeito a atrasos, o objetivo da busca extremal é estimar o vetor de equilíbrio de Nash θ^* , sem a necessidade de compartilhamento de qualquer informação entre os jogadores. Cada jogador precisa medir apenas o valor de sua própria função de *payoff* (10). Nesse sentido, somos capazes de formular o sistema de malha fechada de uma maneira *descentralizada*, onde não é necessário nenhum conhecimento sobre o lucro ou a ação do outro jogador.

4.1 Preditor Descentralizado

Para compensar os atrasos, propomos as seguintes leis de controle baseadas em preditores

$$\begin{cases} \dot{U}_1(t) = -c_1 U_1(t) + c_1 k_1 \left(\hat{G}_1(t) + \hat{H}_1(t) \int_{t-D_1}^t U_1(\tau) d\tau \right) \\ \dot{U}_2(t) = -c_2 U_2(t) + c_2 k_2 \left(\hat{G}_2(t) + \hat{H}_2(t) \int_{t-D_2}^t U_2(\tau) d\tau \right) \end{cases}, \quad (23)$$

para constantes positivas k_1, k_2, c_1 e c_2 .

4.2 Propriedades ISS das EDP

Por uma questão de simplicidade, assume-se que $c_1, c_2 \rightarrow +\infty$ em (23) resultando na seguinte expressão geral:

$$U_i(t) = k_i \left(\hat{G}_i(t) + \hat{H}_i(t) \int_{t-D_i}^t U_i(\tau) d\tau \right), \quad (24)$$

de modo que o sistema atrasado em malha fechada, (22) e (24), possa ser representado na forma de EDP de transporte,

$$\dot{\bar{G}}^{\text{av}}(t) = H u^{\text{av}}(0, t), \quad (25)$$

$$\partial_t u^{\text{av}}(x, t) = \partial_x u^{\text{av}}(x, t), \quad x \in (0, D), \quad (26)$$

$$u^{\text{av}}(D, t) = U^{\text{av}}(t). \quad (27)$$

A relação entre u^{av} e U^{av} é dada por $u^{\text{av}}(x, t) = U^{\text{av}}(x+t-D)$.

No contexto da *Abordagem por Redução* (Artstein, 1982), ou *Atribuição de Espectro Finito*, emprega-se a transformação

$$\begin{aligned} \bar{G}^{\text{av}}(t) &= \hat{G}^{\text{av}}(t) + H \int_{t-D}^t U^{\text{av}}(\tau) d\tau \\ &= \hat{G}^{\text{av}}(t) + H \int_0^D u^{\text{av}}(\xi, t) d\xi. \end{aligned} \quad (28)$$

Por meio de algumas manipulações matemáticas, não é difícil enxergar que \bar{G} satisfaz

$$\dot{\bar{G}}^{\text{av}}(t) = H U^{\text{av}}(t). \quad (29)$$

Agora, após adicionar e subtrair os termos representados em azul e vermelho em (24), chega-se a

$$U_1(t) = k_1 \left(\hat{G}_1(t) + \hat{H}_1(t) \int_{t-D_1}^t U_1(\tau) d\tau + \epsilon H_{12}^1 \int_{t-D_2}^t U_2(\tau) d\tau \right) - k_1 \epsilon H_{12}^1 \int_{t-D_2}^t U_2(\tau) d\tau, \quad (30)$$

$$U_2(t) = k_2 \left(\hat{G}_2(t) + \hat{H}_2(t) \int_{t-D_2}^t U_2(\tau) d\tau + \epsilon H_{21}^2 \int_{t-D_1}^t U_1(\tau) d\tau \right) - k_2 \epsilon H_{21}^2 \int_{t-D_1}^t U_1(\tau) d\tau, \quad (31)$$

cujas versão média compacta é

$$U^{\text{av}}(t) = K \bar{G}^{\text{av}}(t) + \epsilon K \phi^{\text{av}}(D, t). \quad (32)$$

na qual a variável auxiliar $\phi(D, t)$ é definida como

$$\begin{aligned} \phi(D, t) &:= - \left[H_{12}^1 \int_{t-D_2}^t U_2(\tau) d\tau, H_{21}^2 \int_{t-D_1}^t U_1(\tau) d\tau \right]^T, \\ \phi(1, t) &:= - \left[H_{12}^1 \int_0^1 D_2 u_2(\xi, t) d\xi, H_{21}^2 \int_0^1 D_1 u_1(\xi, t) d\xi \right]^T. \end{aligned} \quad (33)$$

Então, é possível encontrar uma forma compacta para o jogo médio a partir das equações (29) e (32), como

$$\dot{\bar{G}}^{\text{av}}(t) = H K \bar{G}^{\text{av}}(t) + \epsilon H K \phi^{\text{av}}(1, t), \quad (34)$$

$$\partial_t u^{\text{av}}(x, t) = D^{-1} \partial_x u^{\text{av}}(x, t), \quad x \in (0, 1), \quad (35)$$

$$u^{\text{av}}(1, t) = K \bar{G}^{\text{av}}(t) + \epsilon K \phi^{\text{av}}(1, t). \quad (36)$$

De (34), fica claro que a EDO que rege a dinâmica da variável $\bar{G}^{\text{av}}(t)$ é ISS (Karafyllis and Krstić, 2018) com relação a EDP que rege $u(x, t)$ por meio da função $\phi^{\text{av}}(1, t)$. Além disso, o subsistema (35) é ISS (estável em tempo finito) (Karafyllis and Krstić, 2018) com relação a $\bar{G}^{\text{av}}(t)$ sob a condição de contorno $u^{\text{av}}(1, t)$.

4.3 Análise de Estabilidade

Nesta seção, mostraremos que, se ϵ for escolhido suficientemente pequeno, a malha fechada EDP-EDO (34)–(36) é localmente estável. Para esse fim, assumimos a seguinte condição para jogos não cooperativos Frihauf et al. (2012):

Hipótese 1. A matriz Hessiana H dada por (8) é estritamente diagonal dominante, ou seja,

$$\begin{cases} |\epsilon H_{12}^1| < |H_{11}^1| \\ |\epsilon H_{21}^2| < |H_{22}^2| \end{cases}. \quad (37)$$

Pela Hipótese 1, o Equilíbrio de Nash, θ^* , existe e é único, pois matrizes estritamente diagonais dominantes são não-singulares pelo Teorema de Levy-Desplanques (Horn and Johnson, 1985).

O próximo teorema resume as propriedades de estabilidade e convergência da busca extremal no sistema de malha fechada de um jogo não-cooperativo de dois jogadores com atrasos nas ações.

Teorema 1. Considere o sistema de malha fechada (25)–(27) sob a Hipótese 1, atrasos de entrada D_1 e D_2 em um jogo do tipo duopólio com funções de *payoff* (1) e (2), e leis de controle $U_1(t)$ e $U_2(t)$ definidas por (23). Exitem constantes $c > 0$ e $\omega > 0$ suficientemente grandes bem como $\epsilon > 0$ suficientemente pequena tal que o sistema de malha fechada com estado $\hat{\theta}_i(t-D_i), U_i(\tau), \forall \tau \in [t-D_i, t]$ e $\forall i \in 1, 2$ possui

apenas uma solução periódica exponencialmente estável em t de período Π , denotada por $\tilde{\theta}_i^\Pi(t - D_i)$, $U_i^\Pi(\tau)$, $\forall \tau \in [t - D_i, t]$ satisfazendo, $\forall t \geq 0$:

$$\left(\sum_{i=1}^2 [\tilde{\theta}_i^\Pi(t - D_i)]^2 + \int_{t-D_i}^t [U_i^\Pi(\tau)]^2 d\tau \right)^{1/2} \leq \mathcal{O}(1/\omega). \quad (38)$$

Além do mais,

$$\limsup_{t \rightarrow +\infty} |\theta(t) - \theta^*| = \mathcal{O}(|a| + 1/\omega), \quad (39)$$

onde $a = [a_1, a_2]^T$ e θ^* é o único Equilíbrio de Nash em (9).

Prova. A prova do Teorema segue etapas semelhantes às empregadas para provar os resultados sobre a busca extremal na presença de atrasos (Oliveira et al., 2017). Nesse sentido, simplesmente apontaremos as principais diferenças para o caso dos jogos (não a busca clássica por extremos).

Enquanto em Oliveira et al. (2017), na busca clássica por extremos, foi possível provar localmente a estabilidade exponencial do sistema médio de malha fechada usando funcionais de Lyapunov, uma abordagem diferente é adotada aqui para jogos não-cooperativos. Mostraremos que é possível garantir a estabilidade exponencial local para o sistema médio de malha fechada (34)–(36) por meio de uma análise de pequenos ganhos.

Primeiro, considere a representação equivalente de EDP-EDO (34)–(36) reescrito para cada jogador i , $i \in \{1, 2\}$:

$$\dot{G}_i^{\text{av}}(t) = H_{ii}^i k_i \bar{G}_i^{\text{av}}(t) + \epsilon H_{ii}^i k_i \phi_i^{\text{av}}(1, t), \quad (40)$$

$$\partial_t u_i^{\text{av}}(x, t) = D_i^{-1} \partial_x u_i^{\text{av}}(x, t), \quad x \in (0, 1), \quad (41)$$

$$u_i^{\text{av}}(1, t) = k_i \bar{G}_i^{\text{av}}(t) + \epsilon k_i \phi_i^{\text{av}}(1, t), \quad (42)$$

onde $H_{ii}^i < 0$, $k_i > 0$, $0 < \epsilon < 1$ e $D_i^{-1} > 0$. O sistema médio de malha fechada (40)–(42) satisfaz as hipóteses **(H1)** e **(H2)** do Teorema dos Pequenos Ganhos (Karafyllis and Krstić, 2018, Teorema 8.1, p. 198) para casata EDP-EDO com $n = 1$, $c = 1/D_i^{-1}$, $F(\bar{G}_i^{\text{av}}, u^{\text{av}}, 0) = H_{ii}^i k_i \bar{G}_i^{\text{av}} + \epsilon H_{ii}^i k_i \phi_i^{\text{av}}(1)$, $a(x) = f(x, t) = g(x, \bar{G}_i^{\text{av}}, u^{\text{av}}) = 0$, $\varphi(0, u^{\text{av}}, \bar{G}_i^{\text{av}}) = k_i \bar{G}_i^{\text{av}} + \epsilon k_i \phi_i^{\text{av}}(1)$, $\bar{N} = \max(k_i, \epsilon k_i k_H \sqrt{D_j})$, $L = \max(|H_{ii}^i| k_i, \epsilon |H_{ii}^i| k_i k_H \sqrt{D_j})$, $\gamma_2 = k_i$, $A = \gamma_1 = 0$, $B = \epsilon k_i k_H \sqrt{D_j}$, $\gamma_2 = k_i$, $b_2 = 0$ e $i \neq j$. Com a hipótese **(H1)** satisfeita com $M = 1$, $\gamma_3 = \epsilon |H_{ii}^i| k_i k_H \sqrt{D_j}$ e $\sigma = |H_{ii}^i| k_i$ como pode ser facilmente verificado por meio da fórmula da variação das constantes

$$\bar{G}_i^{\text{av}}(t) = \exp(-|H_{ii}^i| k_i t) \bar{G}_i^{\text{av}}(0) + \int_0^1 \exp(-|H_{ii}^i| k_i (t+s)) \epsilon H_{ii}^i k_i \phi_i^{\text{av}}(1, s) ds,$$

e da aplicação da desigualdade de Cauchy-Schwarz ao termo $\phi_i^{\text{av}}(1, t)$ na equação (33):

$$\begin{aligned} \phi_i^{\text{av}}(1, t) &\leq |H_{ij}^i| \left(\int_0^1 D_j d\tau \right)^{\frac{1}{2}} \times \left(\int_0^1 [u_j^{\text{av}}(\xi, t)]^2 d\xi \right)^{\frac{1}{2}} \\ &\leq k_H \sqrt{D_j} \left(\int_0^1 [u_j^{\text{av}}(\xi, t)]^2 d\xi \right)^{\frac{1}{2}}, \end{aligned} \quad (43)$$

¹ Se o escalar $c < 0$ for considerado em (Karafyllis and Krstić, 2018, Teorema 8.1, p. 198), a direção da convecção deve ser invertida de modo que o limite $u(0, t)$ é substituído por $u(1, t)$ e vice-versa.

uma vez que $|H_{ij}^i| < k_H < \frac{1}{\epsilon} |H_{ii}^i|$, onde k_H é uma constante positiva de ordem $\mathcal{O}(1)$, de acordo com a Hipótese 1.

Portanto a condição de pequeno ganho é mantida desde que $\epsilon \in (0, 1)$ seja suficientemente pequeno. Se tal condição é satisfeita, então invocando-se (Karafyllis and Krstić, 2018, Teorema 8.1, p. 198) conclui-se que existem as constantes $\delta, \Theta > 0$ tal que para cada $u_0^{\text{av}} \in C^0([0, 1])$, $\bar{G}_{i,0}^{\text{av}} \in \mathbb{R}$, a única solução generalizada desse problema de valor inicial/fronteira, com $u^{\text{av}}(x, 0) = u_0^{\text{av}}$ e $\bar{G}_i^{\text{av}}(0) = \bar{G}_{i,0}^{\text{av}}$, satisfaz a seguinte estimativa:

$$|\bar{G}_i^{\text{av}}(t)| + \|u^{\text{av}}(t)\|_\infty \leq \Theta(|\bar{G}_{i,0}^{\text{av}}| + \|u_0^{\text{av}}\|_\infty) \exp(-\delta t). \quad (44)$$

Portanto, concluímos que a origem do sistema de circuito fechado médio (34)–(36) é exponencialmente estável sob a hipótese $\epsilon \in (0, 1)$ suficientemente pequeno. Então, de (21) e (28), podemos concluir os mesmos resultados na norma

$$\left(\sum_{i=1}^2 [\tilde{\theta}_i^{\text{av}}(t - D_i)]^2 + \int_0^{D_i} [u_i^{\text{av}}(\tau)]^2 d\tau \right)^{1/2} \quad (45)$$

já que H é não-singular, i.e., $|\tilde{\theta}_i^{\text{av}}(t - D_i)| \leq |H^{-1}| |\hat{G}^{\text{av}}(t)|$.

Conforme desenvolvido em Oliveira et al. (2017), as próximas etapas para concluir a prova seriam a aplicação da teoria da média local para sistemas de dimensões infinitas em Hale and Lunel (1990) mostrando as soluções periódicas que realmente satisfazem a desigualdade (38) e, em seguida, a conclusão da atratividade do Equilíbrio de Nash θ^* de acordo com (39). \square

5. SIMULAÇÕES COM UM JOGO DO TIPO DUOPÓLIO NA PRESENÇA DE ATRASOS CONSTANTES

Para exemplificar um jogo não-cooperativo com dois jogadores que empregam a estratégia de busca extremal com compensação de atraso proposta, considera-se as seguintes funções de *payoff* (1) e (2) sujeitas a atrasos distintos $D_1 = 20$ e $D_2 = 15$ nas decisões dos jogadores e demais constantes $H_{11}^1 = -10$, $H_{22}^1 = 10$, $h_1^1 = 250$, $h_2^1 = -150$, $c_1 = -3000$, $H_{11}^2 = 10$, $H_{22}^2 = -10$, $h_1^2 = -150$, $h_2^2 = 150$ e $c_2 = 6000$. De acordo com (9), este jogo possui um único equilíbrio de Nash dado por

$$\theta_1^* = \frac{100 + 30\epsilon}{4 - \epsilon^2} \quad \text{e} \quad \theta_2^* = \frac{60 + 50\epsilon}{4 - \epsilon^2}. \quad (46)$$

Para alcançar o Equilíbrio de Nash (46) os jogadores implementam uma estratégia de otimização em tempo real baseada no não-conhecimento do modelo com compensação do atraso (Seção 4). Os jogadores $J1$ e $J2$ definem suas ações $\theta_1(t) = \hat{\theta}_1(t) + S_1(t)$ e $\theta_2 = \hat{\theta}_2(t) + S_2(t)$, com $\hat{\theta}_1$ e $\hat{\theta}_2$ de acordo com (23). A estimativa do gradiente \hat{G}_i para cada jogador e a estimativa \hat{H}_i dos termos da diagonal principal da matriz Hessiana é fornecida respectivamente em (18) e (15), onde os sinais de *dither* $S_i(t)$, $M_i(t)$ e $N_i(t)$ são apresentados em (11), (12) e (14).

Para fins de comparação, exceto pelos atrasos, os parâmetros da planta e do controlador foram escolhidos como Frihauf et al. (2012) em todas as simulações: $a_1 = 0.075$, $a_2 = 0.05$, $k_1 = 2$, $k_2 = 5$, $\omega_1 = 26.75$ rad/s, $\omega_2 = 22$ rad/s e $\hat{\theta}_1(0) = 50$, $\hat{\theta}_2(0) = 110/3$. Além disso, as constantes de tempo dos filtros preditores foram definidas como $c_1 = c_2 = 100$.

Diferente das estratégias clássicas para jogos não-cooperativos Başar and Olsder (1999) (livre de atrasos), ao utilizar o algo-

ritmo de busca extremal proposto, cada jogador precisa medir apenas o seu próprio *payoff*, respectivamente, J_1 e J_2 .

Na Fig. 2(a) e Fig. 2(b), pode-se verificar o desempenho da busca extremal proposta em Frihauf et al. (2012), com $\epsilon = 1$, na presença de atrasos D_1 e D_2 nos sinais de entrada θ_1 e θ_2 , ou seja, sem considerar qualquer tipo de compensação de atraso. A simulação mostra que o jogo entra em colapso com a explosão de suas variáveis.

Por outro lado, Fig. 3(a) e Fig. 3(b) mostram que a estratégia de controle proposta neste artigo garante o alcance do Equilíbrio de Nash em jogos não-cooperativos compensando o efeito dos atrasos por meio de uma busca extremal baseada em preditor. Valores diferentes de $\epsilon = 0,5$ e $\epsilon = 0,1$ são considerados para avaliar a robustez do esquema proposto sob diferentes níveis de acoplamento entre os dois jogadores e o impacto correspondente nas respostas transitórias.

Independentemente dos valores usados para ϵ , as simulações indicam que, mesmo com atrasos múltiplos e distintos nas ações dos jogadores ambos os jogadores conseguiram otimizar suas funções de *payoff* no jogo não-cooperativo, alcançando assim o desejado Equilíbrio de Nash.

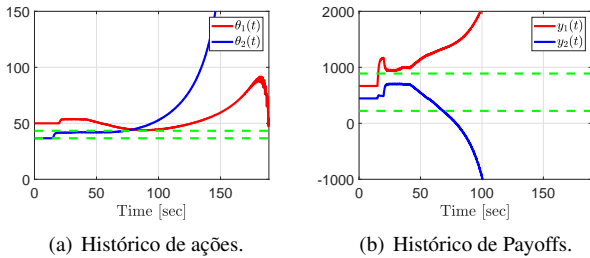


Figura 2. Atrasos Não Compensados: jogo ao implementar a busca de Nash proposta por Frihauf et al. (2012) para $\epsilon = 1$. As linhas tracejadas representam os valores do Equilíbrio de Nash, $\theta_1^* = 43.33$ e $\theta_2^* = 36.67$.

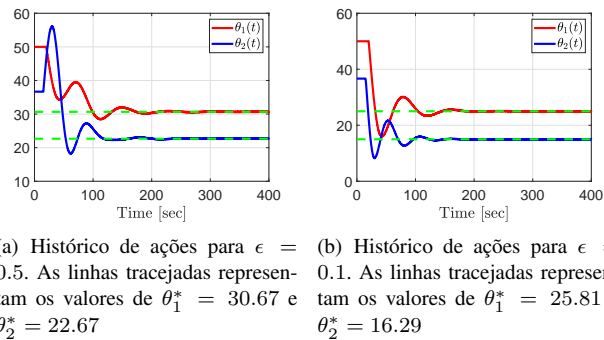


Figura 3. Atrasos compensados: jogo ao implementar a Busca de Nash proposta neste artigo. As linhas tracejadas representam os valores do Equilíbrio de Nash.

6. CONCLUSÕES

Introduzimos uma estratégia baseada no não conhecimento do modelo por meio da busca extremal baseada em preditores para jogos não-cooperativos, do tipo duopólio, com jogadores que possuem funções de *payoff* quadráticas e desconhecidas em um cenário onde há atraso temporal nas ações dos jogadores. A estabilidade e a convergência são garantidas por meio da teoria

da média em dimensões infinitas e uma análise de pequenos ganhos para as malhas de EDO-EDP.

REFERÊNCIAS

- Alasseur, C., Taher, I.B., and Matoussi, A. (2020). An extended mean field game for storage in smart grids. *Journal of Optimization Theory and Applications*, 184, 644–670.
- Alpcan, T. and Başar, T. (2011). *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press.
- Artstein, Z. (1982). Linear systems with delayed controls: A reduction. *IEEE Transactions on Automatic Control*, 27, 869–879.
- Aussel, D. and Svensson, A. (2019). Towards tractable constraint qualifications for parametric optimisation problems and applications to generalised nash games. *Journal of Optimization Theory and Applications*, 182, 404–416.
- Başar, T. and Olsder, G.J. (1999). *Dynamic Noncooperative Game Theory*. SIAM Series in Classics in Applied Mathematics, Philadelphia.
- Başar, T. and Zaccour, G. (2018). *Handbook of Dynamic Game Theory*, volume 2. Springer International Publishing.
- Carmona, R., Fouque, J.P., Mousavi, S.M., and Sun, L.H. (2018). Systemic risk and stochastic games with delay. *Journal of Optimization Theory and Applications*, 179, 366–399.
- Cotrina, J. and Zúñiga, J. (2018). Time-dependent generalized nash equilibrium problem. *Journal of Optimization Theory and Applications*, 179, 1054–1064.
- Frihauf, P., Krstić, M., and Başar, T. (2012). Nash equilibrium seeking in noncooperative games. *IEEE Transactions on Automatic Control*, 57, 1192–1207.
- Fudenberg, D. and Tirole, J. (1991). *Game Theory*. The MIT Press, Cambridge–Massachusetts.
- Ghaffari, A., Krstić, M., and Nesić, D. (2012). Multivariable Newton-based extremum seeking. *Automatica*, 48, 1759–1767.
- Hale, J.K. and Lunel, S.M.V. (1990). Averaging in infinite dimensions. *Journal of Integral Equations and Applications*, 2, 463–494.
- Han, Z., Niyato, D., Saad, W., and Başar, T. (2019). *Game Theory for Next Generation Wireless and Communication Networks: Modeling, Analysis, and Design*. Cambridge University Press.
- Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge Univ. Press.
- Karafyllis, I. and Krstić, M. (2018). *Input-to-State Stability for PDEs*. Springer, Switzerland.
- Khalil, H.K. (2002). *Nonlinear Systems*. Prentice Hall.
- Krstić, M. (2009). *Delay Compensation for Nonlinear, Adaptive, and PDE Systems*. Birkhauser, Boston.
- Krstić, M. and Wang, H.H. (2000). Stability of extremum seeking feedback for general dynamic systems. *Automatica*, 36, 595–601.
- Oliveira, T.R., Krstić, M., and Tsubakino, D. (2017). Extremum seeking for static maps with delays. *IEEE Transactions on Automatic Control*, 62, 1911–1926.
- Pamen, O.M. (2015). Optimal control for stochastic delay systems under model uncertainty: A stochastic differential game approach. *Journal of Optimization Theory and Applications*, 167, 998–1031.
- Wang, W., Sun, H., Van den Brink, R., and Xu, G. (2018). The family of ideal values for cooperative games. *Journal of Optimization Theory and Applications*, 180, 1065–1086.