Mikailalsys Journal of
Mathematics and Statistics

e-ISSN : 3030-816X
p-ISSN : 3030-8399

Index : Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/MJMS.v3i3.7336

A Short Note on: Optimal Control in Matching Pennies Game

Reza Habibi
Iran Banking Institute, Iran
habibi1356@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Aug 25, 2025	Sep 17, 2025	Sep 29, 2025	Oct 3, 2025

Abstract

This paper addresses the need for rigorous stability criteria in fractional integro-differential systems by investigating generalized Ulam–Hyers stability for equations involving a fractional-order derivative. The research objective is to establish sufficient conditions that guarantee generalized Ulam–Hyers stability for Caputo fractional differential equations. Methodologically, the study employs concise mathematical arguments together with an application of the Gronwall inequality to derive the required conditions. The key findings demonstrate that these conditions ensure stability in the generalized Ulam–Hyers sense for the considered class of equations. The paper concludes that the obtained results extend and improve upon existing findings in the literature. The contribution lies in refining the stability theory for fractional-order models by providing tractable criteria grounded in Gronwall-type estimates and Caputo derivatives.

Keywords: Ulam–Hyers Stability; Generalized Stability; Caputo Derivative; Fractional Integro-Differential Equations; Gronwall Inequality; Non-Local Conditions

Introduction

The matching the pennies game is standard example for teaching game theory in which two players throw a coin, separately. If the slides of the coins are matched, player 1 wins one dollar, otherwise his opponent player 2 wins one dollar. It is know that this game does not have a pure Nash equilibrium. However, mixed equilibrium and the rate of convergence of player's learning process such as gradient descent to these equilibriums in a repeated game has been well studied in the existing literature, see Singh *et al*. (2000) and references therein. The bi-matrix payoff of both players are givenThe matching the pennies game is standard example for teaching game theory in which two players throw a coin, separately. If the slides of the coins are matched, player 1 wins one dollar, otherwise his opponent player 2 wins one dollar. It is know that this game does not have a pure Nash equilibrium. However, mixed equilibrium and the rate of convergence of player's learning process such as gradient descent to these equilibriums in a repeated game has been well studied in the existing literature, see Singh *et al*. (2000) and references therein. The bi-matrix payoff of both players are given

		Player 2	
		H	T
Player 1	H	(1,-1)	(-1,1)
	T	(-1,1)	(1,-1)

Suppose that this game is played at $t_i \in (0,1), i = 1, \ldots, n$ where $t_i - t_{i-1} = 1/n$. Let $(\alpha_t, 1-\alpha_t)$ and $(\beta_t, 1-\beta_t)$ be the mixing distributions of player 1 and 2, respectively, in t-th one shot game. Let x_t be the value of t-th game which is 1 with probability of

$$\alpha_t\beta_t + (1-\alpha_t)(1-\beta_t)$$

and -1 with probability of

$$(1-\alpha_t)\beta_t + \alpha_t(1-\beta_t).$$

a) Optimal control. Playing n games, the cumulative value is $\sum_{j=1}^{n} x_j$. Player 1 is interested to maximize $\sum_{j=1}^{n} x_j$ or equivalently, $1/n \sum_{j=1}^{n} x_j$ with respect to α_t. This problem defines an optimal control problem as follows:

Integral objective function. Optimal control is a condition of dynamic systems that satisfy design objectives. Optimal control is achieved with control laws that execute following defined optimality criteria, see Kamien (2013). In this case, it is easy to see that $1/n \sum_{j=1}^{n} x_j$ converges to $\int_0^1 x_t dt$. One may notice that

$$E\left(\int_0^1 x_t dt\right) = \int_0^1 (2\,\alpha_t - 1)(2\beta_t - 1)dt.$$

Observation equation. Following Singh *et al*. (2000), to maximize $v_1 = E(x_t)$ at t-th one shot game, the gradient descent algorithm implies that

$$\alpha' = \frac{\partial v_1}{\partial \alpha} = 2(2\beta - 1).$$

Solution. To solve the optimal control problem, *i.e.*, to maximize

$$I = \int_0^1 (2\,\alpha_t - 1)(2\beta_t - 1)dt$$

with respect to $\alpha' = 2(2\beta - 1)$, it is seen that

$$I = \int_0^1 (2\,\alpha_t - 1)\frac{\alpha'}{2}dt = \frac{\alpha_1 - \alpha_0}{2}(\alpha_1 + \alpha_0 - 1).$$

It is seen that I is minimized if and only if $\alpha_0 = 0$. Computing v_2, the expected payoff of player 2, the gradient descent implies that $\beta' = 2(2\alpha - 1)$ and $\beta_0 = 0.5$. Thus,

$$\begin{cases} \alpha' = 2(2\beta - 1), \alpha_0 = 0, \\ \beta' = 2(2\alpha - 1), \beta_0 = 0.5. \end{cases}$$

Solving the above system of differential equations, it is seen that

$$\alpha_t = \frac{1-\cos(4t)}{2}, \beta_t = \frac{1-\sin(4t)}{2}.$$

Remark 1. Considering Hamilton objective function as

$$H = (2\beta_t - 1)(2\alpha_t - 1) + \lambda(2\alpha_t - 1),$$

and using equations $\frac{\partial H}{\partial \beta} = 0$ and $\frac{\partial H}{\partial \alpha} = -\lambda'$, yields that $\alpha' = 2(2\beta - 1)$ is the optimal path.

b) Randomized pure equilibrium. Although, the pure Nash equilibrium does not exist, however, for randomized pure equilibrium, let $\beta_t = 1$ with probability of b_t and 0 with probability of $1 - b_t$. Suppose that player 1 desire to maximize the probability of HH which is $\alpha_t\beta_t$. Let $E(\alpha_t) = a_t$. The gradient descent implies that $a' = b$ and the integral $\int_0^1 (2\,a_t - 1)(2b_t - 1)dt$ is maximized. The Hamilton function is $h = (2a - 1)(2b - 1) + \zeta b$. It is seen that $\zeta = 2(1 - 2a)$ and $-\zeta' = 2b - 1$. Hence,

$$\begin{cases} -\zeta' = -4a' = -4b \\ -\zeta' = 2b - 1 \end{cases}$$

Thus, $b = 1/6$. Indeed,

$$\beta_t = \begin{cases} 1 & \text{probability } 1/6 \\ 0 & \text{probability } 5/6 \end{cases}$$

This type of control is a randomized bang-bang control and $\alpha_t = \int_0^t \beta_s ds$. Indeed,

$$\alpha_{t_i} - \alpha_{t_{i-1}} = \frac{\beta_{t_i}}{n}.$$

That is, $\alpha_{t_i} = 1/n \sum_{j=1}^{k} \beta_{t_j}$ where $t_{i-1} \leq \frac{k}{n} < t_i$.

c) Partial sum payoff. Define the partial sum y_t as the payoff of player 1, up to time t, indeed,

$$y_t = \frac{1}{n}\sum_{j=1}^{k} x_{t_j}, \quad t_k \leq t < t_{k+1}.$$

Notice that, as $n \to \infty$, then

$$E(y_t) \to \int_0^1 \left(\frac{1+\sin(8s)}{2}\right)ds = \frac{1-\cos(8t)+8t}{16}.$$

The expected payoff of player 2 is exactly $-E(y_t)$. Thus, the maximum payoff for both players, considering their opponents payoff, receives at $E(y_t) = 0$ which achieves at $t = 0$. That is, the best strategy for both players is not to play with his opponent.

d) Scenario analysis. In previous section, it was seen that

$$E(x_t) = (2\alpha_t - 1)(2\beta_t - 1).$$

Three cases are proposed as follows:

Case1. As $\alpha_t > 0.5$ and $\beta_t > 0.5$ or $\alpha_t < 0.5$ and $\beta_t < 0.5$, then $E(x_t) > 0$ and the first player wins, probably.

Case2. Conversely, if $\alpha_t > 0.5$ and $\beta_t < 0.5$ or $\alpha_t < 0.5$ and $\beta_t > 0.5$, then the second player will win.

Case3. If $\alpha_t = 0.5$ or $\beta_t = 0.5$, then $E(x_t) = 0$ and y_t is a martingale with respect to natural filtration.

The next section studies the simulation analysis.

2 Simulations. Here, some useful results are proposed via simulation.

a) First exit. The first time the partial sum of one shot games payoff passes a specified threshold is interesting. To this end, let $\alpha_t = 0.5 + u_t$ and $\beta_t = 0.5 + v_t$, where $u_t, v_t \in (-0.5, 0.5)$ then $E(x_t) = 4u_t v_t$ and

$$x_t = \begin{cases} -1 & \text{probability } 0.5 - 2u_t v_t \\ 1 & \text{probability } -0.5 + 2u_t v_t \end{cases}$$

By generating independent random variables $u_t, v_t \in (-0.5, 0.5)$ uniformly, the first point at which y_t passes its mean for $t = 1, \ldots, 100$ is approximated by beta $Beta(0.15, 0.34)$ distribution.

b) Last meet. To simulate the last time at which both players receive to the same level of cumulative payoffs is another interesting fact. By simulating u_t, v_t, the last point at which y_t and z_t, the partial sum payoff of the second player, meet each other is well approximated by beta $Beta(0.36, 0.3)$ distribution.

c) Random bang-bang control. Here, y_t is simulated with $\alpha_{t_i} = 1/n \sum_{j=1}^{k} \beta_{t_j}$ and

$$\beta_t = \begin{cases} 1 & \text{probability } 1/6 \\ 0 & \text{probability } 5/6. \end{cases}$$

The first point at which y_t passes its mean is estimated at 0.46.

Alternative looking to this problem is to find b_t, the probability of $\beta_t$being 1 such that y_1 is maximized. It is seen that $b_t = 0.5$.

References

Kamien, M. I. (2013). *Dynamic Optimization: the Calculus of Variations and Optimal Control in Economics and Management*. Dover Publications. USA.

Singh, S., Kearns, M., and Mansour, Y. (2000). Nash convergence of gradient dynamics in general-sum games. *Technical Report*. AT&T Labs and Tel Aviv University.