Supplementary Material for Online Optimization in Games via Control Theory: Connecting Regret, Passivity and Poincaré Recurrence

A. Volume and Liouville's Formula

Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be a function. Given an ODE system $\dot{\mathbf{z}} = g(\mathbf{z})$ but with a flexibility to choose the starting point, let $\Phi(\mathbf{z}^0, t)$ be the solution of the ODE system at time t with starting point $\mathbf{z}^0$. Given any set A, let $A(t) = \{\Phi(\mathbf{z}^0, t) \mid \mathbf{z}^0 \in A\}$. When A is measurable, under mild conditions on the ODE system, $A(t)$ is measurable and its volume is $\text{vol}[A(t)] = \int_{A(t)} \mathsf{d}v$. Liouville's formula states that the time derivative of the volume $A(t)$ is equal to the integral of the divergence of the ODE system over $A(t)$:

$$\frac{d}{dt}\text{vol}[A(t)] = \int_{A(t)} \text{trace}\left(\frac{\partial g}{\partial \mathbf{z}}\right) \mathsf{d}v,$$

where $\frac{\partial g}{\partial \mathbf{z}}$ is the *Jacobian* of the ODE system. Note that $\text{trace}\left(\frac{\partial g}{\partial \mathbf{z}}\right) = \sum_{j=1}^d \frac{\partial g_j}{\partial z_j}$, where g_j is the j-th component of the function g. This immediately implies volume preservation for divergence-free systems.

B. Missing Proofs

Proof of Proposition 4: It suffices to prove the forward (only if) direction, as the other direction is symmetric. By the definition of passivity, we have $L^a(\mathbf{q}(t)) \le L^a(\mathbf{q}(0)) + \int_0^t \langle(\mathbf{x}(\tau) - \mathbf{x}^{*,a}), \mathbf{p}(\tau)\rangle \, \mathsf{d}\tau$. This implies

$$\begin{aligned} L^a(\mathbf{q}(t)) \;&\le\; L^a(\mathbf{q}(0)) + \int_0^t \left\langle(\mathbf{x}(\tau) - \mathbf{x}^{*,b}), \mathbf{p}(\tau)\right\rangle \mathsf{d}\tau + \int_0^t \left\langle(\mathbf{x}^{*,b} - \mathbf{x}^{*,a}), \mathbf{p}(\tau)\right\rangle \mathsf{d}\tau \\ &=\; L^a(\mathbf{q}(0)) + \int_0^t \left\langle(\mathbf{x}(\tau) - \mathbf{x}^{*,b}), \mathbf{p}(\tau)\right\rangle \mathsf{d}\tau + \left\langle(\mathbf{x}^{*,b} - \mathbf{x}^{*,a}), (\mathbf{q}(t) - \mathbf{q}(0))\right\rangle. \end{aligned}$$

Thus, by setting $L^b(\mathbf{q}) := L^a(\mathbf{q}) - \left\langle(\mathbf{x}^{*,b} - \mathbf{x}^{*,a}), \mathbf{q}\right\rangle + c$, we have $L^b(\mathbf{q}(t)) \le L^b(\mathbf{q}(0)) + \int_0^t \left\langle(\mathbf{x}(\tau) - \mathbf{x}^{*,b}), \mathbf{p}(\tau)\right\rangle \mathsf{d}\tau$, certifying passivity of the operator S^b. □

Proof of Proposition 6: Suppose that for each action j, the learning operator with shift $\mathbf{e}_j$ is finitely lossless via storage function L^j. Then the storage function $\sum_{j=1}^n x_j^* \cdot L^j$ can be used to certify finitely losslessness of the learning operator with shift of the mixed strategy $\mathbf{x}^*$. □

Theorem 8 states that finite passivity implies constant regret. The following proposition states that the converse (constant regret guaranteed implies finite passivity) is also true, if we restrict to lossless learning dynamics.

Proposition 18. *Suppose that a learning dynamic is lossless. Then the learning dynamic guarantees constant regret if and only if it is finitely lossless.*

Proof: ($\Leftarrow$) Done by Theorem 8.

($\Rightarrow$) Suppose the contrary, i.e. the learning algorithm is lossless, but there exists j such that the learning operator with shift $\mathbf{e}_j$ is *not* finitely lossless. Thus, it has a storage function L^j which is not bounded from below, and

$$L^j(\mathbf{q}(t)) \;=\; L^j(\mathbf{q}(0)) + \int_0^t \langle \mathbf{x}(\tau), \mathbf{p}(\tau)\rangle \, \mathsf{d}\tau - \int_0^t \langle \mathbf{e}_j, \mathbf{p}(\tau)\rangle \, \mathsf{d}\tau.$$

Following the calculation in the proof of Theorem 8, the regret w.r.t. action j at time t is *exactly* equal to $L^j(\mathbf{q}(0)) - L^j(\mathbf{q}(t))$. Since L^j is not bounded from below, for any $r < 0$, there exists $\tilde{\mathbf{q}}$ such that $L^j(\tilde{\mathbf{q}}) \le r$. It is easy to construct $\mathbf{p}$ such that $\mathbf{q}(t) = q(0) + \int_0^t \mathbf{p}(\tau) \, \mathsf{d}\tau = \tilde{\mathbf{q}}$; for instance, set $\mathbf{p}(\tau) = (\tilde{\mathbf{q}} - \mathbf{q}(0))/t$ for all $\tau \in [0, t]$. For this choice of $\mathbf{p}$, the regret at time t is $L^j(\mathbf{q}(0)) - L^j(\tilde{\mathbf{q}}) \ge L^j(\mathbf{q}(0)) - r$. Since we can choose arbitrarily negative value of r, the learning dynamic cannot guarantee constant regret, a contradiction. □

Proof of Proposition 10: Due to NC2, $\langle \nabla E(\mathbf{q}), \mathbf{1}\rangle = 1$, thus $\sum_{j=1}^n \nabla_j E(\mathbf{q}) = 1$. This equality and NC1(i) implies $\nabla E(\mathbf{q}) \in \Delta^n$. Now, consider a learning dynamic with conversion function $f = \nabla E$. Then for any function-of-time $\mathbf{p}$ and any $t > 0$, we have $\frac{\mathrm{d}E(\mathbf{q}(t))}{\mathrm{d}t} = \langle \nabla E(\mathbf{q}(t)), \dot{\mathbf{q}} \rangle = \langle f(\mathbf{q}(t)), \mathbf{p}(t) \rangle = \langle \mathbf{x}(t), \mathbf{p}(t) \rangle$. Integrating both sides w.r.t. t shows that the learning dynamic is lossless via E. □

Proof of Proposition 15: To show that the game operator is passive, according to Definition 1 and the input-output choice of S_2 (see Figure 3), it suffices to show that

$$\int_0^t \langle (\hat{\mathbf{x}}(\tau) - \hat{\mathbf{x}}^*), (-\hat{\mathbf{p}}(\tau)) \rangle \ \mathrm{d}\tau \ = \ - \int_0^t \underbrace{\langle \hat{\mathbf{x}}(\tau), \hat{\mathbf{p}}(\tau) \rangle}_{V_1} \ \mathrm{d}\tau + \int_0^t \underbrace{\langle \hat{\mathbf{x}}^*, \hat{\mathbf{p}}(\tau) \rangle}_{V_2} \ \mathrm{d}\tau \ \geq \ 0.$$

Recall the definition of $c^{\{i,k\}}$ in a graphical constant-sum game. Since V_1 is simply the total payoffs to all agents, V_1 is the sum of the constants $c^{\{i,k\}}$ of all edge-games, i.e. $V_1 = \sum_{i=1}^{m-1} \sum_{k=i+1}^{m} c^{\{i,k\}}$. We denote this double summation by V. It remains to show that $V_2 \geq V$ always if we want to show the game operator is passive, and to show that $V_2 = V$ always if we want to show the game operator is lossless.

Let the action set of agent i be S_i. We expand $V_2 = \langle \hat{\mathbf{x}}^*, \hat{\mathbf{p}} \rangle$ as follows:

$$\begin{aligned}
\langle \hat{\mathbf{x}}^*, \hat{\mathbf{p}} \rangle &= \sum_{i=1}^{m} \sum_{j \in S_i} x_{ij}^* \sum_{\substack{k=1 \\ k \neq i}}^{m} [\mathbf{A}^{ik} \mathbf{x}_k]_j \\
&= \sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \neq i}}^{m} (\mathbf{x}_i^*)^\mathsf{T} \mathbf{A}^{ik} \mathbf{x}_k \\
&= \sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \neq i}}^{m} (\mathbf{x}_k)^\mathsf{T} (\mathbf{A}^{ik})^\mathsf{T} \mathbf{x}_i^* \qquad \text{(just taking transpose)} \\
&= \sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \neq i}}^{m} \left[c^{\{i,k\}} - (\mathbf{x}_k)^\mathsf{T} \mathbf{A}^{ki} \mathbf{x}_i^* \right] \qquad \text{(definition of constant-sum edge-game)} \\
&= \sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \neq i}}^{m} c^{\{i,k\}} \ - \ \sum_{k=1}^{m} \sum_{\substack{i=1 \\ i \neq k}}^{m} (\mathbf{x}_k)^\mathsf{T} \mathbf{A}^{ki} \mathbf{x}_i^* \\
&= 2V \ - \ \underbrace{\sum_{k=1}^{m} \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{m} (\mathbf{x}_k)^\mathsf{T} \mathbf{A}^{ki} \mathbf{x}_i^*}_{U_k}}_{W} .
\end{aligned}$$

It remains to bound W_2. Observe that for each agent k, U_k is the payoff to agent k when she chooses the mixed strategy $\mathbf{x}_k$, while every other agent i chooses the mixed strategy $\mathbf{x}_i^*$. Since the mixed strategies $\mathbf{x}_i^*$ are coming from a Nash equilibrium (NE), by the definition of NE, $U_k \leq v_k^*$, where v_k^* is the payoff to agent k at the NE. Thus, $W \leq \sum_{k=1}^{m} v_k^*$, where the RHS is the total payoffs to all agents at the NE. Since the game is constant-sum, we have $\sum_{k=1}^{m} v_k^* = V$. Hence, $V_2 = \langle \hat{\mathbf{x}}^*, \hat{\mathbf{p}} \rangle \geq 2V - V = V$.

When the NE is fully-mixed, we have the following extra property: at the NE, for the agent k, her payoff from each of her actions is the same, and equals to v_k^*. Thus, U_k exactly equals to v_k^*, hence $W = \sum_{k=1}^{m} v_k^*$ and $V_2 = V$. □

C. Poincaré Recurrence

We first formally state the following corollary of Proposition 15 and Theorem 2.

Corollary 19. *The FIC system which corresponds to a dynamical game system, in which S_1 is any finitely lossless MLO and S_2 is any game operator which corresponds to a graphical constant-sum game with a fully-mixed Nash equilibrium, is finitely lossless. The storage function that demonstrates finitely losslessness of the FIC system is the same as the storage function of S_1. When the external input $\mathbf{r}$ is the zero function, the storage function becomes a constant-of-motion.*

To complete the proof of Theorem 17, we need to show the second property required by the principled approach of (Mertikopoulos et al., 2018). It relies crucially on the following lemma. Recall that we have defined the following in the main paper, which converts the storage function for the original learning operator to the storage function of the new learning operator of (10).

$$\overline{L}(q_1', q_2', \cdots, q_{n-1}') = L(q_1', q_2', \cdots, q_{n-1}', 0), \tag{11}$$

Lemma 20 (Adapted from (Mertikopoulos et al., 2018), Appendix D)**.** *For any continuous FTRL dynamic and for any $\mathbf{x}^* \in \Delta^n$, let L be its finitely lossless storage function defined in* (8)*, and let $\overline{L}$ be the function defined on $\mathbb{R}^{n-1}$ as in* (11)*. Then any level set of $\overline{L}$ is bounded in $\mathbb{R}^{n-1}$, i.e. for any real number $\bar{c}$, the set below is bounded:*

$$\{\ (q_1', \cdots, q_{n-1}') \mid \overline{L}(q_1', \cdots, q_{n-1}') \leq \bar{c}\ \}.$$

Recall the definition of FTRL and Theorem 7. For each agent i, suppose she uses a convex combination of ℓ_i FTRL dynamics indexed by $i1, i2, \cdots, i\ell_i$. Let the storage functions of these FTRL dynamics be $L^{i1}, L^{i2}, \cdots, L^{i\ell_i}$. Also, let $\mathbf{q}'^{,i}$ denote a vector in $\mathbb{R}^{n_i - 1}$ for agent i. Then the storage function of the whole dynamical game system is

$$\sum_{i=1}^{m} \sum_{j=1}^{\ell_i} \alpha_{ij} \cdot \overline{L}^{ij}(\mathbf{q}'^{,i}), \quad \text{where } \alpha_{ij} > 0, \text{ and } \forall i, \ \sum_{j=1}^{\ell_i} \alpha_{ij} = 1.$$

Due to Corollary 19, this storage function is a constant-of-motion when $\mathbf{r} \equiv 0$, and thus is bounded by certain constant $\bar{c}$ when the starting point is already given. Since every L^{ij} and hence $\overline{L}^{ij}$ has infimum zero, we must have: for each agent i, $\alpha_{i1} \cdot \overline{L}^{i1}(\mathbf{q}'^{,i}) \leq \bar{c}$, and hence $\overline{L}^{i1}(\mathbf{q}'^{,i}) \leq \bar{c}/\alpha_{i1}$. Then by Lemma 20, for each agent i, $\mathbf{q}'^{,i}(t)$ remains bounded for all t, and thus the overall vector $\hat{\mathbf{q}}'(t) = (\mathbf{q}'^{,1}(t), \mathbf{q}'^{,2}(t), \cdots, \mathbf{q}'^{,m}(t))$ also remains bounded for all t.

D. Escort Learning Dynamics

An escort learning dynamic (Harper, 2011) is a system of differential equations on variable $\mathbf{x} \in \Delta^n$: for each $1 \leq j \leq n$,

$$\dot{x}_j \ = \ \phi_j(x_j) \cdot \left[p_j - \frac{\sum_{\ell=1}^n \phi_\ell(x_\ell) \cdot p_\ell}{\sum_{\ell=1}^n \phi_\ell(x_\ell)} \right],$$

where each ϕ_j is a positive function on domain $(0, 1)$. Note that when $\phi_j(x_j) = x_j$, this is Replicator Dynamic.

Proposition 21. *Suppose a learning dynamic has the following property: if it starts at a point in the interior of Δ^n, then it stays in the interior forever. We have: the learning dynamic is FTRL via a separable strictly convex regularizer function $h(\mathbf{x}) = \sum_{i=1}^n h_i(\mathbf{x}_i)$ if and only if it is an escort replicator dynamic.*

Proof: If the specified learning dynamic is FTRL, recall that the conversion function is $f(\mathbf{q}) = \arg\max_{\mathbf{x} \in \Delta^n} \{\langle \mathbf{q}, \mathbf{x} \rangle - h(\mathbf{x})\}$. Let $\bar{x}_j = 1/h_j''(x_j)$ and $H := \sum_j \bar{x}_j$. When $\mathbf{x}$ is in the interior of Δ^n, from Appendix D of (Cheung & Piliouras, 2019), we have

$$\frac{\partial x_j}{\partial q_j} \ = \ \bar{x}_j - \frac{[\bar{x}_j]^2}{H} \qquad \text{and} \qquad \forall \ell \neq j, \ \ \frac{\partial x_j}{\partial q_\ell} \ = \ -\frac{\bar{x}_j \bar{x}_\ell}{H}.$$

By the chain rule,

$$\dot{x}_j \ = \ \left[\bar{x}_j - \frac{[\bar{x}_j]^2}{H} \right] \cdot p_j \ + \ \sum_{\ell \neq j} \left[-\frac{\bar{x}_j \bar{x}_\ell}{H} \right] p_\ell \ = \ \bar{x}_j \left(p_j - \frac{\sum_{\ell=1}^n \bar{x}_\ell p_\ell}{\sum_{\ell=1}^n \bar{x}_\ell} \right).$$

By recognizing $\phi(x_j)$ as $\bar{x}_j$, the FTRL dynamic is an escort replicator dynamic. Precisely, we set $\phi_j(\mathbf{x}) = 1/h_j''(x_j)$. Since h is strictly convex, h'' is a positive function, hence ϕ is a positive function too.

Conversely, if the specified algorithm is an escort learning dynamic with escort function ϕ_j for each j, to show that it is a FTRL dynamic with some strictly convex regularizer h, we want h to be separable, and for each j, $h_j''(x_j) = 1/\phi_j(x_j)$. Thus, it suffice to set h_j to be any double anti-derivative of $1/\phi_j$. Since $h_j''(x_j) = 1/\phi_j(x_j)$ is positive, each h_i is strictly convex, and hence h is strictly convex. □

E. Some Plots Illuminating Poincaré Recurrences

We present more plots that illuminate Poincaré recurrences of learning in games.

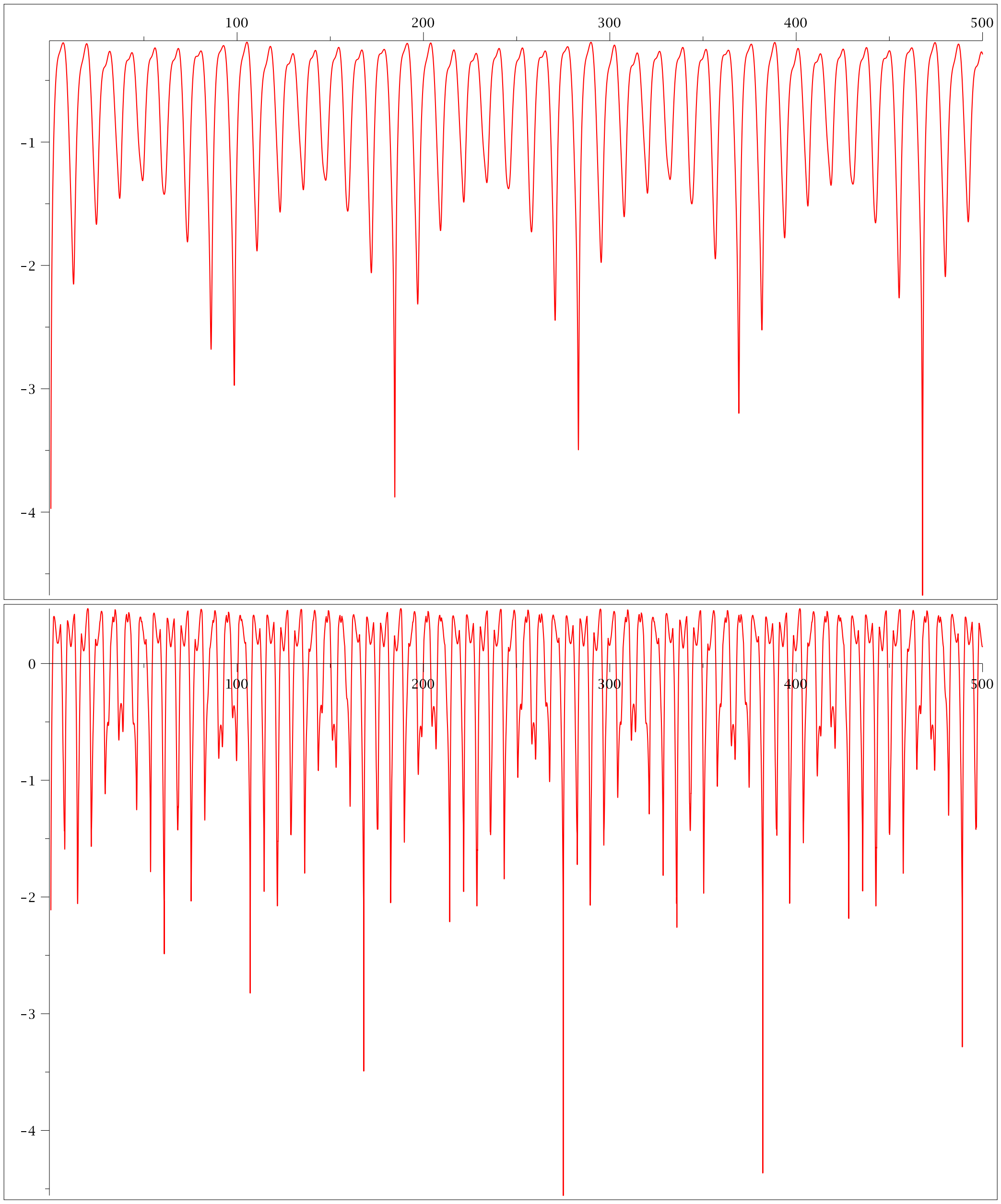

Figure 6. Poincaré recurrences of Replicator Dynamic (RD) and Online Gradient Descent (OGD; bottom) in the classical two-player Rock-Paper-Scissors game. Player 1 starts with mixed strategy $\mathbf{x}_1(0) = (0.5, 0.25, 0.25)$, while Player 2 starts with mixed strategy $\mathbf{x}_2(0) = (0.6, 0.3, 0.1)$. The two graphs plot the logarithm of the Euclidean distance between $(\mathbf{x}^1(t), \mathbf{x}^2(t))$ and $(\mathbf{x}^1(0), \mathbf{x}^2(0))$, from $t = 0.1$ to $t = 500$. Every downward spike corresponds to a moment where the flow gets back close to the starting point. In both cases, the distances drops below 10^{-3} for multiple times.

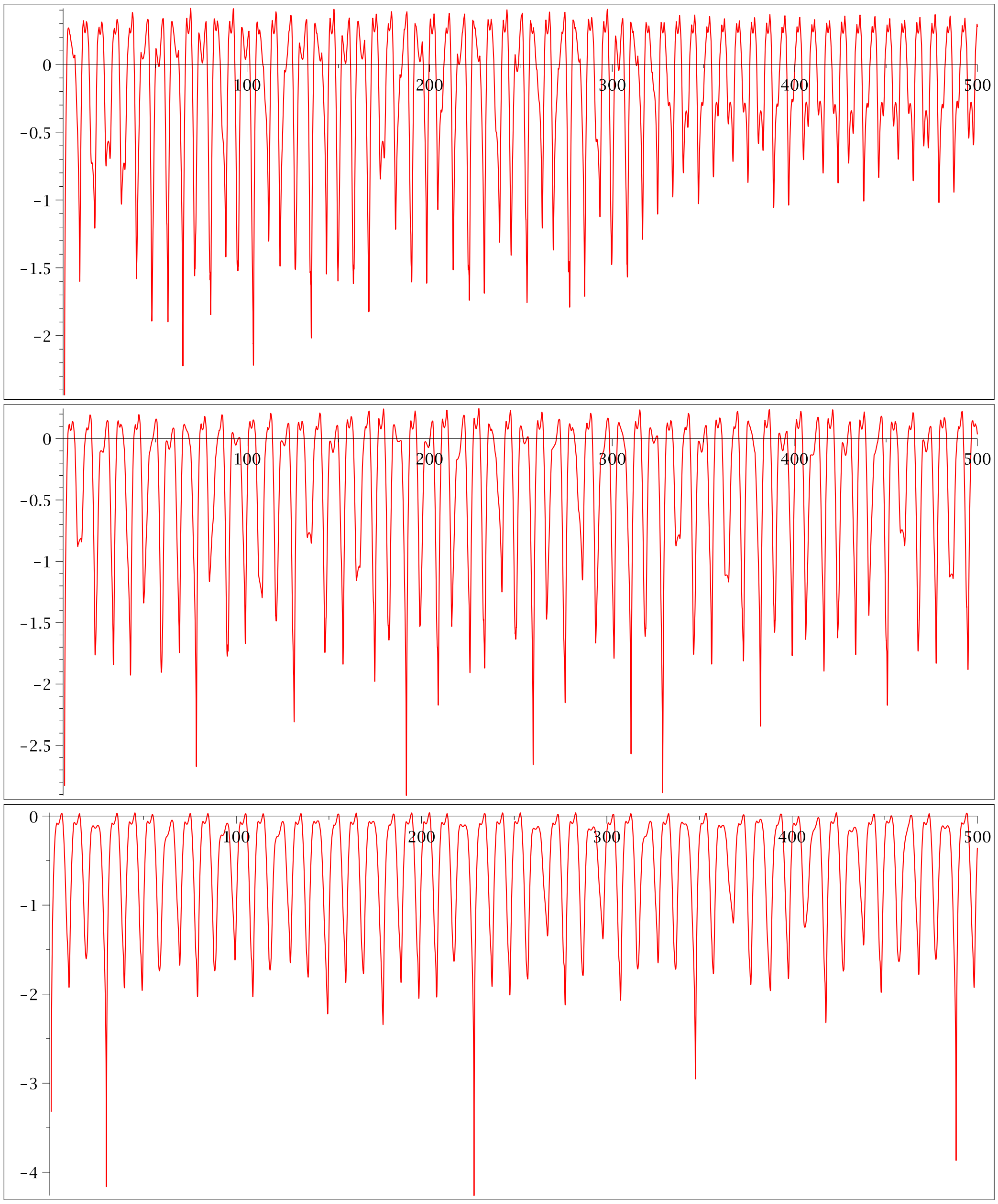

Figure 7. Poincaré recurrence of $\alpha \cdot \mathrm{RD} + (1-\alpha) \cdot \mathrm{OGD}$ in the classical Rock-Paper-Scissors game, for $\alpha = 1/4$ (top), $\alpha = 1/2$ (middle) and $\alpha = 3/4$ (bottom). The starting point is $(\mathbf{x}_1(0), \mathbf{x}_2(0)) = ((0.5, 0.25, 0.25), (0.6, 0.3, 0.1))$. The graphs plot the logarithm of the Euclidean distance between $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ and $(\mathbf{x}_1(0), \mathbf{x}_2(0))$, from $t = 0.1$ to $t = 500$.