Using Reinforcement Learning to Encrypt Communications

Faris Sbahi[1] and Barrett Ames[2]

[1] Primary

faris.sbahi@duke.edu

[2] Consultant

barrett.ames@duke.edu

Abstract. Here we explore the ability of reinforcement learning (RL) agents to cooperate to establish a secure communication protocol in the face of an adversary without explicit communication. Hence, we construct a general sum game to measure multi-agent communication and encryption efficacy. We compare our results with parallel experiments involving generative adversarial networks. Furthermore, we consider the practical implications and viability of RL cryptography–particularly in the context of quantum computing.

Keywords: reinforcement learning, multi-agent, communication, cryptography, generative adversarial networks, quantum computing

1 Introduction

We concern ourselves with the question as to whether

1.1 Autonomous Systems

In this section, we will consider the case when the Hamiltonian $H(x)$ is autonomous. For the sake of simplicity, we shall also assume that it is C^1.

We shall first consider the question of nontriviality, within the general framework of (A_∞, B_∞)-subquadratic Hamiltonians. In the second subsection, we shall look into the special case when H is $(0, b_\infty)$-subquadratic, and we shall try to derive additional information.

The General Case: Nontriviality. We assume that H is (A_∞, B_∞)-subquadratic at infinity, for some constant symmetric matrices A_∞ and B_∞, with $B_\infty - A_\infty$ positive definite. Set:

$$\gamma := \text{smallest eigenvalue of } \ B_\infty - A_\infty \tag{1}$$

$$\lambda := \text{largest negative eigenvalue of } \ J\frac{d}{dt} + A_\infty \ . \tag{2}$$

Theorem 1 tells us that if $\lambda + \gamma < 0$, the boundary-value problem:

$$\begin{aligned} \dot{x} &= JH'(x) \\ x(0) &= x(T) \end{aligned} \tag{3}$$

has at least one solution $\overline{x}$, which is found by minimizing the dual action functional:

$$\psi(u) = \int_o^T \left[\frac{1}{2}\left(\Lambda_o^{-1}u, u\right) + N^*(-u)\right] dt \tag{4}$$

on the range of Λ, which is a subspace $R(\Lambda)_L^2$ with finite codimension. Here

$$N(x) := H(x) - \frac{1}{2}\left(A_\infty x, x\right) \tag{5}$$

is a convex function, and

$$N(x) \leq \frac{1}{2}\left(\left(B_\infty - A_\infty\right)x, x\right) + c \quad \forall x \ . \tag{6}$$

Proposition 1. *Assume $H'(0) = 0$ and $H(0) = 0$. Set:*

$$\delta := \liminf_{x \to 0} 2N(x) \|x\|^{-2} \ . \tag{7}$$

If $\gamma < -\lambda < \delta$, the solution $\overline{u}$ is non-zero:

$$\overline{x}(t) \neq 0 \quad \forall t \ . \tag{8}$$

Proof. Condition (7) means that, for every $\delta' > \delta$, there is some $\varepsilon > 0$ such that

$$\|x\| \leq \varepsilon \Rightarrow N(x) \leq \frac{\delta'}{2} \|x\|^2 \ . \tag{9}$$

It is an exercise in convex analysis, into which we shall not go, to show that this implies that there is an $\eta > 0$ such that

$$f \|x\| \leq \eta \Rightarrow N^*(y) \leq \frac{1}{2\delta'} \|y\|^2 \ . \tag{10}$$

Fig. 1. This is the caption of the figure displaying a white eagle and a white horse on a snow field

Since u_1 is a smooth function, we will have $\|hu_1\|_\infty \leq \eta$ for h small enough, and inequality (10) will hold, yielding thereby:

$$\psi(hu_1) \leq \frac{h^2}{2} \frac{1}{\lambda} \|u_1\|_2^2 + \frac{h^2}{2} \frac{1}{\delta'} \|u_1\|^2 \ . \tag{11}$$

If we choose δ' close enough to δ, the quantity $\left(\frac{1}{\lambda}+\frac{1}{\delta'}\right)$ will be negative, and we end up with

$$\psi(hu_1)<0 \qquad \text{for} \quad h\neq 0 \quad \text{small} \ . \tag{12}$$

On the other hand, we check directly that $\psi(0)=0$. This shows that 0 cannot be a minimizer of ψ, not even a local one. So $\overline{u}\neq 0$ and $\overline{u}\neq \Lambda_o^{-1}(0)=0$. □

Corollary 1. *Assume H is C^2 and (a_∞, b_∞)-subquadratic at infinity. Let ξ_1, $\ldots, \xi_N$ be the equilibria, that is, the solutions of $H'(\xi)=0$. Denote by ω_k the smallest eigenvalue of $H''(\xi_k)$, and set:*

$$\omega := \text{Min}\ \{\omega_1, \ldots, \omega_k\} \ . \tag{13}$$

If:

$$\frac{T}{2\pi}b_\infty < -E\left[-\frac{T}{2\pi}a_\infty\right] < \frac{T}{2\pi}\omega \tag{14}$$

then minimization of ψ yields a non-constant T-periodic solution $\overline{x}$.

We recall once more that by the integer part $E[\alpha]$ of $\alpha\in\mathbb{R}$, we mean the $a\in\mathbb{Z}$ such that $a<\alpha\leq a+1$. For instance, if we take $a_\infty=0$, Corollary 2 tells us that $\overline{x}$ exists and is non-constant provided that:

$$\frac{T}{2\pi}b_\infty < 1 < \frac{T}{2\pi} \tag{15}$$

or

$$T\in\left(\frac{2\pi}{\omega}, \frac{2\pi}{b_\infty}\right) \ . \tag{16}$$

Proof. The spectrum of Λ is $\frac{2\pi}{T}\mathbb{Z}+a_\infty$. The largest negative eigenvalue λ is given by $\frac{2\pi}{T}k_o+a_\infty$, where

$$\frac{2\pi}{T}k_o+a_\infty<0\leq\frac{2\pi}{T}(k_o+1)+a_\infty \ . \tag{17}$$

Hence:

$$k_o = E\left[-\frac{T}{2\pi}a_\infty\right] \ . \tag{18}$$

The condition $\gamma < -\lambda < \delta$ now becomes:

$$b_\infty - a_\infty < -\frac{2\pi}{T}k_o - a_\infty < \omega - a_\infty \tag{19}$$

which is precisely condition (14). □

Lemma 1. *Assume that H is C^2 on $\mathbb{R}^{2n}\backslash\{0\}$ and that $H''(x)$ is non-degenerate for any $x \neq 0$. Then any local minimizer $\widetilde{x}$ of ψ has minimal period T.*

Proof. We know that $\widetilde{x}$, or $\widetilde{x} + \xi$ for some constant $\xi \in \mathbb{R}^{2n}$, is a T-periodic solution of the Hamiltonian system:

$$\dot{x} = JH'(x) \ . \tag{20}$$

There is no loss of generality in taking $\xi = 0$. So $\psi(x) \geq \psi(\widetilde{x})$ for all $\widetilde{x}$ in some neighbourhood of x in $W^{1,2}\left(\mathbb{R}/T\mathbb{Z};\mathbb{R}^{2n}\right)$.

But this index is precisely the index $i_T(\widetilde{x})$ of the T-periodic solution $\widetilde{x}$ over the interval $(0,T)$, as defined in Sect. 2.6. So

$$i_T(\widetilde{x}) = 0 \ . \tag{21}$$

Now if $\widetilde{x}$ has a lower period, T/k say, we would have, by Corollary 31:

$$i_T(\widetilde{x}) = i_{kT/k}(\widetilde{x}) \geq k i_{T/k}(\widetilde{x}) + k - 1 \geq k - 1 \geq 1 \ . \tag{22}$$

This would contradict (21), and thus cannot happen. □

Notes and Comments. The results in this section are a refined version of [**?**]; the minimality result of Proposition 14 was the first of its kind.

To understand the nontriviality conditions, such as the one in formula (16), one may think of a one-parameter family x_T, $T \in \left(2\pi\omega^{-1}, 2\pi b_\infty^{-1}\right)$ of periodic solutions, $x_T(0) = x_T(T)$, with x_T going away to infinity when $T \to 2\pi\omega^{-1}$, which is the period of the linearized system at 0.

Table 1. This is the example table taken out of *The TEXbook,* p. 246

Year	World population
8000 B.C.	5,000,000
50 A.D.	200,000,000
1650 A.D.	500,000,000
1945 A.D.	2,300,000,000
1980 A.D.	4,400,000,000

Theorem 1 (Ghoussoub-Preiss). *Assume $H(t,x)$ is $(0,\varepsilon)$-subquadratic at infinity for all $\varepsilon > 0$, and T-periodic in t*

$$H(t,\cdot) \qquad \text{is convex} \quad \forall t \tag{23}$$

$$H(\cdot,x) \qquad \text{is} \quad T-\text{periodic} \quad \forall x \tag{24}$$

$$H(t,x) \geq n\left(\|x\|\right) \qquad \text{with} \quad n(s)s^{-1} \to \infty \quad \text{as} \quad s \to \infty \tag{25}$$

$$\forall \varepsilon > 0\ , \quad \exists c\ :\ H(t,x) \leq \frac{\varepsilon}{2}\|x\|^2 + c\ . \tag{26}$$

Assume also that H is C^2, and $H''(t,x)$ is positive definite everywhere. Then there is a sequence x_k, $k \in \mathbb{N}$, of kT-periodic solutions of the system

$$\dot{x} = JH'(t,x) \tag{27}$$

such that, for every $k \in \mathbb{N}$*, there is some* $p_o \in \mathbb{N}$ *with:*

$$p \geq p_o \Rightarrow x_{pk} \neq x_k \ . \tag{28}$$

□

*Example 1 (*External forcing*).* Consider the system:

$$\dot{x} = JH'(x) + f(t) \tag{29}$$

where the Hamiltonian H is $(0, b_\infty)$-subquadratic, and the forcing term is a distribution on the circle:

$$f = \frac{d}{dt}F + f_o \qquad \text{with} \quad F \in L^2\left(\mathbb{R}/T\mathbb{Z}; \mathbb{R}^{2n}\right) \ , \tag{30}$$

where $f_o := T^{-1}\int_o^T f(t)dt$. For instance,

$$f(t) = \sum_{k\in\mathbb{N}} \delta_k \xi \ , \tag{31}$$

where δ_k is the Dirac mass at $t = k$ and $\xi \in \mathbb{R}^{2n}$ is a constant, fits the prescription. This means that the system $\dot{x} = JH'(x)$ is being excited by a series of identical shocks at interval T.

Definition 1. *Let* $A_\infty(t)$ *and* $B_\infty(t)$ *be symmetric operators in* $\mathbb{R}^{2n}$*, depending continuously on* $t \in [0, T]$*, such that* $A_\infty(t) \leq B_\infty(t)$ *for all* t*.*

A Borelian function $H : [0, T] \times \mathbb{R}^{2n} \to \mathbb{R}$ *is called* (A_∞, B_∞)*-subquadratic at infinity if there exists a function* $N(t, x)$ *such that:*

$$H(t, x) = \frac{1}{2}\left(A_\infty(t)x, x\right) + N(t, x) \tag{32}$$

$$\forall t \ , \quad N(t, x) \qquad \text{is convex with respect to} \quad x \tag{33}$$

$$N(t,x) \geq n\left(\|x\|\right) \qquad \text{with} \quad n(s)s^{-1} \to +\infty \quad \text{as} \quad s \to +\infty \tag{34}$$

$$\exists c \in \mathbb{R} \ : \quad H(t,x) \leq \frac{1}{2}\left(B_\infty(t)x, x\right) + c \quad \forall x \ . \tag{35}$$

If $A_\infty(t) = a_\infty I$ and $B_\infty(t) = b_\infty I$, with $a_\infty \leq b_\infty \in \mathbb{R}$, we shall say that H is (a_∞, b_∞)-subquadratic at infinity. As an example, the function $\|x\|^\alpha$, with $1 \leq \alpha < 2$, is $(0, \varepsilon)$-subquadratic at infinity for every $\varepsilon > 0$. Similarly, the Hamiltonian

$$H(t,x) = \frac{1}{2}k\|k\|^2 + \|x\|^\alpha \tag{36}$$

is $(k, k+\varepsilon)$-subquadratic for every $\varepsilon > 0$. Note that, if $k < 0$, it is not convex.

Notes and Comments. The first results on subharmonics were obtained by Rabinowitz in [**?**], who showed the existence of infinitely many subharmonics both in the subquadratic and superquadratic case, with suitable growth conditions on H'. Again the duality approach enabled Clarke and Ekeland in [**?**] to treat the same problem in the convex-subquadratic case, with growth conditions on H only.

Recently, Michalek and Tarantello (see [**?**] and [**?**]) have obtained lower bound on the number of subharmonics of period kT, based on symmetry considerations and on pinching estimates, as in Sect. 5.2 of this article.

References

1. R. Elderman, L. J. Pater, A. S. Thie, M. M. Drugan, and M. Wiering, "Adversarial reinforcement learning in a cyber security simulation.," in *ICAART (2)*, pp. 559–566, 2017.
2. M. Abadi and D. G. Andersen, "Learning to protect communications with adversarial neural cryptography," *arXiv preprint arXiv:1610.06918*, 2016.

3. S. Shiva, S. Roy, and D. Dasgupta, "Game theory for cyber security," in *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*, p. 34, ACM, 2010.
4. C. Finn, P. Christiano, P. Abbeel, and S. Levine, "A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models," *arXiv preprint arXiv:1611.03852*, 2016.
5. D. Pfau and O. Vinyals, "Connecting generative adversarial networks and actor-critic methods," *arXiv preprint arXiv:1610.01945*, 2016.
6. L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, "Robust adversarial reinforcement learning," *arXiv preprint arXiv:1703.02702*, 2017.
7. X. Lin, P. A. Beling, and R. Cogill, "Multi-agent inverse reinforcement learning for zero-sum games," *arXiv preprint arXiv:1403.6508*, 2014.
8. R. Sutton and A. Barto, "Reinforcement learning: an introduction.[internet]," 2016.