Supplementary Material

Randomized Prediction Games for Adversarial Machine Learning

Samuel Rota Bulò, *Member, IEEE,* Battista Biggio, *Member, IEEE,* Ignazio Pillai, *Member, IEEE,* Marcello Pelillo, *Fellow, IEEE* Fabio Roli, *Fellow, IEEE*

***Abstract*—This document provides sufficient conditions for the uniqueness of a Nash equilibrium in randomized prediction games. The provided conditions generalize the ones given in [1].**

This document is devoted to provide fine-grained assumptions that guarantee the positive definiteness of $\overline{J}_{\boldsymbol{r}}$ and, hence, the uniqueness of the Nash equilibrium via Thm. 2. In particular, we will provide a first set of conditions to rewrite the pseudo-Jacobian $\overline{J}_{\boldsymbol{r}}(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)$ of our game in terms of the pseudo-Jacobian $J_{\boldsymbol{r}}(\boldsymbol{w}, \dot{\mathtt{X}})$ of the underlying prediction game, *i.e.*,

$$J_{\boldsymbol{r}}(\boldsymbol{w}, \dot{\mathtt{X}}) = \begin{bmatrix} r_l \nabla^2_{\boldsymbol{w},\boldsymbol{w}} c_l(\boldsymbol{w}, \dot{\mathtt{X}}) & r_l \nabla^2_{\boldsymbol{w},\dot{\mathtt{X}}} c_l(\boldsymbol{w}, \dot{\mathtt{X}}) \\ r_d \nabla^2_{\dot{\mathtt{X}},\boldsymbol{w}} c_d(\boldsymbol{w}, \dot{\mathtt{X}}) & r_d \nabla^2_{\dot{\mathtt{X}},\dot{\mathtt{X}}} c_d(\boldsymbol{w}, \dot{\mathtt{X}}) \end{bmatrix}, \quad (32)$$

and then provide a further set of conditions on $J_r(\boldsymbol{w}, \dot{\mathtt{X}})$, adapted from [1], to ensure the uniqueness of the Nash equilibrium in our game.

The first set of conditions is related to the parametrized probability distributions $p_{l/d}(\cdot; \boldsymbol{\theta}_{l/d})$ of the two players.

Assumption 3. *There exist random matrices $\mathtt{V}_l \in \mathbb{R}^{\mathtt{m} \times \mathtt{s}_l}$ and $\mathtt{V}_d \in \mathbb{R}^{\mathtt{m} \times \mathtt{s}_d}$ (each defined on a probability space) such that*

- *random variable $\boldsymbol{w}$ distributed as $p_l(\boldsymbol{w}; \boldsymbol{\theta}_l)$ is equivalent in distribution to $\mathtt{V}_l \boldsymbol{\theta}_l$ for all $\boldsymbol{\theta}_l \in \Theta_l$,*
- *random variable $\dot{\mathtt{X}}$ distributed as $p_d(\dot{\mathtt{X}}; \boldsymbol{\theta}_d)$ is equivalent in distribution to $\mathtt{V}_d \boldsymbol{\theta}_d$ for all $\boldsymbol{\theta}_d \in \Theta_d$, and*
- *random variables $\mathtt{V}_{l/d}$ do not depend on $\boldsymbol{\theta}_{l/d}$, respectively.*

Intuitively, random variables $\boldsymbol{w}$ and $\dot{\mathtt{X}}$ depend on the parameters $\boldsymbol{\theta}_l$ and $\boldsymbol{\theta}_d$ in a non-linear way via their probability distributions $p_l(\boldsymbol{w}; \boldsymbol{\theta}_l)$ and $p_d(\dot{\mathtt{X}}; \boldsymbol{\theta}_d)$. Assumption 3 paves the way for the application of a reparametrization trick, which moves the dependence on $\boldsymbol{\theta}_{l/d}$ from the probability distribution to the sample space of a new random variable and, at the same time, makes this dependence linear. This shift allows us to reparametrize the expectations in $\bar{c}_{l/d}$ as follows:

$$\bar{c}_{l/d}(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) = \mathbb{E}[c_{l/d}(\mathtt{V}_l \boldsymbol{\theta}_l, \mathtt{V}_d \boldsymbol{\theta}_d)]\,. \quad (33)$$

Note that Assumption 3 is *not* too restrictive, as many known distributions satisfy the provided conditions (*e.g.*, at least all those within the location-scale family such as Gaussian, Laplace, uniform, Cauchy, Weibull, exponential and many others). Indeed, if *e.g.* p_l is in the location-scale family, then $\boldsymbol{w}$ is equivalent in distribution to $\boldsymbol{\mu} + \mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{z}$, where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with diagonal given by the argument, $\boldsymbol{\mu} \in \mathbb{R}^{s_l}$ is the *location* parameter, $\boldsymbol{\sigma} \in \mathbb{R}^{s_d}_{++}$ is the positive, *scale* parameter, and $\boldsymbol{z}$ is a random variable belonging to the same family with standard parametrization (*i.e.* location zero and unit scale). By setting $\boldsymbol{\theta}_l = (\boldsymbol{\mu}^\top, \boldsymbol{\sigma}^\top)^\top$ and $\mathtt{V}_l = [\mathtt{I}, \mathrm{diag}(\boldsymbol{z})]$ we have that $\mathtt{V}_l \boldsymbol{\theta}_l = \boldsymbol{\mu} + \mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{z} = \boldsymbol{\mu} + \mathrm{diag}(\boldsymbol{z})\boldsymbol{\sigma}$, which is equivalent in distribution to $\boldsymbol{w}$ as required by Assumption 3.

In order to be able to rewrite the pseudo-Jacobian $\overline{J}_{\boldsymbol{r}}$ of our game in terms of $J_{\boldsymbol{r}}$, we further require $c_{l/d}$ to be twice differentiable. This also implies the satisfaction of condition (i) in Assumption 1, *i.e.*, the twice differentiability of $\bar{c}_{l/d}$. To satisfy Assumption 1, as required by Theorem 2, we also assume $c_{l/d}$ to be convex. For this to hold, it is sufficient to assume the convexity of regularizers and losses. This, in conjunction with the linearity of $\mathtt{V}_{l/d}\boldsymbol{\theta}_{l/d}$, will then imply (ii-iii) in Assumption 1. These conditions are summarized in the following assumption.

Assumption 4. *For all values of $\boldsymbol{w}$ and $\dot{\mathtt{X}}$ sampled from p_l and p_d, respectively, the following conditions are satisfied:*

- *(i) regularizers $\Omega_{l/d}$ are strongly convex and twice differentiable at $(\boldsymbol{w}, \dot{\mathtt{X}})$*
- *(ii) for all $y \in \mathcal{Y}$ and $i \in \{1, \ldots, \mathtt{n}\}$, loss functions $\ell_{l/d}(\cdot, y)$ are convex in $\mathbb{R}$, and twice differentiable at $\boldsymbol{w}^\top \dot{\boldsymbol{x}}_i$.*

Under Assumptions 3 and 4, we can finally compute the pseudo-Jacobian $\overline{J}_{\boldsymbol{r}}$ of our game in terms of the pseudo-Jacobian $J_{\boldsymbol{r}}$ of the underlying prediction game:

$$\overline{J}_{\boldsymbol{r}}(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) = \mathbb{E}\left[\mathtt{V}^\top J_{\boldsymbol{r}}(\mathtt{V}_l \boldsymbol{\theta}_l, \mathtt{V}_d \boldsymbol{\theta}_d)\mathtt{V}\right]. \quad (34)$$

Here, matrix $\mathtt{V}$ is the result of the application of the chain rule for the derivatives in (7), and it is a block-diagonal *random* matrix defined as

$$\mathtt{V} = \begin{bmatrix} \mathtt{V}_l & \mathtt{0} \\ \mathtt{0} & \mathtt{V}_d \end{bmatrix}.$$

To ensure the positive definiteness of $\overline{J}_{\boldsymbol{r}}$ we require some additional conditions given in Assumption 5, which depend on the following quantities:

$$\lambda_l = \inf_{\boldsymbol{\theta}_l \in \Theta_l} \lambda_{\min}(\mathbb{E}[\mathtt{V}_l^\top \nabla^2 \Omega_l(\mathtt{V}_l \boldsymbol{\theta}_l)\mathtt{V}_l])\,,$$
$$\lambda_d = \inf_{\boldsymbol{\theta}_d \in \Theta_d} \lambda_{\min}(\mathbb{E}[\mathtt{V}_d^\top \nabla^2 \Omega_d(\mathtt{V}_d \boldsymbol{\theta}_d)\mathtt{V}_d])\,,$$
$$Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) = \sum_i \mathbb{E}[\psi_i(\boldsymbol{\theta}_l^\top \mathtt{V}_l^\top \mathtt{V}_d^{(i)} \boldsymbol{\theta}_d)\mathtt{V}_d^{(i)} \mathtt{V}_l^\top]\,,$$

where $\psi_i(z) = \frac{d}{dz}\ell_l(z, y_i) + \frac{d}{dz}\ell_d(z, y_i)$, ∇^2 is the Hessian operator, $\lambda_{\min}(\cdot)$ is the smallest eigenvalue of the matrix given as argument, and $\mathtt{V}_d^{(i)}$ is the submatrix of $\mathtt{V}_d$ corresponding to

S. Rota Bulò is with ICT-Tev, Fondazione Bruno Kessler, Trento, Italy
M. Pelillo is with DAIS, Università Ca' Foscari, Venezia, Italy
B. Biggio, I. Pillai, F. Roli are with DIEE, University of Cagliari, Italy

$\dot{\boldsymbol{x}}_i$, *i.e.* $\mathtt{V}_d^{(i)}\boldsymbol{\theta}_d$ is equivalent in distribution to $\dot{\boldsymbol{x}}_i$. Note that $\lambda_{l/d}$ are finite since $\Theta_{l/d}$ are compact spaces.

Assumption 5.

(i) for all $y \in \mathcal{Y}$, $i \in \{1, \dots, \mathtt{n}\}$, and for almost all values of $\boldsymbol{w}$ and $\dot{\mathtt{X}}$ sampled from p_l and p_d, respectively,

$$\ell_l''(\boldsymbol{w}^\top \dot{\boldsymbol{x}}_i, y) = \ell_d''(\boldsymbol{w}^\top \dot{\boldsymbol{x}}_i, y)\,,$$

(ii) the players' regularization parameters $\rho_{l/d}$ satisfy

$$\rho_l \rho_d > \frac{\tau}{4\lambda_l \lambda_d}\,,$$

$$\textit{where } \tau = \sup_{\boldsymbol{\theta}_{l/d} \in \Theta_{l/d}} \lambda_{max}\left(Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)^\top\right)\,,$$

Here, $\ell_{l/d}''(z, y) = \frac{d^2}{dz^2}\ell_{l/d}(z, y)$ and $\lambda_{\max}(\cdot)$ returns the largest eigenvalue of the matrix given as argument.

The subsequent lemma states that the positive definiteness of $\overline{J}_{\boldsymbol{r}}$ is implied by Assumptions 3-5:

Lemma 3. *If a randomized prediction game satisfies Assumptions 3–5 then the pseudo-Jacobian $\overline{J}_{\boldsymbol{r}}(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)$ is positive definite for all $(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) \in \Theta_l \times \Theta_d$ by taking $\boldsymbol{r} = (1, 1)^\top$.*

Proof: By substituting (32) into (34) and unfolding the derivatives we can rewirte $\overline{J}_{\boldsymbol{r}}(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)$, after simple algebraic manipulations, as the sum of the following matrices

$$\mathtt{J}^{(1)} = \mathbb{E}\left[\sum_i \begin{bmatrix} \ell_{l,i}''\mathtt{I} & \mathtt{0} \\ \mathtt{0} & \ell_{d,i}''\mathtt{I} \end{bmatrix} \begin{bmatrix} \mathtt{V}_l^\top \mathtt{V}_d^{(i)} \boldsymbol{\theta}_d \\ \mathtt{V}_d^{(i)\top} \mathtt{V}_l \boldsymbol{\theta}_l \end{bmatrix} \begin{bmatrix} \mathtt{V}_l^\top \mathtt{V}_d^{(i)} \boldsymbol{\theta}_d \\ \mathtt{V}_d^{(i)\top} \mathtt{V}_l \boldsymbol{\theta}_l \end{bmatrix}^\top\right]$$

$$\mathtt{J}^{(2)} = \begin{bmatrix} \rho_l \mathbb{E}[\mathtt{V}_l^\top \nabla^2 \Omega_l(\mathtt{V}_l\boldsymbol{\theta}_l)\mathtt{V}_l] & \sum_i \mathbb{E}[\ell_{l,i}' \mathtt{V}_l^\top \mathtt{V}_d^{(i)}] \\ \sum_i \mathbb{E}[\ell_{d,i}' \mathtt{V}_d^{(i)\top} \mathtt{V}_l] & \rho_d \mathbb{E}[\mathtt{V}_d^\top \nabla^2 \Omega_d(\mathtt{V}_d\boldsymbol{\theta}_d)\mathtt{V}_d] \end{bmatrix}$$

where we wrote $\ell_{l,i}''$ for $\ell_l''(\boldsymbol{\theta}_l^\top \mathtt{V}_l^\top \mathtt{V}_d^{(i)} \boldsymbol{\theta}_d, y_i)$, and $\ell_{d,i}''$ for $\ell_d''(\boldsymbol{\theta}_l^\top \mathtt{V}_l^\top \mathtt{V}_d^{(i)} \boldsymbol{\theta}_d, y_i)$. Similarly, we wrote $\ell_{l,i}'$ and $\ell_{d,i}'$ for the first-order derivatives.

It is clear from the structure of $\mathtt{J}^{(1)}$ that it is positive semidefinite for any $\boldsymbol{\theta}_{l/d} \in \Theta_{l/d}$ if $\ell_{l,i}'' = \ell_{d,i}''$ holds almost surely. Therefore, it sufficies to show that $\mathtt{J}^{(2)}$ is positive definite to prove that $\overline{J}_{\boldsymbol{r}}(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)$ is positive definite. Consider the following matrix

$$\mathtt{H} = \begin{bmatrix} 2\rho_l \lambda_l \mathtt{I} & Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)^\top \\ Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) & 2\rho_d \lambda_d \mathtt{I} \end{bmatrix}\,.$$

For any $\boldsymbol{t} = (\boldsymbol{t}_l^\top, \boldsymbol{t}_d^\top)^\top \neq \mathbf{0}$ we have

$$\begin{aligned}\boldsymbol{t}^\top \mathtt{H} \boldsymbol{t} &= 2\rho_l \lambda_l \|\boldsymbol{t}_l\|^2 + 2\rho_d \lambda_d \|\boldsymbol{t}_d\|^2 + 2\boldsymbol{t}_d^\top Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) \boldsymbol{t}_l \\ &\leq \boldsymbol{t}^\top (\mathtt{J}^{(2)} + \mathtt{J}^{(2)^\top}) \boldsymbol{t}\,,\end{aligned}$$

where we used the definition of Q and the inequalities

$$\begin{aligned}\lambda_l \|\boldsymbol{t}_l\|^2 &\leq \boldsymbol{t}_l^\top \mathbb{E}[\mathtt{V}_l^\top \nabla^2 \Omega_l(\mathtt{V}_l\boldsymbol{\theta}_l)\mathtt{V}_l] \boldsymbol{t}_l \\ \lambda_d \|\boldsymbol{t}_d\|^2 &\leq \boldsymbol{t}_d^\top \mathbb{E}[\mathtt{V}_d^\top \nabla^2 \Omega_d(\mathtt{V}_d\boldsymbol{\theta}_d)\mathtt{V}_d] \boldsymbol{t}_d\,,\end{aligned}$$

which follow from the definitions of $\lambda_{l/d}$. Accordingly, we can prove the positive definiteness of $\mathtt{J}^{(2)}$ by showing the positive definitness of $\mathtt{H}$. To this end, we proceed by showing that all roots of the characteristic polinomial $\det(\mathtt{H} - \lambda\mathtt{I})$ of $\mathtt{H}$ are positive. By properties of the determinant[1] we have

$$\begin{aligned}\det(\mathtt{H} - \lambda\mathtt{I}) = \det((2\rho_l\lambda_l - \lambda)\mathtt{I}) \\ \cdot \det\left((2\rho_d\lambda_d - \lambda)\mathtt{I} - \frac{\mathtt{S}}{2\rho_l\lambda_l - \lambda}\right)\,,\end{aligned}$$

where $\mathtt{S}$ is a diagonal matrix with the eigenvalues of $Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)^\top$. The roots of the first determinant term are all equal to $2\rho_l\lambda_l$, which is positive because $\rho_l > 0$ by construction and $\lambda_l > 0$ follows from the strong-convexity of Ω_l in Assumption 4-i. As for the second determinant term, take the ith diagonal element $\mathtt{S}_{ii}$ of $\mathtt{S}$. Then two roots are given by the solution of the following quadratic polinomial

$$\lambda^2 - 2\lambda(\rho_l\lambda_l + \rho_d\lambda_d) + 4\rho_l\rho_d\lambda_l\lambda_d - \mathtt{S}_{ii} = 0\,,$$

which are given by

$$\lambda_{1,2}^{(i)} = \rho_l\lambda_l + \rho_d\lambda_d \pm \sqrt{(\rho_l\lambda_l - \rho_d\lambda_d)^2 + \mathtt{S}_{ii}}\,.$$

Among the two, $\lambda_2^{(i)}$ (the one with the minus) is the smallest one, which is strictly positive if $\rho_l\rho_d > \frac{\mathtt{S}_{ii}}{4\lambda_l\lambda_d}$. Since the condition has to hold for any choice of the eigenvalue $\mathtt{S}_{ii}$ in the right-hand-side of the inequality, we take the maximum one $\max_i \mathtt{S}_{ii}$, which coincides with $\lambda_{\max}(Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) Q(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)^\top)$. We further maximize the right-hand-side with respect to $(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d) \in \Theta_l \times \Theta_d$, because we want the result to hold for any parametrization. Therefrom we recover the variable τ and the condition (ii). ∎

We finally use this lemma in conjunction to Theorem 2 to prove the uniqueness of the Nash equilibrium of a randomized prediction game satisfying Assumptions 3-5.

Theorem 3 (Uniqueness)**.** *A randomized prediction game satisfying Assumptions 3–5 has a unique Nash equilibrium.*

Proof: From Assumption 4 it follows that $c_{l/d}$ are twice differentiable and, hence, also $\bar{c}_{l/d}$ is twice-differentiable and admits the pseudo-Jacobian. Moreover, by Lemma 3 the pseudo-Jacobian $\overline{J}_{\boldsymbol{r}}$ is positive definite. It is also easy to see from (33) that $\bar{c}_l(\cdot; \boldsymbol{\theta}_d)$ is convex in Θ_l for all $\boldsymbol{\theta}_d \in \Theta_d$. Indeed, c_l is convex in $\boldsymbol{w}$ in view of Assumption 4, and $\mathtt{V}_l\boldsymbol{\theta}_l$ is linear in $\boldsymbol{\theta}_l$. Therefore, $c_l(\mathtt{V}_l\boldsymbol{\theta}_l, \mathtt{V}_d\boldsymbol{\theta}_d)$ is convex with respect to $\boldsymbol{\theta}_l$, and since expectations preserve convexity, it follows that $\bar{c}_l(\cdot; \boldsymbol{\theta}_d)$ is convex in Θ_l as required. By the same arguments, also the convexity of $\bar{c}_d(\boldsymbol{\theta}_l; \cdot)$ in Θ_d holds. Hence, Assumption 1 holds and Theorem 2 applies to prove the uniqueness of the Nash equilibrium. ∎

References

[1] M. Brückner, C. Kanzow, and T. Scheffer, "Static prediction games for adversarial learning problems," *J. Mach. Learn. Res.*, vol. 13, pp. 2617–2654, September 2012.

[1] $\det\begin{bmatrix} a\mathtt{I} & \mathtt{B}^\top \\ \mathtt{B} & d\mathtt{I} \end{bmatrix} = \det(a\mathtt{I})\det(d\mathtt{I} - \frac{1}{a}\mathtt{B}\mathtt{B}^\top)$ and if $\mathtt{U}\mathtt{S}\mathtt{U}^\top$ is the eigen-decomposition of $\mathtt{B}\mathtt{B}^\top$ then the latter deterimnant becomes $\det(\mathtt{U}(d\mathtt{I} - \frac{1}{a}\mathtt{S})\mathtt{U}^\top) = \det(d\mathtt{I} - \frac{1}{a}\mathtt{S})$