

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 \bar{J}_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 $\bar{J}_r(\theta_l, \theta_d)$ of our game in terms of the pseudo-Jacobian $J_r(\mathbf{w}, \dot{\mathbf{X}})$ of the underlying prediction game, *i.e.*,

$$J_r(\mathbf{w}, \dot{\mathbf{X}}) = \begin{bmatrix} r_l \nabla_{\mathbf{w}, \mathbf{w}}^2 c_l(\mathbf{w}, \dot{\mathbf{X}}) & r_l \nabla_{\mathbf{w}, \dot{\mathbf{X}}}^2 c_l(\mathbf{w}, \dot{\mathbf{X}}) \\ r_d \nabla_{\dot{\mathbf{X}}, \mathbf{w}}^2 c_d(\mathbf{w}, \dot{\mathbf{X}}) & r_d \nabla_{\dot{\mathbf{X}}, \dot{\mathbf{X}}}^2 c_d(\mathbf{w}, \dot{\mathbf{X}}) \end{bmatrix}, \quad (32)$$

and then provide a further set of conditions on $J_r(\mathbf{w}, \dot{\mathbf{X}})$, 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 $p_{l/d}(\cdot; \theta_{l/d})$ of the two players.

Assumption 3. *There exist random matrices $V_l \in \mathbb{R}^{m \times s_l}$ and $V_d \in \mathbb{R}^{n \times s_d}$ (each defined on a probability space) such that*

- *random variable \mathbf{w} distributed as $p_l(\mathbf{w}; \theta_l)$ is equivalent in distribution to $V_l \theta_l$ for all $\theta_l \in \Theta_l$,*
- *random variable $\dot{\mathbf{X}}$ distributed as $p_d(\dot{\mathbf{X}}; \theta_d)$ is equivalent in distribution to $V_d \theta_d$ for all $\theta_d \in \Theta_d$, and*
- *random variables $V_{l/d}$ do not depend on $\theta_{l/d}$, respectively.*

Intuitively, random variables \mathbf{w} and $\dot{\mathbf{X}}$ depend on the parameters θ_l and θ_d in a non-linear way via their probability distributions $p_l(\mathbf{w}; \theta_l)$ and $p_d(\dot{\mathbf{X}}; \theta_d)$. Assumption 3 paves the way for the application of a reparametrization trick, which moves the dependence on $\theta_{l/d}$ 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:

$$\bar{c}_{l/d}(\theta_l, \theta_d) = \mathbb{E}[c_{l/d}(V_l \theta_l, V_d \theta_d)]. \quad (33)$$

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 \mathbf{w} is equivalent in distribution to $\boldsymbol{\mu} + \text{diag}(\boldsymbol{\sigma})\mathbf{z}$, where

$\text{diag}(\cdot)$ 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 \mathbf{z} is a random variable belonging to the same family with standard parametrization (*i.e.* location zero and unit scale). By setting $\theta_l = (\boldsymbol{\mu}^\top, \boldsymbol{\sigma}^\top)^\top$ and $V_l = [\mathbf{I}, \text{diag}(\mathbf{z})]$ we have that $V_l \theta_l = \boldsymbol{\mu} + \text{diag}(\boldsymbol{\sigma})\mathbf{z} = \boldsymbol{\mu} + \text{diag}(\mathbf{z})\boldsymbol{\sigma}$, which is equivalent in distribution to \mathbf{w} as required by Assumption 3.

In order to be able to rewrite the pseudo-Jacobian \bar{J}_r of our game in terms of J_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 $V_{l/d} \theta_{l/d}$, will then imply (ii-iii) in Assumption 1. These conditions are summarized in the following assumption.

Assumption 4. *For all values of \mathbf{w} and $\dot{\mathbf{X}}$ 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 $(\mathbf{w}, \dot{\mathbf{X}})$*
- (ii) *for all $y \in \mathcal{Y}$ and $i \in \{1, \dots, n\}$, loss functions $\ell_{l/d}(\cdot, y)$ are convex in \mathbb{R} , and twice differentiable at $\mathbf{w}^\top \dot{\mathbf{x}}_i$.*

Under Assumptions 3 and 4, we can finally compute the pseudo-Jacobian \bar{J}_r of our game in terms of the pseudo-Jacobian J_r of the underlying prediction game:

$$\bar{J}_r(\theta_l, \theta_d) = \mathbb{E}[\mathbf{V}^\top J_r(V_l \theta_l, V_d \theta_d) \mathbf{V}]. \quad (34)$$

Here, matrix \mathbf{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

$$\mathbf{V} = \begin{bmatrix} V_l & 0 \\ 0 & V_d \end{bmatrix}.$$

To ensure the positive definiteness of \bar{J}_r we require some additional conditions given in Assumption 5, which depend on the following quantities:

$$\begin{aligned} \lambda_l &= \inf_{\theta_l \in \Theta_l} \lambda_{\min}(\mathbb{E}[V_l^\top \nabla^2 \Omega_l(V_l \theta_l) V_l]), \\ \lambda_d &= \inf_{\theta_d \in \Theta_d} \lambda_{\min}(\mathbb{E}[V_d^\top \nabla^2 \Omega_d(V_d \theta_d) V_d]), \\ Q(\theta_l, \theta_d) &= \sum_i \mathbb{E}[\psi_i(\theta_l^\top V_l^\top V_d^{(i)} \theta_d) V_d^{(i)} V_l^\top], \end{aligned}$$

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

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

Assumption 5.

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

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

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

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

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

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 \bar{J}_r is implied by Assumptions 3-5:

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

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

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

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

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

It is clear from the structure of $\mathbf{J}^{(1)}$ 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 suffices to show that $\mathbf{J}^{(2)}$ is positive definite to prove that $\bar{J}_r(\boldsymbol{\theta}_l, \boldsymbol{\theta}_d)$ is positive definite. Consider the following matrix

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

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

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

where we used the definition of Q and the inequalities

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

which follow from the definitions of $\lambda_{l/d}$. Accordingly, we can prove the positive definiteness of $\mathbf{J}^{(2)}$ by showing the positive definiteness of \mathbf{H} . To this end, we proceed by showing that all

roots of the characteristic polynomial $\det(\mathbf{H} - \lambda \mathbf{I})$ of \mathbf{H} are positive. By properties of the determinant¹ we have

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

where \mathbf{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 i th diagonal element S_{ii} of \mathbf{S} . Then two roots are given by the solution of the following quadratic polynomial

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

which are given by

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

Among the two, $\lambda_2^{(i)}$ (the one with the minus) is the smallest one, which is strictly positive if $\rho_l \rho_d > \frac{S_{ii}}{4\lambda_l \lambda_d}$. Since the condition has to hold for any choice of the eigenvalue S_{ii} in the right-hand-side of the inequality, we take the maximum one $\max_i S_{ii}$, 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 \bar{J}_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 \mathbf{w} in view of Assumption 4, and $\mathbf{V}_l \boldsymbol{\theta}_l$ is linear in $\boldsymbol{\theta}_l$. Therefore, $c_l(\mathbf{V}_l \boldsymbol{\theta}_l, \mathbf{V}_d \boldsymbol{\theta}_d)$ 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 $\det \begin{bmatrix} a\mathbf{I} & \mathbf{B}^\top \\ \mathbf{B} & d\mathbf{I} \end{bmatrix} = \det(a\mathbf{I}) \det(d\mathbf{I} - \frac{1}{a} \mathbf{B} \mathbf{B}^\top)$ and if $\mathbf{U} \mathbf{S} \mathbf{U}^\top$ is the eigen-decomposition of $\mathbf{B} \mathbf{B}^\top$ then the latter determinant becomes $\det(\mathbf{U}(d\mathbf{I} - \frac{1}{a} \mathbf{S})\mathbf{U}^\top) = \det(d\mathbf{I} - \frac{1}{a} \mathbf{S})$