Supplementary Material for Learning in Nonzero-Sum Stochastic Games with Potentials

The Supplementary material is arranged as follows: first, in Sec. A we give a description of the experimental details and report the hyperparameter values used in our experiments. In Sec. B, we give a detailed discussion of our ablation studies. In Sec. C, we give results of our analysis on a static noncooperative game namely Cournot duopoly problem and in Sec. D we perform an study of the problem and verify our solution analytically. In Sec. E, we give additional details on our supervised learning method to compute the potential function. In Sec. F, we give additional details on our distributed learning method using consensus optimisation required to compute the potential function distributively. In Sec. G, we outline some of the additional notation and detail the technical assumptions used in the proofs of our results which are contained in Sec. H which concludes the supplementary material.

A. Experiment Details & Hyperparameter Settings

The settings for all methods are the same, except for the stated cases that use a shared critic. The optimiser is set to Adam for all methods reported. The learning rates for actors and critics are 1e-4 and 1e-3 respectively. Both actors and critics consist of four fully connected layers with dimensions of [64,64,64,n_{act}].

In the table below we report all hyperparameters used in our experiments. Hyperparameter values in square brackets indicate range of values that were used for performance tuning.

Setting	Value
Clip Gradient Norm	1
Discount factor γ_E	0.99
λ	0.95
Learning rate	1x10^{-4} for actor and 1x10^{-3} for critic
Batch size	256
Buffer size	4096
Policy architecture	MLP
Number of parallel actors	1
Optimization algorithm	Adam
Rollout length	1000*[10, 20]

B. Ablation Studies

Our method allows MARL agents to solve noncooperative SGs within the SPG subclass. In this section, we analyse the behaviour of our method compared against existing MARL baselines in scenarios that range from team (cooperative) SG settings to noncooperative games outside of SPGs. In doing so, we examine their performance when the SPG assumptions are violated and show that **SP**ot-**AC** is still able to perform well when the PG condition (Equation (1)) is mildly violated. Additionally, in these settings we also compare the performance of **SP**ot-**AC** in cooperative settings which are the degenerate case of SPGs.

As in Section 6 (within the main body), we consider a stochastic network routing game which has both continuous action and state spaces and is a nonzero sum game (neither team-based nor zero-sum) which represents a challenge for current MARL methods. As in the Network routing games considered in Section 6, we restrict our attention to networks that have efficient NE. In such network structures, playing an NE (best-response) strategy leads to a higher total return for the (self-interested) agent. For these networks, the average return for an agent serves as a measure the performance of the different algorithms.

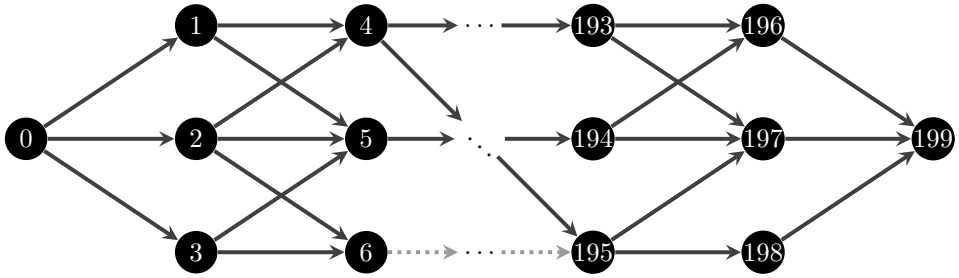

Figure 5. Selfish routing network with 200 nodes.

B.1. Non-Cooperative, Stochastic Potential Game Ablation Study

Fig. 5 demonstrates the large network used in experiments. The class of potential games includes all team-games as a subclass, i.e. every team game is a potential game, and some but not all non-cooperative games are potential. In this ablation study, we examine the performance of **SP**ot-**AC** against other baselines in stochastic potential but non-cooperative games.

To do this, using Lemma C, we know that, in any potential game, the reward function $R_i : \mathcal{S} \times (\times \mathcal{A}_{i\in\mathcal{N}}) \to \mathbb{R}$ for any agent $i \in \mathcal{N}$ can be decomposed into two components: the team game component $J : \mathcal{S} \times (\times \mathcal{A}_{i\in\mathcal{N}}) \to \mathbb{R}$ (i.e. function that all agents seek to maximise) and a strategic (non-cooperative) component which is specific to each agent $L_i : \mathcal{S} \times \left(\times \mathcal{A}_{i\in\mathcal{N}/\{i\}}\right) \to \mathbb{R}$. We now study a set of games in which each agent's reward function has the following form:

$$R_i(s,(a_i,a_{-i})) = \underbrace{J(s,(a_i,a_{-i}))}_{\text{Team game reward}} + c\,\underbrace{L_i(s,a_1,\dots a_{i-1},a_{i+1},\dots,a_N)}_{\text{Non-cooperative part}} \tag{8}$$

The value of the constant $c \in \mathbb{R}$ determines the contribution of the non-cooperative, strategic component. For $c = 0$, the game is a team game and as $c \to \infty$ the non-cooperative component of the game dominates.

As can be seen in the plots, **SP**ot-**AC** has better average return compared to the DDPG-based algorithms, whose performance degrades most due to their team-game requirement. COMIX and COVDN achieve similar levels of performance.

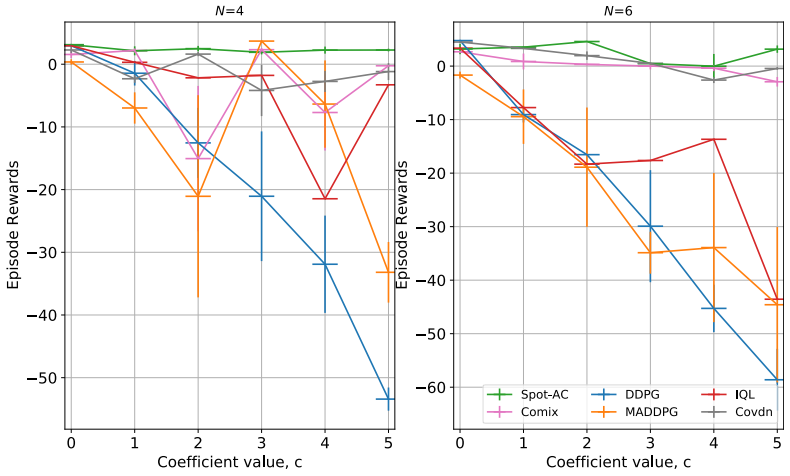

Figure 6. Results of the training curves for the non-cooperative, potential non-atomic routing game when the number of agents $N = 4, 6$, with $c = 0, 1, 2, 3, 4, 5$.

B.2. Non-cooperative, Non-Potential Stochastic Game Ablation Study

Extending the ablation studies of the previous section, we examine the performance of **SP**ot-**AC** in games that are both non-cooperative and not potential. We parameterise the agents' reward function for the congestion game as follows:

$$R_i(s,(a^i,a^{-i})) = \underbrace{r_i(s,(a^i,a^{-i}))}_{\text{potential reward function}} + c \underbrace{J(s,(a^i,a^{-i}))}_{\text{non-potential contribution}}, \qquad \forall s \in \mathcal{S}, \forall (a^i,a^{-i}) \in \mathcal{A}.$$

The functions r_i are those from the original (potential) congestion game (1). J is a generic non-potential reward function. $c = 0$ corresponds to a potential game, and as $c \to \infty$ the non-potential component of the game dominates.

Fig. 7 shows the results of this ablation study in a network routing game with 4 agents. We see that **SP**ot$-$**AC**is able to handle small deviations (small c) from the potential requirements, but performance degrades for larger values. It again outperforms the DDPG baselines, whose performance degrades rapidly with increasing values of the ablation parameter.

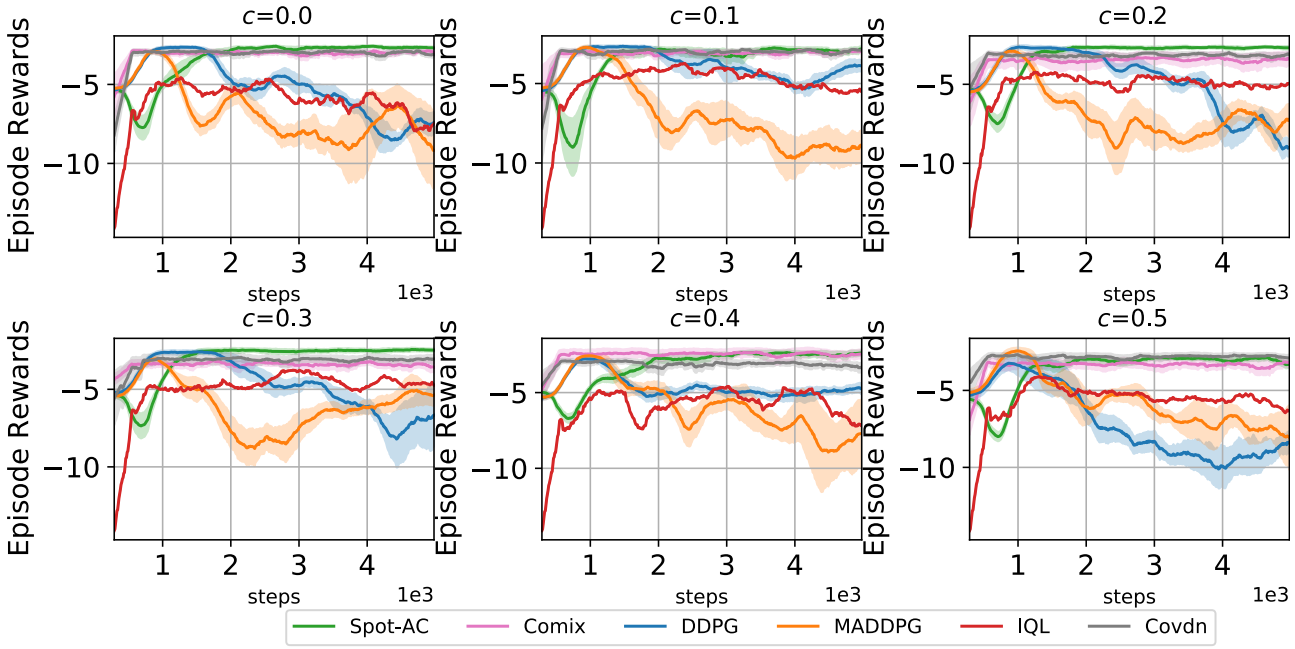

Figure 7. Results of the training curves for non-cooperative, non-potenial, non-atomic routing game when number of agents $N = 4$, coefficients $c = 0.0 - 0.5$.

C. Cournot Duopoly Problem

Cournot Duopoly is a classic static game (Monderer & Shapley, 1996b) that models the imperfect competition in which multiple firms compete in price and production to capture market share. Since the firms' actions are continuous variables, the game is a continuous action setting. It is a nonzero sum game (neither team-based nor zero-sum) which represents a challenge for current MARL methods. Let $a_i \in [-A_i, A_i]$ where $A_i \in \mathbb{R}_{>0}$ which represent the set of actions for Firm $i \in \{1, 2 \ldots, N\} := \mathcal{N}$. Let $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ be given constants, each firm i's reward (profit) is $R_i(a_i, a_{-i}) = a_i(\alpha - \beta \sum_{i \in \mathcal{N}} a_i) - \gamma a_i$. We set $A_i = 1, \forall i \in \mathcal{N}$ and $\alpha = \gamma = 1, \beta = -1$.

D. Analytic Example: Cournot Duopoly

Reward functions

Let $a_1 \in [-A_1, A_1]$ and $a_2 \in [-A_2, A_2]$ where $A_1, A_2 \in \mathbb{R}_{>0}$ which represent the actions for Firm 1 and Firm 2 respectively.

Also let $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ be given constants.

$$R_i(a_1, a_2) = a_i(\alpha - \beta(a_1 + a_2)) - \gamma a_i \tag{9}$$

Cournot Potential Function ($N = 2$ Agents)

$$\phi(a_1, a_2) = \alpha(a_1 + a_2) - \beta(a_1^2 + a_2^2) - \beta a_1 a_2 - \gamma(a_1 + a_2) + k \tag{10}$$

where $k \in \mathbb{R}$ is an arbitrary constant.

D.1. Cournot Duopoly with $N > 2$ Agents

Reward functions

Let $a_i \in [-A_i, A_i]$ where $A_i \in \mathbb{R}_{>0}$ which represent the actions for Firm i, $i \in \{1, 2 \ldots, N\}$.

Also let $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ be given constants.

$$R_i(a_i, a_{-i}) = a_i(\alpha - \beta \sum_{i \in \mathcal{N}} a_i) - \gamma a_i \tag{11}$$

Cournot Potential Function ($N \geq 2$ Agents)

$$\phi(a_i, a_{-i}) = \alpha \sum_{i \in \mathcal{N}} a_i - \beta \sum_{i \in \mathcal{N}} a_i^2 - \beta a_i \sum_{j \in \mathcal{N}/\{i\}} a_j - \gamma \sum_{i \in \mathcal{N}} a_i + k \tag{12}$$

where $k \in \mathbb{R}$ is an arbitrary constant.

Derivatives

$$\frac{\partial R_i(a_i, a_{-i})}{\partial a_i} = \alpha - \beta \sum_{i \in \mathcal{N}} a_i - \beta a_i - \gamma \tag{13}$$

$$\frac{\partial \phi(a_i, a_{-i})}{\partial a_i} = \alpha - 2\beta a_i - \beta \sum_{j \in \mathcal{N}/\{i\}} a_j - \gamma \tag{14}$$

D.2. Analytic Verification of our Method

In this section, we validate that the optimisation in Sec. 5.1 yields the correct results. To do this, we derive analytic expressions for ϕ and show that the solution of the optimisation in Sec. 5.1 coincides with this solution. Recall that our proposition says that:

$$\mathbb{E}_{(a^i, a^{-i}) \sim (\pi_i(\eta^i), \pi_{-i}(\eta^{-i}))} \left[\frac{\partial}{\partial \eta^i} \ln \left[\pi_i(a^i | s; \eta^i) \right] \left(\frac{\partial}{\partial a^i} R_i(s, a^i, a^{-i}) - \frac{\partial}{\partial a^i} \phi(s, a^i, a^{-i}) \right) \right] = \mathbf{0}, \tag{15}$$

We first want to check that any PF in (10) solves (15), indeed:

For $\implies$ we find that

$$\frac{\partial}{\partial a^i} R_i(\cdot, a^i, a^{-i}) = \alpha - \beta(a_1 + a_2) - \beta a_i - \gamma$$

and

$$\begin{aligned}\frac{\partial}{\partial a^i} \phi(\cdot, a^i, a^{-i}) &= \alpha - 2\beta a_i - \beta a_j - \gamma \\ &= \alpha - \beta(a_1 + a_2) - \beta a_i - \gamma\end{aligned}$$

and hence verify:

$$\frac{\partial}{\partial a^i} \phi(\cdot, a^i, a^{-i}) - \frac{\partial}{\partial a^i} R_i(\cdot, a^i, a^{-i}) = 0$$

so that any ϕ in (10) is a candidate solution to (15). Indeed, we observe that

$$\begin{aligned}&\frac{\partial}{\partial a^i} \phi(\cdot, a^i, a^{-i}) - \frac{\partial}{\partial a^i} R_i(\cdot, a^i, a^{-i}) = 0 \\ &\implies \mathbb{E}_{(a^i, a^{-i}) \sim (\pi_i(\eta^i), \pi_{-i}(\eta^{-i}))} \left[\frac{\partial}{\partial \eta^i} \ln \left[\pi_i(a^i | s; \eta^i) \right] \left(\frac{\partial}{\partial a^i} R_i(s, a^i, a^{-i}) - \frac{\partial}{\partial a^i} \phi(s, a^i, a^{-i}) \right) \right] = \mathbf{0},\end{aligned}$$

and hence the forward implication is verified. $\impliedby$

To check the reverse we perform the following optimisation:

$$\text{minimise} \quad \int \left(\frac{\partial}{\partial \eta^i} \ln \left[\pi_i(a^i | \cdot; \eta^i) \right] \left(\frac{\partial}{\partial a^i} R_i(\cdot, a^i, a^{-i}) - \frac{\partial}{\partial a^i} P_\rho(\cdot, a^i, a^{-i}) \right) \right),$$

Consider candidate functions of the following form

$$P_\rho(\cdot, a^i, a^{-i}) = \rho_0 + \rho_{a_1,1} a_1 + \rho_{a_2,1} a_2 + \rho_{a_1,2} a_1^2 + \rho_{a_2,2} a_2^2 + \rho_a a_1 a_2 + c$$

and Gaussian policies: $\pi_i(a^i | \cdot; \eta^i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}(\frac{a^i - \eta}{\sigma})^2}$

Then

$$\frac{\partial}{\partial a^i} P_\rho(\cdot, a^i, a^{-i}) = \rho_{a_i,1} + 2\rho_{a_i,2} a_i + \rho_a a_j$$

$$\text{minimise} \quad -\frac{1}{\sigma^2} \int (a^i - \eta) \left(\alpha - \beta(a_i + a_j) - \beta a_i - \gamma - (\rho_{a_i,1} + 2\rho_{a_i,2} a_i + \rho_a a_j) \right),$$

After matching like terms we find that:

$$\begin{aligned}\alpha - \gamma &= \rho_{a_i,1} \\ -2\beta &= 2\rho_{a_i,2} \\ -\beta &= \rho_a\end{aligned}$$

Hence, we find that

$$\begin{aligned}P_\rho(\cdot, a^i, a^{-i}) &= -(\gamma - \alpha) a_1 - (\gamma - \alpha) a_2 - \beta a_1^2 - \beta a_2^2 - \beta a_1 a_2 + c \\ &= \alpha(a_1 + a_2) - \beta(a_1^2 + a_2^2) - \beta a_1 a_2 - \gamma(a_1 + a_2) = \phi + c\end{aligned}$$

which verifies the reverse.

E. Estimating the Potential Function: Algorithm 2

The following algorithm computes the potential function of the SPG using the supervised learning method described in Sec. 5.1. We illustrate the convergence of the method in Sec. E.1.

Estimating the Potential Function

1: Generate set of random points $((\boldsymbol{\eta}_i^k, \boldsymbol{\eta}_{-i}^k), s^k) \in \boldsymbol{E} \times \mathcal{S}$ for $k = 1, 2, \ldots$ according to the probability density ν.
2: For each $(\boldsymbol{\eta}^k, s^k) \equiv ((\eta_i^k, \eta_{-i}^k), s^k)$, evaluate $g^i(s^k, \boldsymbol{\eta}^k; \boldsymbol{\rho}^k)$ (or for stochastic policies approximate expectation $\mathbb{E}_{(a_i, a_{-i}) \sim (\pi_i, \pi_{-i})} \left[g^i(s^k, \boldsymbol{\eta}^k; \boldsymbol{\rho}^k) \right]$ by MC sampling of actions).
3: Calculate the squared error $g^2(s^k, \boldsymbol{\eta}^k; \boldsymbol{\rho}^k) = \sum_{i \in \mathcal{N}} \left[g^i(s, \boldsymbol{\eta}^k; \boldsymbol{\rho}^k) \right]^2$ using step 2.
4: Take a descent step at $((\boldsymbol{\eta}_i^k, \boldsymbol{\eta}_{-i}^k), s^k)$, compute $\boldsymbol{\rho}^k = \boldsymbol{\rho}^{k-1} - \alpha \nabla_{\boldsymbol{\rho}} g^2(s^k, \boldsymbol{\eta}^k, \boldsymbol{\rho})|^{\boldsymbol{\rho} = \boldsymbol{\rho}^{k-1}}$.
5: Repeat until convergence criterion is satisfied.

E.1. Convergence of The Potential Function

Fig. 8 gives the learning curves for computing the potential function for the selfish routing games using Algorithm 2, corresponding to the method in Section 5.1. The potential function defines the team game which agents jointly seek to maximise. For the training of potential function ϕ, we use a batch size of 2. The learning converges after around 200 iterations, and can easily handle settings with large numbers of agents.

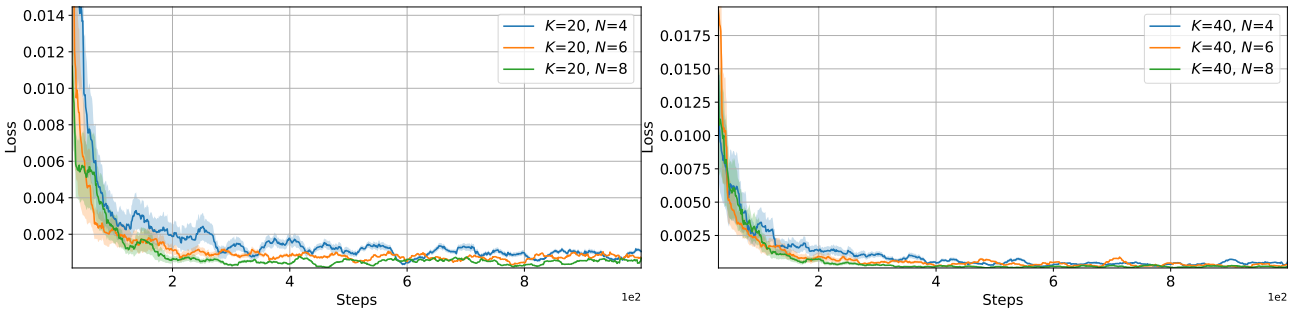

Figure 8. Results of the training curves for potential approximation. Non atomic routing when number of agents $N = 4, 6, 8$, number of nodes $K = 20, 40$. The y-axis is $\|\Delta\phi - \Delta R_i\|^2$ (minimising this quantity gives a candidate function for the potential function c.f. Prop. 1).

F. Consensus Optimisation

In what follows, we denote by $\mathscr{C}$ the set of continuously differentiable functions and by $\mathscr{H}$ the set of measurable functions.

To perform the consensus optimisation step, we use the following update processes:

$$\boldsymbol{\rho}_{l+1}^i = \sum_{j \in \mathcal{N}} c_l(i,j) \rho_l^j - \alpha \cdot \kappa_l^i \tag{16}$$

$$\kappa_{l+1}^i = \sum_{j \in \mathcal{N}} c_l(i,j) \cdot \kappa_l^j + \nabla g^i(s, \boldsymbol{\rho}_{l+1}^i) - \nabla g^i(s, \boldsymbol{\rho}_l^i) \tag{17}$$

where $\alpha > 0$ is the stepsize and $\boldsymbol{C}_l = [c_l(i,j)]_{N \times N}$ is the consensus matrix at iteration $l \in \{0, 1, \ldots\}$.

G. Assumptions

Given a pair of metric spaces (X_1, d_1) and (X_1, d_2), we say that a function $f : X_1 \to X_2$ is Lipschitz if the constant defined by $L_f := \sup_{x_1 \in X_1, x_2 \in X_2} \frac{d_2(f(x_1), f(x_2))}{d_1(x_1, x_2)}$ is finite. The constant L_f is called the Lipschitz constant. We denote by $L_{R\infty} := \max\{L_{R_1}, \ldots L_{R_N}\}$ and similarly for the L-Lipschitz gradients $L_{\frac{\partial R_\infty}{\partial \pi}} := \max\left\{L_{\frac{\partial R_1}{\partial \pi}}, \ldots L_{\frac{\partial R_N}{\partial \pi}}\right\}$.

Consensus update

Require: Parametric function class $\mathscr{H}$, stepsize $\alpha > 0$, initial consensus matrix $\boldsymbol{C}_0 = [c_0(i,j)]_{N\times N}$, initial parameter $\boldsymbol{\rho}_0^i \in \mathrm{T}, \kappa_0^i = \nabla g^i(s, \boldsymbol{\rho}_0^i)$ for all $i \in \mathcal{N}$.

for $l \in \{0, \dots, L-1\}$ **do**
  for agent $i \in \mathcal{N}$ **do**
    $\boldsymbol{\rho}_{l+1}^i = \sum_{j\in\mathcal{N}} c_l(i,j)\rho_l^j - \alpha \cdot \kappa_l^i$
    $\kappa_{l+1}^i = \sum_{j\in\mathcal{N}} c_l(i,j) \cdot \kappa_l^j + \nabla g^i(s, \boldsymbol{\rho}_{l+1}^i) - \nabla g^i(s, \boldsymbol{\rho}_l^i)$
  end for
end for
Output: The vector of functions $[P_{\rho_i}]_{i\in\mathcal{N}}$ for all agent $i \in \mathcal{N}$.

Given a pair of metric spaces (X_1, d_1) and (X_1, d_2), we say that a function $f : X_1 \to X_2$ is Lipschitz if the constant defined by $L_f := \sup_{x_1\in X_1, x_2\in X_2} \frac{d_2(f(x_1),f(x_2))}{d_1(x_1,x_2)}$ is finite. The constant L_f is called the Lipschitz constant. We denote by $L_{R\infty} := \max\{L_{R_1}, \dots L_{R_N}\}$ and similarly for the L-Lipschitz gradients $L_{\frac{\partial R_\infty}{\partial \pi}} := \max\left\{L_{\frac{\partial R_1}{\partial \pi}}, \dots L_{\frac{\partial R_N}{\partial \pi}}\right\}$.

The results of the paper are built under the following assumptions:

Assumption A.2. For any $\theta \in \Theta$, the functions $(R_{i,\theta})_{i\in\mathcal{N}}$ are bounded, measurable functions in the action inputs.
Assumption A.3. The functions $\{R_i\}_{i\in\mathcal{N}}$ are Lipschitz and have L-Lipschitz continuous gradients in $\theta \in \Theta$ that is, for any $i \in \mathcal{N}$, there exist constants $L_{R_i} > 0$ and $L_{\nabla_{\boldsymbol{\theta}} R_i} > 0$ s.th. for any $s, s' \in \mathcal{S}$ and $\forall \boldsymbol{a} \in \boldsymbol{\mathcal{A}}, \theta_a, \theta_b, \theta_c, \theta_d \in \Theta$ we have that:

$$\begin{aligned}&\left\|\nabla_{\boldsymbol{\theta}} R_{i_{\theta_a}}(s'; s, \boldsymbol{a}) - \nabla_{\boldsymbol{\theta}} R_{0_{\theta_b}}(s'; s, \boldsymbol{a})\right\| + \left\|R_{i_{\theta_c}}(s'; s, \boldsymbol{a}) - R_{i_{\theta_d}}(s'; s, \boldsymbol{a})\right\| \\ &\leq L_{\nabla_{\boldsymbol{\theta}} R_i} \|\theta_a - \theta_b\| + L_{R_i} \|\theta_c - \theta_d\| .\end{aligned}$$

Assumption A.4. The functions $\{R_i\}_{i\in\mathcal{N}}$ is continuously differentiable in the state and action inputs.
Assumption A.5. The set of policies $\{\pi\}_{i\in\mathcal{N},\boldsymbol{\eta}\in\boldsymbol{E}}$ is differentiable w.r.t. the policy parameter $\boldsymbol{\eta}$.

Assumption A.2. is rudimental and in general required in optimisation and stochastic approximation theory. Assumptions A.3. and A.4. is typical in Q-learning proofs see pg 27 in (Szepesvári & Munos, 2005) (there in fact the transition function is also assumed to be Lipschitz), assumption A7 in (Antos et al., 2008), in (Szepesvári & Munos, 2005) it is assumed that the transition function and reward function are smooth (see pg 21). In this setting, the assumptions are required to construct approximations of ϕ in terms of a differential equation. Assumption A.1. is fundamental to the structure of a state-based PG. In particular, it extends the notion of potentiality to the state input. The assumption is used in the proof of Theorem 1. Assumption A.5. is standard and required within the framework of policy gradient and actor-critic methods (Sutton et al., 2000; Silver et al., 2014).

H. Proof of Theoretical Results

H.1. Auxiliary Results

Let us denote by $(\mathcal{V}, \|\|)$ any normed vector space.

Lemma A. *For any $f : \mathcal{V} \times \mathcal{V} \to \mathbb{R}, g : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$, we have that:*

$$\left\|\sup_{a\in\mathcal{V}} f(a) - \sup_{a\in\mathcal{V}} g(a)\right\| \leq \sup_{a\in\mathcal{V}} \|f(a) - g(a)\| . \tag{18}$$

Proof.

$$f(a) \leq \|f(a) - g(a)\| + g(a) \tag{19}$$

$$\begin{aligned}\implies \sup_{a\in\mathcal{V}} f(a) &\leq \sup_{a\in\mathcal{V}} \{\|f(a) - g(a)\| + g(a)\} \\ &\leq \sup_{a\in\mathcal{V}} \|f(a) - g(a)\| + \sup_{a\in\mathcal{V}} g(a).\end{aligned} \tag{20}$$

Deducting $\sup_{a\in\mathscr{V}} g(a)$ from both sides of (20) yields:

$$\sup_{a\in\mathscr{V}} f(a) - \sup_{a\in\mathscr{V}} g(a) \leq \sup_{a\in\mathscr{V}} \|f(a) - g(a)\|. \tag{21}$$

After reversing the roles of f and g and redoing steps (19) - (20), we deduce the desired result since the RHS of (21) is unchanged. □

The proof of the Theorem 1 is established through the following results:

Lemma B. *For any c-SPG, the state transitivity conditions holds whenever the reward functions takes the following form:*

$$R_i(s,(a^i,a^{-i})) = g(s,(a^i,a^{-i})) + k(s)h_i(a^i,a^{-i}),$$

for any functions g,k,h for which k^{-1} exists.

Proof. To prove the result, we show that the class of games can be rescaled accordingly. Indeed, we first note that $k^{-1}(s)R_i(s,(a^i,a^{-i})) - k^{-1}(s')R_i(s',(a^i,a^{-i})) = k^{-1}(s)g(s,(a^i,a^{-i})) - k^{-1}(s')g(s',(a^i,a^{-i}))$. Using the invertibility of k, we now consider a rescaled game
$\mathscr{M}(s) = \langle (\mathscr{A}_i)_{i\in\mathcal{N}}, (V_i(s))_{i\in\mathcal{N}}, \mathcal{N}\rangle$ where $V_i := k^{-1}R_i$, then it is easy to see that for these games we have: $V_i(s,(a^i,a^{-i})) - V_i(s',(a^i,a^{-i})) = L(s,(a^i,a^{-i})) - L(s',(a^i,a^{-i}))$ where $L := k^{-1}g$ and hence the state transitivity assumption is satisfied. □

Lemma C. *For any PG, there exists a function $B : \mathbf{\Pi}\times\mathcal{S}\to\mathbb{R}$ ($B\in\mathscr{H}$) such that the following holds for any $i\in\mathcal{N}, \forall(a_t^i,a_t^{-i})\in\boldsymbol{\mathcal{A}}, \forall s\in\mathcal{S}$ $R_i(s,a_t^i,a_t^{-i}) = \phi(s,a_t^i,a_t^{-i}) + F_i(s,a_t^{-i})$ where F_i satisfies $F_i(s,a_t^{-i}) = F_i(s',a_t^{-i})$.*

The result generalises dummy-coordination separability known in PGs to a state-based setting (Slade, 1994; Ui, 2000).

Proof of Lemma C. To establish the forward implication, we make the following observation which is straightforward:

$$\begin{aligned}
&R_i(s,a_t^i,a_t^{-i}) - R_i(s,a_t'^i,a_t^{-i})\\
&= \phi(s,a_t^i,a_t^{-i}) + F_i(s,a_t^{-i}) - \left(\phi(s,a_t^i,a_t^{-i}) + F_i(s,a_t^{-i})\right)\\
&= \phi(s,a_t^i,a_t^{-i}) - \phi(s,a_t^i,a_t^{-i})
\end{aligned}$$

To prove the reverse assume that the game is a state-based potential game. Let us now define the function $T_i(s,a_t^i,a_t^{-i}) := R_i(s,a_t^i,a_t^{-i}) - \phi(s,a_t^i,a_t^{-i})$, then we observe that: $R_i(s,a_t^i,a_t^{-i}) - R_i(s,a_t'^i,a_t^{-i}) = \phi(s,a_t^i,a_t^{-i}) - \phi(s,a_t'^i,a_t^{-i}) \iff R_i(s,a_t^i,a_t^{-i}) - \phi(s,a_t^i,a_t^{-i}) = R_i(s,a_t'^i,a_t^{-i}) - \phi(s,a_t'^i,a_t^{-i})$ and hence $T_i(s,a_t^i,a_t^{-i}) = T_i(s,a_t'^i,a_t^{-i})$ which implies that $T_i(s,a_t^i,a_t^{-i}) \equiv K_i(s,a_t^{-i})$. In a similar way, writing $T_i(s,a_t^i,a_t^{-i}) := R_i(s,a_t^i,a_t^{-i}) - \phi(s,a_t^i,a_t^{-i})$ and using the state transitive property, we deduce that $T_i(s',a_t^i,a_t^{-i}) = T_i(s,a_t^i,a_t^{-i})$ which settles the proof. □

H.2. Proof of Main Results

Proof of Theorem 1

Proposition 4. *There exists a function $B:\mathbf{\Pi}\times\mathcal{S}\to\mathbb{R}$ ($B\in\mathscr{H}$) and the following holds for any $i\in\mathcal{N}$*

$$\mathbb{E}_{s\sim P(\cdot|)}\left[v_i^{\boldsymbol{\pi}}(s) - v_i^{\boldsymbol{\pi}'}(s)\right] = \mathbb{E}_{s\sim P(\cdot|)}\left[B^{\boldsymbol{\pi}}(s) - B^{\boldsymbol{\pi}'}(s)\right]. \tag{22}$$

Proof. We prove the proposition in two parts beginning with the finite case then extending to the infinite horizon case.

Hence, we first seek to show that for any joint strategy $(\pi_i,\pi_{-i})\in\mathbf{\Pi}$, define by $v_{i,k}$ the value function for the finite horizon game of length $k\in\mathbb{N}$ (i.e.
$v_{i,k}^{\boldsymbol{\pi}}(s) := \mathbb{E}_{s_t\sim P,\pi_i,\pi_{-i}}\left[\sum_{t=0}^{k}\gamma^t R_i(s_t,\boldsymbol{a}_t)|s\equiv s_0\right]$ for any $i\in\mathcal{N}$ and $k<\infty$). Then there exists a function $B_k : \mathbf{\Pi}\times\mathcal{S}\to\mathbb{R}$ such that the following holds for any $i\in\mathcal{N}$ and $\forall\pi_i,\pi_i'\in\Pi_i, \forall\pi_{-i}\in\Pi_{-i}$:

$$\mathbb{E}_{s\sim P(\cdot|)}\left[v_{i,k}^{\boldsymbol{\pi}}(s) - v_{i,k}^{\boldsymbol{\pi}'}(s)\right] = \mathbb{E}_{s\sim P(\cdot|)}\left[B_k^{\boldsymbol{\pi}}(s) - B_k^{\boldsymbol{\pi}'}(s)\right]. \tag{23}$$

For the finite horizon case, the result is proven by induction on the number of time steps until the end of the game. Unlike the infinite horizon case, for the finite horizon case the value function and policy have an explicit time dependence.

We consider the case of the proposition at time $T-1$ that is we evaluate the value functions at the penultimate time step. In the following, we employ the shorthands $\boldsymbol{a}_k \equiv (a_k^i, a_k^{-i})$ and by $\boldsymbol{a'}_k \equiv (a_k'^i, a_k^{-i})$ for any $k \in \mathbb{N}$ and similarly $\boldsymbol{\pi}(\cdot) \equiv \prod_{j\in\mathcal{N}} \pi_j$ and $\boldsymbol{\pi'}(\cdot) \equiv \prod_{j\in\mathcal{N}/\{i\}} \pi_j(\cdot)\cdot\pi_i'(\cdot)$. We will also use the shorthands $F^{\boldsymbol{\pi}} \equiv F^{(\pi_i,\pi_{-i})}$ and $F^{\boldsymbol{\pi'}} \equiv F^{(\pi_i',\pi_{-i})}$ given some function F.

In what follows and for the remainder of the script, we employ the following shorthands:

$$\mathcal{P}_{ss'}^{\boldsymbol{a}} =: P(s';\boldsymbol{a},s), \quad \mathcal{P}_{ss'}^{\boldsymbol{\pi}} =: \int_{\mathcal{A}} d\boldsymbol{a}_t \boldsymbol{\pi}(\boldsymbol{a}_t|s)\mathcal{P}_{ss'}^{\boldsymbol{a}_t}, \quad \mathcal{R}_i^{\boldsymbol{\pi}}(s_t) := \int_{\mathcal{A}} d\boldsymbol{a}_t \boldsymbol{\pi}(\boldsymbol{a}_t|s_t) R_i(s_t,\boldsymbol{a}_t)$$

$$\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_t) := \int_{\mathcal{A}} d\boldsymbol{a}_t \boldsymbol{\pi}(\boldsymbol{a}_t|s_t)\phi(s_t,\boldsymbol{a}_t), \quad \boldsymbol{F}_i^{\boldsymbol{\pi}}(s_t) := \int_{\mathcal{A}} d\boldsymbol{a}_t \boldsymbol{\pi}(\boldsymbol{a}_t|s_t) F_i(s_t,\boldsymbol{a}_t), \quad \boldsymbol{v}_{i,k}^{\boldsymbol{\pi}}(s_t) := \int_{\mathcal{A}} d\boldsymbol{a}_t \boldsymbol{\pi}(\boldsymbol{a}_t|s_t) v_{i,k}^{\boldsymbol{\pi}}(s_t,\boldsymbol{a}_t)$$

In this case, we have that:

$$\begin{aligned}
&\mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[v_{i,T-1}^{\boldsymbol{\pi}}(s_{T-1}) - v_{i,T-1}^{\boldsymbol{\pi'}}(s_{T-1})\right] \\
&= \mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[\mathcal{R}_i^{\boldsymbol{\pi}}(s_T) + \gamma\int_{\mathcal{S}} ds_T \int_{\mathcal{A}} \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} v_i^{\boldsymbol{\pi}}(s_T) - \left(\mathcal{R}_i^{\boldsymbol{\pi'}}(s_T) + \gamma\int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} v_i^{\boldsymbol{\pi'}}(s_T)\right)\right] \\
&= \mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_T) - \boldsymbol{\phi}^{\boldsymbol{\pi'}}(s_T) + \gamma\int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} v_i^{\boldsymbol{\pi}}(s_T) - \gamma\int_{\mathcal{S}} ds \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} v_i^{\boldsymbol{\pi'}}(s_T)\right] \\
&= \mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_{T-1}) - \boldsymbol{\phi}^{\boldsymbol{\pi'}}(s_{T-1}) + \gamma\mathbb{E}_{s_T\sim P(\cdot|)}\left[\boldsymbol{v}_i^{\boldsymbol{\pi}}(s_T) - \boldsymbol{v}_i^{\boldsymbol{\pi'}}(s_T)\right]\right].
\end{aligned}$$

We now observe that for any $\pi_i \in \Pi_i$ and for any $\pi_{-i} \in \Pi_{-i}$ we have that $\forall i \in \mathcal{N},\ v_i^{\pi^i,\pi^{-i}}(s_T) = \mathbb{E}_{s_T\sim P(\cdot|)}\left[\mathcal{R}_i^{\boldsymbol{\pi}}(s_T)\right]$, moreover we have that for any $\pi_i, \pi_i' \in \Pi_i$ and for any $i \in \mathcal{N}$, we have

$$\begin{aligned}
&\mathbb{E}_{s_T\sim P(\cdot|)}\left[\mathcal{R}_i^{\boldsymbol{\pi}}(s_T) - \mathcal{R}_i^{\boldsymbol{\pi'}}(s_T)\right] \\
&= \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} \mathcal{R}_i^{\boldsymbol{\pi}}(s_T) - \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} \mathcal{R}_i^{\boldsymbol{\pi'}}(s_T) \\
&= \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} \left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_T) + \boldsymbol{F}_i(a_T^{-i})\right] - \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} \left[\boldsymbol{\phi}^{\boldsymbol{\pi'}}(s_T) + \boldsymbol{F}_i(a_T^{-i})\right] \\
&= \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} \boldsymbol{\phi}^{\boldsymbol{\pi}}(s_T) - \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} \boldsymbol{\phi}^{\boldsymbol{\pi'}}(s_T) + \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} \boldsymbol{F}_i(a_T^{-i}) - \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} \boldsymbol{F}_i(a_T^{-i}) \\
&= \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} \boldsymbol{\phi}^{\boldsymbol{\pi}}(s_T) - \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} \boldsymbol{\phi}^{\boldsymbol{\pi'}}(s_T) \\
&= \mathbb{E}_{s_T\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_T) - \boldsymbol{\phi}^{\boldsymbol{\pi'}}(s_T)\right]
\end{aligned}$$

Since

$$\begin{aligned}
&\int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi}} \boldsymbol{F}_i(a_T^{-i}) - \int_{\mathcal{S}} ds_T \mathcal{P}_{s_T s_{T-1}}^{\boldsymbol{\pi'}} \boldsymbol{F}_i(a_T^{-i}) \\
&= \int_{\mathcal{A}_{-i}} \pi_i(da^i, s_{T-1}) F_i(a^{-i}) \int_{\mathcal{A}_i} \pi_{-i}(da^{-i}, s_{T-1}) \int_{\mathcal{S}} ds_T P(s_T; s_{T-1}, \boldsymbol{a}_T) \\
&\qquad - \int_{\mathcal{A}_{-i}} \pi_i'(da'^i, s_{T-1}) F_i(a^{-i}) \int_{\mathcal{A}_i} \pi_{-i}(da^{-i}, s_{T-1}) \int_{\mathcal{S}} ds_T P(s_T; s_{T-1}, \boldsymbol{a'}_T) \\
&= \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, s_{T-1}) F_i(a^{-i}) \left\{\int_{\mathcal{A}_i} \pi_i(da^i, s_{T-1}) - \int_{\mathcal{A}_i} \pi_i'(da'^i, s_{T-1})\right\}
\end{aligned}$$

$$= 0$$

Hence, we find that

$$\begin{aligned}
&\mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[v^{\boldsymbol{\pi}}_{i,T-1}(s_{T-1}) - v^{\boldsymbol{\pi}'}_{i,T-1}(s_{T-1})\right] \\
&= \mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_{T-1}) - \boldsymbol{\phi}^{\boldsymbol{\pi}'}(s_{T-1}) + \gamma\int_{\mathcal{S}} ds_T \mathcal{P}^{\boldsymbol{\pi}}_{s_T s_{T-1}}\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_T) - \gamma\int_{\mathcal{S}} ds_T \mathcal{P}^{\boldsymbol{\pi}'}_{s_T s_{T-1}}\boldsymbol{\phi}^{\boldsymbol{\pi}'}(s_T)\right] \\
&= \mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_{T-1}) - \boldsymbol{\phi}^{\boldsymbol{\pi}'}(s_T) + \gamma\mathbb{E}_{s_T\sim P(s_T|\cdot)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_{T-1}) - \boldsymbol{\phi}^{\boldsymbol{\pi}'}(s_T)\right]\right] \\
&= \mathbb{E}_{s_{T-1}\sim P(\cdot|)}\left[B^{\boldsymbol{\pi}}_{T-1}(s_{T-1}) - B^{\boldsymbol{\pi}'}_{T-1}(s_{T-1})\right],
\end{aligned}$$

using the iterated law of expectations in the last line and where

$$B^{\boldsymbol{\pi}}_T(s) := \mathbb{E}_{s_t\sim P,\pi_i,\pi_{-i}}\left[\sum_{t=0}^{T}\gamma^t\phi(s_t,\boldsymbol{a}_t)|s\equiv s_0\right]. \tag{24}$$

Hence, we have succeeded in proving that the expression (22) holds for $T-k$ when $k=1$.

Our next goal is to prove that the expression holds for any $0<k\leq T$.

Note that for any $T\geq k>0$, we can write (24) as

$$B^{\boldsymbol{\pi}}_{T-k}(s) = \mathbb{E}_{\pi_i,\pi_{-i}}\left[\phi(s,\boldsymbol{a}_k) + \gamma\int_{\mathcal{S}} ds' P(s';s,\boldsymbol{a}_k)B^{\pi^i,\pi^{-i}}_{T-(k+1)}(s')\cdot 1_{k\leq T}\right].$$

Now we consider the case when we evaluate the expression (22) for any $0<k\leq T$. Our inductive hypothesis is the expression holds for some $0<k\leq T$, that is for a $0<k\leq T$ we have that:

$$\mathbb{E}_{s_{T-k}\sim P(\cdot|)}\left[v^{\boldsymbol{\pi}}_{i,k}(s_{T-k}) - v^{\boldsymbol{\pi}'}_{i,k}(s_{T-k})\right] = \mathbb{E}_{s_{T-k}\sim P(\cdot|)}\left[B^{\boldsymbol{\pi}}_k(s_{T-k}) - B^{\boldsymbol{\pi}'}_k(s_{T-k})\right]. \tag{25}$$

It is easy to see that given (25) and Lemma C, it must be the case that:

$$\mathbb{E}_{s_{T-k}\sim P(\cdot|)}\left[v^{\boldsymbol{\pi}}_{i,k}(s_{T-k})\right] = \mathbb{E}_{s_{T-k}\sim P(\cdot|)}\left[B^{\boldsymbol{\pi}}_k(s_{T-k}) + G^{\pi_{-i}}_{i,k}(s_{T-k})\right]. \tag{26}$$

where $G^{\pi_{-i}}_{i,k}(s) := \mathbb{E}_{P,\pi_{-i}}\left[\sum_{t=0}^{k}\gamma^t F_{-i}(s,a^{-i}_t)\right]$.

Moreover, we recall that F_{-i} satisfies the condition $F_{-i}(s,a^{-i}_t) = F_{-i}(s',a^{-i}_t)$, hence $G^{\pi_{-i}}_{i,k}(s) = G^{\pi_{-i}}_{i,k}(s')$ so from now on we drop the dependence on s and write $G^{\pi_{-i}}_{i,k}$.

It remains to show that the expression holds for $k+1$ time steps prior to the end of the horizon. The result can be obtained using the dynamic programming principle and the base case ($k=1$) result, indeed we have that

$$\begin{aligned}
&\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[v^{\boldsymbol{\pi}}_{i,k+1}(s_{T-(k+1)}) - v^{\boldsymbol{\pi}'}_{i,k+1}(s_{T-(k+1)})\right] \\
&= \mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\Bigg[\mathcal{R}^{\boldsymbol{\pi}}_i(s_{T-(k+1)}) + \gamma\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-(k+1)}s_{T-k}}\boldsymbol{v}^{\boldsymbol{\pi}}_{i,k}(s_{T-k}) \\
&\qquad\qquad - \mathcal{R}^{\boldsymbol{\pi}'}_i(s_{T-(k+1)}) - \gamma\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-(k+1)}s_{T-k}}\boldsymbol{v}^{\boldsymbol{\pi}'}_{i,k}(s_{T-k})\Bigg] \\
&= \mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[\mathcal{R}^{\boldsymbol{\pi}}_i(s_{T-(k+1)}) - \mathcal{R}^{\boldsymbol{\pi}'}_i(s_{T-(k+1)})\right] \\
&+ \gamma\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[\int_{\mathcal{S}} ds\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-(k+1)}s_{T-k}}\boldsymbol{v}^{\boldsymbol{\pi}}_{i,k}(s_{T-k}) - \mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-(k+1)}s_{T-k}}\boldsymbol{v}^{\boldsymbol{\pi}'}_{i,k}(s_{T-k})\right].
\end{aligned}$$

Studying the terms under the first expression, we observe that by construction, we have that:

$$\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[\mathcal{R}_i^{\boldsymbol{\pi}}(s_{T-(k+1)})-\mathcal{R}_i^{\boldsymbol{\pi}'}(s_{T-(k+1)})\right]=\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_{T-(k+1)})-\boldsymbol{\phi}^{\boldsymbol{\pi}'}(s_{T-(k+1)})\right]. \tag{27}$$

We now study the terms within the second expectation.

Using (25) (i.e. the inductive hypothesis), we find that:

$$\begin{aligned}
&\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{v}^{\boldsymbol{\pi}}_{i,k}(s_{T-k})-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{v}^{\boldsymbol{\pi}'}_{i,k}(s_{T-k})\\
&=\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\left[\boldsymbol{B}^{\boldsymbol{\pi}}_{k}(s_{T-k})+\boldsymbol{G}^{\pi_{-i}}_{i,k}\right]-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\left[\boldsymbol{B}^{\boldsymbol{\pi}'}_{k}(s_{T-k})+\boldsymbol{G}^{\pi_{-i}}_{i,k}\right]\\
&=\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{B}^{\boldsymbol{\pi}}_{k}(s_{T-k})-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{B}^{\boldsymbol{\pi}'}_{k}(s_{T-k})\\
&\qquad+\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{G}^{\pi_{-i}}_{i,k}-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{G}^{\pi_{-i}}_{i,k}
\end{aligned}$$

We now observe that

$$\begin{aligned}
&\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{G}^{\pi_{-i}}_{i,k}-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{G}^{\pi_{-i}}_{i,k}\\
&=\int_{\mathcal{S}} ds_{T-k}\int_{\boldsymbol{\mathcal{A}}}\left[\pi_i(da^i_{T-(k+1)},s_{T-(k+1)})-\pi'_i(da^i_{T-(k+1)},s_{T-(k+1)})\right]\\
&\qquad\cdot P(s_{T-k};s_{T-(k+1)},\boldsymbol{a}_{T-(k+1)})\pi_{-i}(da^{-i}_{T-(k+1)},s_{T-(k+1)})G^{\pi_{-i}}_{i,k}\\
&=\int_{\mathcal{S}} ds_{T-k}\int_{\mathcal{A}_{-i}}\pi_{-i}(da^{-i}_{T-(k+1)},s_{T-(k+1)})\\
&\qquad\cdot\left(P(s_{T-k};s_{T-(k+1)},\pi_i,a^{-i}_{T-(k+1)})-P(s_{T-k};s_{T-(k+1)},\pi'_i,a^{-i}_{T-(k+1)})\right)G^{\pi_{-i}}_{i,k}\\
&=0
\end{aligned}$$

We now find that:

$$\begin{aligned}
&\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{v}^{\boldsymbol{\pi}}_{i,k}(s_{T-k})-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{v}^{\boldsymbol{\pi}'}_{i,k}(s_{T-k})\\
&=\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{B}^{\boldsymbol{\pi}}_{k}(s_{T-k})-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{B}^{\boldsymbol{\pi}'}_{k}(s_{T-k})
\end{aligned}\tag{28}$$

Now combining (27) and (28) leads to the fact that:

$$\begin{aligned}
&\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[v^{\boldsymbol{\pi}}_{i,k+1}(s_{T-(k+1)})-v^{\boldsymbol{\pi}'}_{i,k+1}(s_{T-(k+1)})\right]\\
&=\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}}_{s_{T-k}s_{T-k-1}}\boldsymbol{B}^{\boldsymbol{\pi}}_{k}(s_{T-k})-\int_{\mathcal{S}} ds_{T-k}\mathcal{P}^{\boldsymbol{\pi}'}_{s_{T-k}s_{T-k-1}}\boldsymbol{B}^{\boldsymbol{\pi}'}_{k}(s_{T-k})\right]\\
&\qquad+\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[\boldsymbol{\phi}^{\boldsymbol{\pi}}(s_{T-(k+1)})-\boldsymbol{\phi}^{\boldsymbol{\pi}'}(s_{T-(k+1)})\right]
\end{aligned},$$

from which we immediately deduce that

$$\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[v^{\boldsymbol{\pi}}_{i,k+1}(s_{T-(k+1)})-v^{\boldsymbol{\pi}'}_{i,k+1}(s_{T-(k+1)})\right]=\mathbb{E}_{s_{T-(k+1)}\sim P(\cdot|)}\left[B^{\boldsymbol{\pi}}_{k+1}(s_{T-(k+1)})-B^{\boldsymbol{\pi}'}_{k+1}(s_{T-(k+1)})\right],$$

where

$$B^{\boldsymbol{\pi}}_{k}(s)=\mathbb{E}_{\pi_i,\pi_{-i}}\left[\phi(s_k,\boldsymbol{a}_k)+\gamma\int_{\mathcal{S}} ds'P(s';s,\boldsymbol{a}_k)B^{\pi^i,\pi^{-i}}_{k-1}(s')\right],$$

from which we deduce the result. □

Thus far we have established the relation (23) holds only for the finite horizon case. We now extend the coverage to the infinite horizon case in which we can recover the use of stationary strategies. Before doing so, we require the following results:

Lemma D. *For any $t' < \infty$, define by $B^{\boldsymbol{\pi}}_{t'} : \mathbb{N} \times \Pi_i \times \Pi_{-i} \times \mathcal{S} \to \mathbb{R}$ the following function:*

$$B^{\boldsymbol{\pi}}_{t'}(s) := \mathbb{E}_{s_t \sim P, \pi_i, \pi_{-i}} \left[\sum_{t=0}^{T} \gamma^t \phi(s_t, \boldsymbol{a}_t) | s \equiv s_0 \right].$$

then $\exists B^{\boldsymbol{\pi}} : \Pi_i \times \Pi_{-i} \times \mathcal{S} \to \mathbb{R}$ s.t. $\forall s \in \mathcal{S}$ and for any $\pi_i \in \Pi_i, \pi_{-i} \in \Pi_{-i}$,

$$\lim_{t \to \infty} B^{\boldsymbol{\pi}}_t(s) = B^{\boldsymbol{\pi}}(s),$$

where for any $\pi_i \in \Pi_i, \pi_{-i} \in \Pi_{-i}$, the function $B^{\boldsymbol{\pi}}$ is given by:
$B^{\boldsymbol{\pi}}(s) := \mathbb{E}_{s_t \sim P, \pi_i, \pi_{-i}} [\sum_{t=0}^{\infty} \gamma^t \phi(s_t, \boldsymbol{a}_t) | s \equiv s_0]$.

Proof. We prove the result by showing that the sequence $B^{\boldsymbol{\pi}}_n, B^{\boldsymbol{\pi}}_{n+1}, \ldots$ converges uniformly, that is the sequence is a Cauchy sequence. In particular, we show that $\forall \epsilon > 0, \exists T' > 0$ s.th. $\forall t', t'' > T'$ and for any $\pi_i \in \Pi_i, \pi_{-i} \in \Pi_{-i}$

$$\| B^{\boldsymbol{\pi}}_{t'} - B^{\boldsymbol{\pi}}_{t''} \| < \epsilon.$$

Firstly, we deduce that the function ϕ is bounded since each R_i is bounded also (c.f. (43)). Now w.log., consider the case when $t' \geq t''$. We begin by observing the fact that

$$\begin{aligned}
&B^{\boldsymbol{\pi}}_{t'}(s) - B^{\boldsymbol{\pi}}_{t''}(s) \\
&= \mathbb{E}_{s_t \sim P(\cdot; s_{t-1}, a_{t-1}), \pi_i, \pi_{-i}} \left[\sum_{t=0}^{t'} \gamma^t \phi_t(s_t, \boldsymbol{a}_t) - \sum_{t=0}^{t''} \gamma^t \phi_t(s_t, \boldsymbol{a}_t) \right] \\
&= \mathbb{E}_{s_t \sim P(\cdot; s_{t-1}, a_{t-1}), \pi_i, \pi_{-i}} \left[\sum_{t=t''}^{t'} \gamma^t \phi_t(s_t, \boldsymbol{a}_t) \right].
\end{aligned}$$

Hence, we find that

$$\begin{aligned}
&|B^{\boldsymbol{\pi}}_{t'}(s) - B^{\boldsymbol{\pi}}_{t''}(s)| \\
&= \left| \mathbb{E}_{s_t \sim P(\cdot; s_{t-1}, a_{t-1}), \pi_i, \pi_{-i}} \left[\sum_{t=t''}^{t'} \gamma^t \phi_t(s_t, \boldsymbol{a}_t) \right] \right| \\
&\leq \sum_{t=t''}^{t'} \gamma^t \|\phi\|_\infty \leq |\gamma| \frac{\left| \gamma^{t''} - \gamma^{t'} \right|}{1-\gamma} \|\phi\|_\infty \\
&\leq |\gamma^{t''}| \frac{\left| 1 - \gamma^{t'-t''} \right|}{1-\gamma} \|\phi\|_\infty \\
&\leq \frac{|\gamma^{t''}|}{1-\gamma} \|\phi\|_\infty = e^{t'' \ln \gamma} \frac{\|\phi\|_\infty}{1-\gamma} \\
&= e^{-t'' |\ln \gamma|} \left(\frac{\|\phi\|_\infty}{1-\gamma} \right) \leq e^{-T' |\ln \gamma|} \left(\frac{\|\phi\|_\infty}{1-\gamma} \right),
\end{aligned}$$

using Cauchy-Schwarz and since $t' \geq t'' > T'$ and $\gamma \in [0, 1[$. The inequality of the proposition is true whenever T' is chosen to satisfy

$$T' \geq \left| \ln(\epsilon) (\ln(\gamma) \left(\frac{\|\phi\|_\infty}{1-\gamma} \right)^{-1} \right|,$$

hence the result is proven. □

We are now in a position to extend the dynamic potential property (23) to the infinite horizon case:

Proof. The result is proven by contradiction.

To this end, let us firstly assume there exists a constant $c \neq 0$ s.th.

$$\mathbb{E}_{s\sim P(\cdot|)}\left[v_i^{\boldsymbol{\pi}}(s) - v_i^{\boldsymbol{\pi}'}(s)\right] - \mathbb{E}_{s\sim P(\cdot|)}\left[B_i^{\boldsymbol{\pi}}(s) - B_i^{\boldsymbol{\pi}'}(s)\right] = c.$$

Let us now define the following quantities for any $s \in \mathcal{S}$ and for each $\pi_i \in \Pi_i$ and $\pi_{-i} \in \Pi_{-i}$ and $\forall \theta \in \Theta, \forall i \in \mathcal{N}$:

$$v_{i,T'}^{\boldsymbol{\pi}}(s) := \sum_{t=1}^{T'} \int_{\mathcal{S}} ds_{j+1}\mu(s_0)\pi_i(a_0^i, s_0)\pi_{-i}(a_0^{-i}, s_0)\prod_{j=1}^{t-1}\gamma^j P(s_{j+1}|s_j, a_j^i, a_j^{-i}) \cdot\pi_i(a_j^i|s_j)\pi_{-i}(a_j^{-i}|s_j)R_i(s_j, a_j^i, a_j^{-i}),$$

and

$$B_{T'}^{\boldsymbol{\pi}}(s) := \sum_{t=1}^{T'} \int_{\mathcal{S}} ds_{j+1}\mu(s_0)\pi_i(a_0^i, s_0)\pi_{-i}(a_0^{-i}, s_0)\prod_{j=1}^{t-1}\gamma^j P(s_{j+1}|s_j, a_j^i, a_j^{-i})\pi_i(a_j^i|s_j)\pi_{-i}(a_j^{-i}|s_j) \cdot\phi(s_j, a_j^i, a_j^{-i}),$$

so that the quantity $v_{i,T'}^{\boldsymbol{\pi}}(s)$ measures the expected cumulative return until the point $T' < \infty$.

Hence, we straightforwardly deduce that

$$v_i^{\boldsymbol{\pi}}(s) \equiv v_{i,\infty}^{\boldsymbol{\pi}}(s) = v_{i,T'}^{\boldsymbol{\pi}}(s) + \gamma^{T'}\int_{\mathcal{S}} ds_{j+1}\mu(s_0)\pi_i(a_0^i, s_0)\pi_{-i}(a_0^{-i}, s_0)\prod_{j=1}^{T'} P(s_{j+1}|s_j, a_j^i, a_j^{-i})\pi_i(a_j^i|s_j) \cdot\pi_{-i}(a_j^{-i}|s_j)v_i^{\boldsymbol{\pi}}(s_{T'}).$$

Our first task is to establish that the quantity $\left|\lim_{t\to\infty}\mathbb{E}_{s\sim P(\cdot|)}\left[B_{i,t}^{\boldsymbol{\pi}}(s) - B_{i,t}^{\boldsymbol{\pi}'}(s)\right]\right|$ is in fact, well-defined for any $s \in \mathcal{S}$ and $\forall i \in \mathcal{N}$.

This is true since by (25) for any $t > 0$ we have that

$$\left|\mathbb{E}_{s\sim P(\cdot|)}\left[B_{i,t}^{\boldsymbol{\pi}}(s) - B_{i,t}^{\boldsymbol{\pi}'}(s)\right]\right| = \left|\mathbb{E}_{s\sim P(\cdot|)}\left[v_{i,t}^{\boldsymbol{\pi}}(s) - v_{i,t}^{\boldsymbol{\pi}'}(s)\right]\right|, \tag{29}$$

and hence we have that

$$\left|\mathbb{E}_{s\sim P(\cdot|)}\left[B^{\boldsymbol{\pi}}(s) - B^{\boldsymbol{\pi}'}(s)\right]\right| < \infty.$$

To see this, we firstly observe that by the boundedness of R_i, $\exists c > 0$ s.th. $\forall t \in \mathbb{N}, \forall i \in \mathcal{N}$ and for any $\pi_i \in \Pi_i, \pi_{-i} \in \Pi_{-i}$

$$\left|v_{i,t}^{\boldsymbol{\pi}}(s) - v_{i,t}^{\boldsymbol{\pi}'}(s)\right| < c.$$

This is true since for any $k < \infty$ we have

$$v_{i,k}^{\boldsymbol{\pi}}(s) - v_{i,k}^{\boldsymbol{\pi}'}(s) = \mathbb{E}_{s_t\sim P,\pi_i,\pi_{-i}}\left[\sum_{t=0}^{k}\gamma^t R_i(s_t, \boldsymbol{a}_t)\right] - \mathbb{E}_{s_t\sim P,\pi_i',\pi_{-i}}\left[\sum_{t=0}^{k}\gamma^t R_i(s_t, \boldsymbol{a}_t)\right]$$

$$\leq \left| \mathbb{E}_{s_t \sim P, \pi_i, \pi_{-i}} \left[\sum_{t=0}^{k} \gamma^t R_i(s_t, \boldsymbol{a}_t) \right] - \mathbb{E}_{s_t \sim P, \pi_i', \pi_{-i}} \left[\sum_{t=0}^{k} \gamma^t R_i(s_t, \boldsymbol{a}_t) \right] \right|$$
$$\leq 2 \sum_{t=0}^{k} \gamma^t \|R_i\|_\infty = 2 \frac{1-\gamma^k}{1-\gamma} \|R_i\|_\infty.$$

Therefore, by the bounded convergence theorem we have that

$$\lim_{t \to \infty} \left| \mathbb{E}_{s \sim P(\cdot|)} \left[v_{i,t}^{\boldsymbol{\pi}}(s) - v_{i,t}^{\boldsymbol{\pi}'}(s) \right] \right| < \infty. \tag{30}$$

Now, using (29), we deduce that for any $\epsilon > 0$, the following statement holds:

$$\left| \mathbb{E}_{s \sim P(\cdot|)} \left[B_{i,t}^{\boldsymbol{\pi}}(s) - B_{i,t}^{\boldsymbol{\pi}'}(s) \right] \right| < \left| \mathbb{E}_{s \sim P(\cdot|)} \left[v_{i,t}^{\boldsymbol{\pi}}(s) - v_{i,t}^{\boldsymbol{\pi}'}(s) \right] \right| + \epsilon,$$

which after taking the limit as $t \to \infty$ and using (30), Lemma D and the dominated convergence theorem, we find that

$$\lim_{t \to \infty} \left| \mathbb{E}_{s \sim P(\cdot|)} \left[B_{i,t}^{\boldsymbol{\pi}}(s) - B_{i,t}^{\boldsymbol{\pi}'}(s) \right] \right| < \infty.$$

Next we observe that:

$$\begin{aligned}
c &= \mathbb{E}_{s \sim P(\cdot|)} \left[\left(v_i^{\boldsymbol{\pi}} - v_i^{\boldsymbol{\pi}'} \right)(s) \right] - \mathbb{E}_{s \sim P(\cdot|)} \left[\left(B^{\boldsymbol{\pi}} - B^{\boldsymbol{\pi}'} \right)(s) \right] \\
&= \mathbb{E}_{s \sim P(\cdot|)} \left[\left(v_{i,T'}^{\boldsymbol{\pi}} - v_{i,T'}^{\boldsymbol{\pi}'} \right)(s) \right] - \mathbb{E}_{s \sim P(\cdot|)} \left[\left(B_{T'}^{\boldsymbol{\pi}} - B_{T'}^{\boldsymbol{\pi}'} \right)(s) \right] \\
&+ \gamma^{T'} \mathbb{E}_{s_{T'} \sim P(\cdot|)} \Bigg[\int_{\mathcal{S}} ds_{j+1} \mu(s_0) \pi_i(a_0^i, s_0) \pi_{-i}(a_0^{-i}, s_0) \prod_{j=1}^{T'} P(s_{j+1}|s_j, a_j^i, a_j^{-i}) \pi_i(a_j^i|s_j) \pi_{-i}(a_j^{-i}|s_j) \\
&\qquad \cdot \left(v_i^{\boldsymbol{\pi}}(s_{T'}) - B^{\boldsymbol{\pi}}(s_{T'}) \right) \\
&+ \int_{\mathcal{S}} ds_{j+1} \mu(s_0) \pi_i'(a_0'^i, s_0) \pi_{-i}(a_0^{-i}, s_0) \prod_{j=1}^{T'} P(s_{j+1}|s_j, a_j'^i, a_j^{-i}) \pi_i(a_j'^i|s_j) \pi_{-i}(a_j^{-i}|s_j) \\
&\qquad \cdot \left(v_i^{\boldsymbol{\pi}'}(s_{T'}) - B^{\boldsymbol{\pi}'}(s_{T'}) \right) \Bigg].
\end{aligned}$$

Considering the last expectation and its coefficient and denoting it by κ, we observe the following bound:

$$|\kappa| \leq 2\gamma^{T'} \left(\|v_i\| + \|B\| \right).$$

Since we can choose T' freely and $\gamma \in]0, 1[$, we can choose T' to be sufficiently large so that

$$\gamma^{T'} \left(\|v_i\| + \|B\| \right) < \frac{1}{4}|c|.$$

This then implies that

$$\left| \mathbb{E}_{s \sim P(\cdot|)} \left[\left(v_{i,T'}^{\boldsymbol{\pi}} - v_{i,T'}^{\boldsymbol{\pi}'} \right)(s) - \left(B_{T'}^{\boldsymbol{\pi}} - B_{T'}^{\boldsymbol{\pi}'} \right)(s) \right] \right| > \frac{1}{2}c,$$

which is a contradiction since we have proven that for any finite T' it is the case that

$$\mathbb{E}_{s \sim P(\cdot|)} \left[\left(v_{i,T'}^{\boldsymbol{\pi}} - v_{i,T'}^{\boldsymbol{\pi}'} \right)(s) - \left(B_{T'}^{\boldsymbol{\pi}} - B_{T'}^{\boldsymbol{\pi}'} \right)(s) \right] = 0,$$

and hence we deduce the thesis. □

Proof of Lemma 1. The result is proven after a straightforward extension of the static case (Lemma 2.7. in (Monderer & Shapley, 1996b)). □

Proposition 5. *There exists a function $B : \mathcal{S} \times \mathbf{\Pi} \to \mathbb{R}$ such that $\forall s \in S$ we have that*

$$\boldsymbol{\pi} \in \arg\sup_{\boldsymbol{\pi'} \in \mathbf{\Pi}} B^{\boldsymbol{\pi'}}(s) \implies \boldsymbol{\pi} \in NE\{\mathcal{G}\}.$$

Proof of Prop. 5. We do the proof by contradiction. Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N) \in \arg\sup_{\boldsymbol{\pi'} \in \mathbf{\Pi}} v^{\boldsymbol{\pi'}}(s)$. Let us now therefore assume that $\boldsymbol{\pi} \notin NE\{\mathcal{G}\}$, hence there exists some other strategy profile $\boldsymbol{\pi'} = (\pi_1, \dots, \pi'_i, \dots, \pi_N)$ which contains at least one profitable deviation by one of the agents so that $\pi'_i \neq \pi_i$ for $i \in \mathcal{N}$ i.e. $v_i^{\boldsymbol{\pi'}}(s) > v_i^{\boldsymbol{\pi}}(s)$ (using the preservation of signs of integration). Prop. 4 however implies that $v_i^{\boldsymbol{\pi'}}(s) - v_i^{\boldsymbol{\pi}}(s) > 0$ which is a contradiction since $\boldsymbol{\pi}$ is a maximum of B. □

Proof of Theorem 1. Combining Prop. 5 with Prop. 4 proves Theorem 1. □

Proof of Prop. 1. Since the functions $(R_i)_{i \in \mathcal{N}}$ are differentiable in the action inputs, we first we note the following $R_i(s, a^i, a^{-i}) - R_i(s, a'^i, a^{-i}) = \int_{a^i}^{a'^i} \frac{\partial R_i(s,a,a^{-i})}{\partial a} da$ and $\phi(s, a^i, a^{-i}) - \phi(s, a'^i, a^{-i}) = \int_{a^i}^{a'^i} \frac{\partial \phi(s,a,a^{-i})}{\partial a} da$. We then deduce that $\frac{\partial R_i(s,a,a^{-i})}{\partial a} = \frac{\partial \phi(s,a,a^{-i})}{\partial a}$. Considering actions sampled from stochastic policies, we find that

$$\begin{aligned}
&\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[R_i(s, a^i, a^{-i}) \right] - \mathbb{E}_{\pi_i(\boldsymbol{\eta}'_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[R_i(s, a'^i, a^{-i}) \right] \\
&= \mathbb{E}_{\pi_i(\boldsymbol{\eta}_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[\phi(s, a^i, a^{-i}) \right] - \mathbb{E}_{\pi'_i(\boldsymbol{\eta}'_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[\phi(s, a'^i, a^{-i}) \right].
\end{aligned}$$

Now suppose $a = h(s, \boldsymbol{\eta}^i)$ then

$$\int_{a^i}^{a'^i} \frac{\partial R_i(s, a, a^{-i})}{\partial a} da = \int_{\boldsymbol{\eta}_i = h^{-1}(a^i)}^{\boldsymbol{\eta}'_i = h^{-1}(a'^i)} \frac{\partial R_i(s, h(s, \boldsymbol{\eta}^i), a^{-i})}{\partial a} \frac{dh}{d\boldsymbol{\eta}} d\boldsymbol{\eta} \tag{31}$$

Similarly we find that for $a^i \sim \pi_{i,\epsilon}(\cdot, |s, \eta)$ we have that

$$\begin{aligned}
&\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[R_i(s, a^i, a^{-i}) \right] - \mathbb{E}_{\pi_i(\boldsymbol{\eta}'_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[R_i(s, a'^i, a^{-i}) \right] \\
&= \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, \boldsymbol{\eta}_{-i}) \int_{\boldsymbol{\eta}_i(a^i)}^{\boldsymbol{\eta}'_i(a'^i)} \int_{\mathcal{A}_i} \frac{\partial}{\partial \boldsymbol{\eta}_i} \pi_i(da^i, s; \boldsymbol{\eta}_i) \frac{\partial R_i \left(s, \pi_{i,\epsilon}(\cdot, |s, \eta), a^{-i} \right)}{\partial a^i} d\boldsymbol{\eta}_i \qquad (32) \\
&= \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, \boldsymbol{\eta}_{-i}) \int_{\boldsymbol{\eta}_i(a^i)}^{\boldsymbol{\eta}'_i(a'^i)} \int_{\mathcal{A}_i} \frac{\partial}{\partial \boldsymbol{\eta}_i} \pi_i(da^i, s; \boldsymbol{\eta}_i) \frac{\partial R_i(s, a^i, a^{-i})}{\partial a^i} d\boldsymbol{\eta}_i \\
&= \int_{\boldsymbol{\eta}_i(a^i)}^{\boldsymbol{\eta}'_i(a'^i)} \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, \boldsymbol{\eta}_{-i}) \int_{\mathcal{A}_i} \frac{\partial}{\partial \boldsymbol{\eta}_i} \pi_i(da^i, s; \boldsymbol{\eta}_i) \frac{\partial R_i(s, a^i, a^{-i})}{\partial a^i} d\boldsymbol{\eta}_i \qquad (33)
\end{aligned}$$

By the same reasoning as above we find that

$$\begin{aligned}
&\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[\phi(s, a^i, a^{-i}) \right] - \mathbb{E}_{\pi_i(\boldsymbol{\eta}'_i), \pi_{-i}(\boldsymbol{\eta}_{-i})} \left[\phi(s, a'^i, a^{-i}) \right] \\
&= \int_{\boldsymbol{\eta}_i(a^i)}^{\boldsymbol{\eta}'_i(a'^i)} \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, \boldsymbol{\eta}_{-i}) \int_{\mathcal{A}_i} \frac{\partial}{\partial \boldsymbol{\eta}_i} \pi_i(da^i, s; \boldsymbol{\eta}_i) \frac{\partial \phi(s, a^i, a^{-i})}{\partial a^i} d\boldsymbol{\eta}_i
\end{aligned} \tag{34}$$

Putting (33) and (34) together, we deduce that

$$\begin{aligned}
&\int_{\boldsymbol{\eta}_i(a^i)}^{\boldsymbol{\eta}'_i(a'^i)} \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, \boldsymbol{\eta}_{-i}) \int_{\mathcal{A}_i} \frac{\partial}{\partial \boldsymbol{\eta}_i} \pi_i(da^i, s; \boldsymbol{\eta}_i) \frac{\partial R_i(s, a^i, a^{-i})}{\partial a^i} d\boldsymbol{\eta}_i \\
&= \int_{\boldsymbol{\eta}_i(a^i)}^{\boldsymbol{\eta}'_i(a'^i)} \int_{\mathcal{A}_{-i}} \pi_{-i}(da^{-i}, \boldsymbol{\eta}_{-i}) \int_{\mathcal{A}_i} \frac{\partial}{\partial \boldsymbol{\eta}_i} \pi_i(da^i, s; \boldsymbol{\eta}_i) \frac{\partial \phi(s, a^i, a^{-i})}{\partial a^i} d\boldsymbol{\eta}_i
\end{aligned}$$

Which implies that

$$\int_{\mathcal{A}_{-i}}\int_{\mathcal{A}_i}\pi_{-i}(da^{-i},\boldsymbol{\eta}_{-i})\frac{\partial}{\partial\boldsymbol{\eta}_i}\pi_i(da^i,s;\boldsymbol{\eta}_i)\frac{\partial R_i(s,a^i,a^{-i})}{\partial a^i}$$
$$=\int_{\mathcal{A}_{-i}}\int_{\mathcal{A}_i}\pi_{-i}(da^{-i},\boldsymbol{\eta}_{-i})\frac{\partial}{\partial\boldsymbol{\eta}_i}\pi_i(da^i,s;\boldsymbol{\eta}_i)\frac{\partial \phi(s,a^i,a^{-i})}{\partial a^i}$$

Moreover we also find that

$$\int_{\mathcal{A}_{-i}}\pi_{-i}(da^{-i},\boldsymbol{\eta}_{-i})\int_{\mathcal{A}_i}\frac{\partial}{\partial\boldsymbol{\eta}_i}\pi_i(da^i,s;\boldsymbol{\eta}_i)\frac{\partial R_i(s,a^i,a^{-i})}{\partial a^i}d\boldsymbol{\eta}_i$$
$$=\int_{\mathcal{A}_i}\int_{\mathcal{A}_{-i}}\pi_{-i}(da^{-i}|s;\boldsymbol{\eta}^{-i})\pi_i(da^i|s;\boldsymbol{\eta}^i)\frac{\partial}{\partial\boldsymbol{\eta}^i}\ln\left[\pi_i(da^i|s;\boldsymbol{\eta}^i)\right]\frac{\partial R_i(s,a^i,a^{-i})}{\partial a^i}$$
$$=\mathbb{E}_{\boldsymbol{\pi}(\boldsymbol{\eta})}\left[\frac{\partial}{\partial\boldsymbol{\eta}^i}\ln\left[\pi_i(a^i|s;\boldsymbol{\eta}^i)\right]\frac{\partial R_i(s,a^i,a^{-i})}{\partial a^i}\right].$$

Hence, using the linearity of the expectation and the derivative we arrive at

$$\mathbb{E}_{(a_i,a_{-i})\sim(\pi_i,\pi_{-i})}\left[\frac{\partial}{\partial\boldsymbol{\eta}^i}\ln\left[\pi_i(a^i|s;\boldsymbol{\eta}^i)\right]\left(\frac{\partial R_i(s,a^i,a^{-i})}{\partial a^i}-\frac{\partial\phi(s,a^i,a^{-i})}{\partial a^i}\right)\right]=\mathbf{0} \tag{35}$$

In a similar way we observe that for any c-SPG in which the state transitive assumption holds, we have that $R_i(s',a^i,a^{-i})-R_i(s,a^i,a^{-i})=\int_{s'}^{s}\frac{\partial R_i(s,a,a^{-i})}{\partial s}ds$ and $\phi(s',a^i,a^{-i})-\phi(s',a^i,a^{-i})=\int_{s'}^{s}\frac{\partial\phi(s,a,a^{-i})}{\partial s}ds$. We then find that $\frac{\partial R_i(s,a,a^{-i})}{\partial s}=\frac{\partial\phi(s,a,a^{-i})}{\partial s}$. By identical reasoning as above we deduce that

$$\mathbb{E}_{(a_i,a_{-i})\sim(\pi_i,\pi_{-i})}\left[\frac{\partial}{\partial\boldsymbol{\eta}^i}\ln\left[\pi_i(a^i|s;\boldsymbol{\eta}^i)\right]\left(\frac{\partial R_i(s,a^i,a^{-i})}{\partial s}-\frac{\partial\phi(s,a^i,a^{-i})}{\partial s}\right)\right]=\mathbf{0}. \tag{36}$$

Putting the two statements together leads to the expression:

$$\mathbb{E}_{(a_i,a_{-i})\sim(\pi_i,\pi_{-i})}\left[\frac{\partial}{\partial\boldsymbol{\eta}^i}\ln\left[\pi_i(a^i|s;\boldsymbol{\eta}^i)\right]D[R_i,\phi](s,a^i,a^{-i})\right]=\mathbf{0} \tag{37}$$

which concludes the proof. □

Lemma E. *For any c-SPG the following expression holds* $\forall(\boldsymbol{\eta}^i,\boldsymbol{\eta}^{-i})\in\boldsymbol{E}^{ps},\forall s\in\mathcal{S}$

$$\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i),\pi_{-i}(\boldsymbol{\eta}_{-i})}\left[R_i\right]-\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i'),\pi_{-i}(\boldsymbol{\eta}_{-i})}\left[R_i\right]=\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i),\pi_{-i}(\boldsymbol{\eta}_{-i})}\left[\phi\right]-\mathbb{E}_{\pi_i'(\boldsymbol{\eta}_i'),\pi_{-i}(\boldsymbol{\eta}_{-i})}\left[\phi\right]. \tag{38}$$

Proof of Lemma E. The forward implication is straightforward.

Indeed, assume that (1) holds, that is:

$$R_i(s_t,(a_t^i,a_t^{-i}))-R_i(s_t,(a_t'^i,a_t^{-i}))=\phi(s_t,(a_t^i,a_t^{-i}))-\phi(s_t,(a_t'^i,a_t^{-i})) \tag{39}$$

Define by $\Delta F[s_t](a_t^i,a_t'^i),a_t^{-i}):=F(s_t,(a_t^i,a_t^{-i}))-F(s_t,(a_t'^i,a_t^{-i}))$ for any $a_t'^i\in\mathcal{A}_i$ then for any $F:\mathcal{S}\times\mathcal{A}_i\times\mathcal{A}_{-i}\to\mathbb{R}$ ($F\in\mathscr{H}$) we have that

$$\int_{a_i\in\mathcal{A}_i}\int_{a_i'\in\mathcal{A}_i}\int_{a_{-i}\in\mathcal{A}_{-i}}\pi_i(da^i|\cdot)\pi_i(da'^i|\cdot)\pi_{-i}(da^{-i}|\cdot)\Delta F[s_t](a_t^i,a_t'^i),a_t^{-i})$$
$$=\int_{a_i\in\mathcal{A}_i}\int_{a_{-i}\in\mathcal{A}_{-i}}\pi_i(da^i|\cdot)\pi_{-i}(da^{-i}|\cdot)F(a_t^i,a_t'^i),a_t^{-i})-\int_{a_i'\in\mathcal{A}_i}\int_{a_{-i}\in\mathcal{A}_{-i}}\pi_i(da'^i|\cdot)\pi_{-i}(da^{-i}|\cdot)F(s_t,(a_t'^i,a_t^{-i}))$$
$$=\mathbb{E}_{(\pi_i,\pi_{-i})}\left[F(s_t,(a_t^i,a_t^{-i}))\right]-\mathbb{E}_{(\pi_i',\pi_{-i})}\left[F(s_t,(a_t'^i,a_t^{-i}))\right]$$

This immediately suggests that we can get the result by multiplying (39) by
$\int_{a_i \in \mathcal{A}_i} \int_{a_i' \in \mathcal{A}_i} \int_{a_{-i} \in \mathcal{A}_{-i}} \pi_i(da^i|\cdot)\pi_i(da'^i|\cdot)\pi_{-i}(da^{-i}|\cdot)$.

For the reverse (in pure strategies) we first consider the case in which the pure strategy is a linear map from some parameterisation. We now readily verify that the reverse holds indeed:

$$\begin{aligned}
&\mathbb{E}_{\pi_i(\boldsymbol{\eta}_i),\pi_{-i}(\boldsymbol{\eta}_{-i})}\left[F(s,a^i,a^{-i})\right] - \mathbb{E}_{\pi_i(\boldsymbol{\eta}_i'),\pi_{-i}(\boldsymbol{\eta}_{-i})}\left[F(s,a^i,a^{-i})\right] \\
&= \int_{a' \in \mathcal{A}^i} \int_{a'' \in \mathcal{A}^{-i}} \pi_i(da'|s;\boldsymbol{\eta}_i)\pi_{-i}(da''|s;\boldsymbol{\eta}_{-i})F_i(s,(a',a'')) \\
&\quad - \int_{a' \in \mathcal{A}^i} \int_{a'' \in \mathcal{A}^{-i}} \pi_i(da'|s;\boldsymbol{\eta}_i')\pi_{-i}(da''|s;\boldsymbol{\eta}_{-i})F_i(s,(a',a'')) \\
&= \int_{a' \in \mathcal{A}^i} \int_{a'' \in \mathcal{A}^{-i}} da'da''\delta_i(a' - a^i(s;\boldsymbol{\eta}_i))\delta_{-i}(a'' - a^{-i}(s;\boldsymbol{\eta}_{-i}))F_i(s,(a',a'')) \\
&\quad - \int_{a' \in \mathcal{A}^i} \int_{a'' \in \mathcal{A}^{-i}} da'da''\delta_i(a' - a^i(s;\boldsymbol{\eta}_i'))\delta_{-i}(a'' - a^{-i}(s;\boldsymbol{\eta}_{-i}))F_i(s,(a',a'')) \\
&= \Delta F_i(s,(a^i(s;\boldsymbol{\eta}_i),a^{-i}(s;\boldsymbol{\eta}_{-i})))
\end{aligned}$$

which proves the statement in the linear case.

For the general case, consider a strategy which is defined by a map $h : \mathcal{S} \times E \to \mathcal{A}$. We note that for any $\beta \in \mathbb{R}/\{0\}$ and $x \in X \subset \mathbb{R}$ we have that

$$\int_{\mathbb{R}} \delta(\beta x) = \frac{1}{|\beta|} \int_{\mathbb{R}} \delta(\beta x)dx \tag{40}$$

This is true since for any $\beta \in \mathbb{R}/\{0\}$ we can construct the delta function in the following way:

$$\delta(\beta x) = \lim_{m \to \infty} \frac{1}{|m|\sqrt{\pi}} e^{-(\beta x/m)^2}$$

Now define $n := m\beta^{-1}$, then

$$\begin{aligned}
&\lim_{m \to \infty} \frac{1}{|m|\sqrt{\pi}} e^{-(\beta x/m)^2} \\
&= \lim_{n \to \infty} \frac{1}{|\beta n|\sqrt{\pi}} e^{-(x/n)^2} \\
&= \frac{1}{|\beta|} \lim_{n \to \infty} \frac{1}{|n|\sqrt{\pi}} e^{-(x/n)^2} = \frac{1}{|\beta|}\delta(x)
\end{aligned}$$

In the following, we use the coarea formula (for geometric measures) (Simon et al., 1983; Nicolaescu, 2011) which says that for any open set $X \subset \mathbb{R}^n$ and for any Lipschitz function $f : X \to \mathbb{R}$ on X and for any L^1 function g the following expression holds:

$$\int_X g(x)|\nabla f(x)|dx = \int_{\mathbb{R}} \left(\int_{f^{-1}(s)} g(x)dH_{n-1}(x) \right) ds$$

where H_{n-1} is the $(n-1)-$dimensional Hausdorff measure.

Let us now define $k(s,\boldsymbol{\eta},a) := h^{-1}(s,\boldsymbol{\eta}) - a$.

Now

$$\int_{a \in \mathcal{A}} \pi_{\epsilon=0}(a;s,\boldsymbol{\eta})F_i(s,\boldsymbol{a})da = \int_{a \in \mathcal{A}} \delta(k(s,a,h(\boldsymbol{\eta})))F_i(s,\boldsymbol{a})da$$

By Taylor's theorem, expanding about the point y where y is defined by $k(s,\boldsymbol{\eta},y) = 0$ implies that

$$\int_{a \in \mathcal{A}} \delta(k(s,a,\boldsymbol{\eta}))F_i(s,\boldsymbol{a})da \approx \int_{a \in \mathcal{A}} \delta\left(k'(s,y,\boldsymbol{\eta})(a-y)\right) F_i(s,\boldsymbol{a})da$$

Moreover

$$\begin{aligned}&\int_{a\in\mathcal{A}}\delta(k'(s,y,\boldsymbol{\eta}))(a-y)F_i(s,\boldsymbol{a})da\\&=\int_{y\in k^{-1}(0)}\left(\int_{y-\epsilon}^{y+\epsilon}\delta\left(k'(s,y,\boldsymbol{\eta})(a-y)\right)F_i(s,\boldsymbol{a})da\right)d\sigma(y)\end{aligned}$$

where σ is a Minkowski content measure.

Define $x:=a-y$ so $dx:=da$

$$\begin{aligned}&\int_{y\in k^{-1}(0)}\left(\int_{y-\epsilon}^{y+\epsilon}\delta\left(k'(s,y,g(\boldsymbol{\eta}))(a-y)\right)F_i(s,\boldsymbol{a})da\right)d\sigma(y)\\&=\int_{y\in k^{-1}(0)}\left(\int_{x-\epsilon}^{x+\epsilon}\delta\left(k'(s,y,\boldsymbol{\eta})x\right)F_i(s,x+y)dx\right)d\sigma(y)\\&=\int_{y\in k^{-1}(0)}\left(\int_{x-\epsilon}^{x+\epsilon}\delta(x)\frac{F(s,x+y)}{|k'(s,y,\boldsymbol{\eta})|}dx\right)d\sigma(y)\\&=\int_{y\in k^{-1}(0)}\frac{F_i(s,y)}{|k'(s,y,\boldsymbol{\eta})|}d\sigma(y)\\&=\int_{y\in k^{-1}(0)}F_i(s,y)d\sigma(y)\\&=F_i(s,a)\end{aligned}$$

where we have used (40) in the second step.

Hence we complete the proof by noting that

$$\begin{aligned}&\int_{a\in\mathcal{A}}\pi_{\epsilon=0}(a;s,\boldsymbol{\eta}')F_i(s,\boldsymbol{a})da-\int_{a\in\mathcal{A}}\pi_{\epsilon=0}(a;s,\boldsymbol{\eta}'')F_i(s,\boldsymbol{a})da\\&=F_i(s,a(\eta))-F_i(s,a(\eta'))\end{aligned}$$

as required. □

Proof of Lemma 2. Recall $\Delta F(s_t,(a_t^i,a_t'^i),a_t^{-i}):=F(s_t,(a_t'^i,a_t^{-i}))-F(s_t,(a_t^i,a_t^{-i}))$ define also by $\Delta F(s_t,(\pi_i,\pi_i'),\pi_{-i})):=\mathbb{E}_{(\pi_i,\pi_{-i})}\left[F(s_t,(a_t^i,a_t^{-i}))\right]-\mathbb{E}_{(\pi_i',\pi_{-i})}\left[F(s_t,(a_t'^i,a_t^{-i}))\right]$. We wish to bridge the two cases by proving the following:

$$\left|\Delta F(s_t,(a_t^i,a_t'^i),a_t^{-i})-\Delta F(s_t,(\pi_{i,\epsilon},\pi_{i,\epsilon}'),a^{-i})\right|\leq c\bar{\sigma}_\epsilon^2. \tag{41}$$

where $\bar{\sigma}_\epsilon=\max\{\sigma_\epsilon,\sigma_\epsilon'\}$ and $\sigma_\epsilon,\sigma_\epsilon'$ are the variances of the policies π_ϵ and π_ϵ' respectively. Indeed,

$$\begin{aligned}&\left|\mathbb{E}\left[F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right]\right|\\&\leq\mathbb{E}\left[\left|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right|\right]\\&=\mathbb{E}\Big[\left(1_{|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})|>\gamma}+1_{|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})|\leq\gamma}\right)\\&\quad\cdot\left(F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right)\Big]\end{aligned}$$

Now since F is bounded and continuous, we deduce that

$$\begin{aligned}&\mathbb{E}\left[\left(1_{|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})|>\gamma}\right)\left(F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right)\right]\\&\leq\|F\|_\infty\mathbb{E}\left[\left(1_{|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})|>\gamma}\right)\right]\\&=\|F\|_\infty\mathbb{P}\left(\left|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right|>\gamma\right)\end{aligned}$$

using the properties of the indicator function. Now by the continuity and boundedness of F we deduce that there exists $\delta > 0$ such that $a - b > \delta$ whenever $F(a) - F(b) > \gamma$ applying this result we then find that

$$\begin{aligned}
&\mathbb{E}\Big[\left(1_{|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})|>\gamma} + 1_{|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i})-F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})|\leq\gamma}\right) \\
&\cdot \left(F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i}) - F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right)\Big] \\
&\leq \|F\|_\infty \mathbb{P}\left(\left|F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i}) - F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right| > \gamma\right) + \gamma \\
&\leq \|F\|_\infty \mathbb{P}\left(|\pi_{i,\epsilon=0}(\cdot,|s,\eta) - \pi_{i,\epsilon}(\cdot,|s,\eta)| > \delta\right) + \gamma \\
&\leq \delta^{-2}\|F\|_\infty \sigma_\epsilon^2 + \gamma
\end{aligned}$$

where we have used Tschebeyshev's inequality in the last line.

Now since γ is arbitrary we deduce that

$$\left|\mathbb{E}\left[F(s,\pi_{i,\epsilon=0}(\cdot,|s,\eta),a^{-i}) - F(s,\pi_{i,\epsilon}(\cdot,|s,\eta),a^{-i})\right]\right| \leq \delta^{-2}\|F\|_\infty \sigma_\epsilon^2 \tag{42}$$

Moreover

$$\begin{aligned}
&\left|\Delta F(s_t,(a_t^i,a_t'^i),a_t^{-i}) - \Delta F(s_t,(\pi_i,\pi_i'),\pi_{-i}))\right| \\
&= \left|F(s_t,(a_t'^i,a_t^{-i})) - F(s_t,(a_t^i,a_t^{-i})) - \left(\mathbb{E}_{(\pi_i,\pi_{-i})}\left[F(s_t,(a_t^i,a_t^{-i}))\right] - \mathbb{E}_{(\pi_i',\pi_{-i})}\left[F(s_t,(a_t'^i,a_t^{-i}))\right]\right)\right| \\
&\leq \left|F(s_t,(a_t^i,a_t^{-i})) - \mathbb{E}_{(\pi_i,\pi_{-i})}\left[F(s_t,(a_t^i,a_t^{-i}))\right]\right| + \left|F(s_t,(a_t'^i,a_t^{-i})) - \mathbb{E}_{(\pi_i',\pi_{-i})}\left[F(s_t,(a_t'^i,a_t^{-i}))\right]\right| \\
&\leq c\|F\|_\infty \bar{\sigma}_\epsilon^2
\end{aligned}$$

We deduce the last statement by applying the result to the sequence of ϵ/n and by the sandwich theorem. □

Lemma F. *The function B is given by the following expression for $s \in \mathcal{S}, \forall \boldsymbol{\pi} \in \boldsymbol{\Pi}$:*

$$B^{\boldsymbol{\pi}}(s) - B^{\boldsymbol{\pi'}}(s) = \mathbb{E}_{s_t\sim P}\left[\sum_{t=0}^{\infty}\sum_{i\in\mathcal{N}}\gamma^t \int_0^1 \boldsymbol{\gamma}'(z)\frac{\partial R_i}{\partial \pi_i}(s_t,\boldsymbol{\gamma}(z))\Big| s = s_0\right],$$

where $\boldsymbol{\gamma}(z)$ is a continuous differentiable path in $\boldsymbol{\Pi}$ connecting two strategy profiles $\boldsymbol{\pi} \in \boldsymbol{\Pi}$ and $\boldsymbol{\pi'} \in \boldsymbol{\Pi}$.

Proof. We note that from (38), using the gradient theorem of vector calculus, it is straightforward to deduce that the potential function ϕ can be computed from the reward functions $(R_i)_{i\in\mathcal{N}}$ via the following expression (Monderer & Shapley, 1996b):

$$\phi^{\boldsymbol{\pi}}(s) = \phi^{\boldsymbol{\pi'}}(s) + \sum_{i\in\mathcal{N}}\int_0^1 \boldsymbol{\gamma}'(z)\frac{\partial R_i}{\partial \pi_i}(s_t,\boldsymbol{\gamma}(z)), \tag{43}$$

where $\boldsymbol{\gamma}(z)$ is a continuous differentiable path in $\boldsymbol{\Pi}$ connecting two strategy profiles $\boldsymbol{\pi} \in \boldsymbol{\Pi}$ and $\boldsymbol{\pi'} \in \boldsymbol{\Pi}$.

We then deduce the result (in the finite case) after inserting (43) into (24). □

Proof of Prop. 2. Recall the following definitions:

$$F_i(s,\boldsymbol{\eta},\rho) := \int_{\mathcal{A}} \boldsymbol{\pi}_\epsilon(\boldsymbol{da},\boldsymbol{\eta},s)\frac{\partial}{\partial \eta_i}\pi_{i,\epsilon}(a^i,\eta_i,s)\nabla F_\rho(s,\boldsymbol{a}) \tag{44}$$

and

$$U(s,\boldsymbol{\eta},\rho) := \int_{\mathcal{A}} \boldsymbol{\pi}_\epsilon(\boldsymbol{da},\boldsymbol{\eta},s)\frac{\partial}{\partial \eta_i}\pi_{i,\epsilon}(a^i,\eta_i,s)\nabla R_i(s,\boldsymbol{a}), \tag{45}$$

where we have used the shorthand: $\boldsymbol{\pi}_\epsilon(\boldsymbol{a},\boldsymbol{\eta},s) := \pi_i(a,s,\eta_i)\pi_{-i}(a_i,s,\eta_{-i})$

Since U and F_i are locally Lipschitz continuous each can have at most polynomial growth. By the Hölder inequality we find that:

$$
\begin{aligned}
&\int_\Omega \sum_{i\in\mathcal{N}} |F_i(s,\boldsymbol{\eta},\rho) - u(s,\boldsymbol{\eta})|^2 \, d\nu_1(s,\boldsymbol{\eta}) \\
&= \int_\Omega \sum_{i\in\mathcal{N}} \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla F_\rho(s,\boldsymbol{a}) - \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla R_i(s,\boldsymbol{a}) \right|^2 d\nu_1(s,\boldsymbol{\eta}) \\
&\leq \int_\Omega \sum_{i\in\mathcal{N}} \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \left(\frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla F_\rho(s,\boldsymbol{a}) - \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla R_i(s,\boldsymbol{a}) \right) \right|^2 d\nu_1(s,\boldsymbol{\eta}) \\
&\leq \int_\Omega \left(\sum_{i\in\mathcal{N}} \max_{\boldsymbol{a}\in\mathbb{A}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla F_\rho(s,\boldsymbol{a}) \right|^{q_i} + \max_{\boldsymbol{a}\in\mathbb{A}} N \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla R_i(s,\boldsymbol{a}) \right|^r \right) \qquad (46) \\
&\cdot \left(\sum_{i\in\mathcal{N}} \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \left(\frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla F_\rho(s,\boldsymbol{a}) - \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \nabla R_i(s,\boldsymbol{a}) \right) \right|^2 \right) d\nu_1(s,\boldsymbol{\eta}) \\
&\leq N \int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^{q_i \wedge r} \max_{\boldsymbol{a}\in\mathbb{A}} \left(|\nabla F_\rho(s,\boldsymbol{a})|^{q_i} + |\nabla R_i(s,\boldsymbol{a})|^r \right) \right) \\
&\cdot \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^2 \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \left(\nabla F_\rho(s,\boldsymbol{a}) - \nabla R_i(s,\boldsymbol{a}) \right) \right|^2 \right) d\nu_1(s,\boldsymbol{\eta}) \\
&\leq N \left(\int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^{q_i \wedge r} \max_{\boldsymbol{a}\in\mathbb{A}} \left(|\nabla F_\rho(s,\boldsymbol{a})|^{q_i} + |\nabla R_i(s,\boldsymbol{a})|^r \right) \right)^{r_1} d\nu_1(s,\boldsymbol{\eta}) \right)^{1/r_1} \\
&\quad \cdot \left(\int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^2 \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \left(\nabla F_\rho(s,\boldsymbol{a}) - \nabla R_i(s,\boldsymbol{a}) \right) \right|^2 \right)^{r_2} d\nu_1(s,\boldsymbol{\eta}) \right)^{1/r_2} \\
&\leq cN \left(\int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^{q_i \wedge r} \max_{\boldsymbol{a}\in\mathbb{A}} \left(|\nabla F_\rho(s,\boldsymbol{a}) - \nabla R_i(s,\boldsymbol{a})|^{q_i} + |\nabla R_i(s,\boldsymbol{a})|^{r\wedge q_i} \right) \right)^{r_1} d\nu_1(s,\boldsymbol{\eta}) \right)^{1/r_1} \\
&\quad \cdot \left(\int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^2 \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \left(\nabla F_\rho(s,\boldsymbol{a}) - \nabla R_i(s,\boldsymbol{a}) \right) \right|^2 \right)^{r_2} d\nu_1(s,\boldsymbol{\eta}) \right)^{1/r_2} \\
&\leq cN \left(\int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^{q_i \wedge r} \max_{\boldsymbol{a}\in\mathbb{A}} \left(|\nabla \left[F_\rho(s,\boldsymbol{a}) - R_i(s,\boldsymbol{a}) \right]|^{q_i} + |\nabla R_i(s,\boldsymbol{a})|^{r\wedge q_i} \right) \right)^{r_1} d\nu_1(s,\boldsymbol{\eta}) \right)^{1/r_1} \\
&\quad \cdot \left(\int_\Omega \left(\sum_{i\in\mathcal{N}} \left| \frac{\partial}{\partial \eta_i} \pi_{i,\epsilon}(a^i,\eta_i,s) \right|^2 \left| \int_{\mathbb{A}} \boldsymbol{\pi}_\epsilon(d\boldsymbol{a},\boldsymbol{\eta},s) \left(\nabla \left[F_\rho(s,\boldsymbol{a}) - R_i(s,\boldsymbol{a}) \right] \right) \right|^2 \right)^{r_2} d\nu_1(s,\boldsymbol{\eta}) \right)^{1/r_2} \qquad (47)
\end{aligned}
$$

Now for any $l \in \mathbb{R}_{>0}$ we have that

$$
\begin{aligned}
&|\nabla \left[F_\rho(s,\boldsymbol{a}) - R_i(s,\boldsymbol{a}) \right]|^l \\
&\leq |\nabla \left[F_\rho(s,\boldsymbol{a}) - \phi(s,\boldsymbol{a}) + \phi(s,\boldsymbol{a}) - R_i(s,\boldsymbol{a}) \right]|^l \\
&\leq |\nabla \left[F_\rho(s,\boldsymbol{a}) - \phi(s,\boldsymbol{a}) \right]|^l + |\nabla \left[\phi(s,\boldsymbol{a}) - R_i(s,\boldsymbol{a}) \right]|^l \\
&\leq |\nabla \left[F_\rho(s,\boldsymbol{a}) - \phi(s,\boldsymbol{a}) \right]|^l \qquad (48)
\end{aligned}
$$

using the potentiality property.

Inserting (48) into (47) yields

$$
\begin{aligned}
&\int_\Omega \sum_{i\in\mathcal{N}} \left|F_i(s,\boldsymbol{a},\rho)-u(s,\boldsymbol{a})\right|^2 d\nu_1(s,\boldsymbol{\eta}) \\
&\le cN\left(\int_\Omega\left(\sum_{i\in\mathcal{N}}\left|\frac{\partial}{\partial\eta_i}\pi_{i,\epsilon}(a^i,\eta_i,s)\right|^{q_i\wedge r}\max_{\boldsymbol{a}\in\mathcal{A}}\left(\left|\nabla\left[F_\rho(s,\boldsymbol{a})-\phi(s,\boldsymbol{a})\right]\right|^{q_i}+\left|\nabla R_i(s,\boldsymbol{a})\right|^{r\wedge q_i}\right)\right)^{r_1} d\nu_1(s,\boldsymbol{\eta})\right)^{1/r_1} \\
&\quad\cdot\left(\int_\Omega\left(\sum_{i\in\mathcal{N}}\left|\frac{\partial}{\partial\eta_i}\pi_{i,\epsilon}(a^i,\eta_i,s)\right|^2\left|\int_{\mathcal{A}}\boldsymbol{\pi}_\epsilon(\boldsymbol{da},\boldsymbol{\eta},s)\left(\nabla\left[F_\rho(s,\boldsymbol{a})-\phi(s,\boldsymbol{a})\right]\right)\right|^2\right)^{r_2} d\nu_1(s,\boldsymbol{\eta})\right)^{1/r_2} \\
&\le cN^2\left(\max_{i\in\mathcal{N}}\epsilon^{q_i}+\max_{i\in\mathcal{N}}\|\bar{R}_\nabla\|_\infty^{m\wedge q_i}\right)\epsilon^2
\end{aligned}
$$

where the last line follows from the boundedness of $\nabla R_i \le \bar{R}_\nabla$ and $\frac{\partial\pi_{i,\epsilon}}{\partial\eta_i}$ and Lemma 6 in (Bertsekas & Tsitsiklis, 2000). □

Proof of Prop. 3. We first show that the Bellman operator is a contraction. Indeed, for any bounded $F, F' \in \mathscr{H}$ we have

$$\|[T_\phi F]-[T_\phi F']\| \le \gamma d$$

where $d := \|F-F'\|_\infty$ using the fact that T_ϕ is monotonic.

We now observe that

$$
\begin{aligned}
\left\|\left[T_\phi^k F\right]-\left[T_{\phi_\epsilon}^k F'\right]\right\| &\le \|\phi-\phi_\epsilon\|_\infty+\gamma\left\|T_\phi^{k-1}F-T_{\phi_\epsilon}^{k-1}F'\right\|_\infty \\
&\le \epsilon+\sum_{j=0}^{m-1}\gamma^j\|\phi-\phi_\epsilon\|_\infty+\gamma^m\left\|T_\phi^{k-m}F-T_{\phi_\epsilon}^{k-m}F'\right\|_\infty \\
&\le \epsilon+\sum_{j=0}^{k-1}\gamma^j\|\phi-\phi_\epsilon\|_\infty+\gamma^k d \\
&\le \epsilon\left(1+\frac{1-\gamma^k}{1-\gamma}\right)+\gamma^k d
\end{aligned}
$$

and hence $\lim_{k\to\infty}\left\|\left[T_\phi^k F\right]-\left[T_{\phi_\epsilon}^k F'\right]\right\| \le c\epsilon$ for some $c>0$ from which we deduce the result. □