Some further observations on good strategies with n-bit memory

May 23, 2023

Reminder of the notation. We assume the players' decisions only depend on the outcome of the previous n rounds. An *n-history for player p* is a string $h^p = (a^p_{-1}, \ldots, a^p_{-n}) \in \{C, D\}^n$. An entry a^p_{-k} corresponds to player p's action k rounds ago. Let H^p denote the space of all n-histories of player p. Analogously, we define H^q as the set of n-histories h^q of player q. A pair $h=(h^p, h^q)$ is called an *n-history of the game*. The space of all such histories is $H = H^p \times H^q$. A *memory-n* strategy is a vector $\mathbf{p} = (p_h)_{h\in H}$. One special case of such a memory-n strategy is the *round-k-repeat strategy* for some $1 \leq k \leq n$. Player p uses a *round-k-repeat strategy* $\mathbf{p}^{k-\text{Rep}}$ if in any given round, the player chooses the same action as k rounds ago. That is, if the game's n-history is such that $a^p_{-k}=C$, then $p^{k-\text{Rep}}_h=1$; otherwise $p^{k-\text{Rep}}_h=0$.

If the two players use memory-n strategies $\mathbf{p}$ and $\mathbf{q}$, one can represent the interaction as a Markov chain with transition matrix M. Let $\mathbf{v} = (v_h)_{h\in H}$ be an invariant distribution of this Markov chain. Akin's lemma says that for each k with $1 \leq k \leq n$, the invariant distribution $\mathbf{v}$ satisfies the following relationship,

$$\mathbf{v} \cdot (\mathbf{p} - \mathbf{p}^{k-\text{Rep}}) = \sum_{h\in H} v_h (p_h - p^{k-\text{Rep}}_h) = 0. \tag{1}$$

Based on the invariant distribution $\mathbf{v}$, we can also compute the players' payoffs. To this end, let $\mathbf{S}^k = (S^k_h)_{h\in H}$ denote the vector that returns for each h the one-shot payoff that player p obtained k rounds ago,

$$S^k_h = \begin{cases} b-c & \text{if } a^p_{-k} = C \text{ and } a^q_{-k} = C \\ -c & \text{if } a^p_{-k} = C \text{ and } a^q_{-k} = D \\ b & \text{if } a^p_{-k} = D \text{ and } a^q_{-k} = C \\ 0 & \text{if } a^p_{-k} = D \text{ and } a^q_{-k} = D \end{cases} \tag{2}$$

Then we can define player p's repeated-game payoff $s_{\mathbf{p}}$ as

$$s_{\mathbf{p}} = \mathbf{v} \cdot \mathbf{S}^1 = \mathbf{v} \cdot \mathbf{S}^2 = \ldots = \mathbf{v} \cdot \mathbf{S}^n. \tag{3}$$

Let $\tilde{\mathbf{S}}^k = (\tilde{S}^k_h)_{h\in H}$ denote the analogous vector that returns for each h the one-shot payoff that player q obtained k rounds ago. Then player q's payoff is defined analogously, $s_{\mathbf{q}} = \mathbf{v} \cdot \tilde{\mathbf{S}}^{\mathbf{1}} = \ldots = \mathbf{v} \cdot \tilde{\mathbf{S}}^{\mathbf{n}}$.

Zero-determinant strategies. Based on Akin's Lemma, we can derive a theory of zero-determinant strategies analogous to the case of memory-one strategies. In the following, we say a memory-n strategy $\mathbf{p}$ is a zero-determinant strategy if there are k_1, k_2, k_3 and α, β, γ such that $\mathbf{p}$ can be written as

$$\mathbf{p} = \alpha \mathbf{S}^{k_1} + \beta \tilde{\mathbf{S}}^{k_2} + \gamma \mathbf{1} + \mathbf{p}^{k-\text{Rep}}, \tag{4}$$

where $\mathbf{1}$ is the vector for which every entry is 1. By Akin's Lemma and the definition of payoffs,

$$0 = \mathbf{v} \cdot (\mathbf{p} - \mathbf{p}^{k-\text{Rep}}) = \mathbf{v} \cdot (\alpha \mathbf{S}^{k_1} + \beta \tilde{\mathbf{S}}^{k_2} + \gamma \mathbf{1}) = \alpha s_{\mathbf{p}} + \beta s_{\mathbf{q}} + \gamma. \tag{5}$$

That is, payoffs satisfy a linear relationship.

One interesting special case arises if $k_1 = k_2 = k_3 =: k$ and $\alpha = -\beta = 1/(b+c)$ and $\gamma = 0$. In that case, the formula (4) yields the strategy

$$p_h = \begin{cases} 1 & \text{if } \ a^q_{-k} = C \\ 0 & \text{if } \ a^q_{-k} = D \end{cases} \tag{6}$$

That is, this strategy implements Tit-for-Tat (for $k=1$) or delayed versions thereof (for $k>1$). By Eq. (5), the enforced payoff relationship is $s_{\mathbf{p}} = s_{\mathbf{q}}$ (in particular, these strategies are *good*).

Another interesting special case arises if $k_1 = k_2 = k_3 =: k$ and $\alpha = 0$, $\beta = -1/b$, $\gamma = 1 - c/b$. In that case Eq. (4) yields the strategy

$$p_h = \begin{cases} 1 & \text{if } \ a^q_{-k} = C \\ 1 - c/b & \text{if } \ a^q_{-k} = D \end{cases} \tag{7}$$

That is, the generated strategy is GTFT (if $k=1$), or delayed versions thereof (for $k>1$). By Eq. (5), the enforced payoff relationship is $s_{\mathbf{q}} = b - c$. In particular, these strategies are not *good*, but they satisfy the notion of being *Nash-type* [Akin, 2016].

Proving the conjecture by considering the corner cases. Consider the following subset of 2-bit reactive strategies,

$$\mathcal{N} = \Big\{ \hat{\mathbf{p}} = (\hat{p}_{CC}, \hat{p}_{CD}, \hat{p}_{DC}, \hat{p}_{DD}) \, \Big| \, \hat{p}_{CC} = 1, \ \hat{p}_{CD} + \hat{p}_{DC} \leq 2 - c/b, \ \hat{p}_{DD} \leq 1 - c/b \Big\}. \tag{8}$$

Then one may phrase our conjecture as saying: an agreeable $\hat{\mathbf{p}}$ is of Nash type if and only if $\hat{\mathbf{p}} \in \mathcal{N}$. The set $\mathcal{N}$ is the convex hull of 10 corner points (in the following we use $p^* := 1 - c/b$),

$$\begin{array}{ll} (1,0,0,0) \ [\mathit{GRIM}] & (1,0,0,p^*) \\ (1,0,1,0) \ [\mathit{TFT}] & (1,0,1,p^*) \\ (1,1,0,0) \ [\mathit{Delayed\ TFT}] & (1,1,0,p^*) \\ (1,p^*,1,0) & (1,p^*,1,p^*) [\mathit{GTFT}] \\ (1,1,p^*,0) & (1,1,p^*,p^*) [\mathit{Delayed\ GTFT}] \end{array} \tag{9}$$

One way how to prove our conjecture is thus to prove (*i*) All 10 corner points are of Nash type, and (*ii*) the set of strategies that are of Nash-type is convex. Again, numerical computations suggest that the 10 corner points are indeed of Nash type. We have a rigorous proof (above) for the 4 corner points *TFT*, *Delayed TFT*, *GTFT*, and *Delayed GTFT*. Moreover, a proof that *GRIM* is of Nash type seems doable. We do not know yet how to do a proof for the other 5 strategies (for example, we checked that they are not zero-determinant strategies). One approach that might work is to show that the following auxiliary conjecture is true: If $(1, \hat{p}_{CD}, \hat{p}_{DC}, \hat{p}_{DD})$ is of Nash-type and $\hat{p}'_{CD} \leq \hat{p}_{DC}$, $\hat{p}'_{DC} \leq \hat{p}_{DC}$, $\hat{p}'_{DD} \leq \hat{p}_{DD}$, then the strategy $(1, \hat{p}'_{CD}, \hat{p}'_{DC}, \hat{p}'_{DD})$ is of Nash-type. If that auxiliary conjecture is true, the 10 corner strategies are of Nash type because they can all be derived from *GTFT* or *Delayed GTFT* by decreasing some of the entries.

References

E. Akin. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems*, pages 77–107, 2016.