MIT 14.123 (2009) by Peter Eso
Lecture 12: Repeated Games

1. Finitely Repeated Games
2. Perfect Folk Theorem
3. Renegotiation Proofness

Read: FT 5.1, 5.2, 5.4; Farrell & Maskin (GEB 1989)
1. Repeated Prisoners’ Dilemma

- Unique Nash equilibrium \((D,D) \rightarrow (0,0)\).
  Pareto-optimal \((C,C)\) is not an equilibrium.

- Finite repetition, \(t = 1, \ldots, T\): The only Nash outcome is \((D,D)\) in every period.

  - By induction (similar, not \(\leftrightarrow\) to backward induction, SPE).
    In any Nash equilibrium \(\sigma^*\), both players play \(D\) in period \(T\).
    Hence for any history that has positive probability up to \(T-1\),
    player \(i\) has no incentive to play \(C\) at \(T-1\), because no matter
    what he does his opponent plays \(D\) in period \(T\) anyway.
    Induction on the number of periods gives the result. ■

- In experiments (with humans or in Axelrod’s tournament) we see cooperation.
  “Tit-for-tat” does well in reality with \(T < \infty\).
Single-Deviation Principle

•Repeated games belong to the class of multi-stage games with observable actions ("almost-perfect information games").

•**THM**: A strategy profile of a multi-stage game with observable actions (finite-horizon or infinite-horizon with continuity at $\infty$) is a subgame-perfect equilibrium (SPE) iff the following holds:

For any history $h^t$ (=the play up to, not including $t$) and $i$ assume

– at $t$ and thereafter everybody except for $i$ plays according to the proposed equilibrium strategy profile, and

– at $t+1$ and thereafter $i$ plays the proposed strategy profile;

then $i$ does not have an incentive to deviate at $h^t$. 

SPE with Finite Repetition

• Set of SPE may expand even with finite repetition (not in PD).
• Ingredients: Multiple equilibria that the players rank differently, sufficiently long time horizon, and patience.
• THM (Benoit and Krishna, 1985); two players, no discounting.
  
  Suppose \((v_1', v_2')\) and \((v_1'', v_2'')\) are stage-game Nash eqm payoffs with \(v_1' > v_1''\) and \(v_2'' > v_2'\).
  
  \(\forall (v_1, v_2)\) feasible & in the shaded area, \(\forall \varepsilon > 0\), there is \(T < \infty\) such that \(G^T\) with \(T \geq T\) has SPE with average payoffs within \(\varepsilon\) of \((v_1, v_2)\).
Proof

- Choose $t^*$ such that

$$t^*(v_1' - v_1')/2 > w_1 \equiv \max_a g_1(a),$$
$$t^*(v_2' - v_2'')/2 > w_2 \equiv \max_a g_2(a).$$

- Proposed SPE, at least $2t^*$ periods before the end of the game:
  
  A. Play $(v_1, v_2)$ until time $T - t^*$ unless someone deviates.
  
  B. If no deviation in (A), then in the final $t^*$ periods alternate
    between $(v_1', v_2')$ and $(v_1'', v_2'').$
    
  C. If P1 deviates in (A), then play $(v_1', v_2')$ to the end.
    If P2 deviates in (A), then play $(v_1'', v_2'')$ to the end.

- Indeed approximates $(v_1, v_2)$ for $T$ sufficiently large.
Proof, continued

• Why SPE?

• Denote $t > t^*$ the remaining time.

• If no-one deviated before, P1 gets payoff $(t - t^*)v_1 + t^*(v_1' + v_1'')/2$ if conforms, at most $w_1 + (t-1)v_1' < w_1 + tv_1'$ if deviates.
  Difference: $(t - t^*)(v_1 - v_1') + t^*(v_1'' - v_1')/2 - w_1 > 0$ for conform.

• Same goes for P2 if no-one deviated before.

• If anyone deviated already, then Nash equilibrium is played in every period, subgame perfect.

• In the final $t^*$ periods, alternate over two Nash equilibria: SPE.
2. Infinite Repetition

- Repetition without known bound (can be finite in expectation) expands the set of equilibria even in the Prisoners’ Dilemma.

- **THM**: Infinitely repeated PD, discounted payoffs with $\delta > \frac{1}{2}$: “Grim Trigger” (=play C as long as both play C, play D forever if any player ever plays D) is SPE and yields $(C,C)$, $\forall t$.

  - Equilibrium payoff is 1 per period. Single-period deviation yields payoff 2, and 0 from then on. $1/(1-\delta) > 2$ for $\delta > \frac{1}{2}$.

- This construction is rather special: In the Prisoners’ Dilemma players can punish with stage game Nash equilibrium. This makes the infinitely repeated game equilibrium SPE.

- If the punishment is itself not an equilibrium (=not credible), then the repeated-game equilibrium is only Nash, not SPE.
Destruction By Repetition

- In the one-shot game the unique Nash equilibrium is \((A,A)\) because \(A\) is strictly dominant.

<table>
<thead>
<tr>
<th></th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>2,2</td>
<td>2,1</td>
<td>0,0</td>
</tr>
<tr>
<td>(B)</td>
<td>1,2</td>
<td>1,1</td>
<td>-1,0</td>
</tr>
<tr>
<td>(C)</td>
<td>0,0</td>
<td>0,-1</td>
<td>-1,-1</td>
</tr>
</tbody>
</table>

- \((A,A)\) in all periods is SPE for finite or infinite repetition.

- **Claim**: Infinite repetition with \(\delta > \frac{1}{2}\): \((B,B)\) \(\forall t\) is SPE outcome.

  - **Strategy** \(s^* = \{\text{Play } B \text{ at } t = 1 \text{ and } \forall t \text{ such that both players played } s^* \text{ in period } t-1; \text{ play } C \text{ at } t \text{ if someone deviated from } s^* \text{ at } t-1\}\).

  - If \(s^*(h_t) = B\): Using \(s^*\) get \(K + \delta^t + \delta^{t+1} + \ldots\); one-shot deviation to \(A\) yields \(K + 2\delta^t - \delta^{t+1} + \delta^{t+2} + \ldots\). Gain is \(\delta^t (1 - 2\delta) < 0\).

  - If \(s^*(h_t) = C\): Using \(s^*\) get \(K - \delta^t + \delta^{t+1} + \ldots\); one-shot deviation to \(A\) or \(B\) yields \(K + 0\delta^t - \delta^{t+1} + \delta^{t+2} + \ldots\). Gain is \(\delta^t (1 - 2\delta) < 0\).
General Notation

- Each period play stage game $g$; infinitely repeated game is $g^\infty$. In $g$, players are $N = \{1, \ldots, n\}$, actions $a_i \in A_i$ for $i = 1, \ldots, n$.
- $g_i(\alpha)$ is $i$’s stage game payoff given a (mixed) action profile $\alpha$.
- $\sigma_i$ is infinitely-repeated game strategy for player $i$. Specifies (mixed) action $a_i$ for all histories $h^t = (a_0^t, \ldots, a_{t-1}^t)$, $\forall t \geq 0$.
- $v_i(\sigma) = (1-\delta)\sum_{t\geq0} \delta^t \sigma(h^t) g_i(a^t|\sigma, h^t)$ is average discounted payoffs of strategy-profile $\sigma$. Comparable to per-period payoff.
- If the period-0 actions are already known, one can rewrite this as $v_i(\sigma) = (1-\delta)g_i(a^0) + \delta v_i(\sigma^c(a^0))$, where $\sigma^c(a^0)$ is the strategy profile in periods $t = 1, 2, \ldots$ induced by $\sigma$ given period-0 actions $a^0$.
- $S(\sigma) =$ set of continuation profiles of $\sigma$ after every finite history.
  Note: $\sigma$ is SPE of $g^\infty$ iff all $\sigma' \in S(\sigma)$ is SPE of $g^\infty$. 
Payoff Constraints In Any NE

- Here are two results regarding on the set of average discounted payoffs that may be the result of a Nash equilibrium of $g^T(\delta)$:

- **OBS 1**: Feasibility. If $(v_1,\ldots,v_n)$ are the average discounted payoffs in a Nash equilibrium, then

  $$(v_1,\ldots,v_n) \in \text{co}\{(x_1,\ldots,x_n) \mid \exists (a_1,\ldots,a_n) \text{ with } x_i = g_i(a_1,\ldots,a_n), \forall i\}.$$ 

- **DEF**: Minmax payoff, $v_i = \min_{\sigma_{-i}} \max_{\sigma_i(\sigma_{-i})} g_i(\sigma_i(\sigma_{-i}), \sigma_{-i}).$

- **OBS 2**: Individual Rationality. If $(v_1,\ldots,v_n)$ are the average discounted payoffs in a Nash equilibrium, then $v_i \geq v_i$ for all $i$.

- Suppose $(\sigma^*_i,\sigma^*_{-i})$ is NE of $g^T$, and construct $\sigma_i$ so that $\sigma_i(h')$ is a best-response to $\sigma^*_{-i}(h')$ at every history $h'$. Then,

  $$U_i(\sigma^*_i,\sigma^*_{-i}) \geq U_i(\sigma_i,\sigma^*_{-i}) \geq (1-\delta)/(1-\delta^{T+1}) \left(\sum_t \delta^t v_i\right) = v_i.$$
Nash Folk Theorem For $g^\infty$

- **THM**: If $(v_1,\ldots,v_n)$ is feasible & strictly individually rational, then there exists $\delta < 1$ such that $\forall \delta \geq \delta$, there is a NE of $g^\infty(\delta)$ with average payoffs $(v_1,\ldots,v_n)$.

Assume for simplicity, $\exists (a_1,\ldots,a_n) \in A$ with $g_i(a_1,\ldots,a_n) = v_i$.

- Denote $m^i_{-i}$ the strategy-profile of players other than $i$ that hold player $i$ to his minmax payoff and $m^i_i$ a best response to $m^i_{-i}$.

- Proposed equilibrium strategies: Each $i$ plays
  - $a_i$ at $h_t$ such that $(a_1,\ldots,a_n)$ has been played $\forall t' < t$.
  - $m^i_i$ if player $j$ was the first player to have deviated (or, if multiple players deviated first, simultaneously, then the lowest numbered one among them).
Proof, continued

- If player $i$ follows this strategy, then his average payoff is $v_i$.

- If player $i$ deviates in period $t$, then his average payoff is at most
  
  $$(1-\delta)(v_i + \ldots + \delta^{t-1}v_i + \delta^t w_i + \delta^{t+1}v_i + \delta^{t+2}v_i + \ldots),$$

  where $w_i = \max_{a \in A} g_i(a)$ is $i$’s highest feasible payoff in $G$.

- Deviation is not worth it if
  
  $$ (w_i - v_i) \leq \delta/(1-\delta) \ (v_i - v_i).$$

- Choose $\delta$ such that $\delta/(1-\delta) \geq \max_i (w_i - v_i) / (v_i - v_i).$ ■

- The theorem is useful as it characterizes the set of all Nash equilibria of $g^\infty(\delta)$, at least for high enough $\delta$. 

**Why Go Beyond Nash**

- Nash equilibrium is not a particularly appropriate concept for dynamic games. Reason: Incredible punishment threats.

- We can sustain \((C,C)\) in the infinitely-repeated game by P2 punishing P1 forever in case P1 ever deviates to \(D\).

<table>
<thead>
<tr>
<th></th>
<th>(C)</th>
<th>(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C)</td>
<td>1, 1</td>
<td>0, -10</td>
</tr>
<tr>
<td>(D)</td>
<td>2, 1</td>
<td>0, -10</td>
</tr>
</tbody>
</table>

- But the punishment hurts P2 more than it hurts P1; P2 may not want to carry it out.

- The example calls for requiring subgame perfection.
Perfect Folk Theorem

- THM Fudenberg and Maskin (1986). Let $V^*$ be the set of feasible and strictly IR payoffs of $G$. Assume $\dim(V^*) = n$. Then, for any $(v_1, \ldots, v_n) \in V^*$ there exists $\delta < 1$ such that for all $\delta \geq \delta$, there is a SPE of $g^\infty(\delta)$ with average payoffs $(v_1, \ldots, v_n)$.

- Wlog denote $v_i = 0$, moreover assume $\exists a \in A: g_i(a) = v_i$ for all $i$.

  - Pick $v' \in \text{int}(V^*)$ with $v' < v_i$ for all $i$.
    Let $T$ such that $Tv'_i > w_i = \max_{a \in A} g_i(a)$.

  - Pick $\varepsilon > 0$ so that for each $i$,
    $v_i(\varepsilon) = (v'_i + \varepsilon) \in V^*$ and $v'_i + \varepsilon \leq v_i$.
    Let $a^i$ such that $g_i(a^i) = v_i(\varepsilon)$. 
- Denote $m^i$ the strategy-profile that minmaxes player $i$. Assume that $m^i$ is either pure, or mixing probs can be detected.

- Here is the proposed SPE. Each player $i$ plays the following strategy, which prescribes behavior for three “phases”.

  I. Play $(a_1,\ldots,a_n)$ as long as no-one deviates from $(a_1,\ldots,a_n)$.
     If player $j$ deviates from phase I then go to phase $\Pi_j$.

  $\Pi_j$. Play $m^j_i$ for $T$ periods, then go to phase $\Pi_{j'}$ if no-one deviates.
     If player $k$ deviates in $\Pi_j$, then start over $\Pi_k$.

  $\Pi_{j'}$. Play $a'_i$ as long as no-one deviates from $\Pi_{j'}$.
     If player $k$ deviates in $\Pi_{j'}$, then go to phase $\Pi'_k$. 

II. Play $m^j_i$ for $T$ periods, then go to phase $\Pi_{j'}$ if no-one deviates.
     If player $k$ deviates in $\Pi_j$, then start over $\Pi_k$.

III. Play $a'_i$ as long as no-one deviates from $\Pi_{j'}$.
     If player $k$ deviates in $\Pi_{j'}$, then go to phase $\Pi'_k$. 


Proof, finished

- Single-deviation principle in each phase.

- In phase I, deviating once yields at most \((1-\delta)w_i + \delta^{T+1}v'_i\) which is less than \(v_i = (1-\delta^{T+1})v_i + \delta^{T+1}v_i\) if \(\delta\) is close to 1, e.g., \(\delta > (1+1/T)^{1/T}\), as \((1-\delta^{T+1})v_i = (1-\delta)(1+\delta+...+\delta^T)v_i > (1-\delta)Tv_i > (1-\delta)w_i\).

- In phase II, deviation by \(i\) postpones everything by \(T\), not worth it.

- In II\(_j\), if \(i\) deviates, he gets \((1-\delta)w_i + \delta^{T+1}v'_i\); if he conforms when \(K\) periods are still left of II\(_j\), he gets \((1-\delta^{T+1-K})g_i(m^j) + \delta^{T+1-K} (v'_i + \varepsilon)\). Conform iff \(\delta^{T+1}\varepsilon \geq (1-\delta)w_i + (1-\delta^{T+1-K})g_i(m^j) + (\delta^{T+1-K}-\delta^{T+1})(v'_i + \varepsilon)\), which holds as \(\delta\) approaches 1 (LHS \(\rightarrow \varepsilon\), RHS \(\rightarrow 0\)).

- In phase III\(_i\) or III\(_j\) the proof is like in phase I: Deviation provides gains for one period, loss for \(T\) periods, not worth it. ■
3. Renegotiation Proofness

- Criticism of repeated-game SPE with “punishment phases”: Players may want to renegotiate, if both are hurt by punishment. Farrell & Maskin GEB’89 propose to consider the following.

- **DEF**: An SPE of \( g^\infty, \sigma \), is Weakly Renegotiation Proof (WRP), if \( \forall \sigma', \sigma'' \in S(\sigma), \sigma' \) does not strictly Pareto-dominate \( \sigma'' \).

- Think of \( S(\sigma) \), all possible infinite strategy profiles induced by \( \sigma \), as “the plays we agree are in the playbook”. If \( \sigma' \in S(\sigma) \) strictly Pareto-dominates \( \sigma'' \in S(\sigma) \), then the players renegotiate \( \sigma' \) to \( \sigma'' \).

- In PD, “(D,D) forever” has unique continuation, hence it is WRP. “Grim Trigger” is not WRP; it dominates continuation after (D,D).

- Internal consistency, not comparison across SPE’s.
Theorem (Farrell & Maskin ’89)

• Consider two players; normalize minmax payoffs to 0 and let $V^*$ denote all feasible, IR payoffs.
• Suppose $(v_1, v_2) \in V^*$. If there exist actions $(a_1^1, a_2^1), (a_1^2, a_2^2)$ such that
  (1) $c_1 \equiv \max_x g_1(x, a_1^2) < v_1, g_2(a_1^1) > v_2$
  (2) $c_2 \equiv \max_x g_2(a_2^2, x) < v_2, g_1(a_2^2) > v_1$
  then for $\delta$ near 1 there is a WRP equilibrium with payoffs $(v_1, v_2)$.
  
  Conversely, if $\sigma$ is WRP equilibrium with payoffs $(v_1, v_2)$, then there exist action-pairs $a_1$ and $a_2$ satisfying (1) & (2) weakly.
Proof

• First, we construct a WRP equilibrium if (1) and (2) hold.
• Suppose \((v_1, v_2) = g(a_1, a_2)\). Propose WRP equilibrium as follows:

  (I): Play \((a_1, a_2)\) until \(i\) deviates; then go to \((II_i)\).

  (II_i): Play \(a_i\) for \(t_i\) periods, such that \(t_i g_i(a_i) + w_i < (t_i + 1) v_i\).
  Then go back to (I). If \(j\) deviates from \(II_i\) then (re)start \(II_j\).

• \(t_i\) exists by \(g_i(a_i) < v_i\) and makes deviation from (I) unprofitable.
• Set \(\delta\) high enough so that \(p_i = (1 - \delta^{t_i}) g_i(a_i) + \delta^{t_i} v_i\) satisfies
  \[ p_i > c_i \quad \text{and} \quad (1 - \delta)w_i + \delta p_i < v_i. \]
  Possible because \(g_i(a_i) \leq c_i < v_i\).

• Claim: Proposed strategies form WRP eqm for such high \(\delta\).
Illustration for $i = 1$

- If $P1$ deviates from (I), then $(\Pi_1)$ prescribes $t_1$ periods of $g(a_1)$, and then $v$ forever; payoffs are
  
  $z = (1 - \delta^t_1)g(a_1) + \delta^t_1v_1$.

- During $(\Pi_1)$, slide to $v$.

- If $P1$ deviates in $(\Pi_1)$, he gets $(1 - \delta)c_1 + \delta p_1 < p_1$.

- $P2$ does not deviate from $(\Pi_1)$ because $g(a_1) > v_2$.

- Continuation payoffs lie between $z$ and $v$, Pareto-unranked, WRP!
*Proof Still Not Over*

- **Second:** Given \( \delta \), if WRP eqm with payoffs \((v_1, v_2)\) exists, then there are actions \(a^1, a^2\) such that (1) and (2) hold weakly.
- We show that \(a^1\) satisfying (1) weakly exists; \(a^2\) & (2) analogous.
- Let \(\sigma\) be the WRP eqm given \(\delta\). If there is an action-pair \(a\) such that \(g_1(a) = v_1\) and \(g_2(a) \geq v_2\), and in addition, \(\max_x g_1(x, a_2) \leq v_1\) as well, then \(a\) itself satisfies (1).
- Otherwise, consider \(\sigma^1\), the worst continuation of \(\sigma\) for P1 after period 1 (prompted by a first-period action \(a'\) with \(g_1(a') \geq v_1\)). If there are multiple worst-continuations of \(\sigma\), then take the one that is best for P2.
- \(a^1 = \) initial action of \(\sigma^1\). We claim it satisfies (1) weakly.
The worst continuation of $\sigma$ for $P_1$, $\sigma^1$, satisfies $g_1^*(\sigma^1, \delta) \leq v_1$ and $g_2^*(\sigma^1, \delta) \geq v_2$. (The former by def, the latter by WRP.)

$g_2(a^1) \geq g_2^*(\sigma^1, \delta) \geq v_2$, establishing the second inequality in (1), because $g_2(a^1) < g_2^*(\sigma^1, \delta)$ would imply that $\sigma^1$, the continuation of $\sigma^1$ after $a^1$, satisfies $g_2^*(\sigma^1, \delta) > g_2^*(\sigma^1, \delta)$, hence by WRP $g_1^*(\sigma^1, \delta) \leq g_1^*(\sigma^1, \delta)$, contradicting that $\sigma^1$ is the worst continuation for $P_1$.

The first inequality in (1), weakly, is that $\max_x g_1(x, a^1_2) \leq v_1$. We show a bit more: $\max_x g_1(x, a^1_2) \leq g_1^*(\sigma^1, \delta) \leq v_1$.

If $\max_x g_1(x, a^1_2) > g_1^*(\sigma^1, \delta)$, then $P_1$ could profitably deviate in the first period of playing $\sigma^1$, and since his continuation payoff cannot be lower than $g_1^*(\sigma^1, \delta)$, by definition of $\sigma^1$, the deviation would be profitable overall. Hence $\max_x g_1(x, a^1_2) \leq g_1^*(\sigma^1, \delta)$.