Notes on Trembling Hand Perfection
Course Econ 201B

Rodolfo G. Campos
UCLA

January 21, 2005

1 Definition and intuition

In a Nash equilibrium each player's equilibrium strategy is a best response to the other player's equilibrium strategies. In a Trembling Hand Perfect (THP) equilibrium, there must also be arbitrarilly small perturbations of all players' strategies such that every pure strategy gets strictly positive probability and each player's equilibrium strategy is still a best response to the other players' perturbed strategies. The definition of a THP equilibrium is the following.

Definition 1 *Strategy profile σ is a trembling hand perfect (THP) equilibrium if there exists a sequence of totally mixed strategy profiles $\sigma^n \to \sigma$ such that, for all i, $u_i(\sigma_i, \sigma_{-i}^n) \geq u_i(s_i, \sigma_{-i}^n)$ for all $s_i \in S_i$.*

1.1 Intuition

In a THP equilibrium, the optimality of a player's strategy choice does not depend on an assumption that some pure strategies are getting zero probability in an equilibrium. Thus, THP helps to get rid of strange equilibria, such as (T,L) in the example below, in which a player is playing a weakly dominated strategy.

1.2 Example

Take the following normal form (constructed from FT Figure 8.10)

	L	R
T	2,2	2,2
B	1,0	3,1

There are two pure Nash equilibria: (T,L) and (B,R) and a continuum of semi-mixed equilibria in which player 1 plays T and player 2 randomizes with $\sigma_2(L) \geq \frac{1}{2}$.

1.3 A trembling hand perfect equilibrium

Take the equilibrium where (B,R) is played. Write the candidate for THP as $\sigma = (0,1,0,1)$. To prove that it is THP construct $\sigma^n = (\sigma_1^n(T), \sigma_1^n(B), \sigma_2^n(L), \sigma_2^n(R)) = \left(\frac{1}{2n}, 1-\frac{1}{2n}, \frac{1}{2n}, 1-\frac{1}{2n}\right), n = 1, 2, ...$ Clearly σ^n is totally mixed and $\sigma^n \to \sigma$.

First check if it is optimal for player 1 to play pure strategy B against this tremble:

$u_1\left(0,1,\frac{1}{2n},1-\frac{1}{2n}\right) = \frac{1}{2n} + 3\left(1-\frac{1}{2n}\right) = 3 - \frac{1}{n}$ and $u_1\left(1,0,\frac{1}{2n},1-\frac{1}{2n}\right) = 2\frac{1}{2n} + 2\left(1-\frac{1}{2n}\right) = 2$

$3 - \frac{1}{n} \geq 2$ as long as $n \geq 1$.

Next check if it is optimal for player 2 to play pure strategy R against the tremble:

$u_2\left(\frac{1}{2n},1-\frac{1}{2n},0,1\right) = 2\frac{1}{2n} + 1\left(1-\frac{1}{2n}\right) = 1 + \frac{1}{2n}$ and $u_2\left(\frac{1}{2n},1-\frac{1}{2n},1,0\right) = 2\frac{1}{2n} = \frac{1}{n}$

$1 + \frac{1}{2n} \geq \frac{1}{n}$ which also holds for $n \geq 1$.

Thus, we have that (B,R) is THP.

1.4 Example of a Nash Equilibrium that is not a trembling hand perfect equilibrium

Consider the same game and the Nash equilibrium (T,L). We want to show that it is not a trembling hand perfect equilibrium. Take any totally mixed $\sigma^n = (\sigma_1^n(T), \sigma_1^n(B), \sigma_2^n(L), \sigma_2^n(R)) \to (1,0,1,0)$.

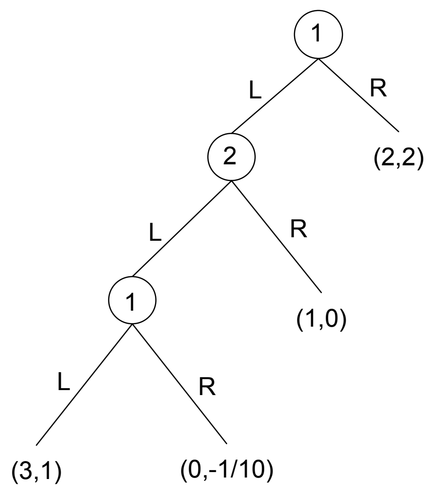
It is easy to check that as long as $\sigma_1^n(B) > 0$ it is not optimal for player 2 to play L. Compute $u_2(\sigma_1^n(T), \sigma_1^n(B), 1, 0) = 2\sigma_1^n(T)$ and $u_2(\sigma_1^n(T), \sigma_1^n(B), 0, 1) = 2\sigma_1^n(T) + \sigma_1^n(B)$. We have that $u_2(\sigma_1^n(T), \sigma_1^n(B), 0, 1) > u_2(\sigma_1^n(T), \sigma_1^n(B), 1, 0)$ as long as $\sigma_1^n(B) > 0$. Hence, (T,L) is not THP.

By the same argument it can be shown that all the equilibria in which player 2 mixes are not THP.

2 Comparison with subgame perfection/sequential equilibrium

2.1 Example of a THP that is not subgame perfect

Consider the following extensive form (Figure 8.11 in FT).



	L_2	R_2
R_1	2,2	2,2
L_1L_1	3,1	1,0
L_1R_1	$0,-\frac{1}{10}$	1,0

The only subgame perfect equilibrium is (L_1, L_2, L_1) ((L_1L_1, L_2) in the normal form). However, (R_1, R_2, R_1) ((R_1, R_2) in the normal form) is THP. Consider the tremble $\sigma^n = (\sigma_1^n(R_1), \sigma_1^n(L_1L_1), \sigma_1^n(L_1R_1), \sigma_2^n(L_2), \sigma_2^n(R_2)) = \left(1 - \frac{1}{50n} - \frac{1}{4n}, \frac{1}{50n}, \frac{1}{4n}, \frac{1}{2n}, 1 - \frac{1}{2n}\right), n = 1, 2, ...$

R_1 is a best response for player 1 since the tremble assignes a probability of at least $\frac{1}{2}$ to R_2.

The relevant calculation for player 2 is[1]

$u_2(L_2) = 2\left(1 - \frac{1}{50n} - \frac{1}{4n}\right) + \frac{1}{50n} - \frac{1}{40n}$

$u_2(R_2) = 2\left(1 - \frac{1}{50n} - \frac{1}{4n}\right)$

$u_2(R_2) \geq u_2(L_2)$ iff $\frac{1}{50n} \leq \frac{1}{40n}$ iff $\frac{4}{5} \leq 1$, which is true regardless of n. Therefore, R_2 is a best response to the tremble.

Thus, (R_1, R_2) is THP.

[1] With some abuse of notation.

3 Technical stuff (FT Ch 8.4)

DISCLAIMER: Do not read this unless you are terribly bored!

3.1 Three equivalent definitions of trembling hand perfection

Definition 2 *(A) Strategy profile σ is a trembling hand perfect equilibrium if there exists a sequence of totally mixed strategy profiles $\sigma^n \to \sigma$ such that, for all i, $u_i(\sigma_i, \sigma_{-i}^n) \geq u_i(s_i, \sigma_{-i}^n)$ for all $s_i \in S_i$.*

Feature: Does not explicitly mention ε

Definition 3 *(B) An ε-constrained equilibrium is a totally mixed strategy profile σ^ε such that, for all i, σ_i^ε solves $\max_{\sigma_i} u_i\left(\sigma_i, \sigma_{-i}^\varepsilon\right)$ subject to $\sigma_i\left(s_i\right) \geq \varepsilon\left(s_i\right)$ for all s_i, for some $\left\{\varepsilon\left(s_i\right)\right\}_{s_i \in S_i, i \in N}$ where $0 < \varepsilon\left(s_i\right) < \varepsilon$. A trembling hand perfect equilibrium is any limit of ε-constrained equilibria σ^ε as $\varepsilon \to 0$.*

Features: Several $\varepsilon\left(s_i\right)$ minimum mixing. One ε is maximal tremble.

Definition 4 *(C) Strategy profile σ^ε is an ε-perfect equilibrium if it is completely mixed, and, for all i and s_i, if there exists s_i' with $u_i\left(s_i, \sigma_{-i}^\varepsilon\right) < u_i\left(s_i', \sigma_{-i}^\varepsilon\right)$, then $\sigma_i^\varepsilon\left(s_i\right) < \varepsilon$. A trembling hand perfect equilibrium is any limit of ε-perfect strategy profiles σ^ε for some sequence ε of positive numbers that converges to zero*

Features: One ε
No optimization
Must put less than ε weight on strategies that are not best responses.

Theorem 5 *All three definitions: (A), (B) and (C) are equivalent.*

Proof. $B \Rightarrow C \Rightarrow A \Rightarrow B$.

1) $B \Rightarrow C$: The sequence σ^ε defined to be an ε-constrained equilibrium in (B) is by construction an ε-perfect equilibrium as defined in (C). Because $\varepsilon\left(s_i\right) > 0$ every profile is completely mixed. Because $\varepsilon\left(s_i\right) < \varepsilon$ there is room to put a sufficiently low weight on strategies that are not a best response. If σ_i^ε

solves $\max_{\sigma_i} u_i\left(\sigma_i, \sigma^{\varepsilon}_{-i}\right)$ subject to $\sigma_i\left(s_i\right) \geq \varepsilon\left(s_i\right)$ and s_i is not a best response, then it is optimal to choose $\sigma_i\left(s_i\right) = \varepsilon\left(s_i\right) < \varepsilon$. Hence $B \Rightarrow C$.

2) $C \Rightarrow A$: Suppose that σ satisfies condition (C). Then there is a sequence $\sigma^{\varepsilon} \to \sigma$ and a constant $d > 0$ with $\sigma_i^{\varepsilon}\left(s_i\right) > d$ for every s_i in the support of σ_i. (Why? The sequence σ_i^{ε} is completely mixed, so $\sigma_i^{\varepsilon}\left(s_i\right) > 0$ for each ε. For $s_i \in$ support $\left(\sigma_i\right)$, $\sigma_i\left(s_i\right) > 0$ by definition. This means that for a pure strategy s_i that belongs to the support, $\sigma_i^{\varepsilon}\left(s_i\right)$ is a sequence of strictly positive terms with a strictly positive limit. Therefore, $d > 0$ with $\sigma_i^{\varepsilon}\left(s_i\right) > d$ must exist.[2]) Since ε tends to zero, eventually $\varepsilon < d$, which implies that all s_i in the support are best responses. Then definition (A) is satisfied.

3) $A \Rightarrow B$: Suppose that σ satisfies condition (A) and call the hypothesized totally mixed strategy profile $\sigma^n \to \sigma$. Now define $\varepsilon_i^n\left(s_i\right)$ in the following way: if s_i is not in the support of σ_i then $\varepsilon_i^n\left(s_i\right) \equiv \sigma_i^n\left(s_i\right)$, and if s_i is in the support of σ_i then $\varepsilon_i^n\left(s_i\right) \equiv \frac{1}{n}$. Consider the following maximization program: $\left\{\max_{\sigma_i} u_i\left(\sigma_i, \sigma^{\varepsilon}_{-i}\right) \text{ subject to } \sigma_i\left(s_i\right) \geq \varepsilon_i^n\left(s_i\right), \forall s_i \in S_i\right\}$. Because σ_i is a best response to σ^n_{-i} (by assumption), one of the corresponding ε-constrained equilibria (i.e. solutions to the maximization problem) has the following form: $\sigma_i^{\varepsilon}\left(s_i\right) = \varepsilon^n\left(s_i\right)$ for $s_i \notin$ support $\left(\sigma_i\right)$ and $\sigma_i^{\varepsilon}\left(s_i\right) = \frac{\sigma_i\left(s_i\right)}{1+\sum_{\left\{s_i; \sigma_i\left(s_i\right)=0\right\}} \sigma_i^n\left(s_i\right)}$ for $s_i \in$ support $\left(\sigma_i\right)$ so that $\sum_{s_i} \sigma_i^{\varepsilon}\left(s_i\right) = 1$. (The book erroneously proposes $\sigma_i^{\varepsilon}\left(s_i\right) = \sigma_i\left(s_i\right)$. But in this case $\sum_{s_i} \sigma_i^{\varepsilon}\left(s_i\right) = \sum_{\left\{s_i; \sigma_i\left(s_i\right)>0\right\}} \sigma_i\left(s_i\right) + \sum_{\left\{s_i; \sigma_i\left(s_i\right)=0\right\}} \varepsilon^n\left(s_i\right) = 1 + \sum_{\left\{s_i; \sigma_i\left(s_i\right)=0\right\}} \sigma_i^n\left(s_i\right) > 1$ since $\sigma_i^n\left(s_i\right) > 0$ for all s_i and all n). Also, notice that this solution is not unique. Define $\varepsilon^n \equiv \max\left\{\varepsilon^n\left(s_i\right)\right\}$. Notice that $\lim_{n\to\infty} \varepsilon^n = 0$ since $\varepsilon^n\left(s_i\right)$ can either be $\frac{1}{n}$ (which tends to zero) or $\sigma_i^n\left(s_i\right)$ if s_i is not in the support of σ_i, which also tends to zero since $\sigma_i^n\left(s_i\right) \to \sigma_i\left(s_i\right) = 0$ for $s_i \notin$ support $\left(\sigma_i\right)$. In addition, $\sigma^{\varepsilon} \to \sigma$ as $n \to \infty$. Hence, (B) is satisfied. ■

3.2 Non-existence of a "truly trembling hand perfect equilibrium"

The definition of trembling hand perfection requires only that there exists **one** sequence such that the proposed strategy profile is a best response to that sequence. The problem with requiring a stronger concept (such that the strategy profile is a best response to all possible sequences) is that in this case such an equilibrium may fail to exist. In what follows, "truly trembling hand perfection" is defined, and an example where such an equilibrium does not exist is presented.

[2] The way to show this is the following: if $\sigma_i^{\varepsilon}\left(s_i\right) \to \sigma_i\left(s_i\right)$ then we can find an arbitrarily small neighborhood of $\sigma_i\left(s_i\right)$ containing all but finitely many $\sigma_i^{\varepsilon}\left(s_i\right)$. Choose a neighborhood such that its lower bound is strictly positive. Then, take the minimum over the elements of the sequence that are outside that neighborhood. This minimum is well defined because there are finitely many elements and is strictly positive because all elements are strictly positive. Now take the smaller number between this minimum and the lower bound of the neighborhood of $\sigma_i\left(s_i\right)$. Divide that number by two and call it d.

Definition 6 *Strategy profile σ is a truly trembling hand perfect equilibrium if for every sequence of totally mixed strategy profiles $\sigma^n \to \sigma$, it is true that for all i, $u_i(\sigma_i, \sigma_{-i}^n) \geq u_i(s_i, \sigma_{-i}^n)$ for all $s_i \in S_i$.*

(More on this in FT chapter 11, page 444.)

Take the example of Figure 11.2 in FT

	L	R
U	3,2	2,2
M	1,1	0,0
D	0,0	1,1

This game has two pure strategy Nash equilibria: (U, L) and (U, R). Player 1 playing U and player 2 randomizing between L and R with any probabilities is a Nash equilibrium as well. However, none of these equilibria are truly trembling hand perfect. We must consider any $\sigma_1^n = (\sigma_1^n(U), \sigma_1^n(M), \sigma_1^n(D)) \to (1, 0, 0)$ with $\sigma_1^n(U), \sigma_1^n(M), \sigma_1^n(D) > 0$. If we consider any tremble such that $\sigma_1^n(M) \neq \sigma_1^n(D)$ player 2 does not want to mix, therefore, the Nash equilibria where player 2 mixes are not truly trembling hand perfect.

If $\sigma_1^n(M) > \sigma_1^n(D)$ then player $2'$s best response is L. If $\sigma_1^n(M) < \sigma_1^n(D)$ then player $2'$s best response is R. Therefore, (U, L) and (U, R) cannot be trembling hand perfect because they depend on which sequence σ_1^n is chosen.