ECAI 2014
T. Schaub et al. (Eds.)

© 2014 The Authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-419-0-201

Practical Performance of Refinements of Nash Equilibria in Extensive-Form Zero-Sum Games

Jiří Čermák, Branislav Bošanský, Viliam Lisý[1]

Abstract. Nash equilibrium (NE) is the best known solution concept used in game theory. It is known that NE is particularly weak even in zero-sum extensive-form games since it can prescribe irrational actions to play that do not exploit mistakes made by an imperfect opponent. These issues are addressed by a number of refinements of NE that strengthen the requirements for equilibrium strategies. However, a thorough experimental analysis of practical performance of the Nash equilibria refinement strategies is, to the best of our knowledge, missing. This paper aims to fill this void and provides the first broader experimental comparison of the quality of refined Nash strategies in zero-sum extensive-form games. The experimental results suggest that (1) there is a significant difference between the best and the worst NE strategy against imperfect opponents, (2) the existing refinements outperform the worst NE strategy, (3) they typically perform close to the best possible NE strategy, and (4) the difference in performance of all compared refinements is very small.

1 Introduction

Game theory presents a widely used mathematical framework for modeling interaction of self-interested agents. The algorithmic results in game theory have led to a number of real-world applications (e.g., in security domain [15], in games like Poker [12], or in auctions and trading agents [18]). Optimal strategies to play in a game are described by a solution concept. Among all, Nash equilibrium (NE) is the best known solution concept that prescribes the optimal behavior of players under the assumption of rationality.

Due to the assumption of rationality, however, NE strategies can be weak when used against opponents that make mistakes. In zero-sum games, NE strategies guarantee at least the gain expected against a perfectly rational opponent (*value of the game*), but they do not exploit the mistakes of the opponent. These situations arise in sequential games where we distinguish mistakes made by the opponent in the past, or in the future, relative to the current state of the game. The mistakes in the past occur when the game reaches a state that NE player had not anticipated since it is preceeded by an irrational action (e.g., unconditionally giving a gift to the NE player). NE strategies do not force the player to further maximize the expected outcome in such state, because the value of the game has already been reached. Moreover, the NE player assumes that the opponent will not make a mistake in the future, e.g., NE does not prefer an action leading to opponent's decision between winning and losing actions compared to an action leading to decision between winning actions only.

To address these issues, a number of refinements of the Nash equilibrium were introduced over the years posing further restrictions on strategies (see [17] for a comprehensive survey). The simplest refinement is the *undominated NE*, where the equilibrium strategies cannot contain dominated strategies. *Sequential equilibrium* [5] uses the notion of beliefs to choose actions optimal against mistakes of opponents in the past. There are two separate groups of equilibra that follow. First group focuses on optimality against both types of mistakes of the opponents. It contains *quasi-perfect equilibrium* [16] and more restrictive *normal-form proper equilibrium* [9], which makes further assumption about the conditional probabilities of these mistakes. Second group contains equilibria that are optimal against mistakes of both players. This group includes *perfect equilibrium* [13] and *proper equilibrium* [11], where the latter again poses certain assumptions on the conditional probabilities of the mistakes.

In all existing works, the reasoning behind the NE refinements is presented on small, tailored examples in order to compactly and properly describe the advantages of newly defined solution concepts. The practical benefits of refinements in more realistic games are typically not analyzed. The exception is the work by Ganzfried et al. [2], which shows that the performance of a complex Poker player is improved if endgames are solved for the undominated equilibrium instead of using a not refined NE. In this paper we present, to the best of our knowledge, the first thorough experimental comparison of different refinements of NE on various domains.

We focus on refined strategies that exploit mistakes of the opponent and compare their expected outcome against imperfect algorithmic opponents (not fully converged solutions from anytime algorithms), as well as imperfect human opponents (behavioral solution concepts). The performance of the refined strategies is compared on a set of zero-sum extensive-form games with imperfect-information. We define and calculate the worst and the best possible NE strategies against these imperfect opponents and analyze the performance of the refinements within these bounds. On one hand the results show that there is a significant difference between these bounds and that the existing refinements typically achieve much better results than the worst possible Nash strategy. On the other hand they show that all the refinements are very close to the best possible Nash strategy suggesting that there is no other refinement of NE that would perform much better in zero-sum games. Finally, our experiments show that the quality of all the refined strategies is similar and even the simplest refinement is close to the most advanced refinement.

2 Technical background

This section introduces zero-sum extensive-form games (EFGs). We assume perfect recall, i.e., the players perfectly remember the history of their actions and all observed information.

A two player EFG consists of a set of players $\mathcal{P} = \{1, 2\}$. We use i to denote a player and $-i$ to denote the opponent of i. An EFG can

[1] Agent Technology Center, Faculty of Electrical Engineering, Czech Technical University in Prague

be visualized as a game tree. Set $\mathcal{H}$ contains all the states of the game represented as nodes in the tree. We denote $P : \mathcal{H} \to \mathcal{P} \cup \{c\}$ the function associating a player from set $\mathcal{P}$ or the chance player c with every state of the game. $\mathcal{Z} \subseteq \mathcal{H}$ is a set of terminal states (leafs in the tree). $u_i(z)$ is a utility function assigning to each leaf the value of preference for player i; $u_i : \mathcal{Z} \to \mathbb{R}$. For zero-sum games it holds that $u_i(z) = -u_{-i}(z), \forall z \in \mathcal{Z}$. The imperfect information is defined using the information sets. $\mathcal{I}_i$ is a partitioning of all $\{h \in \mathcal{H} : P(h) = i\}$ into these information sets. All states h contained in one information set $I \in \mathcal{I}_i$ are indistinguishable to player $i = P(h)$. Due to the assumption of perfect recall, all these states share the same history for player i and a set of available actions $A(h)$. We overload the notation and use $A(I)$ as actions available in I.

A pure strategy s_i for player i is a mapping $\mathcal{I}_i \to A(I_i)$. $\mathcal{S}_i$ is a set of all pure strategies for player i. A mixed strategy δ_i is a distribution over $\mathcal{S}_i$, set of all mixed strategies of i is denoted as Δ_i. A strategy profile is a set of strategies, one strategy for each player. We overload the notation and use u_i also to denote the expected utility of player i when the players play according to (pure or mixed) strategies. We say that strategy s_i weakly dominates s_i' iff $\forall s_{-i} \in \mathcal{S}_{-i} : u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$ and $\exists s_{-i} \in \mathcal{S}_{-i} : u_i(s_i, s_{-i}) > u_i(s_i', s_{-i})$.

In EFGs we can represent the strategies as behavioral strategies b_i for i, which assign probability distribution over $A(I_i)$ for each I_i. $\mathcal{B}_i$ is a set of all behavioral strategies for i. As shown in [6], for all games with perfect recall, behavioral strategies have the same expressive power as mixed strategies. Finally, we can use sequence-form representation for games with perfect recall [4]. A sequence σ_i is a list of actions of player i ordered by their occurrence on the path from root of the game tree to some node. The strategy is then formulated as realization plan r_i that for a sequence σ_i represents the probability of playing actions in σ_i assuming the other players play such that the actions of σ_i can be executed. Realization plan r_i has to satisfy network flow property; i.e., $r_i(\sigma) = \sum_{a \in A(I)} r_i(\sigma \cdot a)$, where I is an information set reached by sequence σ and $\sigma \cdot a$ stands for σ extended by action a.

3 Solution Concepts

This section introduces the Nash equilibrium solution concept together with the refinements that aim to exploit the mistakes of the opponent. We use two EFGs depicted in Figure 1 to illustrate the differences between the solution concepts. For the sake of simplicity we use a simple perfect-information game (1a) to describe Nash, undominated, and sequential equilibrium, and a more complex imperfect-information game (1b) to distinguish quasi-perfect and normal-form proper equilibrium. Our focus is on imperfect-information games; hence, we do not specifically define a well-known subgame-perfect equilibrium that is useful only in perfect-information games. Moreover, the sequential equilibrium is in fact the generalization of the subgame perfection to the games with imperfect information. Since we are in the zero-sum setting, only the utility of player 1 is depicted in the game trees and player 2 aims to minimize this utility value.

Nash equilibrium To define Nash equilibrium we need to introduce the concept of best responses. A best response to a strategy of the opponent δ_{-i} is a strategy δ_i^* for which $u_i(\delta_i^*, \delta_{-i}) \geq u_i(\delta_i, \delta_{-i}),\ \forall \delta_i \in \Delta_i$. $BR_i(\delta_{-i})$ is a set of the best responses to the strategy δ_{-i}. The strategy profile $\{\delta_i, \delta_{-i}\}$ is a Nash equilibrium iff $\delta_i \in BR_i(\delta_{-i}),\ \forall i \in P$. A game can have more than one Nash equilibrium. All Nash equilibrium strategy profiles have the same expected value in zero-sum games, called *the value of game*.

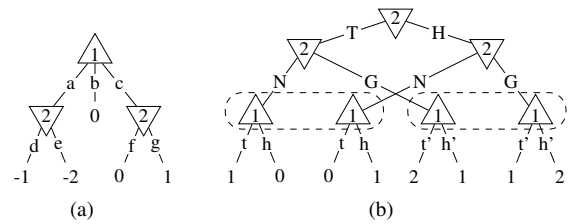

Figure 1: (a) A game with different Nash, undominated, and sequential equilibria. (b) Matching pennies on Christmas morning.

Let us discuss equilibria of a game depicted in Figure 1a. Value of this game is 0, achievable by action b of player 1 or by actions c and f. This game has 4 pure strategy Nash equilibria $\{(c), (f, e)\}$, $\{(b), (f, e)\}$, $\{(c), (f, d)\}$ and $\{(b), (f, d)\}$. It is easy to see, that strategy b for player 1 can be considered insensible, since c weakly dominates b and so player 1 can only gain by playing c. Player 1 can exploit possible mistakes of the opponent in the future (player 2 can play action g by mistake) by preferring c over b. Strategy d is also dominated by e, thus the strategy profiles containing d can be considered irrational. These profiles are NE only because player 2 does not expect player 1 to play a and so there are no restrictions on d and e, and represent the mistakes of the opponent in the past.

Sequential equilibrium Solution concept due to Kreps et al. [5] optimal against mistakes of opponent in the past. This equilibrium uses the notion of beliefs μ, which are probability distributions over the states in information sets symbolizing the likelihood of being in a state when the information set is reached. Assessment is a pair (μ, b), that is beliefs for all players and a behavioral strategy profile. A sequential equilibrium is an assessment which is consistent (μ is generated by b) and the sequential best response against itself (b is created by best responses when considering μ). The set of sequential equilibria forms a non-empty subset of Nash equilibria [5].

Let us again analyze a game from Figure 1a. There are 2 pure strategy sequential equilibria $\{(c), (f, e)\}$ and $\{(b), (f, e)\}$, since beliefs of being in a state after playing action a is equal to 1 and player 2 has to choose the best action in this state. Insisting on sequentiality thus removed 2 insensible equilibria, since the insensibility was caused by not considering the mistake of opponent in the past (playing a).

Undominated equilibrium An undominated equilibrium consists only of undominated strategies in the sense of weak dominance. For two-player zero-sum extensive-form games it holds that every undominated equilibrium of extensive-form game G is a perfect equilibrium (as defined by Selten [13]) of its corresponding normal-form game G' and therefore forms a normal-form perfect equilibrium of G. Furthermore as shown in [13], the set of undominated equilibria forms a non-empty subset of Nash equilibria.

Thanks to the fact that all irrational equilibria of the game from Figure 1a contain dominated strategies, only the sensible equilibrium $\{(c), (f, e)\}$ is undominated.

Quasi-perfect equilibrium Informally speaking, quasi-perfect equilibrium is a solution concept due to van Damme [16] that requires that each player at every information set takes a choice which is optimal against mistakes of all other players. The set of all quasi-perfect equilibria forms a non-empty subset of sequential and undominated equilibria [16].

Since quasi-perfect equilibrium refines undominated equilibrium,

there is also only one pure strategy quasi-perfect equilibrium of game from Figure 1a $\{(c),(f,e)\}$. Now consider the game from Figure 1b. This game is a modification of Matching pennies, called Matching pennies on Christmas morning [10]. In the original Matching pennies, player 2 hides a penny with heads or tails on the top. After that, player 1 guesses what player 2 did. If he guesses correctly, player 1 receives payoff of 1; if not, both players receive 0. The variant on Christmas morning adds an option for player 2, to give player 1 a gift before guessing, which increases his payoff by 1 no matter what. This game has an infinite number of quasi-perfect equilibria, namely all the equilibria which have $\delta_2(TN) = \delta_2(HN) = \delta_1(t) = \delta_1(h) = 0.5$, $\delta_2(TG) = \delta_2(HG) = 0$. There are no restrictions on $\delta(h')$ and $\delta(t')$, because neither of these strategies dominates other (so there is no restriction from the normal-form perfection) and when considering that player 2 makes a mistake G with the same probability in both states then $\mu(TG) = \mu(HG)$, there is no restriction from sequential equilibrium either. Event though these equilibria are not strictly insensible, one might want player 1 to play $\delta_1(h') = \delta_1(t') = 0.5$ because it is more robust to the deviations of opponent.

Normal-form proper equilibrium An equilibrium in behavioral strategies of an extensive-form game is said to be normal-form proper [10] if it is behaviorally equivalent to a proper equilibrium of the corresponding normal-form game. This equilibrium is optimal against mistakes of opponent in the past and in the future. Furthermore, the solution concept assumes that these mistakes are made in a certain manner, meaning that the more costly mistakes are made with exponentially smaller probability than the less costly ones. Finally, as shown in [10], every normal-form proper equilibrium is quasi-perfect and the set of normal-form proper equilibria of every extensive-form game is non-empty.

Since every normal-form proper equilibrium is also quasi-perfect there is again only one normal-form proper equilibrium $\{(c),(f,e)\}$ of the game in Figure 1a. There is also only one normal-form proper equilibrium in Matching pennies on Christmas morning, specifically the one where $\delta_1(h') = \delta_1(t') = 0.5$.

3.1 Algorithms for Computing Equilibria

In order to experimentally compare different Nash equilibrium refinements, we need to compute the refined strategies. This section describes the algorithms for finding these strategies.

Nash equilibrium We first describe the algorithm for computing Nash equilibrium that exploits the sequence form due to Koller et al. [4]. In eqs. (1) to (4) we present a linear program (LP) for solving two player zero-sum EFGs. Matrix A is a utility matrix with rows corresponding to sequences of player 1 and columns to sequences of player 2. Each entry of A corresponds to utility value of a game state reached by the sequence combination assigned to this entry, weighted by the probability of occurrence of this state considering nature. If the reached state is non-terminal, or if the sequence combination is incompatible, the entry is 0. Matrices E and F define the structure of the realization plans for player 1 and 2 respectively. Columns of these matrices are labeled by sequences and rows by information sets. Row for information set I contains -1 on a position corresponding to a sequence leading to I, 1 for the sequences leading from I and zeros otherwise. First row, corresponding to artificial information set has 1 only on position for empty sequence. These matrices ensure that for every information set I_i the probability with which we play a sequence leading to I_i is equal to sum of probabilities of sequences leaving I_i according to r_i. Vectors e, f are indexed by sequences of players and consist of 0, with 1 on the first position. q is a vector of variables representing values in information sets of the opponent. The constraint in (3) enforces the structure of realization plan and the constraint in (2) tightens the upper bound on value in each of the opponent's information sets I_2 for every sequence leaving I_2.

$$\max_{r_1,q} \quad f^\top q \tag{1}$$

$$s.t. \quad -A^\top r_1 + F^\top q \leq 0 \tag{2}$$

$$Er_1 = e \tag{3}$$

$$r_1 \geq 0 \tag{4}$$

Undominated equilibrium Undominated equilibrium is defined as a Nash equilibrium in undominated strategies. It can be computed using 2 LPs. First LP depicted in eqs. (1) to (4) solves the game for Nash equilibrium. The value of the game computed by this LP is then supplied to the second LP presented in eqs. (5) to (8) via constraint (6) to ensure, that the resulting realization plan r_1 is a Nash equilibrium. Second modification of this LP is in the objective, using uniform realization plan for the minimizing player r_2^m.

$$\max_{r_1,q} \quad r_1^\top A r_2^m \tag{5}$$

$$s.t. \quad f^\top q = v_0 \tag{6}$$

$$-A^\top r_1 + F^\top q \leq 0 \tag{7}$$

$$Er_1 = e; \qquad r_1 \geq 0 \tag{8}$$

The restriction to undominated strategies is enforced by the objective (5). The best response to a fully mixed strategy cannot contain dominated strategies and thus we have that r_1 is undominated and therefore normal-form perfect for two-player zero-sum games [17].

Quasi-perfect equilibrium Quasi-perfect equilibrium is a restriction of Nash equilibrium, which prescribes optimal play against mistakes of the opponent in the past and in the future. In eqs. (9) to (12) we present LP due to Miltersen et al. [9]. The main idea of this approach is to use symbolic perturbations of strategies, with ϵ as a parameter, and then use a parameterizable simplex algorithm to solve this LP optimally. The results of such an algorithm are strategies expressed as polynomials in epsilon, which are then used to reconstruct the realization plans even in those information sets which are not reachable when considering a rational opponent (see [9] for the details of this transformation). Vectors $l(\epsilon)$ and $k(\epsilon)$ are indexed by sequences and contain above mentioned symbolic perturbations forcing this LP to create a quasi-perfect equilibrium. Vector v contains slack variables forcing the player to exploit the weak strategies of the opponent, matrices A, E and F, and vectors e,f, and q are as before.

$$\max_{r_1,v,q} \quad q^\top f + v^\top l(\epsilon) \tag{9}$$

$$s.t. \quad F^\top q \leq A^\top r_1 - v \tag{10}$$

$$r_1 \geq k(\epsilon) \tag{11}$$

$$Er_1 = e; \qquad v \geq 0 \tag{12}$$

Even though Miltersen et al. argue in [9] that it is possible to solve this LP using a non-symbolic perturbation, the ϵ required for such a computation can be too small for floating point arithmetics. Therefore, one either needs to use an unlimited precision arithmetics, or the parameterizable simplex algorithm to compute the equilibrium, which limits the scalability.

Normal-form proper equilibrium Normal-form proper equilibrium is a Nash equilibrium optimal against mistakes of opponent, while assuming that the probability of the mistakes depends on the potential loss for such a mistake. The algorithm for computing normal-form proper equilibria of extensive-form zero-sum games is due to Miltersen et al. [10] and it is based on an iterative computation of LP pairs V and W shown in eqs. (13) to (21). In the k-th iteration the LP V generates a strategy that exploits all marked exploitable sequences. The LP uses a set of vectors $\{m_1, ..., m_k\}$, where $m_i \in \{0,1\}^{|f|}$ labels exploitable sequences based on the results of $W^{(i-1)}$, and set $\{v^{(1)}, ..., v^{(k-1)}\}$, where $v^{(i)}$ is a value of $V^{(i)}$; t is a scalar which is used in further iterations as $v^{(k)}$. The constraint (15) ensures, that the computed strategy is a Nash equilibrium.

$$V^{(k)}: \quad \max_{r_1,q,t} \quad t \tag{13}$$

$$s.t. \quad -A^\top r_1 + F^\top q + m^{(k)} t \leq -\sum_{0<i<k} m^{(i)} v^{(i)} \tag{14}$$

$$f^\top q = v^{(0)} \tag{15}$$

$$Er_1 = e; \qquad r_1 \geq 0; \qquad t \geq 0 \tag{16}$$

LP W in k-th iteration marks sequences, which are still exploitable, given previous iterations and $V^{(k)}$. Vector u is used to identify exploitable sequences and variable d is used as an auxiliary scalar for scaling purposes. This algorithm is initialized by $V^{(0)}$ which is a LP generating Nash equilibrium from eqs. (1) to (4) and $W^{(0)}$ which is equal to $W^{(k)}$ only with the sum from constraint (18) omitted, since there are no results from previous iterations.

$$W^{(k)}: \quad \max_{r_1,q,u,d} \quad 1^\top u \tag{17}$$

$$s.t. \quad -A^\top r + F^\top q + u \leq -\sum_{0<i\leq k} m^{(i)} v^{(i)} d \tag{18}$$

$$Er_1 - ed = 0 \tag{19}$$

$$f^\top q - v^{(0)} d = 0 \tag{20}$$

$$0 \leq u; \qquad r_1 \geq 0; \qquad d \geq 1 \tag{21}$$

Although this procedure runs in polynomial time since the number of LP pairs is bounded by the number of sequences of the opponent, in practice this approach can suffer from numerical precision errors when used for solving larger games. The primary reason of this instability is the error that cumulates in equation (18).

4 Experimental Comparison

This section compares the practical performance of the different variants of refinements of Nash equilibrium (NE) strategies. Since all the compared strategies are NE strategies, they cannot be exploited, and thus we are interested in the expected value of these strategies against imperfect opponents. First, we describe the set of methods for creating not perfectly rational opponents, following by the set of domains, on which we compared the quality of the strategies.

4.1 Imperfect Opponents

We use two types of imperfect opponents: (1) we use not fully converged strategies from anytime algorithms used for solving extensive-form games in practice, and (2) we use a game-theoretic model that simulates the decisions made by human opponents, where a player is more likely to make less costly mistakes rather than choosing a completely incorrect move.

We use 2 algorithms for generating the imperfect opponents of the first type: counter-factual regret minimization (CFR) algorithm [19] and Monte-Carlo tree search (MCTS). CFR is used in its basic form which iteratively traverses the whole game tree, updating the strategy in every information set according to a regret minimizing rule. MCTS is used in a most typical game-playing variant: UCB algorithm [3] is used as the selection method and it is used in each information set (this variant is termed Information Set MCTS [1]). An additional modification made to MCTS is nesting—MCTS algorithm runs for certain number of iterations, and then advances to each of succeeding information sets and repeats the whole procedure. This ensures reasonable strategies in all parts of the game tree. The behavioral strategy over actions in each information set corresponds to the frequencies, with which the MCTS algorithm selects the actions in this information set. The first algorithm provably converges to NE, while there are no guarantees for convergence of this variant of MCTS in imperfect-information games. The iterative manner of these algorithms allows us to sample the strategies before the full convergence to generate different variants of imperfect opponents.

The opponents of the second type correspond to quantal-response equilibrium (QRE) [8]. Calculation of QRE is based on a logit function, which prescribes the probability for every action in every information set as follows.

$$B(I,a) = \frac{e^{\lambda u(I,a|B)}}{\sum_{a' \in A(I)} e^{\lambda u(I,a'|B)}} \tag{22}$$

$B(I,a)$ stands for the probability of occurence of the action a in the information set I given B, $u(I,a'|B)$ is the expected utility when playing a in the information set I and according to B otherwise. We can sample the strategies for specific values of the λ parameter. By setting $\lambda = 0$ we get uniform fully mixed strategy and we get an approximative sequential NE strategy when λ is very large.

4.2 Experimental Domains

The performance of refined strategies is compared on Leduc holdem, imperfect-information variant of the card game Goofspiel, and randomly generated extensive-form games. These games were chosen, because they differ in the cause of imperfect information; for Leduc holdem poker the uncertainty is caused by the unobservable action of nature at the beginning of the game, while in imperfect information variant of Goofspiel and Random games the uncertainty is caused by partial observability of opponents moves. The size of evaluated games correspond to the maximal sizes of games, for which we were able to compute quasi-perfect and normal-form proper equilibrium in reasonable time and without numerical precision errors.

Leduc holdem Poker Leduc holdem Poker is a variant of simplified Poker using only 6 cards, namely $\{J, J, Q, Q, K, K\}$. The game starts with a non-optional bet of 1 called ante, after which each of the players receives a single card and a first betting round begins. In this round player 1 decides to either *bet*, adding 1 to the pot, or to *check*. If he bets, second player can either *call*, adding 1 to the pot, *raise* adding 2 to the pot or *fold* which automatically ends the game in the favor of player 1. If player 1 checks, player 2 can either *check* or *bet*. If player 2 raises after a *bet*, player 1 can either *call* or *fold* ending the game in the favor of player 2. This round ends either by *call* or by *check* from both players. After the end of this round, one card is dealt on the table, and a second betting round with the same rules begins. After the second betting round ends, the outcome of the game

is determined. A player wins if (1) her private card matches the table card, or (2) none of the players' cards matches the table card and her private card is higher than the private card of the opponent. If no player wins, the game is a draw and the pot is split.

Goofspiel A card game with three identical packs of cards, two for players and one randomly shuffled and placed in the middle. In our variant both players know the order of the cards in the middle pack. The game proceeds in rounds. Every round starts by revealing the top card of the middle pack. Both players proceed to simultaneously bet on it using their own cards, which are discarded after the bet. Player with higher bet wins the card. After the end of the game, player with higher sum of values of cards collected wins. In an imperfect-information version of Goofspiel, the players do not observe the bet of the opponent and after a turn they only learn whether they have won, lost, or if there was a tie.

Randomly Generated Games Finally, we used randomly generated games without nature, in which we altered several characteristics: the depth of the game (number of moves for each player) and the branching factor representing the number of actions the players can make in each information set. Moreover, each action of a player generates some observation signal (a number from a limited set) for the opponent – the states that share the same history and the same sequence of observations belong to the same information set. Therefore, by changing the amount of possible observation signals we change the number of information sets in the game (e.g., if there is only a single observation signal, neither of the players can observe the actions of the opponent). The utility is calculated as follows: each action is assigned a random integer value uniformly selected from the interval $-l, +l$ for some $l > 0$ and the utility value in a leaf is a sum of all values of actions on the path from the root of the game tree to the leaf. This method for generating the utility values is based on random *T-games* [14] that create more realistic games using the intuition of good and bad moves.

4.3 Experimental Settings

We have implemented the algorithms for computing Nash, undominated[2], and normal-form proper equilibrium, and we use IBM CPLEX 12.5 for solving LPs, Gtf framework[3] for computing quasi-perfect equilibrium and Gambit [7] for computing quantal-response equilibrium. The Gtf framework uses simplex with symbolic perturbations, which limits it's scalability.

We analyze the performance of the refined strategies within an interval determined by the worst and best possible NE strategy against a specific opponent strategy. These bounds are computed via the LPs used for the undominated equilibrium. To compute the best NE against a strategy, we use this strategy in the objective of the second LP. Moreover, if we change the objective to $\min$ in such modified LP, we compute the worst NE.

4.4 Results

The overall results are depicted in Figure 2, the interval between the worst and best NE is the grey area; SQF denotes NE computed using sequence-form LP; UND denotes undominated equilibrium; QPE quasi-perfect; and NFP normal-form proper equilibrium.

[2] We use fully mixed uniform strategy of the opponent as the input.
[3] Available at `http://www.cs.duke.edu/~trold/gtf.html`

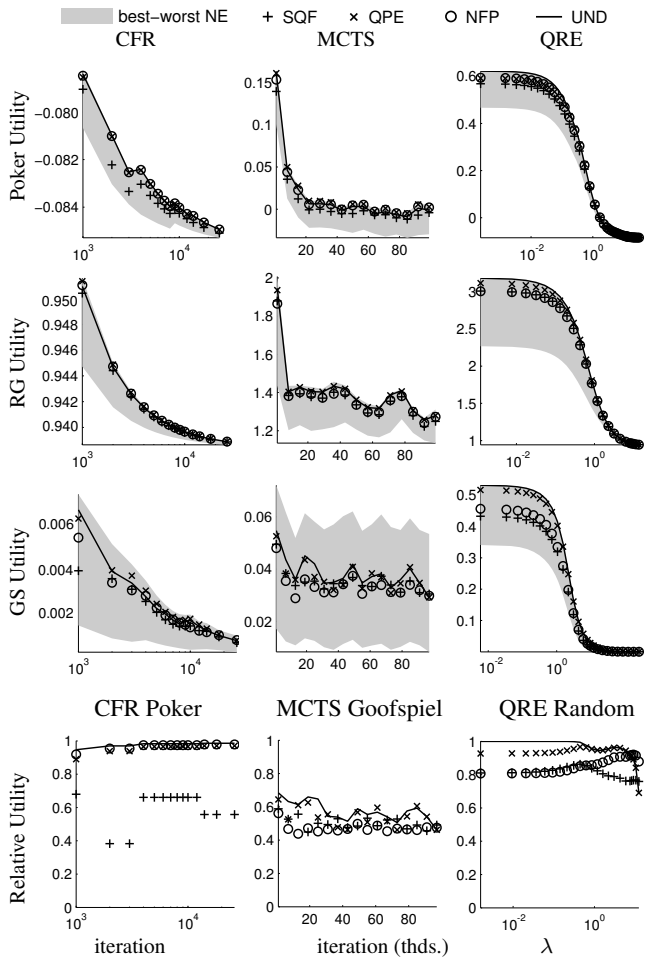

Figure 2: Overview of the utility value for different equilibrium strategies. Results for a single type of imperfect opponent are depicted in columns (CFR, MCTS, QRE), the results for a single domain are depicted in rows (Poker, Goofspiel, Random Games); the last row shows the relative performance on selected domains.

The first row shows the absolute utility values gained by different refinements against different opponents on Leduc holdem from the perspective of player 1 (note the logarithmic scale of x-axis in case of CFR and QRE). The results show that all the refinements have similar performance against all opponents and they all outperform SQF strategy. The similarity of NFP, QPE, and UND refinements can be demonstrated by the maximal difference in absolute utility values between refinements that is equal to 0.03—this occurs against QRE and it is caused by near-optimal performance of UND against QRE for small λ. This is expected since the QRE strategy for small λ is similar to uniformly mixed strategy, to which UND computes the best NE strategy. Besides that the absolute differences were mostly marginal: $6 \cdot 10^{-5}$ for CFR and $8 \cdot 10^{-3}$ for MCTS. To better distinguish the performance of the refinements we depict the relative utility gain for CFR in the interval between the worst and the best NE strategy (leftmost graph in the last row of Figure 2).

Results on the random games with branching factor 3, depth 3 and 3 possible observations are shown in second row of Figure 2. These results are computed as an average over 10 different random games generated with the selected properties but different structure of informations sets and utility values. The results in absolute utility values are very similar as in poker, but the difference between the refine-

ments and SQF decreased. The relative utility gain for QRE opponent is for clarity depicted in the rightmost graph in the last row of Figure 2. It confirms that for smaller λ the UND outperforms every other equilibrium, however with increasing λ the undominated equilibrium gets worse and both QPE and NFP improves their performance as the QRE converges to more rational strategies. Moreover, we performed a different set of experiments by varying the size of the observation set. When set to 1, the game degenerates to a very specific imperfect-information game, where every action of a player is unobservable to the opponent. Interestingly, in this setting all NE collapsed, there was no difference between the worst and the best NE strategy, and thus neither between the refined strategies.

Finally, we present results in absolute utility values on imperfect information Goofspiel with 4 cards in the third row of Figure 2. The middle graph in the last row of Figure 2 depicts the relative performance against MCTS. The results are again computed as means of 10 different random orderings of the middle pack of cards. Again, there is a very similar pattern of behavior against CFR and QRE opponents. Against the MCTS, however, the difference between the refinements and the best NE strategy slightly increased. This is caused by the fact that the MCTS reaches an irrational strategy composed of the correct pure strategies, however, incorrectly mixed. This type of mistakes does not follow the model assumed in QPE and NFP, and neither UND can optimally exploit this strategy. This setting present the only case where further improvements in exploiting the mistakes of the opponent are possible.

The results on all domains and against all imperfect opponents offer several conclusions. First of all, all the refinements typically perform very well and close to optimal NE strategy. This indicates that it is unlikely that a new refinement with dramatically better performance can be defined. Secondly, the performance of all the refinements is very similar in practice (especially against CFR and MCTS) regardless of their theoretical properties. This is interesting and suggest that the situations assumed by these refinements are not that common in real-world games. Moreover, even though NFP considers likeliness of mistakes of the opponent with respect to the potential loss, its performance in practice was similar to QPE against this type of opponent. Finally, the presented results show that in practical applications, it is sufficient to use UND refinement: (1) the quality of strategies is very similar to QPE and NFP, and (2) it is much easier to compute compared to more advanced solution concepts, since it does not require iterative process or unlimited precision arithmetic. We have performed additional measurements for UND, SQF and best-worst NE on bigger domains such as Goofspiel with 5 cards in every deck or Leduc holdem with two possible values of bets and raises. The results obtained on these domains were consistent with the results presented in Figure 2, implying that our reasoning holds with the increasing domain size.

5 Conclusion

This paper presents a thorough comparison of the refinements of Nash equilibrium on a set of zero-sum, extensive-form games with imperfect information. We compare three different refinements against imperfect opponents that simulate mistakes made by an algorithm, or a human opponent. The experimental results show that the existing refinements typically achieve much better results than the worst possible Nash strategy, and confirm the usefulness of using refined solution concepts in practice. Moreover, by comparing these refinements to the best value achievable by any refinement we show that it is unlikely that some other refinement of Nash equilibrium might exist that can dramatically outperform the existing refinements in zero-sum games. Finally, we show that the quality of all the refined strategies is similar and using the simplest undominated equilibrium can be sufficient for many real-world cases.

Presented experimental work offers several directions for future work. A deeper analysis of the presented results should offer a new theoretical insights on the mistakes of the opponent and their optimal exploitation in practice by approximating the best Nash equilibrium strategy. Moreover a similar experimental analysis should be performed for general-sum games as well to illustrate the differences in the more generic model.

ACKNOWLEDGEMENTS

This research was supported by the Czech Science Foundation (grant no. P202/12/2054).

REFERENCES

[1] Peter I Cowling, Edward J Powley, and Daniel Whitehouse, 'Information set monte carlo tree search', *Computational Intelligence and AI in Games, IEEE Transactions on*, **4**, 120–143, (2012).
[2] Sam Ganzfried and Tuomas Sandholm, 'Improving performance in imperfect-information games with large state and action spaces by solving endgames', in *Computer Poker and Imperfect Information Workshop at the National Conference on Artificial Intelligence (AAAI)*, (2013).
[3] Levente Kocsis, Csaba Szepesvári, and Jan Willemson, 'Improved Monte-Carlo Search', (2006).
[4] Daphne Koller, Nimrod Megiddo, and Berhard von Stengel, 'Fast algorithms for finding randomized strategies in game trees', *Proceedings of the 26th annual ACM symposium on Theory of computing*, (1994).
[5] David M. Kreps and Robert Wilson, 'Sequential equilibria', *Econometrica*, (1982).
[6] Harold W. Kuhn, 'Extensive games and the problem of information', *Annals of Mathematics Studies*, (1953).
[7] Richard D McKelvey, Andrew M McLennan, and Theodore L Turocy, 'Gambit: Software tools for game theory', (2010).
[8] Richard D McKelvey and Thomas R Palfrey, 'Quantal response equilibria for normal form games', *Games and economic behavior*, **10**(1), 6–38, (1995).
[9] Peter Bro Miltersen and Troels Bjerre Sørensen, 'Computing a quasi-perfect equilibrium of a two-player game', *Economic Theory*, (2008).
[10] Peter Bro Miltersen and Troels Bjerre Sørensen, 'Fast algorithms for finding proper strategies in game trees', *In proceedings of 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, 874–883, (2008).
[11] Roger B. Myerson, 'Refinements of the nash equilibrium concept', *Game Theory*, **7**, 73–80, (1978).
[12] Tuomas Sandholm, 'The State of Solving Large Incomplete-Information Games, and Application to Poker ', *AI Magazine*, 13–32, (Winter 2010).
[13] Reinhard Selten, 'Reexamination of the perfectness concept for equilibrium points in extensive games', *International Journal of Game Theory*, **4**, 25–55, (1975).
[14] S.J.J. Smith and D.S. Nau, 'An analysis of forward pruning', in *Proceedings of the National Conference on Artificial Intelligence*, pp. 1386–1386, (1995).
[15] Milind Tambe, *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*, Cambridge University Press, 2011.
[16] Eric van Damme, 'A relation between perfect equilibria in extensive form games and proper equilibria in normal form games', *Game Theory*, **13**, 1–13, (1984).
[17] Eric van Damme, *Stability and Perfection of Nash Equilibria*, Springer-Verlag, 1991.
[18] Michael P. Wellman, *Trading Agents*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Pub., 2011.
[19] Martin Zinkevich, Michael Bowling, and Neil Burch, 'A new algorithm for generating equilibria in massive zero-sum games', *In Proceednigs of the 22nd National Conference on Artificial Intelligence*, (2007).