

Approximation of Values and Equilibria in Stochastic Games: Complexity and Algorithms

by

Ali Asadi

August, 2026

*A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Master of Science*

Committee in charge:

Krishnendu Chatterjee, Supervisor
Thomas A. Henzinger



The thesis of Ali Asadi, titled *Approximation of Values and Equilibria in Stochastic Games: Complexity and Algorithms*, is approved by:

Supervisor: Krishnendu Chatterjee, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Thomas A. Henzinger, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

© by Ali Asadi, August, 2026

CC BY 4.0 The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution 4.0 International License. Under this license, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author.

ISTA Master's Thesis, ISSN: 2791-4585

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I accept full responsibility for the content and factual accuracy of this work, including the data and their analysis and presentation, and the text and citation of other work.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Ali Asadi
August, 2026

Signed page is on file

Abstract

We study the computational aspects of the two central solution concepts of stochastic games played on finite graphs for an infinite number of steps: the *value* in two-player zero-sum games, and *Nash equilibria* in multi-player games. In both settings, exact solutions are intractable. For two-player zero-sum concurrent stochastic games, values can be irrational, and the associated decision problem is SQRT-SUM hard. For multi-player turn-based stochastic games, the equilibrium probabilities can be irrational, and deciding the constrained existence of a Nash equilibrium is undecidable in the presence of memory, and $\exists\mathbb{R}$ -complete when the players are restricted to stationary strategies. Hence, in both settings, the natural computational problem is the *approximation* of the solution within an arbitrary additive error. In this thesis, we present improved complexity results and algorithms for such approximation problems.

The first part of the thesis considers two-player zero-sum concurrent stochastic games with two classic objectives: (a) *stateful-discounted* objectives, which are similar to the classic discounted-sum objectives, but states are associated with different discount factors rather than a single discount factor; and (b) *parity* objectives, which are a canonical representation for ω -regular objectives. For stateful-discounted objectives, given an ordering of the discount factors, the limit value is the limit of the value of the stateful-discounted objectives, as the discount factors approach zero according to the given order. The value-approximation problem is known to be in EXPSPACE for the limit value of stateful-discounted objectives and in PSPACE for parity objectives, and the best-known algorithms for both problems are at least exponential time, with an exponential dependence on the number of states and actions. Our main results for the value-approximation problem for the limit value of stateful-discounted objectives and parity objectives are as follows: (a) we establish TFNP^{NP} complexity; and (b) we present algorithms that improve the dependency on the number of actions in the exponent from linear to logarithmic. In particular, if the number of states is constant, our algorithms run in polynomial time.

The second part of the thesis considers multi-player turn-based stochastic games with terminal payoffs, where the players are restricted to stationary strategies. We study the *approximate constrained existence problem* of Nash equilibria: given a game that has a Nash equilibrium in which the payoff of every player lies in a given interval, compute a strategy profile that is an ε -Nash equilibrium and that satisfies the same constraints up to an additive error ε . Our main results are as follows: (a) we present an FNP^{NP} algorithm for the problem; and (b) we establish that its decision version is NP-hard, already with two players. For the upper bound, we show that if a constrained Nash equilibrium exists, then there exists one in which the non-zero probabilities are at least an inverse of a double-exponential in the

input size, and that such a strategy profile can be represented, up to an additive error ε , with floating-point numbers of polynomial size. We complement the upper bound with a lower bound on this approach: we present a game with only five players in which every constrained Nash equilibrium requires double-exponentially small probabilities.

The two parts share a common technical toolkit: bounds on solutions of systems of polynomial inequalities that arise from real algebraic geometry, the floating-point representation of doubly-exponentially small numbers, and guess-and-verify procedures that resolve single-player games with an NP oracle.

Acknowledgements

I thank my supervisor, Krishnendu Chatterjee, for his guidance throughout this work. I thank Thomas A. Henzinger for serving on my thesis committee. I also thank my co-authors, Léonard Brice, Raimundo Saona, Jakub Svoboda, and K. S. Thejaswini, for the collaboration that led to the two papers this thesis is based on.

The research reported in Part I was partially supported by the European Research Council (ERC) Consolidator Grant 863818 (ForM-SMArt), by the Austrian Science Fund (FWF) 10.55776/COE12, and by the French Agence Nationale de la Recherche (ANR) grant ANR-21-CE40-0020 (CONVERGENCE project). The research reported in Part II is a part of project VAMOS, which has received funding from the European Research Council (ERC) under grant agreement No. 101020093.

About the Author

Ali Asadi received his B.Sc. in Computer Engineering from Sharif University of Technology in Tehran in 2021, and joined the Institute of Science and Technology Austria (ISTA) in the same year as a Ph.D. student, advised by Krishnendu Chatterjee.

His research concerns reliable decision-making under uncertainty: how an agent should act when its environment is stochastic, partially observable, adversarial, or not known exactly. He studies these questions on stochastic games, Markov decision processes (MDPs), partially observable MDPs, and robust MDPs, along three lines: the mathematical structure of optimal strategies, equilibria, and long-run objectives; efficient algorithms and tight complexity classifications for the associated decision problems; and robust decision-making under incomplete or misspecified models. The common aim is to design agents that come with provable guarantees despite incomplete information, changing environments, and strategic interaction.

The results of this thesis appeared at the 44th and 45th IARCS Annual Conferences on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2024 and FSTTCS 2025). His further work on games, MDPs, and partially observable models has appeared at LICS, COLT, AAI, UAI, CONCUR, PLDI, and ATVA.

List of Collaborators and Publications

This thesis reuses the following two publications. Following the convention of theoretical computer science, the authors of both papers are listed in alphabetical order and contributed equally to the work.

Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Jakub Svoboda. “Concurrent Stochastic Games with Stateful-Discounted and Parity Objectives: Complexity and Algorithms.” In: *44th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2024*. Vol. 323. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, 5:1–5:17. doi: 10.4230/LIPIcs.FSTTCS.2024.5

Reused in full in Part I, that is, in Chapters 4 to 6.

Ali Asadi, Léonard Brice, Krishnendu Chatterjee, and K. S. Thejaswini. “ ε -Stationary Nash Equilibria in Multi-Player Stochastic Graph Games.” In: *45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2025, BITS Pilani, K K Birla Goa Campus, India, December 17-19, 2025*. Vol. 360. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, 9:1–9:17. doi: 10.4230/LIPIcs.FSTTCS.2025.9

Reused and extended in Part II, that is, in Chapters 7 to 9. An extended preprint is available as [11].

The unified introduction, preliminaries, overview of results, and conclusion were written for this thesis.

Table of Contents

Abstract	i
Acknowledgements	iii
About the Author	iv
List of Collaborators and Publications	v
Table of Contents	vii
List of Figures	ix
List of Tables	ix
List of Algorithms	ix
List of Major Symbols	xi
1 Introduction	1
2 Preliminaries	9
2.1 Basic Notations	9
2.2 Concurrent Stochastic Games	10
2.3 Multi-player Turn-based Stochastic Graph Games	13
2.4 Computational Problems and Complexity Classes	14
2.5 The Existential Theory of the Reals	16
2.6 Floating-point Representation	17
3 Overview of Results	21
3.1 Previous Results for Value Approximation	21
3.2 Our Results for Value Approximation	22
3.3 Previous Results for Constrained Equilibria	23
3.4 Our Results for Constrained Equilibria	23
3.5 Roadmap of the Proofs	24

I	Concurrent Stochastic Games with Stateful-discounted and Parity Objectives	27
4	Mathematical Properties of Stateful-discounted and Limit Values	29
4.1	Notations and Selected Results from Literature	29
4.2	Bounds on Roots of Polynomials with Integer Coefficients	30
4.3	Characterization of Stateful-discounted Value	33
4.4	Characterization of Limit Value	37
4.5	Approximation of Limit Value	39
5	Algorithms for LIMITVALUE and PARITYVALUE	43
5.1	Selected Algorithms from Literature	43
5.2	Algorithm for Approximate Stateful-discounted Value	43
5.3	Algorithms for Approximate Limit and Parity Values	45
6	Complexity of LIMITVALUE and PARITYVALUE	47
6.1	Stateful-discounted Value Approximation in MCs	47
6.2	Stateful-discounted Value Approximation in MDPs	49
6.3	Stateful-discounted, Limit, and Parity Values Approximation in CSGs	50
II	Constrained Approximate Nash Equilibria in Multi-player Stochastic Graph Games	55
7	Computing Constrained Approximate Nash Equilibria	57
7.1	Double-exponentially Small Probabilities Suffice	58
7.2	Floating-point Representation of Approximate Equilibria	59
7.3	The FNP ^{NP} Procedure	60
8	On (Very) Small Probabilities	63
8.1	The Game $\mathcal{G}^n$	63
8.2	Small Probabilities are Necessary	64
8.3	Small Probabilities Suffice	66
8.4	Approximate Equilibria in the Game $\mathcal{G}^n$	67
9	A Lower Bound: NP-hardness	69
9.1	The Reduction	69
9.2	From Satisfying Assignments to Nash Equilibria	71
9.3	From Approximate Nash Equilibria to Satisfying Assignments	71
10	Conclusion	75
	References	77
A	Declaration of the Use of Generative AI and AI Tools	83

List of Figures

8.1	A game where very small probabilities are necessary	64
9.1	A reduction from 3SAT	70

List of Tables

1.1	Complexity upper bounds for value approximation	5
1.2	Algorithmic upper bounds for value approximation	6
1.3	Complexity of the constrained existence of Nash equilibria	6

List of Algorithms

1	APPROXDISCOUNTED	44
2	APPROXLIMIT	45

List of Major Symbols

The following symbols are used throughout the thesis. The symbols of the first group are common to both parts; the symbols of the second group are specific to the first part, which considers concurrent stochastic games (Part I); and the symbols of the third group are specific to the second part, which considers multi-player turn-based stochastic graph games (Part II).

Common symbols

n	Number of states of a game, resp. of vertices of a graph game
B	Bit-size of the numbers of the input
ε	Additive error
ℓ	Precision of the floating-point representation
$\mathbb{P}, \mathbb{E}$	Probability measure and its expectation
$\Delta(\mathcal{X})$	Set of probability distributions over the set $\mathcal{X}$
$\mathbf{d}$	A probability distribution
$\text{supp}(\mathbf{d})$	Support of a probability distribution
$\text{bit}(\cdot)$	Bit-size of a number
$\text{rel}(\cdot, \cdot)$	Relative distance of two numbers
$\mathcal{F}(\ell)$	Set of floating-point numbers with precision ℓ
$\mathcal{D}(\ell)$	Set of floating-point probability distributions with precision ℓ
$\oplus, \ominus, \otimes, \odot$	Finite precision arithmetic operations
$\mathcal{P}$	Markov decision process (MDP)
$\mathcal{C}$	Markov chain (MC)
val	Value, resp. maximal expected payoff
Reach_T	Reachability objective for the target set T
$\mathcal{P}, \mathcal{Q}$	Sets of polynomials
D	Degree of a polynomial

Concurrent stochastic games

G	Concurrent stochastic game (CSG)
$\mathcal{S}, s$	Set of states and a state
$\mathcal{A}, \mathcal{B}$	Sets of actions of Player 1 and Player 2
a, b	Actions of Player 1 and Player 2
m	Number of actions per player and state
δ	Transition probability function
Ω, ω	Set of plays and a play
σ, τ	Strategies of Player 1 and Player 2
Σ, Γ	Sets of strategies of Player 1 and Player 2
G_σ, G_τ	MDPs induced by fixing a strategy
$G_{\sigma, \tau}$	MC induced by fixing both strategies
d	Number of discount factors, resp. index of the parity objective
λ	Discount factor
Λ	Discount function
χ	Assignment function of the discount factors to the states
r	Reward function
p	Priority function
Disc_Λ	Stateful-discounted objective
Parity_p	Parity objective
val_Λ	Stateful-discounted value
val_χ	Limit value of the stateful-discounted objectives
val_p	Parity value
$\nu^{\sigma, \tau}$	Stateful-discounted payoff of a pair of strategies
∇_Λ	Determinant function for stationary strategies
$W_\Lambda^s(z)$	Auxiliary matrix game
$F_\chi^s(z)$	Limit function
Id	Identity matrix

Multi-player turn-based stochastic graph games

$\mathcal{G}$	Game
V, E	Sets of vertices and edges of the underlying graph
Π	Set of players
V_i	Set of vertices controlled by player i
$V_?$	Set of stochastic vertices
$\mathbf{p}$	Probability function of the stochastic vertices
T	Set of terminal vertices
μ, μ_i	Payoff function, and payoff function of player i
v_0	Initial vertex
σ_i	Stationary strategy of player i
$\bar{\sigma}, \bar{\tau}$	Strategy profiles
$\bar{\sigma}_{-i}$	Strategy profile of all players except player i
$\bar{x}, \bar{y}$	Threshold vectors of the payoff constraints

Chapter 1

Introduction

In this thesis, we present improved complexity results and algorithms for the approximate computation of the two central solution concepts of stochastic games played on graphs: the *value* in two-player zero-sum games, and *Nash equilibria* in multi-player games. Below we present the models of stochastic games, the relevant objectives, the solution concepts, the computational problems, previous results, and finally our contributions.

Stochastic games. Stochastic games are games played on finite-state graphs for an infinite number of steps, where the transitions are governed by a probabilistic transition function. These games were introduced in the seminal work of Shapley [47] and are a fundamental model in game theory. We consider the following two models in this thesis.

- *Concurrent stochastic games.* In each step of a concurrent stochastic game, both players simultaneously and independently of the other player choose an action. Given the current state and the chosen actions, the next state is obtained according to a stochastic transition function. An infinite number of such steps results in a play, which is an infinite sequence of states and actions. Concurrent stochastic games have been widely studied in the literature from the mathematical perspective [28, 29, 43, 47], and from the algorithmic and computational complexity perspective, including: complexity for reachability objectives [3, 27, 31, 37], algorithms for limit-average objectives [36, 45], complexity for qualitative solutions for ω -regular objectives [2], complexity for quantitative solutions for ω -regular objectives [8, 22], and in the context of temporal logic [9]. In particular, in the analysis of reactive systems, concurrent games provide the appropriate model for reactive systems with components that interact synchronously [5, 6, 9].
- *Multi-player turn-based stochastic graph games.* In a turn-based stochastic graph game, every vertex is either controlled by one of the players or is a stochastic vertex. In each step, the player who controls the current vertex chooses the successor vertex, and in a stochastic vertex the successor is chosen according to a probability distribution that is part of the game. Every player has their own payoff, and thus the game is not zero-sum in general. Such games provide a modelling framework for a wide range of domains, including epidemic processes [40], formal verification [30], learning theory [1], cyber-physical systems [49], distributed and probabilistic programs [7], and probabilistic

planning [51]. They have been extensively studied, with several results characterizing their computational complexity [24, 54], and others studying restrictions and payoff functions that yield tractable fragments [20, 54].

Objectives and payoffs. An objective is a measurable function that assigns to every play a real-valued reward. In the first part of the thesis, we consider the following two objectives. The classic discounted-sum objective is as follows: every transition is assigned a reward and the objective assigns to a play the discounted-sum of the rewards. While the classic objective has a single discount factor, the *stateful-discounted* objective has multiple discount factors. In the stateful-discounted objective, each state is associated with a discount factor, and, in the objective, the discount at a step depends on the current state. We also consider the boolean *parity* objectives, which are a canonical form to express all ω -regular objectives [52], which provide a robust specification for all properties that arise in verification. For example, all LTL formulas can be converted to deterministic parity automata. In parity objectives, every state is associated with an integer priority, and a play is winning for (or satisfies) the objective if the minimum priority visited infinitely often is even. In the second part of the thesis, we consider *terminal payoffs*: every terminal vertex of the game is associated with a payoff for each player, and a play that reaches a terminal vertex assigns to every player the corresponding payoff. Since the payoffs of the players are unrelated, this simple quantitative objective already captures the interaction of agents with individual goals.

Strategies. Strategies are recipes that define the choice of the players. In general, they are functions that, given a game history, return a probability distribution over the available choices. A *stationary* strategy depends on the history only through the current state (or vertex), i.e., it uses randomness but no memory. For zero-sum concurrent stochastic games, stationary strategies suffice for stateful-discounted objectives, whereas for parity objectives even ε -optimal strategies require infinite memory [2]. For multi-player games, the memory of the players is a critical parameter. If the players are allowed unbounded memory, then the basic decision problems about Nash equilibria are undecidable [54]. Moreover, from a practical perspective, strategies that require large memory lead to complex controllers and are harder to implement. Since finite-memory strategies can be encoded in the state space of the game, it suffices, in many cases, to consider stationary strategies, which require no memory at all. Thus, in the second part of the thesis, we restrict all players to stationary strategies.

Values and equilibria. Given a two-player zero-sum concurrent stochastic game and an objective, the *value* of Player 1 at a state is the maximal expectation that the player can guarantee for the objective against all strategies of Player 2. For stateful-discounted objectives, given an ordering of the discount factors, the *limit value* at a state is the limit of the value function for the discounted objective as the discount factors approach zero in the given order. In multi-player games, the classic solution concept is the *Nash equilibrium* [44]: a strategy profile is a Nash equilibrium if no player can unilaterally deviate to achieve a strictly better payoff. A strategy profile is an ε -*Nash equilibrium* if no player can gain more than ε by unilaterally deviating from their strategy. While Nash equilibria always exist in the games we consider, the equilibria that arise in the existence proofs are often adversarial in nature and offer poor outcomes for all players [24, 53]. Hence, the more compelling question is whether

one can decide, and compute if possible, a Nash equilibrium whose payoff vector satisfies given constraints.

Computational problems. The computational problems we consider in this thesis are as follows. For two-player zero-sum concurrent stochastic games, the main computational problems are: (a) the *value-decision* problem, given a state and a threshold α , asks whether the value at the state is at least α ; and (b) the *value-approximation* problem, given a state and an error $\varepsilon > 0$, asks to compute an approximation of the value for the state within an additive error of ε . We consider the above problems for the limit value of stateful-discounted objectives and the value for parity objectives. For multi-player turn-based stochastic games, the classic problem is the *constrained existence* problem, which, given a game and two threshold vectors, asks whether there exists a Nash equilibrium in which the payoff of every player lies in the given interval. We consider the following approximate variant of this problem: given a game that is guaranteed to have a Nash equilibrium satisfying the constraints, compute a strategy profile that is an ε -Nash equilibrium and satisfies the same constraints up to an additive error ε . The decision version of this problem is a promise problem: it asks to distinguish the case that there is a Nash equilibrium satisfying the constraints from the case that there is not even an ε -Nash equilibrium satisfying the constraints up to ε . Instances that fall outside these two cases are considered indeterminate, and the output may be arbitrary in those cases.

Motivation. The motivation to study the limit of the stateful-discounted objective is as follows. First, this limit generalizes the classic limit-average objectives. Second, it characterizes the value for the parity objectives in concurrent stochastic games [4, 34], where the order of limit corresponds to the order of importance of priorities in parity objectives. Third, the limit value has been shown to correspond to the value for other objectives such as priority mean-payoff for various subclasses of concurrent stochastic games [33]. The motivation to study approximate solutions, as opposed to exact ones, is that in both settings the exact problems are intractable, and the intractability is a consequence of the numbers involved rather than of the combinatorial structure of the games. We elaborate on this point below.

- *Values.* For concurrent games, even for special classes of objectives such as reachability and safety, values can be irrational, and the decision problem related to exact value is SQRT – SUM hard [27]. The SQRT – SUM problem is a classic problem in computational geometry, and whether SQRT – SUM belongs to NP has been a long-standing open problem.
- *Equilibria.* For multi-player games, the constrained existence problem is undecidable when the players are allowed unbounded memory: it is already known that checking whether the payoff of a player can lie in a given interval is undecidable with just ten agents [54, Theorem 4.9]. Restricting players to deterministic (pure) strategies does not help, as the problem remains undecidable. When the players are restricted to stationary strategies, the problem becomes decidable, but it is $\exists\mathbb{R}$ -complete [38], where $\exists\mathbb{R}$ is the class of problems that can be reduced to the satisfiability of a formula in the existential theory of the reals. Moreover, for the functional version of the problem, the probabilities of the stationary strategies can be irrational [54, Theorem 4.6], even for games with just four players. Such strategies cannot be executed in real-world scenarios, since it is hard to simulate choosing an edge with an irrational probability.

In addition, the relaxation of the equilibrium notion is well-motivated on its own. The standard definition of Nash equilibrium requires that no player can improve their payoff by any non-zero amount by unilaterally deviating from their strategy, even if the improvement is arbitrarily small [44]. This strict notion of optimality contributes significantly to the computational complexity of deciding whether a constrained Nash equilibrium exists. From a practical perspective, however, it is reasonable to assume that players are unlikely to deviate from their current strategies for only negligible gains. Hence, ε -Nash equilibria are a well-studied relaxation of the classic notion [25, 41].

Previous results for values. For a single discount factor, the limit value corresponds to the value of the well-studied mean-payoff or long-run average objectives [43], and, for parity objectives, the computational problems admit a linear reduction to the limit value of stateful-discounted objectives [4, 34]. The value-decision problem for concurrent stochastic games is SQRT – SUM hard [27]: this result holds for reachability objectives, and hence also for parity objectives and the limit value for even a single discount factor. The complexity upper bounds for the value-approximation problem of concurrent stochastic games are as follows: (a) EXPSpace for the limit value of stateful-discounted objectives; and (b) PSPACE for parity objectives [21, 22]. The above result for the limit value follows from a reduction to the theory of reals, where the number of discount factors corresponds to the number of quantifier alternations. For the special class of reachability objectives, the complexity upper bound of TFNP^{NP} for the value-approximation problem has been established in [31], where TFNP^{NP} is the total functional form of the second level of the polynomial hierarchy. The result of [31] has been recently extended to limit-average objectives (which correspond to the limit value of a single discount factor) [19]. To the best of our knowledge, the above complexity upper bounds are the best bounds for the limit value of general stateful-discounted objectives and parity objectives. The best-known algorithms for the value-approximation problem are as follows: (a) double exponential time for the limit value of stateful-discounted objectives; and (b) exponential time for parity objectives, where the exponent is a product that depends at least linearly on the number of states and actions [21, 22] (see Chapter 3 for further details). While iterative approaches are desirable, they neither exist for parity objectives nor guarantee efficiency even in special cases. For example, for reachability and safety objectives, iterative approaches like value-iteration or strategy-iteration have a double-exponential lower bound [35]; and, for parity objectives, iterative approaches like strategy iteration are not known, as strategies require infinite memory [2].

Previous results for equilibria. The constrained existence problem of Nash equilibria in multi-player turn-based stochastic games is undecidable for general strategies [54, Theorem 4.9], and, for stationary strategies, it is NP-hard [54, Theorem 4.4] and was later shown to be $\exists\mathbb{R}$ -complete [38]. The $\exists\mathbb{R}$ -completeness result requires seven players. For approximate solutions, the work of Deligkas, Fearnley, Melissourgos, and Spirakis [26] observed that the class $\varepsilon\text{-}\exists\mathbb{R}$, which captures the complexity of deciding whether approximate solutions exist to systems of real-valued constraints, is polynomial-time reducible to the class $\exists\mathbb{R}$ itself. This equivalence in the complexity of $\exists\mathbb{R}$ and of its approximate version hinges on the fact that solutions to such existential theory of the reals problems can involve values that are double-exponentially large. To address this, they identified a restricted fragment of $\exists\mathbb{R}$ for which ε -approximate solutions can be computed by a quasi-polynomial time approximation

scheme. In a related direction, Bose, Ibsen-Jensen, and Totzke [19] studied ε -Nash equilibria in concurrent and partial-observation settings, also using approximation-based techniques. Their work differs from ours in two key ways. First, they do not address the constrained existence problem for ε -Nash equilibria; the unconstrained variant of the problem is significantly easier in terms of computational complexity. Second, they assume that the number of players is fixed, whereas in our setting the number of players is part of the input. To the best of our knowledge, the results of this thesis are the first analysis of constrained ε -Nash equilibria in multi-player games where the number of players is not fixed. Finally, the approximation techniques that we use have appeared in various forms in the literature, particularly in the context of approximating values in concurrent games. Frederiksen and Miltersen [31] introduced floating-point approximation methods to compute the values of concurrent reachability games up to arbitrary precision, and placed the problem in the complexity class TFNP^{NP} . For that class of games, however, no hardness results such as the NP-hardness we establish here were previously known.

Our contributions. In this thesis, our main contributions are as follows. For the value-approximation problem in concurrent stochastic games: (a) we establish TFNP^{NP} upper bounds, both for the limit value of stateful-discounted objectives and for parity objectives; and (b) we present algorithms which are exponential time and improve the dependency on the number of actions in the exponent from linear to logarithmic. In particular, if the number of states is constant, our algorithms run in polynomial time. The comparison of previous results and our results is summarized in Tables 1.1 and 1.2. For the approximate constrained existence problem of Nash equilibria in multi-player turn-based stochastic games: (a) we present an FNP^{NP} algorithm for the functional problem, which implies an NP^{NP} upper bound for its decision version; (b) we establish that the decision version is NP-hard, already for games with two players; and (c) we present a game with five players in which every constrained Nash equilibrium requires double-exponentially small probabilities, which shows a limit of the techniques we use. The comparison of previous results and our results is summarized in Table 1.3.

	Complexity	
	Previous	Ours
Limit	EXPSpace (Theory of reals)	TFNP^{NP} (Theorem 3.2-Item 1, Theorem 3.3-Item 1)
Parity	PSPACE [21, 22]	

Table 1.1: Complexity upper bounds of the value-approximation in concurrent stochastic games for the limit value of stateful-discounted objectives and parity objectives.

Technical contributions. Our main technical contributions are as follows. For the value-approximation problem, we first present a bound on the roots of multi-variate polynomials with integer coefficients (Section 4.2). Given the bounds on roots of polynomials, we establish new characterizations for the limit and stateful-discounted values (Section 4.3 and Section 4.4), which lead to an approximation of the limit value by the stateful-discounted value when the discount factors are double-exponentially small (Section 4.5). Given this connection, we establish the improved complexities and algorithms for the value-approximation for the limit value of stateful-discounted objectives and parity objectives in Chapter 5 and

	Algorithms	
	Previous	Ours
Limit	$\exp(\mathcal{O}(2^d m^2 n + \log(1/\varepsilon) + \log(B)))$ (Theory of reals)	$\exp\left(\mathcal{O}\left(\begin{matrix} nd \log(m) + \log(B) \\ + \log(\log(1/\varepsilon)) \end{matrix}\right)\right)$ (Theorem 3.2-Item 2, Theorem 3.3-Item 2)
Parity	$\exp\left(\mathcal{O}\left(\begin{matrix} mn + d \log(n) + \log(B) \\ + \log(\log(1/\varepsilon)) \end{matrix}\right)\right)$ [21, 22]	

Table 1.2: Algorithmic upper bounds of the value-approximation in concurrent stochastic games for the limit value of stateful-discounted objectives and parity objectives, where n is the number of states, m is the number of actions, d is the number of discount factors/parity index, B is the bit-size of numbers in the input, ε is the additive error, and $\exp$ is the exponential function.

	Upper bound	Lower bound
Exact	$\exists \mathbb{R}$ [38]	$\exists \mathbb{R}$ -hard [38] (7 players)
Approximate	$\text{FNP}^{\text{NP}}, \text{NP}^{\text{NP}}$ (Theorem 3.5)	NP-hard (Theorem 3.6, 2 players)

Table 1.3: Complexity of the constrained existence problem of Nash equilibria in multi-player turn-based stochastic games, where the players are restricted to stationary strategies. The approximate problem asks for an ε -Nash equilibrium that satisfies the constraints up to an additive error ε , under the promise that an exact constrained Nash equilibrium exists.

Chapter 6. For the approximate constrained existence problem of Nash equilibria, our approach relies on showing that if a constrained stationary equilibrium exists, then there is one in which the probability values used in the strategies are not too small—specifically, they are lower-bounded by the inverse of a doubly exponential function of the input size (Section 7.1). This structural property is established by an explicit encoding of the problem in the existential theory of the reals, along with the bounds on the solutions of systems of polynomial inequalities of Basu, Pollack, and Roy [18], in the spirit of the work of Hansen, Koucký, and Miltersen on zero-sum concurrent reachability games [37]. This bound alone is insufficient, because representing such tiny probabilities explicitly requires exponentially many bits in fixed-point representation. Hence, we show that if a constrained Nash equilibrium exists, then there is also an ε -Nash equilibrium that can be encoded with floating-point numbers of polynomial size (Section 7.2), and we present a coNP procedure that verifies that no player has a profitable deviation (Section 7.3).

A common toolkit. Although the two solution concepts we study are different, the technical approach that leads to the improved complexity bounds is common to both parts of the thesis, and consists of the following three ingredients.

- *Bounds from real algebraic geometry.* In both settings, the numbers that arise in a solution can be doubly exponentially small, and in both settings the required bounds are obtained by writing the problem as a system of polynomial inequalities with integer coefficients. For values, we require a new bound on the region around the origin where a multi-variate polynomial cannot vanish (Lemma 4.4); for equilibria, we require the bound of [18] on the radius of a ball that intersects every cell of the partition induced by a set of polynomials (Lemma 2.8).
- *Floating-point representation.* Since the numbers involved are doubly exponentially small, the fixed-point representation requires exponentially many bits, and hence it does not lead to polynomial-size certificates. In both parts, we use the floating-point representation of numbers and probability distributions, along with the continuity of

the value of Markov chains with respect to the relative distance of the transition probabilities [50], to establish that approximate solutions admit polynomial-size certificates (Section 2.6).

- *Guess-and-verify with an NP oracle.* In both parts, the approximate solution is obtained by guessing a polynomial-size certificate and verifying it, where the verification requires the resolution of single-player games, i.e., Markov decision processes, whose transition probabilities are represented in floating point. This resolution is achieved by an oracle call to NP, which places the value-approximation problem in TFNP^{NP} and the approximate constrained existence problem in FNP^{NP} .

Organization of the thesis. The rest of this thesis is organized as follows. In Chapter 2, we present the definitions of the two models of stochastic games, the objectives, the solution concepts, the computational problems, and the selected results from the literature that both parts of the thesis rely on. In Chapter 3, we present the previous results and the formal statements of our results. The two parts of the thesis are then dedicated to the proofs of these results. The first part considers the value-approximation problem for concurrent stochastic games: in Chapter 4, we establish the mathematical properties of the stateful-discounted and limit values; in Chapter 5, we present the algorithms; and in Chapter 6, we establish the TFNP^{NP} upper bounds. The second part considers the approximate constrained existence problem of Nash equilibria: in Chapter 7, we present the FNP^{NP} algorithm; in Chapter 8, we present the game with five players that requires double-exponentially small probabilities; and in Chapter 9, we establish the NP-hardness of the decision version. Finally, in Chapter 10, we conclude with a discussion of open problems.

Publications. This thesis is based on the following two publications. The results of the first part appeared in the work of Asadi, Chatterjee, Saona, and Svoboda [12], and the results of the second part are based on the work of Asadi, Brice, Chatterjee, and Thejaswini [10]; an extended preprint is available as [11].

Chapter 2

Preliminaries

We present the definitions and the results from the literature that are used throughout the thesis. The chapter is organized as follows. In Section 2.1, we present basic notations. In Section 2.2, we present concurrent stochastic games, which is the model of the first part of the thesis, along with the objectives, the strategies, and the values. In Section 2.3, we present multi-player turn-based stochastic graph games, which is the model of the second part of the thesis, along with the notions of Nash equilibria. In Section 2.4, we present the computational problems we consider and the relevant complexity classes. Finally, in Section 2.5 and Section 2.6, we present the selected results from the literature on real algebraic geometry and on the floating-point representation of numbers, which both parts of the thesis rely on.

2.1 Basic Notations

Sets and distributions. Given a finite set $\mathcal{X}$, a probability distribution over $\mathcal{X}$ is a function $d: \mathcal{X} \rightarrow [0, 1]$ such that $\sum_{x \in \mathcal{X}} d(x) = 1$. The set of all probability distributions over $\mathcal{X}$ is denoted by $\Delta(\mathcal{X})$. For $d \in \Delta(\mathcal{X})$, the support of d is defined as $\text{supp}(d) := \{x \mid d(x) > 0\}$. For a positive integer k , the set of positive integers smaller than or equal to k is defined as $[k] := \{1, \dots, k\}$. Given a real x , we denote 2^x by $\exp(x)$.

Numbers and their representation. Integers are always implicitly assumed to be represented in binary. Given a positive integer k , we define $\text{bit}(k) := \lceil \log_2(k + 1) \rceil$, which is the number of bits required to write k . For a rational k_1/k_2 , we define $\text{bit}(k_1/k_2) := \text{bit}(k_1) + \text{bit}(k_2)$. Given a real x , the sign function is

$$\text{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Moreover, we use the classic arithmetic with infinity, i.e., $x + \infty = \infty$ and $x - \infty = -\infty$. Throughout the thesis, we denote by B the bit-size of the numbers that appear in the input.

Graphs. A directed graph (V, E) consists of a set V of *vertices* and a set E of ordered pairs of vertices, called *edges*. For simplicity, we often write uv for an edge $(u, v) \in E$. For a vertex u , we denote by $E(u) := E \cap (\{u\} \times V)$ the set of outgoing edges of u , and, for a set $X \subseteq V$ of

vertices, we denote by $E(X) := \bigcup_{u \in X} E(u)$. A *path* in the directed graph (V, E) is a (finite or infinite) sequence of vertices from V such that every pair of two consecutive elements is an edge. A *cycle* is a path with distinct vertices such that the last element of the sequence and the first one also form an edge.

Probability spaces. Given a (finite or infinite) set of outcomes Ω and a probability measure $\mathbb{P}$ over Ω , let X be a random variable over Ω , i.e., a mapping $X: \Omega \rightarrow \mathbb{R}$. We write $\mathbb{E}^\mathbb{P}[X]$, or simply $\mathbb{E}[X]$, for the expected value of X , when it is defined.

2.2 Concurrent Stochastic Games

In this section, we present the model of the first part of the thesis, and follow the standard definitions related to concurrent stochastic games.

Concurrent stochastic games. A concurrent stochastic game (CSG) is a two-player finite game $G = (\mathcal{S}, \mathcal{A}, \mathcal{B}, \delta)$ consisting of

- the set of states $\mathcal{S}$, of size n ;
- the sets of actions for each player $\mathcal{A}$ and $\mathcal{B}$, with at most m actions; and
- the stochastic transition function $\delta: \mathcal{S} \times \mathcal{A} \times \mathcal{B} \rightarrow \Delta(\mathcal{S})$.

Steps. Given an initial state $s \in \mathcal{S}$, the game proceeds as follows. In each step, both players choose an action simultaneously, $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Based on both actions (a, b) and current state s , the next state is drawn according to the probability distribution $\delta(s, a, b)$.

Histories and plays. At step k of a CSG, each player possesses information in the form of the finite sequence of the states visited and the actions chosen by both players. A k -history $\omega^{(k)} = \langle s_0, a_0, b_0, s_1, a_1, b_1, \dots, s_k \rangle$ is a finite sequence of states and actions such that, for all steps $0 \leq t < k$, we have $s_{t+1} \in \text{supp}(\delta(s_t, a_t, b_t))$. The set of all k -histories is denoted by $\Omega^{(k)}$. Similarly, a play $\omega = \langle s_0, a_0, b_0, s_1, a_1, b_1, \dots \rangle$ is an infinite sequence of states and actions such that, for all steps $t \geq 0$, we have $s_{t+1} \in \text{supp}(\delta(s_t, a_t, b_t))$. The set of all plays is denoted by Ω . For any state s , the set of all plays starting at s , i.e., $\omega = \langle s_0, a_0, b_0, \dots \rangle$ where $s_0 = s$, is denoted by Ω_s .

Objectives. An objective is a measurable function that assigns a real number to all plays. Player 1 aims to maximize the expectation of the objective, while Player 2 minimizes it. We consider the following objectives.

- *Parity objective.* Given a priority function $p: \mathcal{S} \rightarrow \{0, \dots, d\}$ with d as its index, the *parity* objective is an indicator of the even parity condition on minimal priority visited infinitely often in plays. More formally, we define $\text{Parity}_p: \Omega \rightarrow \{0, 1\}$ as

$$\text{Parity}_p(\omega) := \begin{cases} 1 & \min\{p(s) \mid \forall i \geq 0 \exists j \geq i \ s_j = s\} \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

- **Stateful-discounted objective.** Consider d discount factors $\lambda_1, \dots, \lambda_d \in (0, 1]$. Given an assignment function $\chi: \mathcal{S} \rightarrow [d]$, we define the discount function $\Lambda: \mathcal{S} \rightarrow \{\lambda_1, \dots, \lambda_d\}$ as $\Lambda(s) := \lambda_{\chi(s)}$ for all states $s \in \mathcal{S}$. Given a reward function $r: \mathcal{S} \times \mathcal{A} \times \mathcal{B} \rightarrow [0, 1]$ that assigns a reward value $r(s, a, b)$ for all (s, a, b) , the stateful-discounted objective $\text{Disc}_\Lambda: \Omega \rightarrow [0, 1]$ is defined as, for all $\omega = \langle s_0, a_0, b_0, \dots \rangle$,

$$\text{Disc}_\Lambda(\omega) := \sum_{i \geq 0} \left(r(s_i, a_i, b_i) \Lambda(s_i) \prod_{j < i} (1 - \Lambda(s_j)) \right).$$

Strategies. A strategy is a function that assigns a probability distribution over actions to every finite history and is denoted by $\sigma: \bigcup_k \Omega^{(k)} \rightarrow \Delta(\mathcal{A})$ for Player 1 (resp. $\tau: \bigcup_k \Omega^{(k)} \rightarrow \Delta(\mathcal{B})$ for Player 2). Given strategies σ and τ , the game proceeds as follows. At step k , the current history is some $\omega^{(k)} \in \Omega^{(k)}$. Player 1 (resp. Player 2) chooses an action according to the distribution $\sigma(\omega^{(k)})$ (resp. $\tau(\omega^{(k)})$). The set of all strategies for Player 1 and Player 2 is denoted by Σ and Γ respectively. A *stationary* strategy depends on the past observations only through the current state. A stationary strategy for Player 1 (resp. Player 2) is denoted by $\sigma: \mathcal{S} \rightarrow \Delta(\mathcal{A})$ (resp. $\tau: \mathcal{S} \rightarrow \Delta(\mathcal{B})$). The set of all stationary strategies for Player 1 and Player 2 is denoted by Σ^S and Γ^S respectively. A *pure stationary* strategy $\sigma: \mathcal{S} \rightarrow \mathcal{A}$ (resp. $\tau: \mathcal{S} \rightarrow \mathcal{B}$) for Player 1 (resp. Player 2) is a stationary strategy that maps to Dirac distributions only. The set of all pure stationary strategies for Player 1 and Player 2 is denoted by Σ^{PS} and Γ^{PS} respectively.

Probability space. An initial state s and a pair of strategies (σ, τ) induce a unique probability over Ω_s , endowed with the sigma-algebra generated by the cylinders corresponding to finite histories. We denote by $\mathbb{P}_s^{\sigma, \tau}$ and $\mathbb{E}_s^{\sigma, \tau}$ the probability and the expectation respectively.

We state the determinacy for CSGs with stateful-discounted and parity objectives.

Theorem 2.1 (Parity determinacy [42]). *For all CSGs, states s , and priority functions p ,*

$$\sup_{\sigma \in \Sigma} \inf_{\tau \in \Gamma} \mathbb{E}_s^{\sigma, \tau}[\text{Parity}_p] = \inf_{\tau \in \Gamma} \sup_{\sigma \in \Sigma} \mathbb{E}_s^{\sigma, \tau}[\text{Parity}_p].$$

Theorem 2.2 (Stateful-discounted determinacy [47]). *For all CSGs, states s , reward functions, and discount functions Λ , we have*

$$\sup_{\sigma \in \Sigma^S} \inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma, \tau}[\text{Disc}_\Lambda] = \inf_{\tau \in \Gamma^S} \sup_{\sigma \in \Sigma^S} \mathbb{E}_s^{\sigma, \tau}[\text{Disc}_\Lambda].$$

Values. The above determinacy results imply that switching the quantification order of strategies does not make a difference and leads to the unique notion of value. The stateful-discounted value for a state s is defined as

$$\text{val}_\Lambda(s) := \sup_{\sigma \in \Sigma^S} \inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma, \tau}[\text{Disc}_\Lambda].$$

We define the parity value $\text{val}_p(s)$ for a state s analogously. The limit value for a state s is defined as

$$\text{val}_\chi(s) := \lim_{\lambda_1 \rightarrow 0^+} \cdots \lim_{\lambda_d \rightarrow 0^+} \text{val}_\Lambda(s).$$

ε -optimal strategies. Given $\varepsilon \geq 0$, a strategy σ for Player 1 is ε -optimal for the stateful-discounted objective if, for all states $s \in \mathcal{S}$, we have

$$\inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma, \tau}[\text{Disc}_\Lambda] \geq \text{val}_\Lambda(s) - \varepsilon.$$

We say the strategy is *optimal* if $\varepsilon = 0$. The notion of ε -optimal strategies for Player 2 is defined analogously. Similarly, we define ε -optimal strategies for the parity objectives.

Markov decision processes and Markov chains. For $i \in \{1, 2\}$, a Player- i Markov decision process (Player- i MDP) is a special class of CSGs where the other player has only one action and is denoted by $\mathcal{P} = (\mathcal{S}, \mathcal{A}, \delta: \mathcal{S} \times \mathcal{A} \rightarrow \Delta(\mathcal{S}))$. A Markov chain (MC) is a special class of MDPs where both players have only one action and is denoted by $\mathcal{C} = (\mathcal{S}, \delta: \mathcal{S} \rightarrow \Delta(\mathcal{S}))$. In Markov chains we write $\delta(s, s')$ to denote $\delta(s)(s')$.

Absorbing MCs. We say an MC $\mathcal{C}$ is *absorbing* if there exists a subset of absorbing states $\mathcal{S}_0 \subseteq \mathcal{S}$ such that

- for all $s \in \mathcal{S}_0$, we have $\delta(s, s) = 1$; and
- for all $s_0 \in \mathcal{S} \setminus \mathcal{S}_0$, there exist states $s_1, \dots, s_k$ such that $\delta(s_i, s_{i+1}) > 0$ and $s_k \in \mathcal{S}_0$.

States in $\mathcal{S}_0$ are called absorbing.

MDPs and MCs given stationary strategies in CSGs. Given a stationary strategy σ for Player 1 in a game G , by fixing the strategy σ , we obtain a Player-2 MDP $G_\sigma = (\mathcal{S}, \mathcal{B}, \delta_\sigma)$ where the transition function $\delta_\sigma: \mathcal{S} \times \mathcal{B} \rightarrow \Delta(\mathcal{S})$ is given by

$$\delta_\sigma(s, b)(s') := \sum_{a \in \mathcal{A}} \delta(s, a, b)(s') \cdot \sigma(s)(a),$$

for all $s, s' \in \mathcal{S}$ and $b \in \mathcal{B}$. Analogously, we obtain a Player-1 MDP G_τ by fixing a stationary strategy τ for Player 2. Moreover, by fixing stationary strategies σ and τ for both players, we obtain an MC $G_{\sigma, \tau} = (\mathcal{S}, \delta_{\sigma, \tau})$, where the transition function $\delta_{\sigma, \tau}: \mathcal{S} \rightarrow \Delta(\mathcal{S})$ is given by

$$\delta_{\sigma, \tau}(s)(s') = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \delta(s, a, b)(s') \cdot \sigma(s)(a) \cdot \tau(s)(b),$$

for all $s, s' \in \mathcal{S}$.

Reachability objectives in MCs. Given an MC $\mathcal{C}$ and a target set $T \subseteq \mathcal{S}$, the reachability objective is the indicator function of plays eventually reaching T . More formally, for a play $\omega = \langle s_0, s_1, \dots \rangle$, we define $\text{Reach}_T: \Omega \rightarrow \{0, 1\}$ as

$$\text{Reach}_T(\omega) := \begin{cases} 1 & \exists i \geq 0 \ s_i \in T \\ 0 & \text{otherwise} \end{cases}$$

We define the probability of reaching the target set T from state s as $\text{val}_T(s) := \mathbb{E}_s[\text{Reach}_T]$.

2.3 Multi-player Turn-based Stochastic Graph Games

In this section, we present the model of the second part of the thesis. In contrast to the previous section, the games are turn-based, the number of players is arbitrary, and every player has their own payoff, and hence the games are not zero-sum. Throughout the thesis, we use the word *game* for multi-player simple quantitative turn-based games played on graphs.

Definition 2.3 (Game). A *game* is a tuple $\mathcal{G}$ that consists of:

- a directed graph (V, E) , called the *underlying* graph of $\mathcal{G}$, where the number of vertices is denoted by n ;
- a finite set Π of *players*;
- a partition $(V_i)_{i \in \Pi \cup \{?\}}$ of the set V , where V_i denotes the set of vertices *controlled* by player i , and the vertices in $V_?$ are called *stochastic vertices*;
- a *probability function* $p: E(V_?) \rightarrow [0, 1]$, such that for each stochastic vertex v , the restriction of p to $E(v)$ is a probability distribution;
- a mapping $\mu: T \rightarrow [0, 1]^\Pi$ called *payoff function*, where T is the set of *terminal vertices*, that is, vertices of the graph (V, E) that have no outgoing edges. We also write μ_i , for each player i , for the function that maps a terminal vertex t to the i^{th} coordinate of the tuple $\mu(t)$; and
- a vertex $v_0 \in V$, which is the *initial* vertex of the game.

When referring to a game $\mathcal{G}$, we will often use the notations V for vertices, E for edges, Π for players, and so on without necessarily recalling them.

Definition 2.4 (Markov decision process, Markov chain). A *Markov decision process* is a game with one player. A *Markov chain* is a game with zero players.

Note that the above notions coincide with the ones of Section 2.2, where the states of the MDP (resp. of the MC) are the vertices of the game, and the terminal vertices are absorbing.

Plays. We call *play* a path in the underlying graph that is infinite, or whose last vertex is a terminal vertex. The payoff functions μ and μ_i , defined only for terminal vertices, are extended naturally to plays as follows: for a play of the form ht , with $t \in T$, we define $\mu(ht) = \mu(t)$, and for an infinite play π , we define $\mu(\pi) = (0)_{i \in \Pi}$, i.e., all players receive the payoff 0 in an infinite path where no terminal vertex is reached.

Strategies and strategy profiles. In the second part of the thesis, by *strategy*, we always mean a stationary strategy, that is, a strategy that only depends on the current vertex. Thus, in a game $\mathcal{G}$, a strategy for player i is a mapping σ_i that maps each vertex $v \in V_i$ to a probability distribution over the successors of v . A *strategy profile* for a subset $P \subseteq \Pi$ is a tuple $(\sigma_i)_{i \in P}$, that we usually write $\bar{\sigma}$ if $P = \Pi$, and $\bar{\sigma}_{-i}$ if $P = \Pi \setminus \{i\}$ for some player i . We then write $(\bar{\sigma}_{-i}, \sigma'_i)$ to denote the strategy profile $\bar{\tau}$ defined by $\tau_i = \sigma'_i$ and $\tau_j = \sigma_j$ for $j \neq i$.

A strategy profile $\bar{\sigma}$ for all players in the game $\mathcal{G}$ defines a probability measure $\mathbb{P}_{\bar{\sigma}}$ over plays, which turns the payoff functions μ_i into random variables. We then write $\mathbb{E}(\bar{\sigma})$ for the expectation operator $\mathbb{E}^{\mathbb{P}_{\bar{\sigma}}}$. If the game is a Markov chain, then we just write $\mathbb{E}$ to represent this value. We say the maximum payoff of an MDP is the maximum of all expectations over all stationary strategies, and write $\text{val} := \max_{\sigma} \mathbb{E}(\sigma)$ to denote it.

Equilibria. In such games, we study *equilibria*, that is, strategy profiles that offer some stability guarantees. The most famous notion of equilibrium is the *Nash equilibrium* [44].

Definition 2.5 (Nash equilibrium). A *Nash equilibrium*, or *NE* for short, in the game $\mathcal{G}$ is a strategy profile $\bar{\sigma}$ such that for each player i and every strategy σ'_i , we have

$$\mathbb{E}(\bar{\sigma}_{-i}, \sigma'_i)[\mu_i] \leq \mathbb{E}(\bar{\sigma})[\mu_i].$$

A classical quantitative relaxation of Nash equilibria is that of ε -approximate Nash equilibria, or just ε -Nash equilibria.

Definition 2.6 (ε -Nash equilibrium). Let $\varepsilon > 0$. An ε -*Nash equilibrium*, or ε -*NE* for short, in the game $\mathcal{G}$ is a strategy profile $\bar{\sigma}$ such that for each player i and every strategy σ'_i , we have $\mathbb{E}(\bar{\sigma}_{-i}, \sigma'_i)[\mu_i] - \varepsilon \leq \mathbb{E}(\bar{\sigma})[\mu_i]$.

2.4 Computational Problems and Complexity Classes

In this section, we present the computational problems that we consider in this thesis. In both parts, the problems ask for an approximation of a solution within an additive error ε , and, in both parts, the complexity upper bounds we establish are in the second level of the functional polynomial hierarchy.

Value-approximation problems. The computational problems of the first part of the thesis are the approximation of the limit value of stateful-discounted objectives and of the value of parity objectives in CSGs.

LIMITVALUE. Consider a CSG G , a state s , a reward function r , an assignment function $\chi: \mathcal{S} \rightarrow [d]$, and an additive error ε . The transition function δ and the reward function r are represented by rational numbers using at most B bits. Compute an approximation v of the limit value at state s such that

$$|v - \text{val}_{\chi}(s)| \leq \varepsilon.$$

PARITYVALUE. Consider a CSG G , a state s , a priority function p with index d , and an additive error ε . The transition function δ is represented by rational numbers using at most B bits. Compute an approximation v of the parity value at state s such that

$$|v - \text{val}_p(s)| \leq \varepsilon.$$

Constrained existence problems. The computational problem of the second part of the thesis is the approximate variant of the following classic decision problem about Nash equilibria.

CONSTRAINEDNE. Given a game $\mathcal{G}$ and two vectors $\bar{x}, \bar{y} \in \mathbb{Q}^{\Pi}$, called *threshold vectors*, does there exist an NE $\bar{\sigma}$ in $\mathcal{G}$ that satisfies the inequality $\bar{x} \leq \mathbb{E}(\bar{\sigma})[\mu] \leq \bar{y}$?

The CONSTRAINEDNE problem is undecidable if the players are allowed to use memory [54], which is excluded by our formalism here. Where the players are restricted to *stationary* strategies—as in our definition—this problem is $\exists\mathbb{R}$ -complete [38]. In order to obtain a more tractable problem, we consider here an approximate version defined using the notion of ε -NEs: we wish to obtain, for sure, the answer *yes* when an NE satisfying the constraint exists, and to obtain, for sure, the answer *no* when we are far from having such an NE, because there is not even an ε -NE that satisfies the constraint *up to* ε —but in the limit case, we can accept any answer. Formally, we want an algorithm that solves the following problem.

APPROXCONSTRAINEDNE. Given a game $\mathcal{G}$, two vectors $\bar{x}, \bar{y} \in \mathbb{Q}^{\Pi}$, and a rational number $\varepsilon > 0$ input using fixed-point representation, such that either:

1. there is an NE $\bar{\sigma}$ in $\mathcal{G}$ with $x_i \leq \mathbb{E}(\bar{\sigma})[\mu_i] \leq y_i$ for each player i ; or
2. there is no ε -NE $\bar{\sigma}$ in $\mathcal{G}$ satisfying $x_i - \varepsilon \leq \mathbb{E}(\bar{\sigma})[\mu_i] \leq y_i + \varepsilon$ for each player i ;

holds, then are we in case 1?

We call *functional* APPROXCONSTRAINEDNE problem the functional version of this problem, in which the answer *yes* is replaced by a succinct representation of an ε -NE that satisfies the constraint up to ε . In the sequel, we show an FNP^{NP} upper bound on this functional problem, which implies an NP^{NP} upper bound on the non-functional one; and an NP lower bound on the non-functional approximate problem, which implies an FNP lower bound on the functional one.

Complexity classes. We recall the complexity classes that we use. The class FNP is the class of search problems whose solutions can be verified in polynomial time, and TFNP is the subclass of FNP consisting of the *total* search problems, i.e., the problems for which a solution is guaranteed to exist for every input. For a class $\mathcal{C}$ of search problems, we denote by $\mathcal{C}^{\text{NP}}$ the class of problems that are solvable in $\mathcal{C}$ with access to an NP oracle. The class TFNP^{NP} , which is also written $\text{TFNP}[\text{NP}]$ in the literature, is the total functional form of the second level of the polynomial hierarchy. The value-approximation problems of the first part are total: for every input, an approximation of the value exists, and hence the relevant class is TFNP^{NP} . In contrast, the APPROXCONSTRAINEDNE problem is a promise problem: the output is required to be correct only on the inputs that satisfy the promise, and hence the relevant class is FNP^{NP} . Finally, $\exists\mathbb{R}$ is the class of problems that can be reduced in polynomial time to the satisfiability of a sentence in the existential theory of the reals, which we define in the next section.

2.5 The Existential Theory of the Reals

Both parts of the thesis rely on the encoding of the problem at hand as a system of polynomial inequalities with integer coefficients, and on quantitative bounds on the solutions of such systems. In this section, we present the relevant definitions and the two results from real algebraic geometry that we use.

Sentences in the theory of the reals. A sentence in the *existential theory of the reals* is of the form

$$\varphi = \exists x_1 \exists x_2 \cdots \exists x_k F(x_1, \dots, x_k),$$

where $F(x_1, \dots, x_k)$ is a *quantifier-free* formula in the language of ordered fields over the reals $\mathbb{R}$. More concretely, atomic sub-formulas are

$$P(x_1, \dots, x_k) = 0 \quad \text{or} \quad P(x_1, \dots, x_k) < 0,$$

where $P \in \mathbb{R}[x_1, \dots, x_k]$ is a polynomial over $\mathbb{R}$, and arbitrary formulas are built by Boolean connectives. More generally, we consider sentences with several blocks of alternating quantifiers, as in Theorem 2.7 below.

The following result is a specialization of Theorem 1 of [16], and is the basic algorithmic tool for the resolution of such sentences.

Theorem 2.7 ([16]). *Consider ℓ variables $x_1, \dots, x_\ell$ and a set of polynomials $\mathcal{P} = \{P_1, \dots, P_k\}$ such that, for all $i \in [k]$, the polynomial P_i is a polynomial in $x_1, \dots, x_\ell$ of degree at most D , with integer coefficients of bit-size at most B . Let $X_1, \dots, X_d$ be a partition of the variables into d subsets such that X_i has size ℓ_i . Let*

$$\Phi = (Q_d X_d) \cdots (Q_1 X_1) \quad \varphi(P_1, \dots, P_k)$$

be a sentence with d alternating quantifiers, i.e., $Q_i \in \{\exists, \forall\}$ and $Q_{i+1} \neq Q_i$, where $\varphi(P_1, \dots, P_k)$ is a quantifier-free formula whose atomic formulas are of the form $P_i \bowtie 0$ with $\bowtie \in \{<, >, =\}$. Then, there exists an algorithm to decide the truth of Φ in time

$$k^{\Pi_i \mathcal{O}(\ell_i+1)} \cdot D^{\Pi_i \mathcal{O}(\ell_i)} \cdot \mathcal{O}(\text{len}(\varphi)B^2),$$

where $\text{len}(\varphi)$ is the length of the quantifier-free formula φ .

Along with the above algorithmic result, the following complexity result also follows from [16]: if there is a constant number of quantifier alternations, then the complexity is PSPACE, and in general the complexity is EXPSpace.

We now present the second result, which bounds the magnitude of the solutions of a system of polynomial inequalities. Before we state the result, we recall that a partition of semi-algebraic subsets of $\mathbb{R}^k$ generated by a set of polynomials $\mathcal{P}$ is the collection of all nonempty sets of the form

$$\{x \in \mathbb{R}^k \mid \text{sign}(P(x)) = \varsigma(P) \text{ for all } P \in \mathcal{P}\},$$

where $\varsigma: \mathcal{P} \rightarrow \{-1, 0, 1\}$ ranges over all possible sign assignments. Intuitively, the partition is obtained by slicing $\mathbb{R}^k$ according to which side of each polynomial in $\mathcal{P}$ a point lies on (positive, zero, or negative). The parts of this partition are called the *cells* of $\mathcal{P}$.

Lemma 2.8 ([18, Theorem 1.3.5]). *Given a set $\mathcal{P}$ of polynomials of degree D in k variables with coefficients in $\mathbb{Z}$ of bit-size B , there exists a ball of radius $2^{BD^{\mathcal{O}(k)}}$ that intersects every cell of $\mathcal{P}$.*

2.6 Floating-point Representation

The bounds of the previous section imply that the solutions of the problems we consider involve numbers that are doubly exponentially small in the size of the input. In the fixed-point representation, such numbers require exponentially many bits, and hence they do not yield polynomial-size certificates. In this section, we present the floating-point representation of numbers and of probability distributions, and we recall the results from the literature that allow us to argue that approximate solutions admit polynomial-size certificates in this representation. The definitions and the results of this section are used in both parts of the thesis, and are due to the work of Frederiksen and Miltersen [31] and the work of Solan [50].

Floating-point number representation. We define the set of floating-point numbers with precision ℓ as

$$\mathcal{F}(\ell) := \{w \cdot 2^e \mid w \in \{0, \dots, 2^\ell - 1\}, e \in \mathbb{Z}\}.$$

The floating-point representation of an element $x = w \cdot 2^e \in \mathcal{F}(\ell)$ uses $\text{bit}(w) + \text{bit}(|e|)$ bits. In particular, a number as small as 2^{-2^B} is representable with $\mathcal{O}(B)$ bits, whereas its fixed-point representation requires 2^B bits. We define the relative distance of two positive real numbers $x, \tilde{x}$ as

$$\text{rel}(x, \tilde{x}) := \max \left\{ \frac{x}{\tilde{x}}, \frac{\tilde{x}}{x} \right\} - 1 = \inf \{ \alpha > 0 : x \leq (1 + \alpha)\tilde{x}, \tilde{x} \leq (1 + \alpha)x \}.$$

We set $\text{rel}(0, 0) := 0$, and for non-negative x and $\tilde{x}$, $\text{rel}(x, \tilde{x}) := \infty$ if exactly one of x and $\tilde{x}$ is zero. Intuitively, $\text{rel}(x, \tilde{x})$ describes the multiplicative distance. Observe that if two values are close to 0, but ε apart for some small ε , then their relative distance is larger than if these values are much larger and only ε apart. We say x is (ℓ, i) -close to $\tilde{x}$ if $\text{rel}(x, \tilde{x}) \leq (1 - 2^{1-\ell})^{-i} - 1$, where ℓ is a positive integer and i is a non-negative integer. The notion of (ℓ, i) -closeness measures how far two numbers can differ after i steps of rounding in an ℓ -bit floating-point system: the parameter ℓ controls the precision, i.e., the gaps between representable numbers, and the parameter i allows for cumulative rounding error. Thus, x is (ℓ, i) -close to $\tilde{x}$ if their relative difference is within about $i \cdot 2^{1-\ell}$, that is, within i units of machine precision.

Arithmetic operations. We define $\oplus^\ell, \ominus^\ell, \otimes^\ell, \oslash^\ell$ as finite precision arithmetic operations $+, -, *, /$ respectively by truncating the result of the exact arithmetic operation to ℓ bits. We drop the superscript ℓ if the context is clear.

Floating-point probability distribution representation. We denote by $\mathcal{D}(\ell)$ the set of all floating-point probability distributions with precision ℓ . A probability distribution $d \in \Delta([t])$ belongs to $\mathcal{D}(\ell)$ if there exists $w_1, w_2, \dots, w_t \in \mathcal{F}(\ell)$ such that

- for all $i \in [t]$, we have $d(i) = \frac{w_i}{\sum_{j \in [t]} w_j}$; and
- $\sum_{j \in [t]} w_j$ and 1 are (ℓ, t) -close.

We define the relative distance rel for probability distributions as

$$\text{rel}(d, \tilde{d}) := \max \{ \text{rel}(d(i), \tilde{d}(i)) : i \in [t] \}.$$

We say d is (ℓ, i) -close to $\tilde{d}$ if $\text{rel}(d, \tilde{d}) \leq (1 - 2^{1-\ell})^{-i} - 1$, where ℓ is a positive integer and i is a non-negative integer.

Below we recall the results from the literature related to the floating-point representation, the approximation of probability distributions, and the values of Markov chains.

Lemma 2.9 ([31, Lemma 1]). *Consider three non-negative real numbers x, y, z . If x and y are (ℓ, i) -close, and y and z are (ℓ, j) -close, then x and z are $(\ell, i + j)$ -close.*

Lemma 2.10 ([31, Lemma 4]). *Consider non-negative real numbers x, y . Let $\tilde{x} \in \mathcal{F}(\ell)$ be a number that is (ℓ, i) -close to x and $\tilde{y} \in \mathcal{F}(\ell)$ be a number that is (ℓ, j) -close to y . Then, the following assertions hold.*

1. *The number $\tilde{x} \oplus \tilde{y}$ is $(\ell, \max(i, j) + 1)$ -close to $x + y$;*
2. *the number $\tilde{x} \ominus \tilde{y}$ is $(\ell, \max(i, j) + 1)$ -close to $x - y$;*
3. *the number $\tilde{x} \otimes \tilde{y}$ is $(\ell, i + j + 1)$ -close to $x * y$; and*
4. *the number $\tilde{x} \oslash \tilde{y}$ is $(\ell, i + j + 1)$ -close to x/y .*

Moreover, all arithmetic operations can be computed in polynomial time with respect to ℓ .

Lemma 2.11 ([31, Lemma 5]). *Consider $x_1, \dots, x_t \in \mathcal{F}(\ell)$. Let $d(i) := x_i \oslash \left(\bigoplus_{j=1}^t x_j \right)$. Then, there exists $\tilde{d} \in \mathcal{D}(\ell)$ such that for all i , we have $\tilde{d}(i) = d(i) / \left(\sum_{j=1}^t d(j) \right)$, and d and $\tilde{d}$ are $(\ell, 2t)$ -close.*

Lemma 2.12 ([31, Lemma 6]). *Consider a probability distribution $d \in \Delta([t])$. Then, there exists $\tilde{d} \in \mathcal{D}(\ell)$ such that d and $\tilde{d}$ are $(\ell, 2t + 2)$ -close.*

The following result of Solan bounds the change of the reachability value of a Markov chain in terms of the relative distance of the transition functions. It is the continuity property that allows us to replace the exact probabilities of a solution by their floating-point approximations.

Lemma 2.13 ([50, Thm. 6]). *Consider two absorbing MCs $\mathcal{C}$ and $\tilde{\mathcal{C}}$ with identical state sets and a target set T . We denote by val_T and $\tilde{\text{val}}_T$ the reachability value of $\mathcal{C}$ and $\tilde{\mathcal{C}}$ respectively. Fix $\varepsilon := \max_{s, s'} \text{rel}(\delta(s, s'), \tilde{\delta}(s, s'))$. Then, for all states $s \in S$, we have*

$$|\text{val}_T(s) - \tilde{\text{val}}_T(s)| \leq 4n\varepsilon.$$

The following corollary extends the above result from reachability objectives to objectives that assign an arbitrary payoff in $[0, 1]$ to the absorbing states, which is the form in which we use it in the second part of the thesis.

Corollary 2.14. *Consider two absorbing MCs $\mathcal{C}$ and $\tilde{\mathcal{C}}$ with identical state sets, a set S_0 of absorbing states, and a payoff function $\mu: S_0 \rightarrow [0, 1]$. We denote by val_μ and $\tilde{\text{val}}_\mu$ the expected payoff of $\mathcal{C}$ and $\tilde{\mathcal{C}}$ respectively. Fix $\varepsilon := \max_{s, s'} \text{rel}(\delta(s, s'), \tilde{\delta}(s, s'))$. Then, for all states $s \in S$, we have $|\text{val}_\mu(s) - \tilde{\text{val}}_\mu(s)| \leq 4n\varepsilon$.*

Proof. For all $z \in [0, 1]$, we define the target set $T_z := \{s' \in \mathcal{S}_0 \mid \mu(s') \geq z\}$. Since the MC is absorbing, an absorbing state is reached with probability 1, and hence, for all states s , we have

$$\text{val}_\mu(s) = \int_0^1 \text{val}_{T_z}(s) dz,$$

and analogously for $\widetilde{\mathcal{C}}$. By Lemma 2.13, for all $z \in [0, 1]$, we have $|\text{val}_{T_z}(s) - \widetilde{\text{val}}_{T_z}(s)| \leq 4n\varepsilon$. Therefore,

$$\left| \text{val}_\mu(s) - \widetilde{\text{val}}_\mu(s) \right| \leq \int_0^1 \left| \text{val}_{T_z}(s) - \widetilde{\text{val}}_{T_z}(s) \right| dz \leq 4n\varepsilon,$$

which yields the result. $\square$

Finally, we recall that the reachability value of a Markov chain whose transition function is represented in floating point can be approximated in polynomial time.

Lemma 2.15 ([31, Thm. 4]). *Consider an absorbing MC $\mathcal{C}$ and a target set T . For all $s \in \mathcal{S}$, we have $\delta(s) \in \mathcal{D}(\ell)$ where $\ell \geq 1000n^2$. Then, there exists a polynomial-time algorithm that for all states $s \in \mathcal{S}$, computes an approximation $v \in \mathcal{F}(\ell)$ for the reachability value such that*

$$|v - \text{val}_T(s)| \leq 80n^4 2^{-\ell}.$$

Analogously to Corollary 2.14, we extend the above result from reachability objectives to payoffs on the absorbing states.

Corollary 2.16. *Consider an absorbing MC $\mathcal{C}$, a set $\mathcal{S}_0$ of absorbing states, and a payoff function $\mu: \mathcal{S}_0 \rightarrow [0, 1]$. For all $s \in \mathcal{S}$, we have $\delta(s) \in \mathcal{D}(\ell)$, and, for all $s \in \mathcal{S}_0$, we have $\mu(s) \in \mathcal{F}(\ell)$, where $\ell \geq 1000n^2$. Then, there exists a polynomial-time algorithm that for all states $s \in \mathcal{S}$, computes an approximation v of the expected payoff $\text{val}_\mu(s)$ such that*

$$|v - \text{val}_\mu(s)| \leq 128n^4 2^{-\ell}.$$

Proof. We construct an MC $\mathcal{C}_1$ from $\mathcal{C}$ with a reachability objective and two additional absorbing states $\top$ and $\perp$, so the set of states is $\mathcal{S}_1 := \mathcal{S} \cup \{\top, \perp\}$, and the target set is $T := \{\top\}$. The transition function δ_1 is defined as $\delta_1(s) := \delta(s)$ for all states $s \in \mathcal{S} \setminus \mathcal{S}_0$, and, for all $s \in \mathcal{S}_0$, we have

$$\delta_1(s, \top) := \mu(s), \quad \delta_1(s, \perp) := 1 - \mu(s).$$

Note that $\mathcal{C}_1$ is absorbing, and that, since in $\mathcal{C}$ an absorbing state is reached with probability 1, for all states $s \in \mathcal{S}$, the reachability value of T in $\mathcal{C}_1$ is $\text{val}_\mu(s)$. We define the approximate transition function δ_2 where, for all $s \in \mathcal{S}_0$, we have

$$\delta_2(s, \top) := \mu(s), \quad \delta_2(s, \perp) := 1 \ominus \mu(s),$$

and $\delta_2(s) := \delta_1(s)$ for all other states. By Lemma 2.10, we have that δ_1 and δ_2 are $(\ell, 1)$ -close, and δ_2 is computable in polynomial time. By Lemma 2.11, there exists $\delta_3 \in \mathcal{D}(\ell)$ such that δ_2

and δ_3 are $(\ell, 2n)$ -close, and δ_3 is computable in polynomial time. Therefore, by Lemma 2.9, we have that δ_1 and δ_3 are $(\ell, 2n + 1)$ -close, and hence, for all states s, s' , we have

$$\begin{aligned}
\text{rel}(\delta_1(s, s'), \delta_3(s, s')) &\leq \frac{1}{(1 - 2^{1-\ell})^{2n+1}} - 1 \\
&\leq \frac{1}{1 - (2n + 1)2^{1-\ell}} - 1 && \text{(Bernoulli inequality)} \\
&= \frac{(2n + 1)2^{1-\ell}}{1 - (2n + 1)2^{1-\ell}} && \text{(rearrange)} \\
&\leq (2n + 1)2^{2-\ell} && (\ell \geq 1000n^2) \\
&\leq 12n2^{-\ell}. && (n \geq 1)
\end{aligned}$$

We denote by $\mathcal{C}_2$ the MC derived from $\mathcal{C}_1$ with the transition function δ_3 , and by val_T^2 its reachability value. By Lemma 2.13, for all states $s \in \mathcal{S}$, we have $|\text{val}_\mu(s) - \text{val}_T^2(s)| \leq 48n^22^{-\ell}$. By Lemma 2.15, we compute in polynomial time an approximation v of $\text{val}_T^2(s)$ such that $|v - \text{val}_T^2(s)| \leq 80n^42^{-\ell}$. By combining the two inequalities, we get $|v - \text{val}_\mu(s)| \leq 128n^42^{-\ell}$, which yields the result. $\square$

Chapter 3

Overview of Results

In this chapter, we present the previous results in the literature and the formal statements of our results. The chapter is organized as follows. In Section 3.1 and Section 3.2, we consider the value-approximation problems **LIMITVALUE** and **PARITYVALUE** in concurrent stochastic games. In Section 3.3 and Section 3.4, we consider the **APPROXCONSTRAINEDNE** problem in multi-player turn-based stochastic games. Finally, in Section 3.5, we present the roadmap of the proofs, which occupy the two parts of the thesis.

3.1 Previous Results for Value Approximation

We discuss the previous works on computing the approximation of limit and parity values in CSGs. A natural approach for these computational problems is via the theory of reals, whose main computational result we recalled in Theorem 2.7. We now discuss the algorithms and complexity results from the literature for the limit value of stateful-discounted objectives. The basic computational approach is via the theory of reals. For a single discount factor, the reduction to the theory of reals and dealing with its limit (which corresponds to limit-average objectives) was presented in [23]. In the general case (d discount factors), each limit can be considered as the quantification $\exists \varepsilon_{i'} \forall \varepsilon_i \leq \varepsilon_{i'}$ in the theory of reals. Thus, concurrent stochastic games with the limit value of stateful-discounted objectives can be reduced to the theory of reals with quantifier alternation. This reduction gives a theory of reals sentence with the following parameters:

$$\ell = \mathcal{O}(m^2n), \quad k = \mathcal{O}(m^2n), \quad D = 4, \quad \prod_i (\ell_i + 1) = \mathcal{O}(2^d m^2n).$$

Applying Theorem 2.7 to the reduction we obtain the following result.

Theorem 3.1 (**LIMITVALUE: Previous Result**). *For the **LIMITVALUE** problem, the following assertions hold.*

1. *The problem is in EXPSPACE; and*
2. *the problem can be solved in time $\exp(\mathcal{O}(2^d m^2n + \log(1/\varepsilon) + \log(B)))$.*

For parity objectives, the result of [4, 34] reduces CSGs with parity objectives to CSGs with the limit value of stateful-discounted objectives. The reduction is achieved as follows. Consider the formula $R(a_0, a_1, \dots, a_{2d-1})$ from [4], which is a formula with multiple discount factors. Since for stateful-discounted objectives the mapping is contractive, the fixpoints are unique (least and greatest fixpoints coincide). The last sentence of [4, Theorem 4] states that the limit of $R(a_0, a_1, \dots, a_{2d-1})$ corresponds to the value for parity objectives. The Pre operator of the formula corresponds to the Bellman operator for stateful-discounted objectives, which establishes the connection to stateful-discounted games. This connection is made more explicit in the construction provided in [34, Section 2.2]. This linear reduction and the above theorem lead to similar results for parity objectives. Besides this reduction to the theory of reals, there are two other approaches for the PARITYVALUE problem. First, we can consider the nested fixpoint characterization as provided in [8] and a reduction to the theory of reals with quantifier alternation. However, this does not lead to a better complexity. Second, a different approach is presented in [22] and [21, Chapter 8]. This approach has the following components: (a) it enumerates over all possible subsets of actions for every state; and (b) for each of the enumerations, it requires a solution of a qualitative value problem (or limit-sure winning) in concurrent stochastic games with parity objectives, and the value-approximation for concurrent stochastic games with reachability objectives. This approach gives PSPACE complexity and the algorithmic complexity is $\exp(\mathcal{O}(mn + d \log(n) + \log(\log(1/\varepsilon)) + \log(B)))$.

3.2 Our Results for Value Approximation

Our main results for the value-approximation problems are as follows.

Theorem 3.2 (LIMITVALUE: Complexity and Algorithm). *For the LIMITVALUE problem, the following assertions hold.*

1. *The problem is in TFNP^{NP} ; and*
2. *the problem can be solved in time $\exp(\mathcal{O}(nd \log(m) + \log(B) + \log(\log(1/\varepsilon))))$.*

Theorem 3.3 (PARITYVALUE: Complexity and Algorithm). *For the PARITYVALUE problem, the following assertions hold.*

1. *The problem is in TFNP^{NP} ; and*
2. *the problem can be solved in time $\exp(\mathcal{O}(nd \log(m) + \log(B) + \log(\log(1/\varepsilon))))$.*

The comparison of the above results with the previous results is summarized in Tables 1.1 and 1.2. In particular, the complexity upper bound improves the previous EXPSPACE upper bound for the limit value and the previous PSPACE upper bound for parity objectives, and the algorithms improve the dependency on the number of actions in the exponent from linear to logarithmic. Hence, if the number of states is constant, then our algorithms run in polynomial time.

3.3 Previous Results for Constrained Equilibria

We now discuss the previous works on the constrained existence of Nash equilibria in multi-player turn-based stochastic games. The **CONSTRAINEDNE** problem is undecidable when the players are allowed to use memory: it is already undecidable to check whether there is a Nash equilibrium in which the payoff of a player lies in a given interval, even for games with ten players [54, Theorem 4.9]. The problem remains undecidable when the players are restricted to pure strategies. When the players are restricted to stationary strategies, which is the setting we consider, the problem is decidable, and the following results are known.

Theorem 3.4 (**CONSTRAINEDNE**: Previous Results [38, 54]). *For the **CONSTRAINEDNE** problem, where the players are restricted to stationary strategies, the following assertions hold.*

1. *The problem is NP-hard [54, Theorem 4.4]; and*
2. *the problem is $\exists\mathbb{R}$ -complete [38], where the hardness holds for games with seven players.*

Moreover, for the functional version of the **CONSTRAINEDNE** problem, the probabilities of the stationary strategies of an equilibrium can be irrational [54, Theorem 4.6], even for games with four players. Hence, the exact problem does not admit an algorithm that outputs an equilibrium explicitly, and the approximation of the equilibrium is the natural computational problem. For approximate solutions, no upper bound better than the one for the exact problem was known. Indeed, the work of Deligkas, Fearnley, Melissourgos, and Spirakis [26] showed that the class $\varepsilon\text{-}\exists\mathbb{R}$, which captures the complexity of deciding whether approximate solutions exist to systems of real-valued constraints, is polynomial-time reducible to $\exists\mathbb{R}$ itself, and that this equivalence hinges on the fact that the solutions of such systems can be doubly exponentially large. The techniques we use for the upper bound originate in the work of Frederiksen and Miltersen [31], who established a TFNP^{NP} upper bound for the value-approximation problem in two-player zero-sum concurrent reachability games; for that class of games, no hardness result, such as the NP-hardness we establish here, is known. Finally, the work of Bose, Ibsen-Jensen, and Totzke [19] studied ε -Nash equilibria in concurrent and partial-observation games; they consider the unconstrained problem, which is computationally significantly simpler, and they assume that the number of players is fixed, while in our setting the number of players is part of the input.

3.4 Our Results for Constrained Equilibria

Our main results for the **APPROXCONSTRAINEDNE** problem are as follows. The first result is the upper bound.

Theorem 3.5 (**APPROXCONSTRAINEDNE**: Complexity). *There exists an FNP^{NP} procedure that solves the functional **APPROXCONSTRAINEDNE** problem. Consequently, the **APPROXCONSTRAINEDNE** problem is in NP^{NP} .*

The second result is the hardness of the problem, which holds already for two players, whereas the $\exists\mathbb{R}$ -hardness of the exact problem requires seven players.

Theorem 3.6 (APPROXCONSTRAINEDNE: Hardness). *The APPROXCONSTRAINEDNE problem is NP-hard, even with two players.*

The third result concerns the limits of our approach for the upper bound. The upper bound relies on the fact that if a constrained NE exists, then there is one in which the non-zero probabilities are at least an inverse of a double-exponential in the size of the input (Lemma 7.1). One might suspect that such small probabilities are merely theoretical and are not actually required in equilibria. The following result shows that this is not the case, already for games with five players, whereas the $\exists\mathbb{R}$ -completeness of the exact problem requires seven players.

Theorem 3.7 (Patience of constrained equilibria). *For all $n \in \mathbb{N}$, there exist a game $\mathcal{G}^n$ with five players and two threshold vectors $\bar{x}$ and $\bar{y}$ such that there is a stationary Nash equilibrium $\bar{\sigma}$ in $\mathcal{G}^n$ satisfying $\bar{x} \leq \mathbb{E}(\bar{\sigma})[\mu] \leq \bar{y}$, but where all such stationary Nash equilibria contain probabilities that are double-exponentially small in n .*

The comparison of the above results with the previous results is summarized in Table 1.3. The above results leave a gap between the NP^{NP} upper bound and the NP lower bound for the APPROXCONSTRAINEDNE problem. We remark that Theorem 3.7 concerns exact Nash equilibria only: in the game $\mathcal{G}^n$, there exists an ε -Nash equilibrium that satisfies the constraints and in which all probabilities are representable with polynomially many bits in the fixed-point representation (Lemma 8.8). We conjecture that this is a general phenomenon (Conjecture 8.9), which would close the above gap and establish that the decision version of the problem is NP-complete.

3.5 Roadmap of the Proofs

The two parts of the thesis are dedicated to the proofs of the above results. Although the two solution concepts are different, the proofs of the upper bounds follow the same three steps, which we summarize below.

Bounds on the numbers that arise in a solution. The first step is to write the problem as a system of polynomial inequalities with integer coefficients, and to bound the numbers that arise in its solutions. For the value-approximation problem, we establish a new bound on the roots of multi-variate polynomials with integer coefficients, which shows that there is a region around the origin where such a polynomial cannot vanish (Lemma 4.4 in Section 4.2). Along with new characterizations of the stateful-discounted and limit values (Section 4.3 and Section 4.4), this bound implies that the limit value is approximated by the stateful-discounted value when the discount factors are double-exponentially small (Section 4.5). For the constrained existence problem, we present an explicit encoding of the problem in the existential theory of the reals, and, along with the bound of Basu, Pollack, and Roy (Lemma 2.8), we show that if a constrained NE exists, then there is one whose non-zero probabilities are at least an inverse of a double-exponential in the size of the input (Section 7.1).

Polynomial-size certificates in floating point. The bounds of the first step show that the numbers involved are doubly exponentially small, and hence they cannot be written explicitly in the fixed-point representation. The second step is to establish that approximate solutions

admit polynomial-size certificates in the floating-point representation of Section 2.6. For the value-approximation problem, we show that there exist ε -optimal stationary strategies whose non-zero probabilities are bounded from below, and hence that are representable in polynomial size (Lemma 6.6). For the constrained existence problem, we show that if a constrained NE exists, then there is an ε -NE, satisfying the constraints up to ε , in which all probability distributions are floating-point distributions of polynomial size (Lemma 7.3). In both cases, the argument relies on the continuity of the values with respect to the relative distance of the probabilities.

Guess-and-verify with an NP oracle. The third step is the algorithm: guess the certificate of the second step, and verify it. In both parts, the verification requires the resolution of single-player games, i.e., MDPs, whose transition probabilities are represented in floating point, and, in both parts, this resolution is achieved by an oracle call to NP (Lemma 6.2 and Lemma 7.4). The resulting procedures place the LIMITVALUE and PARITYVALUE problems in TFNP^{NP} (Chapter 6) and the functional APPROXCONSTRAINEDNE problem in FNP^{NP} (Chapter 7).

Algorithms and lower bounds. Besides the above, the first part presents algorithms for the LIMITVALUE and PARITYVALUE problems (Chapter 5), which are obtained by a binary search over the value, where each step requires the resolution of an auxiliary matrix game; and the second part presents the two lower bounds, namely the game with five players that requires doubly exponentially small probabilities (Chapter 8) and the NP-hardness of the APPROXCONSTRAINEDNE problem (Chapter 9).

Part I*

Concurrent Stochastic Games with Stateful-discounted and Parity Objectives

* The content of this part was published in full as [12].

Chapter 4

Mathematical Properties of Stateful-discounted and Limit Values

In this chapter, we present a new approach to the limit value approximation via the stateful-discounted value (Theorem 4.12). We follow the approach of [13] extending it step by step using similar arguments. We use this technical result to improve complexities and algorithmic bounds of computing ε -approximation of the limit and parity values. The chapter is organized as follows. In Section 4.1, we present some useful definitions and previous results in the literature. In Section 4.2, we present a bound on the roots of multi-variate polynomials which is used to establish a connection between the stateful-discounted and limit values. In Sections 4.3 and 4.4, we introduce new characterizations of the stateful-discounted and limit values. In Section 4.5, we finally show Theorem 4.12.

4.1 Notations and Selected Results from Literature

We first present some basic notations and definitions. Recall the notations related to numbers and their bit-size from Section 2.1.

Polynomials. A *uni-variate* polynomial P of degree D with integer coefficients of bit-size B is defined as $P(x) := \sum_{i=0}^D c_i x^i$ where $|c_i| < 2^B$. We define $\|P\|_\infty := \max(|c_0|, \dots, |c_D|)$. A k -variate polynomial P in $x_1, \dots, x_k$ of degree $D_1, \dots, D_k$ with integer coefficients of bit-size B is defined as

$$P(x_1, \dots, x_k) := \sum_{0 \leq i_1 \leq D_1} \dots \sum_{0 \leq i_k \leq D_k} c_{i_1, \dots, i_k} \prod_{j=1}^k x_j^{i_j},$$

where $|c_{i_1, \dots, i_k}| \leq 2^B$. Polynomial P is nonzero if $c_{i_1, \dots, i_k} \neq 0$ for some $i_1, \dots, i_k$. We say α is a root of P if $P(\alpha) = 0$. In this thesis, we only consider real roots.

Matrix notations. Given a square matrix M , we denote the determinant of M by $\det(M)$ and denote the signed sum of all minors of M by $S(M)$. Given two $k \times \ell$ matrices M_1 and M_2 , we denote the Hadamard product of M_1 and M_2 by $M_1 \odot M_2$. Given a positive integer k , often implicitly clear from context, we denote by $\mathbf{1}$ (resp. $\mathbf{0}$) the k -dimensional vector with all elements equal to 1 (resp. 0) and denote by Id the $k \times k$ identity matrix.

Matrix games. A matrix M defines a game played between two opponents, where rows (resp. columns) correspond to possible actions for the row- (resp. column-) player, and the entry $(M)_{i,j}$ is the reward the column-player pays the row-player when the pair of actions (i, j) is chosen. The value of a matrix game, denoted $\text{val } M$, is the maximum amount the row-player can guarantee, i.e., the amount they can obtain regardless of the column-player's strategy.

Below, we recall some useful results from the literature regarding the value of a matrix game (Lemma 4.1) and the determinant of a polynomial matrix (Lemma 4.2).

Lemma 4.1 ([48, Thm. 2]). *Consider a matrix game M . Then, there exists a square sub-matrix M_0 such that $S(M_0) \neq 0$ and $\text{val}(M) = \frac{\det(M_0)}{S(M_0)}$.*

Lemma 4.2 ([17, Prop. 8.12]). *Consider a $k \times k$ matrix M whose entries are polynomials in $x_1, \dots, x_\ell$ of degrees $D_1, \dots, D_\ell$ with integer coefficients of bit-size B . Then, $\det(M)$ is a polynomial in $x_1, \dots, x_\ell$ of degrees $kD_1, \dots, kD_\ell$ with integer coefficients of bit-size $kB + k \text{bit}(k) + \ell \text{bit}(k \max(D_1, \dots, D_\ell) + 1)$.*

Corollary 4.3. *Consider a $k \times k$ matrix M whose entries are polynomials in $x_1, \dots, x_\ell$ of degrees $D_1, \dots, D_\ell$ with integer coefficients of bit-size B . Then, $S(M)$ is a polynomial in $x_1, \dots, x_\ell$ of degrees $kD_1, \dots, kD_\ell$ with integer coefficients.*

Proof. By definition, $S(M)$ is the signed sum of all minors of M . Therefore, the result follows from Lemma 4.2. $\square$

4.2 Bounds on Roots of Polynomials with Integer Coefficients

In this section, we present a bound on the roots of multi-variate polynomials P with integer coefficients (Lemma 4.4). This result shows that there exists a region close to $\mathbf{0}$ within which P does not have a root. We use this technical result to establish a connection between the stateful-discounted value and limit value.

Lemma 4.4. *Consider a nonzero polynomial P in $x_1, \dots, x_\ell$ of degrees $D_1, \dots, D_\ell$ with integer coefficients of bit-size B . Let $D := \max(D_1, \dots, D_\ell)$ and $B_1 := 4\ell \text{bit}(D) + B + 1$. Then,*

$$\forall x_1 \in (0, \exp(-B_1)] \quad \forall x_2 \in (0, (x_1)^{D+1}] \quad \dots \quad \forall x_\ell \in (0, (x_{\ell-1})^{D+1}] \\ |P(x_1, \dots, x_\ell)| \geq \exp(B_1 - \ell) \cdot (x_\ell)^{D+1}.$$

Proof. We proceed with the proof by induction on ℓ .

Base case $\ell = 1$. Consider $P(x_1) = c_0 + c_1x_1 + \dots + c_{D_1}x_1^{D_1}$. Note that P is nonzero. Let k be the smallest index where $c_k \neq 0$. Therefore, for all $x_1 \leq \exp(-B_1)$, we have

$$\begin{aligned}
|P(x_1)| &\geq |c_k|x_1^k - \sum_{j>k} |c_j|x_1^j && \text{(triangle inequality)} \\
&\geq |c_k|x_1^k - x_1^{k+1} \sum_{j>k} |c_j| && (x_1^j \leq x_1^{k+1}) \\
&\geq x_1^k - x_1^{k+1} \sum_{j>k} |c_j| && (|c_k| \geq 1) \\
&\geq x_1^k - D \exp(B) x_1^{k+1} && (|c_j| \leq \exp(B)) \\
&\geq x_1^k (1 - D \exp(B) x_1) && \text{(rearrange)} \\
&\geq x_1^k (\exp(B_1) x_1 - D \exp(B) x_1) && (x_1 \leq \exp(-B_1)) \\
&\geq \exp(B_1 - 1) x_1^{D+1}, && (k \leq D \text{ and } D \exp(B) \leq \exp(B_1 - 1))
\end{aligned}$$

which completes the case.

Induction case $\ell > 1$. We partition P into

$$P(x) = P_0 + P_1(x_1) + \dots + P_\ell(x_1, \dots, x_\ell).$$

Note that P is nonzero. Therefore, let i be the smallest index where $P_i \neq 0$. By the choice of i , for all $j < i$, we have that $P_j = 0$. Fix $x_1, \dots, x_\ell$ such that $x_1 \leq \exp(-B_1)$, and $x_i \leq x_{i-1}^{D+1}$ for $i > 1$. We show that

$$|P_i(x_1, \dots, x_i)| - \sum_{j>i} |P_j(x_1, \dots, x_j)| \geq \exp(B_1 - \ell) x_\ell^{D+1}, \quad (4.1)$$

which concludes the case.

To bound the term $|P_i(x_1, \dots, x_i)|$ of the LHS of Eq. (4.1) from below, we define

$$Q(y) := P_i(x_1, \dots, x_{i-1}, y) = \sum_{j=1}^{D_i} C_j(x_1, \dots, x_{i-1}) y^j,$$

where for all $1 \leq j \leq D_i$, we have

$$C_j(x_1, \dots, x_{i-1}) := \sum_{j_1=0}^{D_1} \dots \sum_{j_{i-1}=0}^{D_{i-1}} c_{j_1, \dots, j_{i-1}, j, 0, \dots, 0} \prod_{k=1}^{i-1} x_k^{j_k}.$$

Note that Q is a univariate polynomial of degree at most D_i , where the absolute values of the coefficients are at most

$$\begin{aligned}
|C_j(x_1, \dots, x_{i-1})| &\leq \exp(B)(D+1)^{i-1} \\
&\leq \exp(B + (i-1) \text{bit}(D+1)) \\
&\leq \exp(B + 2\ell \text{bit}(D)) \\
&\leq \exp(B_1 - \text{bit}(D) - (i+1)), && (4.2)
\end{aligned}$$

where in the first inequality we use $c_{j_1, \dots, j_{i-1}, j, 0, \dots, 0} \leq \exp(B)$, in the second inequality we use $(D+1)^{i-1} \leq \exp((i-1) \text{bit}(D+1))$, in the third inequality we use $i-1 \leq \ell$ and $\text{bit}(D+1) \leq 2 \text{bit}(D)$, and in the fourth inequality we use $B + 2\ell \text{bit}(D) + \text{bit}(D) + (i+1) \leq B_1$. For all $1 \leq j \leq D_i$, note that C_j is a polynomial in $x_1, \dots, x_{i-1}$ of degrees $D_1, \dots, D_{i-1}$ with integer coefficients of bit-size B . Therefore, if C_j is nonzero, then by induction, we have

$$C_j(x_1, \dots, x_{i-1}) \geq \exp(B_1 - (i-1))x_{i-1}^{D+1} \geq \exp(B_1 - (i-1))x_i. \quad (4.3)$$

Consider $Q(y) = c_0 + c_1y + \dots + c_{D_i}y^{D_i}$, where for all j , we have $c_j = C_j(x_1, \dots, x_{i-1})$. Since P_i is nonzero, by the definition of Q , there exists j such that $C_j \neq 0$. Therefore, by Eq. (4.3), we have that Q is also nonzero. Let k be the smallest index where $c_k \neq 0$. Therefore,

$$\begin{aligned} & |P_i(x_1, \dots, x_i)| \\ &= |Q(x_i)| \quad (\text{def. } Q) \\ &\geq |c_k x_i^k| - \sum_{j>k} |c_j x_i^j| \quad (\text{triangle inequality}) \\ &\geq |c_k x_i^k| - x_i^{k+1} \sum_{j>k} |c_j| \quad (x_i^j \leq x_i^{k+1}) \\ &= x_i^k \left(|c_k| - x_i \sum_{j>k} |c_j| \right) \quad (\text{rearrange}) \\ &\geq x_i^k (|c_k| - \exp(B_1 - (i+1))x_i) \quad (\text{Eq. (4.2)}) \\ &\geq x_i^k (\exp(B_1 - (i-1))x_i - \exp(B_1 - (i+1))x_i) \quad (\text{Eq. (4.3)}) \\ &\geq x_i^{D+1} (\exp(B_1 - (i-1)) - \exp(B_1 - (i+1))) \quad (k \leq D) \quad (4.4) \end{aligned}$$

To bound the term $\sum_{j>i} |P_j(x_1, \dots, x_j)|$ of the LHS of Eq. (4.1) from above, consider P_j for all $j > i$. Recall that P_j is a polynomial in $x_1, \dots, x_j$ of degrees at most $D_1, \dots, D_j$ with integer coefficients of bit-size B . Therefore, the number of terms in P_j is at most $(D+1)^j$, and for each term, the degree of x_j is at least 1. Therefore,

$$\begin{aligned} |P_j(x_1, \dots, x_j)| &\leq (D+1)^j \cdot \exp(B) \cdot x_j \quad (c_{i_1, \dots, i_j, 0, \dots, 0} \leq \exp(B)) \\ &\leq (D+1)^\ell \cdot \exp(B) \cdot x_j \quad ((D+1)^j \leq (D+1)^\ell). \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{j>i} |P_j(x_1, \dots, x_j)| &\leq (D+1)^\ell \exp(B) \sum_{j>i} x_j \\ &\leq \ell(D+1)^\ell \exp(B) \cdot x_{i+1} \\ &\leq \exp(\text{bit}(\ell) + \ell \text{bit}(D+1) + B) \cdot x_{i+1} \\ &\leq \exp(\text{bit}(\ell) + 2\ell \text{bit}(D) + B) \cdot x_{i+1} \\ &\leq \exp(B_1 - (i+1))x_{i+1} \\ &\leq \exp(B_1 - (i+1))x_i^{D+1}, \quad (4.5) \end{aligned}$$

where in the second inequality we use $x_j \leq x_{i+1}$ for all $j \geq i+1$, in the third inequality we use $\ell \cdot (D+1)^\ell \leq \exp(\text{bit}(\ell) + \ell \text{bit}(D+1))$, in the fourth inequality we use $\text{bit}(D+1) \leq 2 \text{bit}(D)$,

in the fifth inequality we use $\text{bit}(\ell) + 2\ell \text{bit}(D) + B + (i + 1) \leq B_1$, and in the sixth inequality we use $x_{i+1} \leq x_i^{D+1}$. By combining Eqs. (4.4) and (4.5), we get

$$\begin{aligned} |P_i(x_1, \dots, x_i)| - \sum_{j>i} |P_j(x_1, \dots, x_j)| &\geq \exp(B_1 - i) x_i^{D+1} \\ &\geq \exp(B_1 - \ell) x_i^{D+1} \quad (i \leq \ell) \\ &\geq \exp(B_1 - \ell) x_\ell^{D+1}, \quad (x_\ell \leq x_i) \end{aligned}$$

which concludes the claim, completes the case, and yields the result. $\square$

We notice that Lemma 4.4 may be compared with the minimal distance of two semi-algebraic sets that do not intersect, formalized for example in [46, Corollary 3.8], but these results are incomparable. For example, consider a multivariate polynomial P and the algebraic sets $\{x \in \mathbb{R}^\ell : P(x_1, \dots, x_\ell) = 0\}$ and $\{0\}$. These sets intersect at the origin if $P(0, \dots, 0) = 0$. In this case, there is no minimal distance between these two sets, while Lemma 4.4 provides a region for them. To the best of our knowledge, Lemma 4.4 is a new result.

4.3 Characterization of Stateful-discounted Value

In this section, we introduce a new characterization of the stateful-discounted value in CSGs (Corollary 4.6 and Lemma 4.5), which generalizes the result presented in [13] from a single discount factor to multiple discount factors. In particular, Corollary 4.6 generalizes Theorem 1 of [13].

Stateful-discounted payoff. Consider a CSG G , a state s , a reward function r , and a discount function Λ . Given a pair of stationary strategies (σ, τ) , we define the stateful-discounted payoff as

$$\nu^{\sigma, \tau}(s) := \mathbb{E}_s^{\sigma, \tau}[\text{Disc}_\Lambda].$$

By fixing σ and τ , we obtain a transition function

$$\delta^{\sigma, \tau}(s, s') := \sum_{\substack{a \in \mathcal{A} \\ b \in \mathcal{B}}} \sigma(s)(a) \cdot \tau(s)(b) \cdot \delta(s, a, b)(s'),$$

which is described as a matrix, i.e., $\delta^{\sigma, \tau} \in \mathbb{R}^{n \times n}$. The stage reward function is defined as

$$r^{\sigma, \tau}(s) := \sum_{\substack{a \in \mathcal{A} \\ b \in \mathcal{B}}} \sigma(s)(a) \cdot \tau(s)(b) \cdot r(s, a, b),$$

which is described as a vector, i.e., $r^{\sigma, \tau} \in \mathbb{R}^n$. Therefore, the Bellman operator defined in [47] can be written as a recursive expression:

$$\nu^{\sigma, \tau} = \Lambda \odot r^{\sigma, \tau} + (\mathbf{1} - \Lambda) \odot (\delta^{\sigma, \tau} \nu^{\sigma, \tau}).$$

The matrix $\text{Id} - ((\mathbf{1} - \Lambda) \mathbf{1}^\top) \odot \delta^{\sigma, \tau}$ is strictly diagonally dominant, and therefore, is invertible. By Cramer's rule, we have

$$\nu^{\sigma, \tau}(s) = \frac{\nabla_\Lambda^s(\sigma, \tau)}{\nabla_\Lambda(\sigma, \tau)}, \quad (4.6)$$

where $\nabla_\Lambda(\sigma, \tau) := \det(\text{Id} - ((\mathbf{1} - \Lambda)\mathbf{1}^\top) \odot \delta^{\sigma, \tau})$ and $\nabla_\Lambda^s(\sigma, \tau)$ is the determinant of an $n \times n$ matrix derived by substituting the s -th column of the matrix $\text{Id} - ((\mathbf{1} - \Lambda)\mathbf{1}^\top) \odot \delta^{\sigma, \tau}$ with $\Lambda \odot r^{\sigma, \tau}$.

Auxiliary matrix game $W_\Lambda^s(z)$. We define a matrix game where the actions of each player are the pure stationary strategies in the stochastic game. The payoff of the game is obtained by the linearization of the quotient in Eq. (4.6). More formally, for all parameters $z \in \mathbb{R}$, $\hat{\sigma} \in \Sigma^{PS}$, and $\hat{\tau} \in \Gamma^{PS}$, we define the payoff of the matrix game as

$$W_\Lambda^s(z)[\hat{\sigma}, \hat{\tau}] := \nabla_\Lambda^s(\hat{\sigma}, \hat{\tau}) - z \cdot \nabla_\Lambda(\hat{\sigma}, \hat{\tau}).$$

The value of $W_\Lambda^s(z)$ is denoted by $\text{val}(W_\Lambda^s(z))$.

The following statements (Lemma 4.5 and Corollary 4.6) connect the stateful-discounted value with the value of the matrix game.

Lemma 4.5. *Consider a CSG G , a state s , a reward function, and an assignment function $\chi: \mathcal{S} \rightarrow [d]$. Then, the following assertions hold.*

1. *The map $(z, \lambda_1, \dots, \lambda_d) \mapsto \text{val}(W_\Lambda^s(z))$ is continuous;*
2. *for all discount factors $\lambda_1, \dots, \lambda_d$ and $z_1, z_2 \in \mathbb{R}$ such that $z_1 \leq z_2$, we have that $\text{val}(W_\Lambda^s(z_1)) \geq \text{val}(W_\Lambda^s(z_2)) + (z_2 - z_1)(\min_i \lambda_i)^n$, in particular, $z \mapsto \text{val}(W_\Lambda^s(z))$ is strictly decreasing; and*
3. *for all discount factors $\lambda_1, \dots, \lambda_d$, we have $\text{val}(W_\Lambda^s(\text{val}_\Lambda(s))) = 0$.*

Corollary 4.6. *Consider a CSG G , a state s , a reward function, and a discount function Λ . Then, $\text{val}_\Lambda(s)$ is the unique $z^* \in \mathbb{R}$ such that*

$$\text{val}(W_\Lambda^s(z^*)) = 0.$$

Below we define randomized strategies in the matrix game derived from stationary strategies in concurrent stochastic games.

Strategies for the matrix game. Given a Player-1 (resp. Player 2) stationary strategy σ (resp. τ), we denote by σ (resp. τ) a randomized strategy for the matrix game $W_\Lambda^s(z)$ defined as

$$\sigma(\hat{\sigma}) := \prod_{s \in \mathcal{S}} \sigma(s)(\hat{\sigma}(s)) \quad \forall \hat{\sigma} \in \Sigma^{PS}.$$

Also, we define

$$W_\Lambda^s(z)[\sigma, \tau] := \sum_{\hat{\sigma} \in \Sigma^{PS}} \sum_{\hat{\tau} \in \Gamma^{PS}} \sigma(\hat{\sigma}) \cdot \tau(\hat{\tau}) \cdot W_\Lambda^s(z)[\hat{\sigma}, \hat{\tau}].$$

The following result is instrumental to prove that the value of the matrix game is strictly decreasing (Lemma 4.5-Item 2).

Lemma 4.7. *Consider a $k \times k$ stochastic matrix M and a discount function $\Lambda: [k] \rightarrow \{\lambda_1, \dots, \lambda_d\}$. Then, we have*

$$\det(\text{Id} - ((\mathbf{1} - \Lambda)\mathbf{1}^\top) \odot M) \geq \left(\min_i \lambda_i\right)^k.$$

Proof. Fix $\widehat{M} := \text{Id} - ((\mathbf{1} - \Lambda)\mathbf{1}^\top) \odot M$. We claim that the matrix $\widehat{M}$ is a strictly diagonally dominant matrix. Indeed, M is a stochastic matrix. Therefore, for the i -th row of $\widehat{M}$, we have

$$\begin{aligned} \widehat{M}_{i,i} - \sum_{j \neq i} |\widehat{M}_{i,j}| &= 1 - (1 - \Lambda(i)) \sum_j M_{i,j} && (\text{def. } \widehat{M}) \\ &= \Lambda(i). && (M \text{ is a stochastic matrix}) \end{aligned}$$

Consider the (possibly complex) eigenvalues $\xi_1, \dots, \xi_k$. By Gershgorin circle theorem [32], for all i , we have that $|\xi_i - 1| \leq 1 - \min_i \lambda_i$. Therefore, we have

$$\det(\widehat{M}) = \prod_i \xi_i \geq \left(\min_i \lambda_i \right)^k,$$

which concludes the proof. $\square$

The following result is instrumental to prove that the stateful-discounted value forces the matrix game to have value zero (Lemma 4.5-Item 3).

Lemma 4.8. *Consider a CSG G , a state s , a reward function, a discount function Λ , and a Player-1 stationary strategy σ . Then, for all $\widehat{\tau} \in \Gamma^{PS}$ and $z \in \mathbb{R}$, we have*

$$W_\Lambda^s(z)[\sigma, \widehat{\tau}] = \nabla_\Lambda^s(\sigma, \widehat{\tau}) - z \cdot \nabla_\Lambda(\sigma, \widehat{\tau}),$$

where σ is the strategy for $W_\Lambda^s(z)$ derived by the strategy σ .

Proof. For all $\sigma \in \Sigma^S$ and $\tau \in \Gamma^S$, we define $M(\sigma, \tau) := \text{Id} - ((\mathbf{1} - \Lambda)\mathbf{1}^\top) \odot \delta^{\sigma, \tau}$. We also denote by $\sigma_{s \rightarrow a}$ the stationary strategy derived from σ where Player 1 chooses the action a at the state s . By definition, $\nabla_\Lambda(\sigma, \tau) = \det(M(\sigma, \tau))$. We now show that

$$\nabla_\Lambda(\sigma, \widehat{\tau}) = \sum_{\widehat{\sigma} \in \Sigma^{PS}} \sigma(\widehat{\sigma}) \cdot \nabla_\Lambda(\widehat{\sigma}, \widehat{\tau}).$$

Indeed, the s -th row of $M(\sigma, \widehat{\tau})$ only depends on $\sigma(s)$. Therefore, by multi-linearity of the determinant, we have

$$\begin{aligned} \det(M(\sigma, \widehat{\tau})) &= \det \left(\sum_{a \in \mathcal{A}} \sigma(s)(a) M(\sigma_{s \rightarrow a}, \widehat{\tau}) \right) \\ &= \sum_{a \in \mathcal{A}} \sigma(s)(a) \cdot \det(M(\sigma_{s \rightarrow a}, \widehat{\tau})). \end{aligned}$$

Hence, by induction, we have

$$\nabla_\Lambda(\sigma, \widehat{\tau}) = \det(M(\sigma, \widehat{\tau})) = \sum_{\widehat{\sigma} \in \Sigma^{PS}} \sigma(\widehat{\sigma}) \cdot \det(M(\widehat{\sigma}, \widehat{\tau})) = \sum_{\widehat{\sigma} \in \Sigma^{PS}} \sigma(\widehat{\sigma}) \cdot \nabla_\Lambda(\widehat{\sigma}, \widehat{\tau}).$$

By similar arguments, we can show that

$$\nabla_\Lambda^s(\sigma, \widehat{\tau}) = \sum_{\widehat{\sigma} \in \Sigma^{PS}} \sigma(\widehat{\sigma}) \cdot \nabla_\Lambda^s(\widehat{\sigma}, \widehat{\tau}).$$

Therefore,

$$\begin{aligned}
W_{\Lambda}^s(z)[\sigma, \hat{\tau}] &= \sum_{\hat{\sigma} \in \Sigma^{PS}} \sigma(\hat{\sigma}) \cdot W_{\Lambda}^s(z)[\hat{\sigma}, \hat{\tau}] \\
&= \sum_{\hat{\sigma} \in \Sigma^{PS}} \sigma(\hat{\sigma}) [\nabla_{\Lambda}^s(\hat{\sigma}, \hat{\tau}) - z \cdot \nabla_{\Lambda}(\hat{\sigma}, \hat{\tau})] \\
&= \nabla_{\Lambda}^s(\sigma, \hat{\tau}) - z \cdot \nabla_{\Lambda}(\sigma, \hat{\tau}),
\end{aligned}$$

which completes the proof. $\square$

Proof of Lemma 4.5. We prove the three items as follows.

1. By the continuity of the determinant, the entries of $W_{\Lambda}^s(z)$ depend continuously on parameters $z, \lambda_1, \dots, \lambda_d$. Therefore, the map $(z, \lambda_1, \dots, \lambda_d) \mapsto \text{val}(W_{\Lambda}^s(z))$ is continuous, which yields the item.
2. For all $\hat{\sigma} \in \Sigma^{PS}$ and $\hat{\tau} \in \Gamma^{PS}$, we have

$$\begin{aligned}
W_{\Lambda}^s(z_1)[\hat{\sigma}, \hat{\tau}] - W_{\Lambda}^s(z_2)[\hat{\sigma}, \hat{\tau}] &= (z_2 - z_1) \nabla_{\Lambda}(\hat{\sigma}, \hat{\tau}) && (\text{def. } W_{\Lambda}^s) \\
&= (z_2 - z_1) \det(\text{Id} - ((\mathbf{1} - \Lambda) \mathbf{1}^{\top}) \odot \delta^{\hat{\sigma}, \hat{\tau}}) && (\text{def. } \nabla_{\Lambda}) \\
&\geq (z_2 - z_1) \left(\min_i \lambda_i \right)^n && (\text{Lemma 4.7})
\end{aligned}$$

The result follows from the fact that increasing each entry of a matrix game by at least t increases the value by at least t .

3. By the result of [47], there exist optimal stationary strategies σ^* and τ^* for CSGs with stateful-discounted objectives. Therefore, for all $\hat{\tau} \in \Gamma^{PS}$, we have

$$\nu^{\sigma^*, \hat{\tau}}(s) = \frac{\nabla_{\Lambda}^s(\sigma^*, \hat{\tau})}{\nabla_{\Lambda}(\sigma^*, \hat{\tau})} \geq \text{val}_{\Lambda}(s).$$

Hence,

$$\nabla_{\Lambda}^s(\sigma^*, \hat{\tau}) - \text{val}_{\Lambda}(s) \cdot \nabla_{\Lambda}(\sigma^*, \hat{\tau}) \geq 0.$$

By Lemma 4.8, in the matrix game $W_{\Lambda}^s(\text{val}_{\Lambda}(s))$, we have

$$W_{\Lambda}^s(\text{val}_{\Lambda}(s))[\sigma^*, \hat{\tau}] = \nabla_{\Lambda}^s(\sigma^*, \hat{\tau}) - \text{val}_{\Lambda}(s) \cdot \nabla_{\Lambda}(\sigma^*, \hat{\tau}) \geq 0 \quad \forall \hat{\tau} \in \Gamma^{PS},$$

which guarantees that $\text{val}(W_{\Lambda}^s(\text{val}_{\Lambda}(s))) \geq 0$. By symmetric arguments on τ^* , we get $\text{val}(W_{\Lambda}^s(\text{val}_{\Lambda}(s))) \leq 0$. The result follows from combining these two inequalities. $\square$

Proof of Corollary 4.6. By Lemma 4.5-Item 2, the mapping $z \mapsto \text{val}(W_{\Lambda}^s(z))$ is strictly decreasing. By Lemma 4.5-Item 3, we know that $\text{val}(W_{\Lambda}^s(\text{val}_{\Lambda}(s))) = 0$. Hence, there exists the unique $z^* = \text{val}_{\Lambda}(s) \in \mathbb{R}$ such that

$$\text{val}(W_{\Lambda}^s(z^*)) = 0,$$

which yields the result. $\square$

4.4 Characterization of Limit Value

In this section, we introduce a new characterization of the limit value in CSGs (Corollary 4.10 and Lemma 4.9), which generalizes the result presented in [13] from a single discount factor to multiple discount factors. In particular, Corollary 4.10 generalizes Theorem 2 of [13].

Limit function. Given a CSG G , a reward function, and an assignment function $\chi: \mathcal{S} \rightarrow [d]$, we define the limit function as

$$F_{\chi}^s(z) := \lim_{\lambda_1 \rightarrow 0^+} \cdots \lim_{\lambda_d \rightarrow 0^+} \frac{\text{val}(W_{\Lambda}^s(z))}{(\lambda_d)^n}.$$

Lemma 4.9. Consider a CSG G , a state s , a reward function, and an assignment function $\chi: \mathcal{S} \rightarrow [d]$. Then, the following assertions hold.

1. For all $z \in \mathbb{R}$, the limit $F_{\chi}^s(z)$ exists in $\mathbb{R} \cup \{-\infty, +\infty\}$; and
2. for all $z_1 < z_2$, we have $F_{\chi}^s(z_1) \geq F_{\chi}^s(z_2) + (z_2 - z_1)$; moreover, there exist $z_1, z_2 \in \mathbb{R}$ such that $F_{\chi}^s(z_2) \leq 0 \leq F_{\chi}^s(z_1)$.

Corollary 4.10. Consider a CSG G , a state s , a reward function, and an assignment function $\chi: \mathcal{S} \rightarrow [d]$. Then, $\text{val}_{\chi}(s)$ is the unique $z^* \in \mathbb{R}$ such that

$$\forall z > z^* \quad F_{\chi}^s(z) < 0 \quad \text{and} \quad \forall z < z^* \quad F_{\chi}^s(z) > 0.$$

Below we first present Lemma 4.11, which shows that, given a fixed parameter z , the value of the matrix game is a rational function when discount factors are small enough, which implies the existence of the limit function. We then prove Lemma 4.9 and Corollary 4.10.

Lemma 4.11. Consider a CSG G , a state s , a reward function, an assignment function $\chi: \mathcal{S} \rightarrow [d]$, and a real number $z \in \mathbb{R}$. Then, there exist two polynomials P and Q in $\lambda_1, \dots, \lambda_d$ such that

$$\begin{aligned} & \exists \lambda_1^0 > 0 \quad \forall \lambda_1 \in (0, \lambda_1^0) \quad \cdots \quad \exists \lambda_d^0 > 0 \quad \forall \lambda_d \in (0, \lambda_d^0) \\ \text{s.t.} \quad & Q(\lambda_1, \dots, \lambda_d) \neq 0 \quad \text{and} \quad \text{val}(W_{\Lambda}^s(z)) = \frac{P(\lambda_1, \dots, \lambda_d)}{Q(\lambda_1, \dots, \lambda_d)}. \end{aligned}$$

Proof. By Lemma 4.1, there exists a sub-matrix M in $W_{\Lambda}^s(z)$ such that

$$\text{val}(W_{\Lambda}^s(z)) = \frac{\det(M)}{S(M)}.$$

Since z is fixed, the matrix M is a matrix with polynomial entries in $\lambda_1, \dots, \lambda_d$. Therefore, by Lemma 4.2 and Corollary 4.3, there exist two polynomials P and Q in $\lambda_1, \dots, \lambda_d$ such that

$$\text{val}(W_{\Lambda}^s(z)) = \frac{P(\lambda_1, \dots, \lambda_d)}{Q(\lambda_1, \dots, \lambda_d)}.$$

Since the matrix $W_\Lambda^s(z)$ is finite, the set of all sub-matrices is finite. Therefore, there exist two finite sets of polynomials $\mathcal{P}$ and $\mathcal{Q}$ such that for all $\lambda_1, \dots, \lambda_d$, there exists $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ such that

$$\text{val}(W_\Lambda^s(z)) = \frac{P(\lambda_1, \dots, \lambda_d)}{Q(\lambda_1, \dots, \lambda_d)}.$$

By Lemma 4.5-Item 1, we have $\text{val}(W_\Lambda^s(z))$ is continuous in $\lambda_1, \dots, \lambda_d$. Therefore, as $(\lambda_1, \dots, \lambda_d)$ varies in $(0, 1]^d$, the value of the matrix game can only jump from one rational function to another if the graphs of two rationals intersect. More formally, $\text{val}(W_\Lambda^s(z))$ can jump from $\frac{P_1}{Q_1}$ to $\frac{P_2}{Q_2}$ when discount factors are $\lambda_1, \dots, \lambda_d$, if we have

$$\frac{P_1(\lambda_1, \dots, \lambda_d)}{Q_1(\lambda_1, \dots, \lambda_d)} = \frac{P_2(\lambda_1, \dots, \lambda_d)}{Q_2(\lambda_1, \dots, \lambda_d)}.$$

We now claim that for every $P_1, P_2 \in \mathcal{P}$ and $Q_1, Q_2 \in \mathcal{Q}$, either $\frac{P_1}{Q_1}$ and $\frac{P_2}{Q_2}$ are congruent, or

$$\begin{aligned} & \exists \lambda_1^0 > 0 \quad \forall \lambda_1 \in (0, \lambda_1^0) \quad \dots \quad \exists \lambda_d^0 > 0 \quad \forall \lambda_d \in (0, \lambda_d^0) \\ & \text{s.t.} \quad \frac{P_1(\lambda_1, \dots, \lambda_d)}{Q_1(\lambda_1, \dots, \lambda_d)} \neq \frac{P_2(\lambda_1, \dots, \lambda_d)}{Q_2(\lambda_1, \dots, \lambda_d)}. \end{aligned}$$

Indeed, we define the polynomial $C := P_1 \cdot Q_2 - P_2 \cdot Q_1$.

Case $C = 0$. If $C = 0$, then $\frac{P_1}{Q_1}$ and $\frac{P_2}{Q_2}$ are congruent, which completes the case.

Case $C \neq 0$. If $C \neq 0$, then by Lemma 4.4, we have that

$$\begin{aligned} & \exists \lambda_1^0 > 0 \quad \forall \lambda_1 \in (0, \lambda_1^0) \quad \dots \quad \exists \lambda_d^0 > 0 \quad \forall \lambda_d \in (0, \lambda_d^0) \\ & \text{s.t.} \quad C(\lambda_1, \dots, \lambda_d) \neq 0. \end{aligned}$$

If $C(\lambda_1, \dots, \lambda_d) \neq 0$, then $\frac{P_1(\lambda_1, \dots, \lambda_d)}{Q_1(\lambda_1, \dots, \lambda_d)} \neq \frac{P_2(\lambda_1, \dots, \lambda_d)}{Q_2(\lambda_1, \dots, \lambda_d)}$, which completes the case and concludes the proof. $\square$

Proof of Lemma 4.9. We prove the two items as follows.

1. It is a direct implication of Lemma 4.11.
2. We define $z_1 := \min r(s, a, b)$ and $z_2 := \max r(s, a, b)$. For all $\lambda_1, \dots, \lambda_d$, we have

$$\text{val}(W_\Lambda^s(z_2)) \leq \text{val}(W_\Lambda^s(\text{val}_\Lambda(s))) \leq \text{val}(W_\Lambda^s(z_1)).$$

By dividing both sides by $(\lambda_d)^n$ and taking $\lambda_d, \dots, \lambda_1$ to 0 respectively, we get

$$F_\chi^s(z_2) \leq 0 \leq F_\chi^s(z_1),$$

Finally, fix arbitrary $z_1 < z_2$. By Lemma 4.5-Item 2,

$$\frac{\text{val}(W_\Lambda^s(z_1))}{(\lambda_d)^n} \geq \frac{\text{val}(W_\Lambda^s(z_2))}{(\lambda_d)^n} + (z_2 - z_1) \left(\frac{\min_i \lambda_i}{\lambda_d} \right)^n.$$

With the other discount factors fixed, the last ratio equals 1 for all sufficiently small λ_d . Taking the iterated limits therefore gives $F_\chi^s(z_1) \geq F_\chi^s(z_2) + (z_2 - z_1)$, which yields the result.

□

Proof of Corollary 4.10. Consider the function F_χ^s . The quantitative monotonicity and the two weak sign inequalities in Lemma 4.9-Item 2 imply that there is a unique point z^* such that $F_\chi^s(z) > 0$ for $z < z^*$ and $F_\chi^s(z) < 0$ for $z > z^*$. For all $\varepsilon > 0$, we have that $F_\chi^s(z^* + \varepsilon) < 0$. Hence,

$$\begin{aligned} \exists \lambda_1^0 > 0 \quad \forall \lambda_1 \in (0, \lambda_1^0) \quad \dots \quad \exists \lambda_d^0 > 0 \quad \forall \lambda_d \in (0, \lambda_d^0) \\ \text{s.t.} \quad \text{val}(W_\Lambda^s(z^* + \varepsilon)) < 0. \end{aligned}$$

By Lemma 4.5-Item 2,

$$\begin{aligned} \exists \lambda_1^0 > 0 \quad \forall \lambda_1 \in (0, \lambda_1^0) \quad \dots \quad \exists \lambda_d^0 > 0 \quad \forall \lambda_d \in (0, \lambda_d^0) \\ \text{s.t.} \quad \text{val}_\Lambda(s) < z^* + \varepsilon. \end{aligned}$$

Therefore, for all $\varepsilon > 0$, we get

$$\lim_{\lambda_1 \rightarrow 0^+} \dots \lim_{\lambda_d \rightarrow 0^+} \text{val}_\Lambda(s) \leq z^* + \varepsilon.$$

By taking ε to 0, we have

$$\lim_{\lambda_1 \rightarrow 0^+} \dots \lim_{\lambda_d \rightarrow 0^+} \text{val}_\Lambda(s) \leq z^*.$$

By symmetric arguments on $z^* - \varepsilon$, we get

$$\lim_{\lambda_1 \rightarrow 0^+} \dots \lim_{\lambda_d \rightarrow 0^+} \text{val}_\Lambda(s) \geq z^*,$$

which concludes the proof. □

4.5 Approximation of Limit Value

In this section, we introduce an approach for the approximation of the limit value via the stateful-discounted value (Theorem 4.12). The rest of the section is dedicated to its proof.

Theorem 4.12. Consider a CSG G , a state s , a reward function r , an assignment function $\chi: \mathcal{S} \rightarrow [d]$, and an additive error $\varepsilon > 0$. The transition function δ and the reward function r are represented by rational numbers. Let D be the least common multiple of their denominators, and let B bound the bit-sizes of D and of all numerators. Note that $\text{bit}(D)$ is polynomial in the input size. Fix

$$D := \max(|\Sigma^{PS}|, |\Gamma^{PS}|), \quad B_1 := 11Dn(B + \text{bit}(n) + \text{bit}(D) + \text{bit}(\varepsilon)),$$

and, for all $1 \leq i \leq d$, we set $\lambda_i^0 := \exp(-B_1(nD + 1)^{i-1})$. Then, we have

$$|\text{val}_{\Lambda^0}(s) - \text{val}_\chi(s)| \leq \varepsilon.$$

Remark 4.13 (Novelty). As mentioned previously, our result is a generalization of [13]. The key non-trivial aspect of the generalization relies on the fact that [13] considers uni-variate polynomials, whereas our result requires analysis of multi-variate polynomials. Lemma 4.4 is the key mathematical foundation, and the complete proofs require significant technical generalization.

Below we first show the algebraic properties of $\text{val}(W_\Lambda^s(z))$ (Lemma 4.14) to derive some insights on the asymptotic behavior of the sign of the map $(\lambda_1, \dots, \lambda_d) \mapsto \text{val}(W_\Lambda^s(z))$ as $\lambda_d, \dots, \lambda_1$ go to 0 respectively (Lemma 4.15). We then use Lemma 4.15 to establish a connection between the stateful-discounted value and the limit value in Theorem 4.12.

Lemma 4.14. *Consider a CSG G , a state s , a reward function r , and an assignment function $\chi: \mathcal{S} \rightarrow [d]$. The transition function δ and the reward function r are represented by rational numbers. Let D be the least common multiple of their denominators, and let B bound the bit-sizes of D and of all numerators. Let $D := \max(|\Sigma^{PS}|, |\Gamma^{PS}|)$. Then, there exist two finite sets $\mathcal{P}$ and $\mathcal{Q}$ of nonzero polynomials in $z, \lambda_1, \dots, \lambda_d$ of degrees $D, nD, \dots, nD$ with integer coefficients such that for all $z \in \mathbb{R}$ and $\lambda_1, \dots, \lambda_d$, there exist $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ such that*

$$Q(z, \lambda_1, \dots, \lambda_d) \neq 0, \quad \text{val}(W_\Lambda^s(z)) = \frac{P(z, \lambda_1, \dots, \lambda_d)}{Q(z, \lambda_1, \dots, \lambda_d)}.$$

Moreover, the coefficients of P are of bit-size $7Dn(B + \text{bit}(n) + \text{bit}(D))$.

Proof. By definition, $W_\Lambda^s(z)[\hat{\sigma}, \hat{\tau}] = \nabla_\Lambda^s(\hat{\sigma}, \hat{\tau}) - z \cdot \nabla_\Lambda(\hat{\sigma}, \hat{\tau})$ for all pure stationary strategies $\hat{\sigma}$ and $\hat{\tau}$. Note that $\nabla_\Lambda^s(\hat{\sigma}, \hat{\tau})$ and $\nabla_\Lambda(\hat{\sigma}, \hat{\tau})$ are the determinant of two matrices whose entries are polynomial in $\lambda_1, \dots, \lambda_d$ of degrees $1, 1, \dots, 1$ with rational coefficients whose denominators divide D and whose numerators have bit-size B . Therefore, by Lemma 4.2, the entries of $D^n W_\Lambda^s(z)$ are polynomials in $z, \lambda_1, \dots, \lambda_d$ of degrees $1, n, \dots, n$ with integer coefficients of bit-size

$$nB + n \text{bit}(n) + d \text{bit}(n + 1) \leq nB + 3n \text{bit}(n) =: B'.$$

We now define two sets of nonzero polynomials $\mathcal{P}$ and $\mathcal{Q}$ as

$$\mathcal{P} := \{\det(M) \mid M \text{ is a sub-matrix of } D^n W_\Lambda^s(z)\},$$

$$\mathcal{Q} := \{D^n S(M) \mid M \text{ is a sub-matrix of } D^n W_\Lambda^s(z)\}.$$

By Lemma 4.1, for all z , there exists a sub-matrix M in the matrix game $D^n W_\Lambda^s(z)$ such that

$$\text{val}(W_\Lambda^s(z)) = \frac{\det(M)}{D^n S(M)},$$

where the equality is due to the fact that $\det(D^{-n}M) = D^{-kn} \det(M)$ and $S(D^{-n}M) = D^{-(k-1)n} S(M)$ for the $k \times k$ matrix M . Note that M is a polynomial matrix. Hence, there exist $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ such that for all z and $\lambda_1, \dots, \lambda_d$, we have

$$Q(z, \lambda_1, \dots, \lambda_d) \neq 0, \quad \text{val}(W_\Lambda^s(z)) = \frac{P(z, \lambda_1, \dots, \lambda_d)}{Q(z, \lambda_1, \dots, \lambda_d)}.$$

We now show that $\mathcal{P}$ and $\mathcal{Q}$ satisfy the requirements of the lemma. Let $P := \det(M)$ and $Q := D^n S(M)$, where M is a $k \times k$ sub-matrix of $D^n W_\Lambda^s(z)$. First, by Lemma 4.5-Item 2, we know that $z \mapsto \text{val}(W_\Lambda^s(z))$ is strictly decreasing, therefore P is nonzero. Second, we know that entries of M are polynomials in $z, \lambda_1, \dots, \lambda_d$ of degrees $1, n, \dots, n$ with integer coefficients of bit-size B' . Therefore, by Lemma 4.2, we have that P is a polynomial in $z, \lambda_1, \dots, \lambda_d$ of degrees $k, kn, \dots, kn$ with integer coefficients of bit-size

$$\begin{aligned} kB' + k \text{bit}(k) + (d+1) \text{bit}(kn+1) &\leq DB' + D \text{bit}(D) + (d+1) \text{bit}(nD+1) \\ &\leq D(B' + \text{bit}(D) + 4n \text{bit}(nD)) \\ &= D(nB + 3n \text{bit}(n) + \text{bit}(D) + 4n \text{bit}(nD)) \\ &\leq Dn(B + 3 \text{bit}(n) + \text{bit}(D) + 4 \text{bit}(nD)) \\ &\leq 7Dn(B + \text{bit}(n) + \text{bit}(D)), \end{aligned}$$

where in the first inequality we use $k \leq D$, in the second inequality we use $(d+1) \text{bit}(nD+1) \leq 4nD \text{bit}(nD)$, in the first equality we use $B' = nB + 3n \text{bit}(n)$, in the third inequality we use $\text{bit}(D) \leq n \text{bit}(D)$, and in the fourth inequality we use $\text{bit}(nD) \leq \text{bit}(n) + \text{bit}(D)$.

Similarly, since $S(M)$ is the sum of the entries of the adjugate matrix of M , by Corollary 4.3, we have that Q is a polynomial in $z, \lambda_1, \dots, \lambda_d$ of degrees $D, nD, \dots, nD$ with integer coefficients, which completes the proof. $\square$

Lemma 4.15. Consider a CSG G , a state s , a reward function r , an assignment function $\chi: S \rightarrow [d]$, and a rational number z of bit-size κ . The transition function δ and the reward function r are represented by rational numbers. Let D be the least common multiple of their denominators, and let B bound the bit-sizes of D and of all numerators. Fix

$$D := \max(|\Sigma^{PS}|, |\Gamma^{PS}|), \quad B_1 := 11Dn(B + \text{bit}(n) + \text{bit}(D) + \kappa),$$

and, for all $1 \leq i \leq d$, we have $\lambda_i^0 := \exp(-B_1(nD+1)^{i-1})$. Then,

$$\begin{cases} \text{val}(W_{\Lambda^0}^s(z)) > 0 & \implies F_\chi^s(z) \geq 0, \\ \text{val}(W_{\Lambda^0}^s(z)) < 0 & \implies F_\chi^s(z) \leq 0, \\ \text{val}(W_{\Lambda^0}^s(z)) = 0 & \implies F_\chi^s(z) = 0. \end{cases}$$

Proof. We claim that

$$\begin{aligned} \forall \lambda_1 \in (0, \exp(-B_1)] \quad \dots \quad \forall \lambda_i \in (0, (\lambda_{i-1})^{nD+1}] \quad \dots \quad \forall \lambda_d \in (0, (\lambda_{d-1})^{nD+1}] \\ \text{sign}(\text{val}(W_\Lambda^s(z))) = \text{sign}(\text{val}(W_{\Lambda^0}^s(z))), \end{aligned}$$

which proves the result. By Lemma 4.5-Item 1, the map $(z, \lambda_1, \dots, \lambda_d) \mapsto \text{val}(W_\Lambda^s(z))$ is continuous. Therefore, the necessary condition for changing sign of $(\lambda_1, \dots, \lambda_d) \mapsto \text{val}(W_\Lambda^s(z))$ is that $\text{val}(W_{\Lambda^1}^s(z)) = 0$ for some $\lambda_1^1, \dots, \lambda_d^1$. For the sake of contradiction, assume there exists $\lambda_1^1, \dots, \lambda_d^1$ such that $\lambda_1^1 \leq \exp(-B_1)$ and $\lambda_i^1 \leq (\lambda_{i-1}^1)^{nD+1}$ for all $i > 1$. Let $\mathcal{P}$ and $\mathcal{Q}$ be the finite sets defined in Lemma 4.14. Let $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ such that

$$\text{val}(W_{\Lambda^1}^s(z)) = \frac{P(z, \lambda_1^1, \dots, \lambda_d^1)}{Q(z, \lambda_1^1, \dots, \lambda_d^1)}.$$

Since $\text{val}(W_{\Lambda^1}^s(z)) = 0$, we have that $P(z, \lambda_1^1, \dots, \lambda_d^1) = 0$. Write the fixed rational $z = a/b$ in lowest terms. We define

$$P_z(\lambda_1, \dots, \lambda_d) := b^D \cdot P(z, \lambda_1, \dots, \lambda_d).$$

Polynomial P is a polynomial in $z, \lambda_1, \dots, \lambda_d$ of degrees $D, nD, \dots, nD$ with integer coefficients of bit-size $7Dn(B + \text{bit}(n) + \text{bit}(D))$. Since z is a rational number of bit-size κ , we have that P_z is a polynomial in $\lambda_1, \dots, \lambda_d$ of degrees $nD, \dots, nD$ with integer coefficients of bit-size $7Dn(B + \text{bit}(n) + \text{bit}(D) + \kappa)$. If $P_z = 0$, then on the one hand, we have $\text{val}(W_{\Lambda^0}^s(z)) = 0$, and on the other hand, we have that

$$\begin{aligned} \forall \lambda_1 \in (0, \exp(-B_1)] \quad \dots \quad \forall \lambda_i \in (0, (\lambda_{i-1})^{nD+1}] \quad \dots \quad \forall \lambda_d \in (0, (\lambda_{d-1})^{nD+1}] \\ \text{val}(W_{\Lambda^0}^s(z)) = 0, \end{aligned}$$

which implies that $F_{\chi}^s(z) = 0$ and the proof of this case is done. Therefore, P_z is nonzero. Since $B_1 = 11Dn(B + \text{bit}(n) + \text{bit}(D) + \kappa)$, by Lemma 4.4, we have that $P_z(\lambda_1^1, \dots, \lambda_d^1) \neq 0$, which contradicts the assumption and completes the proof. $\square$

Proof of Theorem 4.12. We set $\kappa := \text{bit}(\varepsilon)$. We define $\mathcal{Z} := \{0, 2^{-\kappa}, 2 \cdot 2^{-\kappa}, \dots, 1\}$. By Lemma 4.5-Item 2, we have that $z \mapsto \text{val}(W_{\Lambda^0}^s(z))$ is strictly decreasing. Therefore, there exists $z \in \mathcal{Z}$ such that $\text{val}(W_{\Lambda^0}^s(z)) \geq 0$ and $\text{val}(W_{\Lambda^0}^s(z + 2^{-\kappa})) \leq 0$. By Lemma 4.5-Item 2, we have

$$z \leq \text{val}_{\Lambda^0}(s) \leq z + 2^{-\kappa}. \quad (4.7)$$

By Lemma 4.15 and Corollary 4.10, we have

$$z \leq \text{val}_{\chi}(s) \leq z + 2^{-\kappa}. \quad (4.8)$$

The result follows from combining Eq. (4.7) and Eq. (4.8). $\square$

Chapter 5

Algorithms for LIMITVALUE and PARITYVALUE

In this chapter, we present algorithms for computing ε -approximation of stateful-discounted, limit, and parity values. The chapter is organized as follows. In Section 5.1, we recall two classical algorithmic procedures that are used in our algorithms. In Section 5.2, we present an algorithm for computing ε -approximate stateful-discounted value. In Section 5.3, we present an algorithm for computing ε -approximate limit value, and as a consequence, we also obtain an algorithm for computing ε -approximate parity value.

5.1 Selected Algorithms from Literature

In this section, we recall classical algorithms for computing the determinant of a matrix and computing the value of a matrix game.

Lemma 5.1 ([15]). *Consider a $k \times k$ matrix M with rational entries of bit-size B . Then, there exists a procedure $\text{DET}(M)$ that computes the determinant of M in time $\tilde{O}(k^4 B^2)$, and $\text{DET}(M)$ is of bit-size $\mathcal{O}(k \log(k)B)$.*

Lemma 5.2 ([39]). *Consider a $k \times k$ matrix game M with rational entries of bit-size B . Then, there exists a procedure $\text{VAL}(M)$ that computes the value of the matrix game M in time $\tilde{O}(k^{3.5}B)$, and $\text{VAL}(M)$ is of bit-size $\mathcal{O}(B)$.*

5.2 Algorithm for Approximate Stateful-discounted Value

In this section, we present an algorithm for computing ε -approximation of the stateful-discounted value in CSGs. Given a CSG G , a reward function, and a discount function Λ , the procedure runs a binary search over the stateful-discounted value of state s . At the beginning, $\underline{z}$ and $\bar{z}$ are the under and over approximation of $\text{val}_\Lambda(s)$. In each step, the algorithm halves the interval $[\underline{z}, \bar{z}]$ by increasing $\underline{z}$ or decreasing $\bar{z}$ based on the sign of $\text{val}(W_\Lambda^s(\frac{\underline{z} + \bar{z}}{2}))$. After $\text{bit}(\varepsilon)$ steps, the algorithm outputs the ε -approximate value $(\bar{z} + \underline{z})/2$. The formal description is shown in Algorithm 1, and the correctness and the time complexity of the algorithm are shown in Lemma 5.3.

Algorithm 1 APPROXDISCOUNTED

Input: Game G , state s , reward function r , a discount function Λ , additive error ε

Output: Approximate stateful-discounted value v such that $|v - \text{val}_\Lambda(s)| \leq \varepsilon$

```
1: procedure APPROXDISCOUNTED( $G, s, r, \Lambda, \varepsilon$ )
2:    $\underline{z} \leftarrow 0$  and  $\bar{z} \leftarrow 1$ 
3:   while  $\bar{z} - \underline{z} > \varepsilon$  do
4:      $z \leftarrow \frac{\underline{z} + \bar{z}}{2}$ 
5:      $\nu \leftarrow \text{val}(W_\Lambda^s(z))$ 
6:     if  $\nu \geq 0$  then
7:        $\underline{z} \leftarrow z$ 
8:     else
9:        $\bar{z} \leftarrow z$ 
10:  return  $\frac{\underline{z} + \bar{z}}{2}$ 
```

Lemma 5.3. Consider a CSG G , a state s , a reward function r , a discount function Λ , and an additive error $\varepsilon > 0$. The transition function δ , the reward function r , and the discount function Λ are represented by rational numbers of bit-size B . Then, Algorithm 1 computes the ε -approximation of the stateful-discounted value of state s . Moreover, the algorithm runs in time $\exp(\mathcal{O}(n \log(m) + \log(B) + \log(\log(1/\varepsilon))))$.

Proof. We first present the proof of correctness and then the time complexity of the algorithm.

Correctness. The procedure is a binary search over the stateful-discounted value of state s . At the beginning, $\underline{z}$ and $\bar{z}$ are the under and over approximation of $\text{val}_\Lambda(s)$. In each step, by Lemma 4.5-Item 2, the algorithm halves the interval $[\underline{z}, \bar{z}]$ by increasing (resp. decreasing) $\underline{z}$ (resp. $\bar{z}$) based on the sign of ν . Therefore, in all steps, $\underline{z} \leq \text{val}_\Lambda(s) \leq \bar{z}$ is invariant. The algorithm terminates after at most $\text{bit}(\varepsilon)$ steps. The correctness of the procedure follows from the invariant $\underline{z} \leq \text{val}_\Lambda(s) \leq \bar{z}$ and $\bar{z} - \underline{z} \leq \varepsilon$.

Time complexity. The procedure executes at most $\text{bit}(\varepsilon)$ iterations. All lines except Line 5 require constant arithmetic operations. Line 5 consists of two parts as follows.

- *Construction of $W_\Lambda^s(z)$.* Recall that

$$W_\Lambda^s(z)[\hat{\sigma}, \hat{\tau}] = \nabla_\Lambda^s(\hat{\sigma}, \hat{\tau}) - z \cdot \nabla_\Lambda(\hat{\sigma}, \hat{\tau}),$$

where $\nabla_\Lambda^s(\hat{\sigma}, \hat{\tau})$ and $\nabla_\Lambda(\hat{\sigma}, \hat{\tau})$ are the determinants of two $n \times n$ matrices. The construction of matrices runs in time $\mathcal{O}(n^2 B)$. The algorithm uses DET to compute the determinants, which runs in $\tilde{\mathcal{O}}(n^4 B^2)$ by Lemma 5.1. The determinants are of bit-size $\mathcal{O}(n \log(n) B)$. The number of entries of $W_\Lambda^s(z)$ is at most m^{2n} . Therefore, its construction runs in time $\tilde{\mathcal{O}}(m^{2n})$. The entries of $W_\Lambda^s(z)$ are of bit-size $\mathcal{O}(n \log(n) B + \log(1/\varepsilon))$.

- *Computation of $\text{val}(W_\Lambda^s(z))$.* The procedure uses VAL to compute the value of the matrix game $W_\Lambda^s(z)$. Therefore, by Lemma 5.2, it runs in time $\tilde{\mathcal{O}}(m^{3.5n} B^2 \log^2(1/\varepsilon))$.

Hence, Line 5 runs in time $\tilde{\mathcal{O}}(m^{3.5n} B^2 \log^3(1/\varepsilon))$, which completes the proof. $\square$

5.3 Algorithms for Approximate Limit and Parity Values

In this section, we present an algorithm for computing ε -approximation of the limit and parity values in CSGs. Given a CSG G , a reward function, and an assignment function χ , the procedure outputs the $\varepsilon/2$ -approximate of the stateful-discounted value of state s for some Λ^0 by calling APPROXDISCOUNTED. By Theorem 4.12, the stateful-discounted value is an $\varepsilon/2$ -approximation of the limit value. Thus, the returned value of the algorithm is indeed an ε -approximate of the limit value. The formal description is shown in Algorithm 2, and the correctness and the time complexity of the algorithm is shown in Lemma 5.4. Since CSGs with parity objectives have a linear-size reduction to CSGs with the limit value of stateful-discounted objectives, as a consequence of the above algorithm, we obtain an algorithm for parity value approximation.

Algorithm 2 APPROXLIMIT

Input: Game G , state s , reward function r , assignment function χ , additive error ε

Output: Approximate limit value v such that $|v - \text{val}_\chi(s)| \leq \varepsilon$

```

1: procedure APPROXLIMIT( $G, s, r, \chi, \varepsilon$ )
2:    $D \leftarrow \text{lcm of the denominators in } \delta \text{ and } r$ 
3:    $B \leftarrow \max(B, \text{bit}(D))$ 
4:    $D \leftarrow m^n$ 
5:    $B_1 \leftarrow 11Dn(B + \text{bit}(n) + \text{bit}(D) + \text{bit}(\varepsilon))$ 
6:   for  $i \leftarrow 1$  to  $d$  do
7:      $\lambda_i^0 \leftarrow \exp(-B_1(nD + 1)^{i-1})$ 
8:    $v \leftarrow \text{APPROXDISCOUNTED}(G, s, r, \Lambda^0, \varepsilon/2)$ 
9:   return  $v$ 

```

Lemma 5.4. *Consider a CSG G , a state s , a reward function r , an assignment function $\chi: \mathcal{S} \rightarrow [d]$, and an additive error $\varepsilon > 0$. The transition function δ and the reward function r are represented by rational numbers of bit-size B . Then, Algorithm 2 computes the ε -approximation of the limit value of state s . Moreover, the algorithm runs in time $\exp(\mathcal{O}(nd \log(m) + \log(B) + \log(\log(1/\varepsilon))))$.*

Proof. We first present the proof of correctness and then the time complexity of the algorithm.

Correctness. The procedure computes the $\varepsilon/2$ -approximate stateful-discounted value v for Λ^0 and outputs v as the approximate limit value. The procedure APPROXDISCOUNTED outputs v such that

$$|v - \text{val}_{\Lambda^0}(s)| \leq \varepsilon/2. \quad (5.1)$$

By Theorem 4.12, we have that

$$|\text{val}_{\Lambda^0}(s) - \text{val}_\chi(s)| \leq \varepsilon/2. \quad (5.2)$$

By combining Eqs. (5.1) and (5.2), we get that v is an ε -approximation of $\text{val}_\chi(s)$.

Time complexity. All lines except Line 8 require at most $\mathcal{O}(d)$ arithmetic operations. In Line 8, the algorithm calls APPROXDISCOUNTED with parameters $(G, s, r, \Lambda^0, \varepsilon/2)$. The bit-size of the discount function Λ^0 is $\tilde{\mathcal{O}}(n^d m^{nd} (B + \text{bit}(\varepsilon)))$. Therefore, this line runs in time $\tilde{\mathcal{O}}(m^{7.5nd} B^2 \log^5(1/\varepsilon))$, which completes the proof. $\square$

Proof of Theorem 3.2-Item 2. It is a direct implication of Lemma 5.4. $\square$

Corollary 5.5. *Consider a CSG G , a priority function p , a state s , and an additive error $\varepsilon > 0$. The transition function δ is represented by rational numbers of bit-size B . Then, there exists an algorithm that computes the ε -approximation of the parity value of state s . Moreover, the algorithm runs in time $\exp(\mathcal{O}(nd \log(m) + \log(B) + \log(\log(1/\varepsilon))))$.*

Proof. By [4, 34], there exists a linear-size reduction from the CSGs with parity objectives to the CSGs with the limit value of stateful-discounted objectives. Therefore, the result follows from Lemma 5.4. $\square$

Proof of Theorem 3.3-Item 2. It is a direct implication of Corollary 5.5. $\square$

Chapter 6

Complexity of LIMITVALUE and PARITYVALUE

In this chapter, we show that the LIMITVALUE and PARITYVALUE problems are in TFNP^{NP} . The chapter is organized as follows. In Sections 6.1 to 6.3, we present algorithms for computing ε -approximate stateful-discounted value in MCs, MDPs, and CSGs. Each algorithm is used in the subsequent algorithm as a procedure. By our technical result on the limit value approximation via the stateful-discounted value (Theorem 4.12), we consequently obtain a TFNP^{NP} procedure for the LIMITVALUE problem. Since there exists a linear-size reduction from CSGs with parity objectives to CSGs with the limit value of stateful-discounted objectives [4, 34], the PARITYVALUE problem is also in TFNP^{NP} . Throughout the chapter, we use the definitions and the results related to Markov chains, Markov decision processes, and the floating-point representation of numbers and probability distributions that we presented in Section 2.2 and Section 2.6.

6.1 Stateful-discounted Value Approximation in MCs

In this section, we present an algorithm for computing ε -approximate stateful-discounted value in MCs by a reduction from MCs with stateful-discounted objectives to MCs with reachability objectives.

Lemma 6.1. *Consider an MC $\mathcal{C}$, a reward function r , and a discount function Λ . For all $s \in S$, we set*

$$\delta(s) \in \mathcal{D}(\ell), \quad r(s) \in \mathcal{F}(\ell), \quad \Lambda(s) \in \mathcal{F}(\ell),$$

where $\ell \geq 1000n^2$. Then, there exists a polynomial-time algorithm that for all states $s \in S$, computes an approximation v for the stateful-discounted value such that

$$|v - \text{val}_\Lambda(s)| \leq 176n^4 2^{-\ell}.$$

Proof. We construct a new MC $\mathcal{C}_1$ from $\mathcal{C}$ with a reachability objective and two additional absorbing states $\top$ and $\perp$ so the set of states is $S_1 := S \cup \{\top, \perp\}$. The target set is $T := \{\top\}$,

and the state $\perp$ is absorbing. The transition function δ_1 is defined as

$$\delta_1(s, s') := \begin{cases} (1 - \Lambda(s)) \cdot \delta(s, s') & s, s' \in \mathcal{S} \\ \sum_{s'' \in \mathcal{S}} \Lambda(s) \cdot r(s) \cdot \delta(s, s'') & s \in \mathcal{S}, s' = \top \\ \sum_{s'' \in \mathcal{S}} \Lambda(s) \cdot (1 - r(s)) \cdot \delta(s, s'') & s \in \mathcal{S}, s' = \perp \\ 1 & s \in \{\top, \perp\}, s' = s \\ 0 & \text{otherwise} \end{cases}$$

We denote by val_Λ the stateful-discounted value in $\mathcal{C}$ and by val_T^1 the reachability value in $\mathcal{C}_1$. Note that $\mathcal{C}_1$ is absorbing. Observe that the Bellman equation for $\mathcal{C}$ with the stateful-discounted objective is the same as the Bellman equation for $\mathcal{C}_1$ with the reachability objective. Therefore, for all $s \in \mathcal{S}$, we have

$$\text{val}_\Lambda(s) = \text{val}_T^1(s). \quad (6.1)$$

We define the approximate transition function as

$$\delta_2(s, s') := \begin{cases} (1 \ominus \Lambda(s)) \otimes \delta(s, s') & s, s' \in \mathcal{S} \\ \bigoplus_{s'' \in \mathcal{S}} \Lambda(s) \otimes r(s) \otimes \delta(s, s'') & s \in \mathcal{S}, s' = \top \\ \bigoplus_{s'' \in \mathcal{S}} \Lambda(s) \otimes (1 \ominus r(s)) \otimes \delta(s, s'') & s \in \mathcal{S}, s' = \perp \\ 1 & s \in \{\top, \perp\}, s' = s \\ 0 & \text{otherwise} \end{cases}$$

By the definition of finite precision arithmetic operators, we have $\delta_2(s, s') \in \mathcal{F}(\ell)$. Also, by Lemma 2.10, we have $\delta_1(s, s')$ and $\delta_2(s, s')$ are $(\ell, n+3)$ -close, and δ_2 is computable in polynomial time. By Lemma 2.11, there exists $\delta_3 \in \mathcal{D}(\ell)$ such that δ_2 and δ_3 are $(\ell, 2n)$ -close. Moreover, δ_3 is computable in polynomial time. Therefore, by Lemma 2.9, we have δ_1 and δ_3 are $(\ell, 3n+3)$ -close. Therefore, for all $s, s' \in \mathcal{S}$, we have

$$\begin{aligned} \text{rel}(\delta_1(s, s'), \delta_3(s, s')) &\leq \frac{1}{(1 - 2^{1-\ell})^{3n+3}} - 1 \\ &\leq \frac{1}{1 - (3n+3)2^{1-\ell}} - 1 && \text{(Bernoulli inequality)} \\ &= \frac{(3n+3)2^{1-\ell}}{1 - (3n+3)2^{1-\ell}} && \text{(rearrange)} \\ &\leq (3n+3)2^{2-\ell} && (\ell \geq 1000n^2) \\ &\leq 24n2^{-\ell}. && (n \geq 1) \end{aligned}$$

We define an MC $\mathcal{C}_2$ derived from $\mathcal{C}_1$ with transition function δ_3 . We denote by val_T^2 the reachability value in $\mathcal{C}_2$. By Lemma 2.13, for all $s \in \mathcal{S}$, we have

$$|\text{val}_T^1(s) - \text{val}_T^2(s)| \leq 96n^2 2^{-\ell}. \quad (6.2)$$

By Lemma 2.15, for all $s \in \mathcal{S}$, we can compute an approximation $v(s)$ of the reachability value of $\mathcal{C}_2$ starting from s in polynomial time such that

$$|v(s) - \text{val}_T^2(s)| \leq 80n^4 2^{-\ell}. \quad (6.3)$$

By combining Eqs. (6.1) to (6.3), for all states $s \in \mathcal{S}$, we have

$$|v(s) - \text{val}_\Lambda(s)| \leq 176n^4 2^{-\ell},$$

which yields the result. $\square$

6.2 Stateful-discounted Value Approximation in MDPs

In this section, we present an NP procedure for the approximate decision problem of the stateful-discounted values in MDPs. This is accomplished by guessing a pure stationary strategy and verifying if the strategy achieves the given threshold. The verification procedure uses the algorithm for computing approximate stateful-discounted values in MCs.

Lemma 6.2. *The problem of deciding if the stateful-discounted value for Player-1 MDPs is below a threshold up to an additive error is in NP where the input is a Player-1 MDP $\mathcal{P}$, a reward function r , a discount function Λ , a state s , a threshold $0 \leq \alpha \leq 1$, an additive error $\varepsilon = 2^{-\kappa}$ and a positive integer ℓ such that, for all $s' \in \mathcal{S}$ and $a \in \mathcal{A}$, we have*

$$\delta(s', a) \in \mathcal{D}(\ell), \quad r(s', a) \in \mathcal{F}(\ell), \quad \Lambda(s') \in \mathcal{F}(\ell), \quad \ell \geq 1000n^2 + \kappa.$$

Note that the numbers α and ε are represented in fixed-point binary and the NP procedure is such that

- *if $\alpha \leq \text{val}_\Lambda(s) - \varepsilon$, then it outputs YES; and*
- *if $\alpha \geq \text{val}_\Lambda(s) + \varepsilon$, then it outputs NO.*

Proof. We first present the NP procedure and then prove its soundness and completeness.

Procedure. The procedure guesses a pure stationary strategy σ for Player 1. Note that the size of the representation of a pure stationary strategy is polynomial with respect to the size of the representation of $\mathcal{P}$. By fixing σ , we obtain an MC $\mathcal{P}_\sigma$. We denote by v_σ its stateful-discounted value. By Lemma 6.1, there exists a polynomial time algorithm that computes an ε -approximation $\tilde{v}_\sigma$ of v_σ . Our procedure outputs YES if $\alpha \leq \tilde{v}_\sigma(s)$. If there exists no such pure stationary strategy, the procedure outputs NO.

Soundness. If $\alpha \leq \text{val}_\Lambda(s) - \varepsilon$, then, by [29], there exists a pure stationary strategy σ such that $\text{val}_\Lambda = v_\sigma$. The procedure non-deterministically guesses σ . By Lemma 6.1, we have $\tilde{v}_\sigma(s) + \varepsilon \geq v_\sigma(s)$. Therefore, we have $\alpha \leq \tilde{v}_\sigma(s)$, and the procedure outputs YES.

Completeness. If $\alpha \geq \text{val}_\Lambda(s) + \varepsilon$, then for all pure stationary strategies σ , we have $\alpha \geq v_\sigma(s) + \varepsilon$. By Lemma 6.1, we have $\tilde{v}_\sigma(s) - \varepsilon \leq v_\sigma(s)$. Therefore, $\alpha \geq \tilde{v}_\sigma(s)$ which implies that the procedure outputs NO and yields the result. $\square$

Lemma 6.2 also holds for Player-2 MDPs by symmetric arguments. More formally, we have the following result.

Corollary 6.3. *The problem of deciding if the stateful-discounted value for Player-2 MDPs is above a threshold up to an additive error is in NP where the input is a Player-2 MDP $\mathcal{P}$, a reward function r , a discount function Λ , a state s , a threshold $0 \leq \alpha \leq 1$, an additive error $\varepsilon = 2^{-\kappa}$ and a positive integer ℓ such that, for all $s' \in \mathcal{S}$ and $b \in \mathcal{B}$, we have*

$$\delta(s', b) \in \mathcal{D}(\ell), \quad r(s', b) \in \mathcal{F}(\ell), \quad \Lambda(s') \in \mathcal{F}(\ell), \quad \ell \geq 1000n^2 + \kappa.$$

6.3 Stateful-discounted, Limit, and Parity Values Approximation in CSGs

In this section, we first present a continuity result for MDPs with the stateful-discounted objectives (Lemma 6.4), which is a generalization of [29, Eq. 4.19]. We then show the existence of ε -optimal stationary strategies for stateful-discounted objectives that are representable in polynomial-size with respect to the size of the game and $\text{bit}(\varepsilon)$ (Lemma 6.6). We finally present a TFNP^{NP} algorithm for CSGs with the stateful-discounted objectives, where the discount function is represented in floating-point (Lemma 6.8). By our technical result on the limit value approximation via the stateful-discounted value (Theorem 4.12), we consequently obtain a TFNP^{NP} procedure for the LIMITVALUE problem. Since there exists a linear-size reduction from CSGs with parity objectives to CSGs with the limit value of stateful-discounted objectives [4, 34], the PARITYVALUE problem is also in TFNP^{NP} .

Lemma 6.4. *Consider $i \in \{1, 2\}$ and two Player- i MDPs $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ with identical state and action sets, a reward function r , and a discount function Λ . We denote by val_Λ and $\widetilde{\text{val}}_\Lambda$ the stateful-discounted value of $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ respectively. Then, for all $s \in \mathcal{S}$, we have*

$$\left| \text{val}_\Lambda(s) - \widetilde{\text{val}}_\Lambda(s) \right| \leq \frac{\|\delta - \widetilde{\delta}\|_\infty}{\min_s \Lambda(s)}.$$

Proof. We only prove for Player-1 MDPs. The proof for Player-2 MDPs is symmetric. By the Bellman equation defined in [47], for all $s \in \mathcal{S}$, we have

$$\begin{aligned} & \left| \text{val}_\Lambda(s) - \widetilde{\text{val}}_\Lambda(s) \right| \\ & \leq \max_a \left| \Lambda(s)r(s, a) + (1 - \Lambda(s)) \sum_{s'} \delta(s, a)(s') \text{val}_\Lambda(s') \right. \\ & \quad \left. - \Lambda(s)r(s, a) - (1 - \Lambda(s)) \sum_{s'} \widetilde{\delta}(s, a)(s') \widetilde{\text{val}}_\Lambda(s') \right| \\ & = (1 - \Lambda(s)) \max_a \left| \sum_{s'} \delta(s, a)(s') (\text{val}_\Lambda(s') - \widetilde{\text{val}}_\Lambda(s')) \right. \\ & \quad \left. + \sum_{s'} (\delta(s, a)(s') - \widetilde{\delta}(s, a)(s')) \widetilde{\text{val}}_\Lambda(s') \right| \\ & \leq (1 - \Lambda(s)) \left(\|\text{val}_\Lambda - \widetilde{\text{val}}_\Lambda\|_\infty + \|\delta - \widetilde{\delta}\|_\infty \|\widetilde{\text{val}}_\Lambda\|_\infty \right) \\ & \leq (1 - \Lambda(s)) \left(\|\text{val}_\Lambda - \widetilde{\text{val}}_\Lambda\|_\infty + \|\delta - \widetilde{\delta}\|_\infty \right) \\ & \leq \left(1 - \min_s \Lambda(s) \right) \left(\|\text{val}_\Lambda - \widetilde{\text{val}}_\Lambda\|_\infty + \|\delta - \widetilde{\delta}\|_\infty \right), \end{aligned} \tag{6.4}$$

where in the first inequality we use the Bellman equation, in the first equality we use algebraic manipulation, in the second inequality we use the definition of the norm and Cauchy–Schwarz

inequality, in the third inequality we use $\|\widetilde{\text{val}}_\Lambda\|_\infty \leq \|r\|_\infty \leq 1$, and in the fourth inequality we use $\min_s \Lambda(s) \leq \Lambda(s)$. Since Eq. (6.4) holds for all states s , we have

$$\|\text{val}_\Lambda - \widetilde{\text{val}}_\Lambda\|_\infty \leq \left(1 - \min_s \Lambda(s)\right) \left(\|\text{val}_\Lambda - \widetilde{\text{val}}_\Lambda\|_\infty + \|\delta - \widetilde{\delta}\|_\infty\right),$$

or equivalently

$$\|\text{val}_\Lambda - \widetilde{\text{val}}_\Lambda\|_\infty \leq \left(\frac{1 - \min_s \Lambda(s)}{\min_s \Lambda(s)}\right) \|\delta - \widetilde{\delta}\|_\infty \leq \frac{\|\delta - \widetilde{\delta}\|_\infty}{\min_s \Lambda(s)},$$

which yields the result. $\square$

Lemma 6.5. *Consider two Player-1 stationary strategies σ and $\tilde{\sigma}$ and a stateful-discounted objective in a CSG with n states. Fix $\alpha := \max_s \text{rel}(\sigma(s), \tilde{\sigma}(s))$. If $\alpha < \infty$, then, for every Player-2 stationary strategy τ and state s ,*

$$|\mathbb{E}_s^{\sigma, \tau}[\text{Disc}_\Lambda] - \mathbb{E}_s^{\tilde{\sigma}, \tau}[\text{Disc}_\Lambda]| \leq 4(n+2)\alpha.$$

Proof. Fix τ and apply the reduction to reachability used in Lemma 6.1 to both induced MCs. Every transition in the two resulting MCs is the same non-negative linear combination of the corresponding probabilities in σ and $\tilde{\sigma}$. Hence, their relative distance is at most α . The result follows from Lemma 2.13, applied to the resulting MCs with $n+2$ states. $\square$

Lemma 6.6. *Consider a CSG G , a reward function, a discount function Λ , and an additive error $\varepsilon = 2^{-\kappa}$. Fix $\underline{\Lambda} := \min_s \Lambda(s)$. Let ℓ be a positive integer such that $\ell \geq 1000(n+m)^2 + \kappa$. Then, there exist ε -optimal stationary strategies σ and τ such that*

$$\begin{aligned} \forall s \in \mathcal{S} \quad & \sigma(s) \in \mathcal{D}(\ell), \quad \tau(s) \in \mathcal{D}(\ell), \\ \forall s \in \mathcal{S}, a \in \mathcal{A} \quad & \sigma(s)(a) \neq 0 \implies \sigma(s)(a) \geq \frac{\underline{\Lambda}\varepsilon}{8m}, \\ \forall s \in \mathcal{S}, b \in \mathcal{B} \quad & \tau(s)(b) \neq 0 \implies \tau(s)(b) \geq \frac{\underline{\Lambda}\varepsilon}{8m}. \end{aligned}$$

Proof. We only prove the existence of σ since the existence of τ follows by symmetric arguments. By [47], there exists the optimal stationary strategy σ^* for Player 1. Fix $\delta := \underline{\Lambda}\varepsilon/(4m)$. For every state s , let B_s contain the actions whose probability under $\sigma^*(s)$ is at most δ , and let $\rho_s := \sum_{a \in B_s} \sigma^*(s)(a)$. Since $\rho_s \leq m\delta \leq 1/4$, choose an action $a_s \notin B_s$. We define σ_1 by

$$\sigma_1(s)(a) := \begin{cases} 0 & a \in B_s, \\ \sigma^*(s)(a) + \rho_s & a = a_s, \\ \sigma^*(s)(a) & \text{otherwise.} \end{cases}$$

By construction, $\sigma_1(s)$ is a probability distribution and $\sum_a |\sigma_1(s)(a) - \sigma^*(s)(a)| = 2\rho_s \leq \underline{\Lambda}\varepsilon/2$. Fixing either strategy yields Player-2 MDPs whose rewards and transition rows differ by at most this quantity. The same Bellman argument as in Lemma 6.4 shows that σ_1 is

$\varepsilon/2$ -optimal. By Lemma 2.12, there exists σ_2 such that for all $s \in \mathcal{S}$, we have $\sigma_2(s) \in \mathcal{D}(\ell)$, and $\sigma_1(s)$ and $\sigma_2(s)$ are $(\ell, 2m+2)$ -close. Fix $\alpha := \max_s \text{rel}(\sigma_1(s), \sigma_2(s))$. Then,

$$\begin{aligned} \alpha &\leq \frac{1}{(1 - 2^{1-\ell})^{2m+2}} - 1 \\ &\leq 16m2^{-\ell} \\ &\leq \frac{\varepsilon}{8(n+2)}. \end{aligned}$$

By Lemma 6.5, replacing σ_1 with σ_2 changes every payoff by at most $\varepsilon/2$. Every nonzero probability of σ_1 is greater than δ ; moreover, finite relative distance preserves support and $\alpha \leq 1$, so every nonzero probability of σ_2 is at least $\delta/(1+\alpha) \geq \underline{\Lambda}\varepsilon/(8m)$. Combining the two approximation steps shows that σ_2 is ε -optimal, completing the proof. $\square$

Remark 6.7. For all CSGs with stateful-discounted objectives, there exist ε -optimal stationary strategies which can be represented in polynomial size with respect to the size of the game and $\text{bit}(\varepsilon)$. Indeed, the preceding lower bound on nonzero probabilities has an exponent whose binary representation is polynomial in the floating-point representation of $\underline{\Lambda}$ and in $\text{bit}(\varepsilon)$, while the required mantissa precision is independent of $\underline{\Lambda}$.

Lemma 6.8. *The problem of computing an ε -approximation of the stateful-discounted value for CSGs is in TFNP^{NP} for inputs CSGs G , reward functions, discount functions Λ , states s , additive errors $\varepsilon = 2^{-\kappa}$, and positive integers ℓ such that, for all states $s' \in \mathcal{S}$, we have $\Lambda(s') \in \mathcal{F}(\ell)$.*

Proof. We first present the procedure and then prove its soundness and completeness.

Procedure. The procedure guesses two $\varepsilon/4$ -optimal stationary strategies σ^* and τ^* and an integer $j \in [0, 2^{\kappa+2}]$. By Remark 6.7, the size of the representation of stationary strategies is polynomial with respect to the size of G and $\text{bit}(\varepsilon)$. By fixing strategy σ^* (resp. τ^*), we obtain a Player-2 MDP G_{σ^*} (resp. a Player-1 MDP G_{τ^*}). We denote by v_{σ^*} (resp. v_{τ^*}) the value of G_{σ^*} (resp. G_{τ^*}). We define $\alpha := j2^{-(\kappa+2)} \in [0, 1]$. The procedure outputs α if $\alpha - 3\varepsilon/4 \leq v_{\sigma^*} - \varepsilon/4$ and $\alpha + 3\varepsilon/4 \geq v_{\tau^*} + \varepsilon/4$, which can be verified by two NP oracles implemented by Lemma 6.2 and Corollary 6.3.

Soundness. We assume that α is not an ε -approximation of the stateful-discounted value of G , i.e., that $\alpha \notin [\text{val}_{\Lambda}(s) - \varepsilon, \text{val}_{\Lambda}(s) + \varepsilon]$, and prove that the procedure does not output α in this case. Without loss of generality, we assume $\alpha \geq \text{val}_{\Lambda}(s) + \varepsilon$. For all stationary strategies σ , we have that $v_{\sigma} \leq \text{val}_{\Lambda}(s)$. Therefore, we have

$$\alpha - 3\varepsilon/4 > \text{val}_{\Lambda}(s) + \varepsilon/4 \geq v_{\sigma} + \varepsilon/4.$$

Hence, the NP procedure defined in Corollary 6.3 successfully decides that the inequality $\alpha - 3\varepsilon/4 \leq v_{\sigma} - \varepsilon/4$ is not true, and therefore the procedure does not output α , which yields the soundness of the procedure.

Completeness. By Remark 6.7, there exist $\varepsilon/4$ -optimal stationary strategies σ^* and τ^* which are polynomial-size representable. The procedure non-deterministically guesses σ^* and τ^* .

Since both strategies are $\varepsilon/4$ -optimal, we have $\text{val}_\Lambda(s) \leq v_{\sigma^*} + \varepsilon/4$ and $v_{\tau^*} - \varepsilon/4 \leq \text{val}_\Lambda(s)$. By the choice of j , there exists α such that $\alpha \in [\text{val}_\Lambda(s) - \varepsilon/4, \text{val}_\Lambda(s) + \varepsilon/4]$. Therefore,

$$\begin{aligned}\alpha - 3\varepsilon/4 &\leq \text{val}_\Lambda(s) - \varepsilon/2 \leq v_{\sigma^*} - \varepsilon/4, \text{ and} \\ \alpha + 3\varepsilon/4 &\geq \text{val}_\Lambda(s) + \varepsilon/2 \geq v_{\tau^*} + \varepsilon/4,\end{aligned}$$

and the procedure outputs α . This yields the completeness of the procedure and completes the proof. $\square$

Proof of Theorem 3.2-Item 1. It is a direct implication of Lemma 6.8 and Theorem 4.12. $\square$

Proof of Theorem 3.3-Item 1. By [4, 34], there exists a linear-size reduction from the PARITYVALUE problem to the LIMITVALUE problem. Therefore, the result follows from Theorem 3.2-Item 1. $\square$

Part II*

Constrained Approximate Nash Equilibria in Multi-player Stochastic Graph Games

* The content of this part is based on [10]; an extended preprint is available as [11].

Chapter 7

Computing Constrained Approximate Nash Equilibria

In this chapter, we present the FNP^{NP} upper bound for the functional APPROXCONSTRAINEDNE problem (Theorem 3.5). Throughout the chapter, we use n for the number of vertices in the game, $|E|$ for the number of edges in the game, and B for the bit-size used to represent the probabilities and the payoffs. The probabilities and the payoffs of the input are represented in fixed-point binary, and hence, for all $\ell \geq B$, they belong to $\mathcal{F}(\ell)$. We assume that we are given an instance $(\mathcal{G}, \bar{x}, \bar{y}, \varepsilon)$ of the functional APPROXCONSTRAINEDNE problem. We call $(\varepsilon\text{-})$ constrained NE an $(\varepsilon\text{-})$ NE in $\mathcal{G}$ such that each player i 's payoff lies in the interval $[x_i, y_i]$ (resp. $[x_i - \varepsilon, y_i + \varepsilon]$).

The rest of the chapter is organized as follows, and follows the three steps that we outlined in Section 3.5. In Section 7.1, we show how the problem of computing a constrained stationary NE is encoded as a sentence in the existential theory of the reals, and, by applying the result of Basu, Pollack, and Roy (Lemma 2.8), we show that any solution to the resulting system of polynomial inequalities lies within a ball of at most double-exponential radius. This implies that, in the worst case, the smallest non-zero probabilities in a stationary equilibrium may be as small as inverse double-exponential in the size of the input. In Section 7.2, we show that it is sufficient to represent the probabilities using floating-point numbers with polynomial precision. The argument relies on the continuity result of Solan (Corollary 2.14): small perturbations in transition probabilities lead to only small changes in the value of the associated Markov chain, and hence approximate equilibria obtained where the probabilities are slightly perturbed still yield approximately correct payoffs. Finally, in Section 7.3, we provide a coNP procedure (Lemma 7.4) to decide whether the maximum payoff in a Markov decision process, where the probabilities are given in floating-point representation, lies above or below a given threshold up to an additive ε ; and we present the FNP^{NP} procedure, which guesses a stationary strategy profile that approximately satisfies the constraints and verifies it using the above results.

7.1 Double-exponentially Small Probabilities Suffice

In this section, we show that if a constrained stationary NE exists, then there is one in which all non-zero probabilities are bounded from below by the inverse of a double-exponential function of the input size.

Lemma 7.1. *If there exists a constrained Nash equilibrium $\bar{\sigma}$, then there exists a constrained Nash equilibrium such that the non-zero probabilities are at least $2^{-B \cdot 2^{\mathcal{O}(|E||\Pi|)}}$.*

Proof. This statement is similar to a result in the work of Hansen, Koucký, and Miltersen [37, Thm. 4]. We write the sentence in the existential theory of the reals for solving the constrained stationary Nash equilibrium problem in the following proposition.

Proposition 7.2. *There exists a sentence in the existential theory of the reals for computing a constrained stationary Nash equilibrium with $\mathcal{O}(|E||\Pi|)$ variables, where all polynomials are of degree at most 2.*

Proof. The following part is similar to the sentence for constrained stationary Nash equilibria found in the work of Ummels and Wojtczak [54, Theorem 4.5]. However, we require the explicit bounds on the number of variables, the degree of the polynomials, and the length of the sentence. Thus, we rewrite the explicit sentence. First, by the assumption, there exists a strategy profile that is a stationary Nash equilibrium $\bar{\sigma}$. Let S be the set of all edges that are taken with non-zero probability in the strategy profile $\bar{\sigma}$. Furthermore, we define the set of vertices V_S from which terminals can be reached with non-zero probability when the edges outside S are removed. Second, we have a formula that states that the variables p_{vw} indeed describe a strategy. We enforce that the non-zero variables belong to the set S . We further ensure that for stochastic vertices, the variable p_{vw} encodes exactly the value dictated by the probability function p of the stochastic vertex. This involves satisfying the following constraints:

1. $p_{vw} > 0$, for each $vw \in S$;
2. $p_{vw} \leq 1$, for each $vw \in S$;
3. $p_{vw} = 0$, for each $vw \notin S$;
4. $\sum_{vw \in E(v)} p_{vw} = 1$, for each non-terminal vertex v ; and
5. $p_{vw} = p(vw)$, for all stochastic vertices v .

The variables r_v^i correspond to the payoff player i receives by starting from the vertex v . The payoffs r_v^i are equal to 0 for all vertices outside V_S , because no terminal is reachable from such vertices. The payoffs for other vertices satisfy the following constraints.

6. $r_t^i = \mu_i(t)$, for $t \in T$;
7. $r_v^i = 0$, for $v \notin V_S$; and
8. $r_v^i = \sum_{vw \in E(v)} p_{vw} r_w^i$, for $v \in V_S \setminus T$.

We now check that there is no unilateral deviation for the players by adding the following constraints. This check is enough because, when fixing the strategies of all players except player i , the resulting structure is an MDP. For MDPs, the Bellman equations characterize the value of each state, and the equation below enforces that the value at a vertex v is at least as large as the value of any successor, thereby ensuring consistency with the Bellman equations of the induced MDP [14, Theorem 10.109].

$$9. \quad r_v^i \geq r_w^i, \text{ for } v \in V_i \text{ and } vw \in E.$$

$$10. \quad x_i \leq r_{v_0}^i \leq y_i, \text{ for each player } i.$$

Finally, we add the auxiliary variables g_{vw} to keep track of the size of p_{vw} , which satisfy the following constraint:

$$11. \quad g_{vw} p_{vw} = 1, \text{ for each } vw \in S.$$

Thus, our sentence is $\exists p \exists r \exists g$ that satisfies constraints 1–11, completing the proof. $\lrcorner$

Recall from Lemma 2.8 that if a system of polynomial inequalities is satisfiable, then there is a valuation satisfying it within a ball whose radius is double-exponentially large with respect to the input size. Note that each of the constraints in the above sentence is a polynomial of degree at most 2, and that the total number of variables is $\mathcal{O}(|E||\Pi|)$. Note also that we had introduced the auxiliary variables g_{vw} and required that $g_{vw} p_{vw} = 1$ for each $vw \in S$. Since this is also a polynomial in the system, using Lemma 2.8, for all $vw \in S$, we have $g_{vw} \leq 2^{B \cdot 2^{\mathcal{O}(|E||\Pi|)}}$. Consequently, we have $p_{vw} \geq 2^{-B \cdot 2^{\mathcal{O}(|E||\Pi|)}}$, which completes the proof. $\square$

7.2 Floating-point Representation of Approximate Equilibria

The bound of the previous section is not sufficient to obtain a polynomial-size certificate, because representing such tiny probabilities explicitly requires exponentially many bits in the fixed-point representation. In this section, we show that the probabilities can be represented succinctly in the floating-point representation of Section 2.6, at the cost of considering ε -Nash equilibria instead of exact ones. While floating-point representations cannot express the exact probabilities required in an exact Nash equilibrium, they are sufficient for representing approximate equilibria.

Lemma 7.3. *If there exists a constrained stationary Nash equilibrium $\bar{\sigma}$, then there exists a $(64n^2 2^{-\ell})$ -constrained stationary Nash equilibrium $\bar{\sigma}'$ such that for all players i and vertices $v \in V_i$, we have $\bar{\sigma}'(v) \in \mathcal{D}(\ell)$, where $\ell \geq 1000n^2$.*

Proof. A stationary strategy consists of a probability distribution over the successors of each vertex. We can truncate a strategy by truncating each of these probability distributions. Let the strategy profile $\bar{\sigma}'$ be the truncation of the strategy profile $\bar{\sigma}$ as defined in Lemma 2.11.

Therefore, for all players i and vertices $v \in V_i$, we have that $\bar{\sigma}(v)$ and $\bar{\sigma}'(v)$ are $(\ell, 2n)$ -close. Consequently, for all vertices $v, w \in V$,

$$\begin{aligned} \text{rel}(\bar{\sigma}(v)(w), \bar{\sigma}'(v)(w)) &\leq \frac{1}{(1 - 2^{1-\ell})^{2n}} - 1 \\ &\leq \frac{1}{1 - 2n2^{1-\ell}} - 1 && \text{(Bernoulli inequality)} \\ &= \frac{2n2^{1-\ell}}{1 - 2n2^{1-\ell}} && \text{(rearrange)} \\ &\leq 8n2^{-\ell}. && (\ell \geq 1000n^2) \end{aligned} \quad (7.1)$$

We now bound the difference of the payoffs of $\bar{\sigma}$ and $\bar{\sigma}'$. By fixing a strategy profile, we obtain an MC in which the terminal vertices are absorbing; moreover, we can assume that the MC is absorbing, since we can redirect the vertices from which no terminal is reachable to a fresh absorbing vertex with payoff 0 for all players, without changing any of the payoffs. Therefore, by Eq. (7.1) and Corollary 2.14, for all players i , we have

$$|\mathbb{E}(\bar{\sigma})[\mu_i] - \mathbb{E}(\bar{\sigma}')[\mu_i]| \leq 32n^22^{-\ell}. \quad (7.2)$$

We fix $\Delta := 32n^22^{-\ell}$ and show that $\bar{\sigma}'$ is a 2Δ -constrained Nash equilibrium. First, since $\bar{\sigma}$ satisfies the payoff constraints, by Eq. (7.2), the strategy profile $\bar{\sigma}'$ satisfies the payoff constraints up to the additive error Δ . Second, consider a player i and a strategy σ_i'' for player i . Since the strategy profiles $(\bar{\sigma}_{-i}, \sigma_i'')$ and $(\bar{\sigma}'_{-i}, \sigma_i'')$ differ only in the strategies of the players other than i , the bound of Eq. (7.1) applies to the MCs they induce, and hence, by Corollary 2.14, we have

$$\mathbb{E}(\bar{\sigma}'_{-i}, \sigma_i'')[\mu_i] \leq \mathbb{E}(\bar{\sigma}_{-i}, \sigma_i'')[\mu_i] + \Delta.$$

Since $\bar{\sigma}$ is a Nash equilibrium, we have $\mathbb{E}(\bar{\sigma}_{-i}, \sigma_i'')[\mu_i] \leq \mathbb{E}(\bar{\sigma})[\mu_i]$, and, by Eq. (7.2), we have $\mathbb{E}(\bar{\sigma})[\mu_i] \leq \mathbb{E}(\bar{\sigma}')[\mu_i] + \Delta$. By combining the three inequalities, we get

$$\mathbb{E}(\bar{\sigma}'_{-i}, \sigma_i'')[\mu_i] \leq \mathbb{E}(\bar{\sigma}')[\mu_i] + 2\Delta,$$

which shows that $\bar{\sigma}'$ is a 2Δ -Nash equilibrium and completes the proof. $\square$

In the above result, we show that it is sufficient to consider floating-point representable strategies to solve the problem. Note that the bit-size of the representation of such a strategy profile is polynomial with respect to the size of the game and $\text{bit}(\varepsilon)$: by Lemma 7.1, the non-zero probabilities are at least $2^{-B \cdot 2^{\mathcal{O}(|E||\Pi|)}}$, and hence, in the floating-point representation, the exponent of each of them requires at most $\mathcal{O}(B|E||\Pi|)$ bits, while the mantissa requires ℓ bits, where ℓ is polynomial in the size of the game and $\text{bit}(\varepsilon)$.

7.3 The FNP^{NP} Procedure

In this section, we present the FNP^{NP} procedure. We first present the coNP procedure that decides whether the maximum payoff of an MDP whose transition probabilities are represented in floating point is below a given threshold, up to an additive error.

Lemma 7.4. *The problem of deciding if the payoff for MDPs is below a threshold up to an additive error is in coNP where the input is an MDP $\mathcal{P}$, a payoff function μ , a vertex v , a threshold $0 \leq \alpha \leq 1$, an additive error $\varepsilon = 2^{-\kappa}$ and a positive integer ℓ such that, for all stochastic vertices $w \in V$ and terminal vertices $t \in T$, we have*

$$p(w, \cdot) \in \mathcal{D}(\ell), \quad \mu(t) \in \mathcal{F}(\ell), \quad \ell \geq 1000n^2 + \kappa.$$

Define $\text{val} := \sup_{\sigma} \mathbb{E}(\sigma)[\mu]$. Note that the numbers α and ε are represented in fixed-point binary and the procedure is such that

- if $\alpha \leq \text{val} - \varepsilon$, then it outputs no; and
- if $\alpha \geq \text{val} + \varepsilon$, then it outputs yes.

Proof. We present the complement of the procedure, which is in NP, and then prove its soundness and completeness.

Procedure. The procedure guesses a pure stationary strategy σ for the player. Note that the size of the representation of a pure stationary strategy is polynomial with respect to the size of the representation of $\mathcal{P}$. By fixing σ , we obtain an MC $\mathcal{C}$, which we can assume to be absorbing, as in the proof of Lemma 7.3. We denote by val_{σ} the value $\mathbb{E}(\sigma)[\mu]$. By Corollary 2.16, there exists a polynomial time algorithm that computes an ε -approximation $\widehat{\text{val}}_{\sigma}$ of val_{σ} . Our procedure outputs *no* if $\alpha \leq \widehat{\text{val}}_{\sigma}$. If there exists no such pure stationary strategy, the procedure outputs *yes*.

Completeness. If $\alpha \leq \text{val} - \varepsilon$, then, by [29], there exists a pure stationary strategy σ such that $\text{val} = \text{val}_{\sigma}$. The procedure non-deterministically guesses σ . By Corollary 2.16, we have $\widehat{\text{val}}_{\sigma} + \varepsilon \geq \text{val}_{\sigma}$. Therefore, we have $\alpha \leq \widehat{\text{val}}_{\sigma}$, and the procedure outputs *no*.

Soundness. If $\alpha \geq \text{val} + \varepsilon$, then for all pure stationary strategies σ , we have $\alpha \geq \text{val}_{\sigma} + \varepsilon$. By Corollary 2.16, we have $\widehat{\text{val}}_{\sigma} - \varepsilon \leq \text{val}_{\sigma}$. Therefore, $\alpha \geq \widehat{\text{val}}_{\sigma}$, which implies that the procedure outputs *yes* and yields the result. $\square$

We now have all the ingredients to present the FNP^{NP} procedure.

Proof of Theorem 3.5. We first present the procedure and then prove its soundness and completeness.

Procedure. Let ℓ be a positive integer that is polynomial in the size of the game and in $\text{bit}(\varepsilon)$, and large enough that $64n^22^{-\ell} \leq \varepsilon/8$ and $\ell \geq B + 1000n^2 + \text{bit}(\varepsilon)$. Since the input probabilities are represented in fixed-point binary, they belong to $\mathcal{D}(\ell)$. The procedure guesses a stationary strategy profile $\bar{\sigma}$. By Lemma 7.1 and Lemma 7.3, the size of the representation of $\bar{\sigma}$ is polynomial with respect to the size of the game and $\text{bit}(\varepsilon)$. By fixing the strategy profile $\bar{\sigma}$, we obtain an MC $\mathcal{C}$. We denote by val_i the payoff of $\mathcal{C}$ where the payoff function is μ_i . By Corollary 2.16, the procedure computes the $\varepsilon/8$ -approximate payoff $\widehat{\text{val}}_i$ in polynomial time. For each player i , by fixing the strategy profile $\bar{\sigma}_{-i}$, we obtain an MDP $\mathcal{P}_i$. We denote by $\widetilde{\text{val}}_i$ the value of $\mathcal{P}_i$ where the payoff function is μ_i . Then, the procedure checks if there exists an ε -unilateral deviation for player i by deciding if $\widetilde{\text{val}}_i$ is at most $\widehat{\text{val}}_i + 3\varepsilon/4$ up to the

additive error $\varepsilon/8$, which is decided by an oracle call to the coNP procedure of Lemma 7.4. The procedure also checks, for every player i , that

$$x_i - 7\varepsilon/8 \leq \widehat{\text{val}}_i \leq y_i + 7\varepsilon/8.$$

If all the checks succeed, it outputs $\bar{\sigma}$.

Soundness. We assume that $\bar{\sigma}$ is not an ε -constrained Nash equilibrium. If $\text{val}_i < x_i - \varepsilon$ for some player i , then $\widehat{\text{val}}_i < x_i - 7\varepsilon/8$; analogously, if $\text{val}_i > y_i + \varepsilon$, then $\widehat{\text{val}}_i > y_i + 7\varepsilon/8$. In either case, the payoff check rejects $\bar{\sigma}$. It remains to consider the case where an ε -unilateral deviation is possible for some player i , equivalently, $\text{val}_i \leq \widetilde{\text{val}}_i - \varepsilon$. Therefore, we have

$$\begin{aligned} \widehat{\text{val}}_i + 3\varepsilon/4 &\leq \text{val}_i + 7\varepsilon/8 && \left(\widehat{\text{val}}_i \text{ is an } \varepsilon/8\text{-approximation of } \text{val}_i \right) \\ &\leq \widetilde{\text{val}}_i - \varepsilon/8. && (\varepsilon\text{-unilateral deviation}) \end{aligned}$$

However, for this case, the coNP procedure defined in Lemma 7.4 successfully outputs *no*. Therefore, our procedure does not output $\bar{\sigma}$ as an ε -constrained stationary Nash equilibrium, which yields the soundness of our procedure.

Completeness. By the promise of the problem, there exists a constrained Nash equilibrium. Hence, by Lemma 7.1 and Lemma 7.3, there exists an $\varepsilon/8$ -constrained stationary Nash equilibrium $\bar{\sigma}$ which is polynomial-size representable. The procedure non-deterministically guesses $\bar{\sigma}$. Since $\bar{\sigma}$ satisfies the payoff constraints up to $\varepsilon/8$ and $|\widehat{\text{val}}_i - \text{val}_i| \leq \varepsilon/8$, it passes the payoff checks. Since the strategy profile is an $\varepsilon/8$ -Nash equilibrium, for all players i , we have $\widetilde{\text{val}}_i \leq \text{val}_i + \varepsilon/8$. Therefore, we get

$$\begin{aligned} \widehat{\text{val}}_i + 3\varepsilon/4 &\geq \text{val}_i + 5\varepsilon/8 && \left(\widehat{\text{val}}_i \text{ is an } \varepsilon/8\text{-approximation of } \text{val}_i \right) \\ &\geq \widetilde{\text{val}}_i + 4\varepsilon/8 && (\widetilde{\text{val}}_i \leq \text{val}_i + \varepsilon/8) \\ &\geq \widetilde{\text{val}}_i + \varepsilon/8. \end{aligned}$$

Thus, the coNP procedure outputs *yes*. Therefore, our procedure successfully decides that $\bar{\sigma}$ is an ε -constrained stationary Nash equilibrium. This yields the completeness of the procedure and completes the proof. $\square$

Chapter 8

On (Very) Small Probabilities

In this chapter, we provide a tight lower bound for Lemma 7.1, by showing that in a constrained NE, players might need to use double-exponentially small probabilities (Theorem 3.7). This result itself is not surprising due to the $\exists\mathbb{R}$ -completeness of the problem [38] even in games with 7 players or more. But we tighten this result by producing an explicit game with only 5 players where such double-exponentially small probabilities are required. The chapter is organized as follows. In Section 8.1, we present the construction of the game $\mathcal{G}^n$. In Section 8.2, we show that every constrained Nash equilibrium in $\mathcal{G}^n$ requires double-exponentially small probabilities, and, in Section 8.3, we show that such a Nash equilibrium indeed exists. Finally, in Section 8.4, we show that the above lower bound does not hold for ε -Nash equilibria in the game $\mathcal{G}^n$, and we present a conjecture that generalizes this observation.

8.1 The Game $\mathcal{G}^n$

Let $n \in \mathbb{N}$. The game $\mathcal{G}^n$ is depicted by Figure 8.1. It contains five players, named $\circ$, $\square$, $\triangle$, $\diamond$, and $\heartsuit$: the shape of each vertex indicates which player controls it, and black vertices are stochastic ones—the probability distribution is then indicated on the outgoing edges. All omitted rewards are 0. For convenience, we allow rewards greater than 1 here, accounting for the fact that all rewards can easily be normalized by dividing them by the greatest of them (namely n). The threshold vectors are $\bar{x}$ and $\bar{y}$, where $x_\circ = y_\circ = 1$, and, for every other player i , we have $x_i = 0$ and $y_i = n$; that is, the constraint requires that player $\circ$'s expected payoff is exactly 1, and leaves the payoffs of the other players unconstrained. Note that the game $\mathcal{G}^n$ has $\mathcal{O}(n)$ vertices and edges. For now, the reader should ignore the notations $\beta_i, \gamma_i, \delta_i, \eta_i$.

Intuitively, the construction is the following: in each gadget def_i , from the vertex r_i , player $\circ$ has to randomize between the terminal vertices t_i and t'_i (in the gadget def_0 the randomization is imposed, because player $\circ$ does not control the vertex r_0). The probability of going to the vertex t_i will be called α_i . We are interested in NEs such that player $\circ$ gets expected payoff 1: that means that the players should never go to those gadgets, nor to the terminal vertices t_{i0} or t_{i1} . But then, deviating and going to those gadgets or vertices must never be a profitable deviation. This imposes strict restrictions on which expected payoffs the players should get in each gadget mul_i , and therefore on how they can randomize their own strategies. In particular, from the vertex c_i , we will see that player $\diamond$ must necessarily

randomize, which is possible in a Nash equilibrium only if both sides are equivalent from their perspective: with this idea, each gadget mul_i binds the values of α_{i-1} and α_i , imposing $\alpha_i = \alpha_{i-1}^2$. Then, by induction, we find $\alpha_n = \frac{1}{2^{2^n}}$.

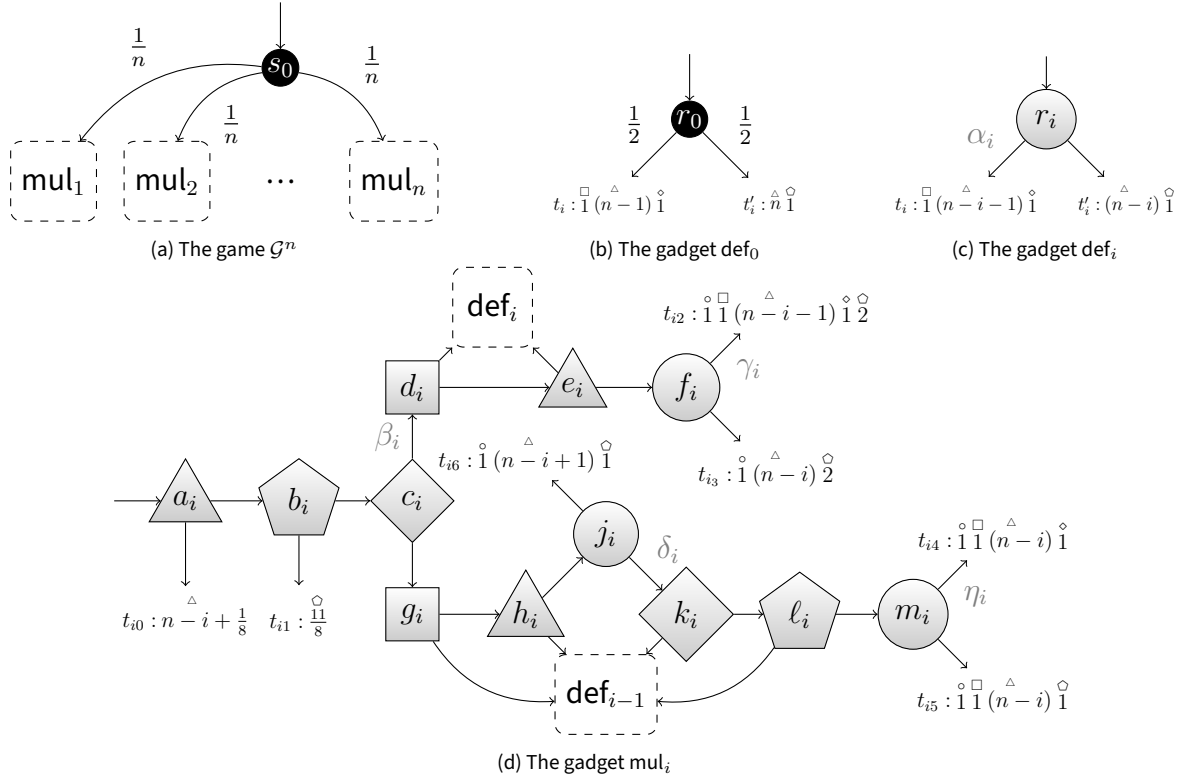


Figure 8.1: A game where very small probabilities are necessary. The rewards corresponding to all the players not explicitly mentioned are zero.

8.2 Small Probabilities are Necessary

In this section, we show that every constrained Nash equilibrium of the game $\mathcal{G}^n$ requires a double-exponentially small probability.

Lemma 8.1. *In the game $\mathcal{G}^n$, if there exists a stationary Nash equilibrium $\bar{\sigma}$ in which player $\bigcirc$ gets expected payoff 1, then we have $\sigma_{\bigcirc}(r_n) = \frac{1}{2^{2^n}}$.*

Proof. Let $\bar{\sigma}$ be such a stationary NE.

For each $i \in \{0, \dots, n\}$, we write $\alpha_i, \beta_i, \gamma_i, \delta_i$, and η_i , for the probabilities $\sigma_{\bigcirc}(r_i)(t_i)$, $\sigma_{\bigcirc}(c_i)(d_i)$, $\sigma_{\bigcirc}(f_i)(t_{i2})$, $\sigma_{\bigcirc}(j_i)(k_i)$, and $\sigma_{\bigcirc}(m_i)(t_{i4})$, respectively. Since the vertex r_0 is stochastic, we fix $\alpha_0 = \frac{1}{2}$. We will now show the following.

Proposition 8.2. *For each i , we have $\alpha_i = \frac{1}{2^{2^i}}$.*

Proof. The proposition is immediately true for $i = 0$.

Let now $i > 0$, and let us assume that we have $\alpha_{i-1} = \frac{1}{2^{2^{i-1}}}$. Let us already note that since player $\bigcirc$ gets expected payoff 1, and since there is no terminal vertex in this game where

they can get more than 1, all terminal vertices where they get less than 1 must be reached with probability 0. That implies for example that from the vertex a_i , which is reached with non-zero probability, player $\triangle$ goes deterministically to b_i , and that from the vertex b_i , player $\diamond$ goes deterministically to the vertex c_i .

This will first help us prove the following claim.

Claim 8.3. *We have $0 < \beta_i < 1$.*

Proof. Since player $\triangle$ and player $\diamond$ go to the terminal vertices t_{i0} and t_{i1} , respectively, with probability 0, and have no profitable deviation in $\bar{\sigma}$, we know that from the vertex c_i , those players must get at least expected payoffs $n - i + \frac{1}{8}$ and $\frac{11}{8}$, respectively. For player $\triangle$, that implies $\beta_i < 1$, since player $\triangle$'s expected payoff is always smaller than or equal to $n - i$ if the vertex d_i is reached. For player $\diamond$ it implies $\beta_i > 0$, since player $\diamond$'s expected payoff is smaller than or equal to 1 if the vertex g_i is reached. $\lrcorner$

Now, since the vertex d_i is reached with non-zero probability, and since no terminal vertex giving player $\square$ the reward 0 must ever be reached, we necessarily have $\sigma_{\square}(d_i)(r_i) = \sigma_{\triangle}(e_i)(r_i) = 0$, and from the vertex d_i player $\square$ and $\triangle$ deterministically go to the vertex f_i . There, we can prove the following claim.

Claim 8.4. *We have $\gamma_i = \alpha_i$.*

Proof. By deviating and going to the gadget def_i , player $\square$ and player $\triangle$ can get the expected payoffs α_i and $n - i - \alpha_i$, respectively. But none of those deviations is profitable, since the strategy profile $\bar{\sigma}$ is an NE. For player $\square$, this means that they get at least the expected payoff α_i from f_i , i.e., that we have $\gamma_i \geq \alpha_i$. For player $\triangle$, this means that they get at least the expected payoff $n - i - \alpha_i$ from f_i , i.e., that we have $\gamma_i \leq \alpha_i$. $\lrcorner$

On the other side, the same argument implies that player $\square$ and $\triangle$ deterministically go to the vertex j_i , and that we have, then, the following result.

Claim 8.5. *We have $\delta_i = \alpha_{i-1}$.*

Proof. Since going to the gadget def_{i-1} is not a profitable deviation for player $\square$, they must get at least the expected payoff α_{i-1} by going to j_i , which implies $\delta_i \geq \alpha_{i-1}$. Since we have $\alpha_{i-1} = \frac{1}{2^{2^{i-1}}} > 0$ by induction hypothesis, we now know that the vertex k_i is reached with positive probability, which implies that from there, player $\diamond$, and then player $\diamond$, deterministically go to the vertex m_i . Thus, if the vertex k_i is reached, player $\triangle$ gets expected payoff $n - i$. Since going to the gadget def_{i-1} is not profitable for them either, they must get at least payoff $n - i + 1 - \alpha_{i-1}$ by going to j_i , that is, we have $\delta_i \leq \alpha_{i-1}$. $\lrcorner$

Now that we know that the vertex k_i is reached with positive probability, and therefore that from there, player $\diamond$ and $\diamond$ go to the vertex m_i , we can apply the same reasoning in that last vertex.

Claim 8.6. *We have $\eta_i = \alpha_{i-1}$.*

Proof. Since going to the gadget def_{i-1} is not a profitable deviation for player $\diamond$, we have $\eta_i \geq \alpha_{i-1}$. Since it is not a profitable deviation for player $\diamond$, we have $\eta_i \leq \alpha_{i-1}$. $\lrcorner$

We can now connect α_{i-1} and α_i . By Claim 8.3, from the vertex c_i , player $\diamond$ proceeds to a strict randomization. Since going deterministically to d_i or to g_i must not be profitable deviations, they must get exactly the same expected payoff on both sides. From d_i , using Claim 8.4, we find that they get expected payoff α_i . From g_i , using Claim 8.5 and Claim 8.6, we find that they get expected payoff α_{i-1}^2 . Using the induction hypothesis, we obtain

$$\alpha_i = \alpha_{i-1}^2 = \frac{1}{2^{2^i}},$$

as desired. ┘

By taking $i = n$, we can conclude. □

8.3 Small Probabilities Suffice

We now know that if a stationary NE satisfying this constraint exists, it necessarily includes a double-exponentially small probability. We now prove that such an NE exists.

Lemma 8.7. *For every $n \in \mathbb{N}$, the game $\mathcal{G}^n$ has a stationary Nash equilibrium where player $\circ$'s expected payoff is 1.*

Proof. The proof of Lemma 8.1 defines almost completely how such a stationary NE must behave. The only probability that was not completely defined was β_i : let us fix it to $\frac{1}{2}$, and let us call this strategy profile $\bar{\sigma}$. Let us now check that no player has a profitable deviation.

Player $\circ$. Player $\circ$ has expected payoff 1 when following $\bar{\sigma}$, and there is no terminal vertex that could give them a better reward: therefore, they have no profitable deviation.

Player $\square$. From each vertex of the form d_i , the play continues deterministically to f_i and then ends in t_{i2} with probability $\gamma_i = \alpha_i$, giving player $\square$ expected payoff α_i . Thus, going to the gadget def_i is not a profitable deviation. From vertices of the form g_i , the situation is analogous.

Player $\triangle$. From each vertex of the form e_i , player $\triangle$ gets expected payoff $n - i - \alpha_i$, hence going to the gadget def_i is not a profitable deviation; and the situation is analogous from the vertices of the form h_i . From each vertex a_i , they get expected payoff

$$\frac{1}{2}(n - i - \alpha_i) + \frac{1}{2}(n - i + 1 - \alpha_{i-1}) = n - i + \frac{1}{2} - \frac{1}{2}\alpha_i - \frac{1}{2}\alpha_{i-1}.$$

Since $\alpha_{i-1} \leq \alpha_0 = \frac{1}{2}$ and $\alpha_i \leq \alpha_1 = \frac{1}{4}$, this expected payoff is greater than or equal to the quantity

$$n - i + \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{2} = n - i + \frac{1}{8},$$

hence going to the terminal vertex t_{i0} is not a profitable deviation either.

Player $\diamond$. From each vertex of the form k_i , player $\diamond$ gets expected payoff α_{i-1} , hence going to the gadget def_{i-1} is not a profitable deviation. From the vertex c_i , player $\diamond$ will get expected payoff $\alpha_i = \alpha_{i-1}^2$ whether they go to d_i or to g_i , hence they have no profitable deviation from c_i either.

Player $\diamond$. From every vertex of the form ℓ_i , player $\diamond$ gets expected payoff $1 - \alpha_{i-1}$, hence going to the gadget def_{i-1} is not a profitable deviation. From a vertex of the form b_i , player $\diamond$ gets expected payoff

$$\frac{1}{2} \cdot 2 + \frac{1}{2} (1 - \alpha_{i-1} + \alpha_{i-1}(1 - \alpha_{i-1})) = \frac{3}{2} - \frac{1}{2} \alpha_{i-1}^2 \geq \frac{3}{2} - \frac{1}{2} \left(\frac{1}{2}\right)^2 = \frac{11}{8},$$

hence they have no profitable deviation from the vertex b_i either. $\square$

Proof of Theorem 3.7. The result follows from combining Lemma 8.1 and Lemma 8.7. $\square$

Theorem 3.7 should be understood as a lower bound on Lemma 7.1, that shows that one cannot expect a better complexity than NP^{NP} using the techniques we use here.

8.4 Approximate Equilibria in the Game $\mathcal{G}^n$

However, let us now note that the above result is true only for exact Nash equilibria. For ε -NEs, our construction does not lead to the same result, as shown by the following lemma.

Lemma 8.8. *Let $n \in \mathbb{N}$ and $\varepsilon > 0$. In the game $\mathcal{G}^n$, there exists a stationary ε -NE $\bar{\tau}$ where player $\circ$'s expected payoff is 1, and where for each player i and every edge uv with $u \in V_i$, the size $\text{bit}(\tau_i(u)(v))$ is bounded by a polynomial of $\text{bit}(n)$ and $\text{bit}(\varepsilon)$.*

Proof. Let I be the smallest positive integer such that we have

$$\frac{1}{2^{2^I+1}n} \leq \varepsilon.$$

Let us consider the strategy profile $\bar{\tau}$, where, for every i :

- from the vertex r_i , player $\circ$ goes to t_i with probability θ_i ;
- from the vertex a_i , the players go deterministically to c_i , where player $\diamond$ randomizes uniformly between d_i and g_i ;
- from the vertex d_i , the players go deterministically to f_i , where player $\circ$ goes to t_{i2} with probability θ_i ;
- from the vertex g_i , the players go deterministically to the vertex j_i , where player $\circ$ goes to k_i with probability θ_{i-1} ; and
- from the vertex k_i , the players go deterministically to the vertex m_i , where player $\circ$ goes to t_{i4} with probability θ_{i-1} ,

where $\theta_0 = \frac{1}{2}$ and, for $1 \leq i \leq n$,

$$\theta_i = \begin{cases} \frac{1}{2^{2^i}} & i < I, \\ 0 & i \geq I. \end{cases}$$

If $I > n$, then $\bar{\tau}$ coincides with the Nash equilibrium from Lemma 8.7, so the result follows. Hence, assume $1 \leq I \leq n$.

Following the same arguments as in the proof of Lemma 8.7, no player has a profitable deviation anywhere in this strategy profile, except player $\diamond$, in the gadget mul_I , who now has an incentive to go deterministically to the vertex g_I . Indeed, if the vertex d_I is reached, they get expected payoff 0, while if g_I is reached, they get $\theta_{I-1}^2 = \frac{1}{2^{2I}}$. Then, since the vertex c_I is itself reached with probability $\frac{1}{n}$, such a deviation is profitable by

$$\frac{1}{n} \cdot \frac{1}{2} \cdot \frac{1}{2^{2I}} = \frac{1}{2^{2I+1}n},$$

which is smaller than ε by hypothesis. Therefore, the strategy profile $\bar{\tau}$ is an ε -NE, and all probabilities can be written with a number of bits that is polynomial in the size of the instance (which includes ε). $\square$

We conjecture that this can be generalized, and that as soon as we consider ε -NEs instead of NEs, double-exponentially small probabilities are no longer required.

Conjecture 8.9. *Let $\mathcal{G}$ be a game, let $\bar{x}, \bar{y}$ be two vectors denoting the constraints, and let $\varepsilon > 0$. If there exists a stationary ε -NE in $\mathcal{G}$ where the expected payoffs are between $\bar{x}$ and $\bar{y}$, then there exists one that has polynomial-size fixed-point representation.*

Proving such a result would lead to an NP upper bound on our approximate problem, since such an ε -NE can then be guessed and checked in polynomial time.

Chapter 9

A Lower Bound: NP-hardness

We dedicate this chapter to show the NP-hardness of the APPROXCONSTRAINEDNE problem (Theorem 3.6), which holds already for games with two players. While it was already known that the CONSTRAINEDNE problem is NP-hard [54, Theorem 4.4], and later shown to be even $\exists\mathbb{R}$ -complete [38], we prove that even the relaxed version of the problem—deciding whether an approximate stationary Nash equilibrium exists that satisfies the given constraints up to ε —is NP-hard.¹

The chapter is organized as follows. In Section 9.1, we present the reduction from the 3SAT problem. In Section 9.2, we show that a satisfying assignment of the formula yields a constrained Nash equilibrium in the constructed game, and, in Section 9.3, we show that even a constrained ε -Nash equilibrium yields a satisfying assignment of the formula.

9.1 The Reduction

We proceed by reduction from the 3SAT problem. Let $\varphi = \bigwedge_{i=1}^m (L_{i1} \vee L_{i2} \vee L_{i3})$ be a CNF formula over the variables $x_1, \dots, x_n$, where each L_{ij} is a literal, that is, a variable x_k or its negation $\neg x_k$. Note that n and m are the parameters of the formula, and that the game $\mathcal{G}^\varphi$ we construct below has $\mathcal{O}(n + m)$ vertices and edges.

Construction of the game $\mathcal{G}^\varphi$. We define the game $\mathcal{G}^\varphi$ as follows.

- There are two players, player $\circ$ and player $\square$.
- Player $\circ$ controls the vertex $\top_L$, for each literal L . Player $\square$ controls the vertex C_i , for each clause C_i , and the vertex L , for each literal L . The stochastic vertices are named s_0 and s_1 . The terminal vertices are named $t_\square$, $t_\emptyset$, $t_{\circ\square}$, and $t_\circ$.
- The initial vertex is the stochastic vertex s_0 , from which the outgoing edges are distributed as follows:
 - the probability of going to the vertex x_k , for each $k \in \{1, \dots, n\}$, is $\frac{1}{2^{k+1}}$;
 - the probability of going to the vertex $\neg x_k$, for each k , is $\frac{1}{2^{k+1}}$;

¹This result is stronger than the one presented in the conference version [10], where the hardness was proven only for an arbitrary number of players.

- the probability of going to the vertex s_1 is $\frac{1}{2^{n+1}}$; and
- the probability of going to the terminal vertex t_o is $\frac{1}{2^{n+1}}$.

Note that these probabilities sum up to 1, since $2 \sum_{k=1}^n \frac{1}{2^{k+1}} = 1 - \frac{1}{2^n}$.

- From the stochastic vertex s_1 , the outgoing edges are distributed as follows: for each $i \in \{1, \dots, m\}$, the probability of going to the vertex C_i is $\frac{1}{m}$.
- From each vertex C_i , player $\square$ can move to the vertices $t_{\square}$, L_{i1} , L_{i2} , and L_{i3} .
- From each vertex L , player $\square$ can move to the vertices $\top_L$ and t_o .
- From each vertex $\top_L$, player $\circ$ can move to the vertices $t_{\circ\square}$ or $t_{\emptyset}$.
- In the terminal vertex $t_{\square}$, player $\circ$ gets the payoff 0 and player $\square$ gets the payoff 1. In the terminal vertex $t_{\emptyset}$, both players receive the payoff 0. In the terminal vertex $t_{\circ\square}$, they both receive 1, and in t_o , player $\circ$ receives 1, and player $\square$ receives 0.

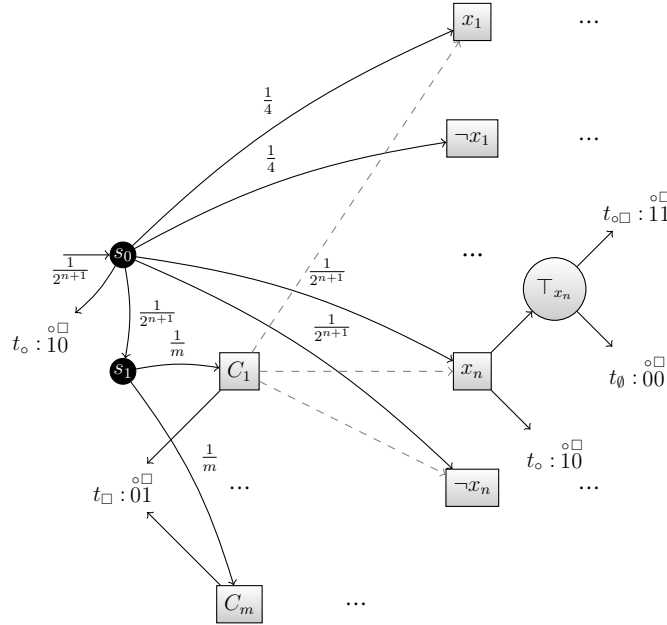


Figure 9.1: A reduction from 3SAT

That game is depicted by Figure 9.1. The gray dashed arrows correspond to the case where $C_1 = x_1 \vee x_n \vee \neg x_n$. For convenience, the vertex t_o is depicted twice.

The instance of the APPROXCONSTRAINEDNE problem. Given the formula φ , the reduction outputs the game $\mathcal{G}^\varphi$, the threshold vectors $\bar{x} := \bar{y} := (1, \frac{1}{2})$, that is, the constraint requires that player $\circ$'s expected payoff is exactly 1 and player $\square$'s expected payoff is exactly $\frac{1}{2}$, and the additive error

$$\varepsilon := 2^{-\kappa}, \quad \text{where } \kappa := n + 3 + \text{bit}(6n + 8m + 1). \quad (9.1)$$

Note that the game, the threshold vectors, and the error are computable in polynomial time with respect to the size of φ , and that $\varepsilon \leq \frac{1}{2^{n+3(6n+8m+1)}}$, since $2^{\text{bit}(k)} \geq k + 1$ for all positive

integers k . The two sections below show that the formula φ is satisfiable if and only if the constructed instance is a positive instance of the APPROXCONSTRAINEDNE problem, which establishes Theorem 3.6.

9.2 From Satisfying Assignments to Nash Equilibria

Proposition 9.1. *If the formula φ is satisfiable, then there exists a stationary Nash equilibrium in the game $\mathcal{G}^\varphi$ such that player $\bigcirc$'s expected payoff is 1, and player $\square$'s expected payoff is $\frac{1}{2}$.*

Proof. We first define the strategy profile $\bar{\sigma}$, then show that it satisfies the constraints, and finally that it is an NE.

Definition of $\bar{\sigma}$. Let $\nu: \{x_1, \dots, x_n\} \rightarrow \{0, 1\}$ be a valuation satisfying the formula φ . We define the stationary strategy profile $\bar{\sigma}$ as follows: from each clause vertex C_i , player $\square$ goes deterministically to a given vertex L such that the literal L is satisfied by the valuation ν . From each vertex of the form L , player $\square$ goes deterministically to the vertex $\top_L$ if the valuation ν satisfies L , and to the vertex $t_\circ$ otherwise. From each vertex of the form $\top_L$, player $\bigcirc$ deterministically goes to the vertex $t_{\circ\square}$ if the valuation ν satisfies L , and to the vertex $t_\emptyset$ otherwise.

The strategy profile $\bar{\sigma}$ satisfies the constraints. First, let us note that this game contains no cycle, hence the probability of never reaching any terminal vertex is 0. This strategy profile never reaches the terminal vertex $t_\square$, nor $t_\emptyset$. Therefore, player $\bigcirc$'s expected payoff is 1.

As concerns player $\square$, they get the payoff 0 only when the vertex $t_\circ$ is reached, i.e., only in plays that either reach $t_\circ$ immediately from s_0 , or that traverse a vertex L such that the literal L is not satisfied. For each k , exactly one of the two literals x_k and $\neg x_k$ is in that situation. That literal is reached with probability exactly $\frac{1}{2^{k+1}}$, since it cannot be reached from s_1 : player $\square$'s strategy targets only satisfied literals. Moreover, the probability of reaching $t_\circ$ immediately is $\frac{1}{2^{n+1}}$. Therefore, the probability of reaching that vertex is

$$\sum_{k=1}^n \frac{1}{2^{k+1}} + \frac{1}{2^{n+1}} = \left(\frac{1}{2} - \frac{1}{2^{n+1}} \right) + \frac{1}{2^{n+1}} = \frac{1}{2},$$

and player $\square$'s expected payoff is $\frac{1}{2}$.

The strategy profile $\bar{\sigma}$ is an NE. Player $\bigcirc$ has no profitable deviation, since they get the expected payoff 1, and could not get more. As for player $\square$, from any vertex of the form C_i , or L where the literal L is satisfied by ν , they also get the expected payoff 1, which is the maximum. From a vertex of the form L where L is not satisfied by ν , they get the payoff 0, but the only deviation available consists in going to the vertex $\top_L$, from which player $\bigcirc$ goes to the vertex $t_\emptyset$, giving them also the payoff 0. $\square$

9.3 From Approximate Nash Equilibria to Satisfying Assignments

For the converse, we show the following stronger result, where the equilibrium is only required to be an ε -NE and to satisfy the constraints up to ε .

Proposition 9.2. Let ε be as in Eq. (9.1). If there is a stationary ε -Nash equilibrium $\bar{\sigma}$ in the game $\mathcal{G}^\varphi$ that satisfies the constraints

- $\mathbb{E}(\bar{\sigma})[\mu_\circ] \geq 1 - \varepsilon$; and
- $\frac{1}{2} - \varepsilon \leq \mathbb{E}(\bar{\sigma})[\mu_\square] \leq \frac{1}{2} + \varepsilon$,

then the formula φ is satisfiable.

Proof. Let $\bar{\sigma}$ be such a stationary ε -NE. For each vertex v and each player $i \in \{\circ, \square\}$, we write $e_i(v)$ for player i 's expected payoff when the strategy profile $\bar{\sigma}$ is played from the vertex v . We first record the following observation, which we use repeatedly. Since the game $\mathcal{G}^\varphi$ is acyclic, every vertex is visited at most once in a play. Therefore, if a vertex v is reached with probability $p > 0$, then, on the one hand, no player has a deviation from the vertex v that is profitable by ε/p or more, since such a deviation would be profitable by ε or more in the whole game; and, on the other hand, since all the payoffs are at most 1, we have $e_\circ(v) \geq 1 - \varepsilon/p$, because $1 - \varepsilon \leq \mathbb{E}(\bar{\sigma})[\mu_\circ] \leq p \cdot e_\circ(v) + (1 - p)$.

We now study the behavior of $\bar{\sigma}$ from the literal vertices.

Claim 9.3. Let L be a literal, either of the form x_k or of the form $\neg x_k$. Let $\delta_k := 2^{k+1}\varepsilon$. Then, we have either $e_\square(L) \leq \delta_k$, or $e_\square(L) \geq 1 - 3\delta_k$.

Proof. The vertex L is reached with probability at least $\frac{1}{2^{k+1}}$. Hence, by the above observation, no player has a deviation from the vertex L that is profitable by δ_k or more, and, moreover, we have $e_\circ(L) \geq 1 - \delta_k$. Let us now write β for the probability $\sigma_\square(L)(\top_L)$, and γ for the probability $\sigma_\circ(\top_L)(t_\square)$.

We have $e_\circ(L) = 1 - \beta + \beta\gamma$. That expected payoff must be larger than or equal to $1 - \delta_k$, which implies

$$\beta - \beta\gamma \leq \delta_k. \quad (9.2)$$

Similarly, we have $e_\square(L) = \beta\gamma$. Since the deviation that consists in going deterministically to $\top_L$ gives player $\square$ the expected payoff γ , and is not profitable by more than δ_k , we obtain

$$\gamma - \beta\gamma \leq \delta_k. \quad (9.3)$$

Note that, by the choice of ε in Eq. (9.1), we have $\delta_k \leq 2^{n+1}\varepsilon \leq \frac{1}{4(6n+8m+1)} < \frac{1}{4}$. We can now separate two cases.

- If $\beta \leq \frac{1}{2}$, then Eq. (9.3) implies $\frac{\gamma}{2} \leq \gamma(1 - \beta) \leq \delta_k$, i.e., $\gamma \leq 2\delta_k$, which implies

$$e_\square(L) = \beta\gamma \leq \frac{1}{2} \cdot 2\delta_k = \delta_k.$$

- If $\beta > \frac{1}{2}$, then Eq. (9.2) implies $\frac{1}{2}(1 - \gamma) \leq \beta(1 - \gamma) \leq \delta_k$, i.e., $\gamma \geq 1 - 2\delta_k$. Now, using Eq. (9.3), we obtain $(1 - 2\delta_k)(1 - \beta) \leq \gamma(1 - \beta) \leq \delta_k$, i.e.,

$$\beta \geq 1 - \frac{\delta_k}{1 - 2\delta_k} = \frac{1 - 3\delta_k}{1 - 2\delta_k},$$

and finally

$$e_\square(L) = \beta\gamma \geq (1 - 2\delta_k) \cdot \frac{1 - 3\delta_k}{1 - 2\delta_k} = 1 - 3\delta_k.$$

Note that the above computations are valid because $\delta_k < \frac{1}{4}$. $\lrcorner$

Let us now call *green* the literals L such that $e_{\square}(L) \geq 1 - 3\delta_k$, and *red* those such that $e_{\square}(L) \leq \delta_k$. By Claim 9.3, every literal is green or red, and, since $\delta_k < \frac{1}{4}$, no literal is both. As one can expect, green and red literals will correspond to true and false literals—but we still need to prove that every clause has a green literal, and that opposite literals have different colors. In the sequel, we fix

$$\delta := 2^{n+1}m\varepsilon,$$

and we note that, by the choice of ε in Eq. (9.1), we have $\delta \leq \frac{m}{4(6n+8m+1)} \leq \frac{1}{32}$, and that $\delta_k \leq \delta$ for all k .

Claim 9.4. *Every clause has a green literal.*

Proof. Let C_i be a clause of φ . The vertex C_i is reached with probability $\frac{1}{2^{n+1}m}$, and hence, by the observation above, no player has a deviation from C_i that is profitable by δ or more, and we have $e_{\circ}(C_i) \geq 1 - \delta$. Since going to the vertex $t_{\square}$ gives player $\circ$ the payoff 0, and all the other successors of C_i give player $\circ$ at most the payoff 1, we have $e_{\circ}(C_i) \leq 1 - \sigma_{\square}(C_i)(t_{\square})$, and therefore

$$\sigma_{\square}(C_i)(t_{\square}) \leq \delta. \quad (9.4)$$

Now, since going deterministically to the vertex $t_{\square}$ gives player $\square$ the payoff 1 and must not be profitable by more than δ , we necessarily have $e_{\square}(C_i) \geq 1 - \delta$. Moreover, by Eq. (9.4), we have

$$e_{\square}(C_i) = \sigma_{\square}(C_i)(t_{\square}) + \sum_{j=1}^3 \sigma_{\square}(C_i)(L_{ij}) \cdot e_{\square}(L_{ij}) \leq \delta + \max_j e_{\square}(L_{ij}),$$

and hence there is a literal L of the clause C_i such that $e_{\square}(L) \geq 1 - 2\delta$. Since $1 - 2\delta > \frac{1}{2} > \delta \geq \delta_k$, the literal L is not red, and hence, by Claim 9.3, it is green. $\lrcorner$

Claim 9.5. *Two opposite literals always have different colors.*

Proof. We first consider a general expression of player $\square$'s expected payoff, namely

$$\mathbb{E}(\bar{\sigma})[\mu_{\square}] = \sum_{k=1}^n \frac{1}{2^{k+1}} (e_{\square}(x_k) + e_{\square}(\neg x_k)) + \frac{1}{2^{n+1}} e_{\square}(s_1), \quad (9.5)$$

where the last term accounts for the plays that traverse the vertex s_1 , and where the plays that reach the vertex $t_{\circ}$ immediately contribute 0. For each literal L , we define $c(L) := 1$ if L is green and $c(L) := 0$ if L is red, and we define

$$S := \sum_{k=1}^n \frac{1}{2^{k+1}} (c(x_k) + c(\neg x_k)) + \frac{1}{2^{n+1}}.$$

We first bound the distance between S and $\mathbb{E}(\bar{\sigma})[\mu_{\square}]$. On the one hand, by Claim 9.3, for all literals L over the variable x_k , we have $|e_{\square}(L) - c(L)| \leq 3\delta_k = 3 \cdot 2^{k+1}\varepsilon$. On the other

hand, by the proof of Claim 9.4, we have $e_{\square}(C_i) \geq 1 - \delta$ for all clauses C_i , and hence $e_{\square}(s_1) = \frac{1}{m} \sum_{i=1}^m e_{\square}(C_i) \geq 1 - \delta$. Therefore, by Eq. (9.5), we have

$$\begin{aligned} |\mathbb{E}(\bar{\sigma})[\mu_{\square}] - S| &\leq \sum_{k=1}^n \frac{1}{2^{k+1}} \cdot 6 \cdot 2^{k+1} \varepsilon + \frac{1}{2^{n+1}} \delta \\ &= 6n\varepsilon + m\varepsilon, \end{aligned}$$

where the last equality uses $\delta = 2^{n+1}m\varepsilon$. Consequently, by the constraint on player $\square$'s expected payoff, we have

$$\left| S - \frac{1}{2} \right| \leq \varepsilon + (6n + m)\varepsilon = (6n + m + 1)\varepsilon \leq 2^{-(n+3)}, \quad (9.6)$$

where the last inequality is by the choice of ε in Eq. (9.1). We now show that Eq. (9.6) implies that, for all k , we have $c(x_k) + c(\neg x_k) = 1$, which is the desired result. Indeed, we define $\theta_k := c(x_k) + c(\neg x_k) - 1 \in \{-1, 0, 1\}$, and note that

$$S - \frac{1}{2} = \sum_{k=1}^n \frac{\theta_k}{2^{k+1}},$$

since $\sum_{k=1}^n \frac{1}{2^{k+1}} + \frac{1}{2^{n+1}} = \frac{1}{2}$. For the sake of contradiction, assume that $\theta_k \neq 0$ for some k , and let k_0 be the smallest such index. Then, we have

$$\left| S - \frac{1}{2} \right| \geq \frac{1}{2^{k_0+1}} - \sum_{k>k_0} \frac{1}{2^{k+1}} = \frac{1}{2^{n+1}},$$

which contradicts Eq. (9.6) and completes the proof of the claim. $\lrcorner$

By Claim 9.5, the valuation ν defined by $\nu(x_k) := 1$ if and only if the literal x_k is green is well defined, and it satisfies exactly the green literals. By Claim 9.4, every clause of φ has a green literal, and hence the valuation ν satisfies the formula φ , which ends the proof. $\square$

We can now conclude the proof of the hardness result.

Proof of Theorem 3.6. Given a 3SAT formula φ , we construct in polynomial time the instance $(\mathcal{G}^\varphi, \bar{x}, \bar{y}, \varepsilon)$ of the APPROXCONSTRAINEDNE problem as described in Section 9.1. We show that the constructed instance satisfies the promise of the problem, and that the instance is positive if and only if the formula φ is satisfiable.

On the one hand, if φ is satisfiable, then, by Proposition 9.1, there exists a stationary Nash equilibrium in $\mathcal{G}^\varphi$ that satisfies the constraints, and hence the instance is in the first case of the promise, i.e., it is a positive instance.

On the other hand, if φ is not satisfiable, then, by the contrapositive of Proposition 9.2, there is no stationary ε -Nash equilibrium in $\mathcal{G}^\varphi$ that satisfies the constraints up to ε , and hence the instance is in the second case of the promise, i.e., it is a negative instance.

Therefore, an algorithm that solves the APPROXCONSTRAINEDNE problem decides the satisfiability of φ , and hence the problem is NP-hard. Since the game $\mathcal{G}^\varphi$ has two players, the hardness holds already for two players, which completes the proof. $\square$

Chapter 10

Conclusion

In this thesis, we presented improved complexity upper bounds and algorithms for the approximate computation of values and of Nash equilibria in stochastic games. We conclude with a discussion of the open questions that our results raise.

Value approximation. For the value-approximation problem for concurrent stochastic games with the limit value of stateful-discounted objectives and with parity objectives, we established TFNP^{NP} upper bounds, and we presented algorithms that improve the dependency on the number of actions in the exponent from linear to logarithmic. There are several interesting directions for future work. First, whether the complexity can be further improved from TFNP^{NP} to TFNP is a major open question, even for reachability objectives. Second, whether for parity objectives, the dependency on d can be improved from linear to logarithmic, retaining the logarithmic dependence on m , is another interesting open question. Finally, the study of priority mean-payoff objectives for concurrent stochastic games and their connection to stateful-discounted objectives is another interesting direction for future work.

Constrained approximate equilibria. For the constrained existence problem of Nash equilibria in multi-player turn-based stochastic games, we studied the following question: assuming the existence of a stationary Nash equilibrium that satisfies given payoff constraints, can we efficiently compute a stationary strategy profile that is an ε -Nash equilibrium and satisfies the constraints up to an additive error ε ? We showed that the decision version of this problem lies in the class NP^{NP} , and the corresponding search problem belongs to FNP^{NP} . On the other hand, we established only an NP-hardness lower bound. This leaves a gap in the known complexity of the problem, and closing this gap remains an open question.

One possible direction is to prove that if an ε -Nash equilibrium exists, then there also exists one in which all probabilities are at least inverse-exponential in the size of the input, assuming the existence of a Nash equilibrium (Conjecture 8.9). Such a result would not contradict our example in Chapter 8, which shows that double-exponentially small probabilities are required in the exact version of the problem. If exponentially small probabilities suffice in the approximate setting, then a strategy profile could be represented using only polynomially many bits in the fixed-point representation. Since such a strategy yields a polynomial certificate, and verification requires only finding the values of MDPs with fixed-point representation, this would yield an NP-completeness result for the decision version.

Alternatively, if such a bound does not hold, and there exists a counterexample where all ε -Nash equilibria satisfying the constraints require double-exponentially small probabilities,

this would suggest an inherent complexity. It would also point to intrinsic limitations in our techniques, analogous to those demonstrated by Hansen, Koucký, and Miltersen [37], who showed that double-exponentially small probabilities are necessary for ε -optimal strategies in concurrent stochastic games. Their result implies that any algorithm that explicitly manipulates these probabilities must use at least exponential space in the worst case, and our results would also imply the same for our problem.

Generality of the model. Although our results for the constrained existence problem focus on turn-based stochastic games with terminal payoffs, we note that the lower bound also holds in more general settings, including concurrent games and broader objectives such as limit-average (mean-payoff) objectives. For the upper bound, we remark that the sentences in the existential theory of the reals that we use for turn-based games with reachability objectives can be naturally adapted to concurrent games with mean-payoff objectives, with only a constant increase in the degree of the polynomials involved. By following similar proof techniques, one can still obtain an FNP^{NP} upper bound for the corresponding constrained ε -NE problem in these more general settings. However, we restricted our presentation to turn-based games with terminal payoffs to maintain clarity.

A common perspective. Finally, we remark that the two parts of the thesis suggest a common perspective on the two solution concepts. In both settings, the exact problem is intractable because of the numbers that a solution requires: the values and the equilibrium probabilities are irrational, and the smallest non-zero probabilities of a solution are doubly exponentially small. In both settings, the approximate problem becomes tractable at the second level of the functional polynomial hierarchy, and the algorithm is the same guess-and-verify procedure, where the certificate is a strategy profile in floating-point representation and the verification is a resolution of single-player games. In both settings, the remaining gap is between the second and the first level of the hierarchy, and, in both settings, closing the gap amounts to the same question: are doubly exponentially small numbers really necessary for approximate solutions? An answer to this question for one of the two solution concepts would likely shed light on the other.

References

- [1] Alekh Agarwal, Nan Jiang, Sham M. Kakade, and Wen Sun. *Reinforcement Learning: Theory and Algorithms*. Book draft. 2021. URL: <https://rltheorybook.github.io/>.
- [2] Luca de Alfaro and Thomas A Henzinger. “Concurrent omega-regular games.” In: *Proceedings Fifteenth Annual IEEE Symposium on Logic in Computer Science*. IEEE. 2000, pp. 141–154.
- [3] Luca de Alfaro, Thomas A Henzinger, and Orna Kupferman. “Concurrent Reachability Games.” In: *Proceedings 39th Annual Symposium on Foundations of Computer Science*. IEEE Computer Society. 1998, pp. 564–575. DOI: 10.1109/SFCS.1998.743507.
- [4] Luca de Alfaro, Thomas A Henzinger, and Rupak Majumdar. “Discounting the future in systems theory.” In: *International Colloquium on Automata, Languages, and Programming*. Springer. 2003, pp. 1022–1037.
- [5] Luca de Alfaro, Thomas A Henzinger, and Freddy YC Mang. “The control of synchronous systems.” In: *International Conference on Concurrency Theory*. Springer. 2000, pp. 458–473.
- [6] Luca de Alfaro, Thomas A Henzinger, and Freddy YC Mang. “The control of synchronous systems, part II.” In: *International Conference on Concurrency Theory*. Springer. 2001, pp. 566–581.
- [7] Luca de Alfaro, Thomas A. Henzinger, and Ranjit Jhala. “Compositional Methods for Probabilistic Systems.” In: *CONCUR 2001 — Concurrency Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 351–365. ISBN: 978-3-540-44685-9.
- [8] Luca de Alfaro and Rupak Majumdar. “Quantitative solution of omega-regular games.” In: *Proceedings of the thirty-third annual ACM symposium on Theory of computing*. 2001, pp. 675–683.
- [9] Rajeev Alur, Thomas A Henzinger, and Orna Kupferman. “Alternating-time temporal logic.” In: *Journal of the ACM (JACM)* 49.5 (2002), pp. 672–713.
- [10] Ali Asadi, Léonard Brice, Krishnendu Chatterjee, and K. S. Thejaswini. “ ϵ -Stationary Nash Equilibria in Multi-Player Stochastic Graph Games.” In: *45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2025, BITS Pilani, K K Birla Goa Campus, India, December 17-19, 2025*. Vol. 360. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, 9:1–9:17. DOI: 10.4230/LIPIcs.FSTTCS.2025.9.

- [11] Ali Asadi, Léonard Brice, Krishnendu Chatterjee, and K. S. Thejaswini. *ε -Stationary Nash Equilibria in Multi-player Stochastic Graph Games*. 2025. arXiv: 2508.15356 [cs.GT]. URL: <https://arxiv.org/abs/2508.15356>.
- [12] Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Jakub Svoboda. “Concurrent Stochastic Games with Stateful-Discounted and Parity Objectives: Complexity and Algorithms.” In: *44th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2024*. Vol. 323. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, 5:1–5:17. DOI: 10.4230/LIPIcs.FSTTCS.2024.5.
- [13] Luc Attia and Miquel Oliu-Barton. “A formula for the value of a stochastic game.” In: *Proceedings of the National Academy of Sciences* 116.52 (2019), pp. 26435–26443.
- [14] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. The MIT Press, 2008. ISBN: 026202649X.
- [15] Erwin H Bareiss. “Sylvester’s identity and multistep integer-preserving Gaussian elimination.” In: *Mathematics of computation* 22.103 (1968), pp. 565–578.
- [16] Saugata Basu. “New results on quantifier elimination over real closed fields and applications to constraint databases.” In: *Journal of the ACM (JACM)* 46.4 (1999), pp. 537–555.
- [17] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in real algebraic geometry*. Vol. 10. Springer, 2006.
- [18] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. “On the combinatorial and algebraic complexity of quantifier elimination.” In: *Journal of the ACM (JACM)* 43.6 (1996), pp. 1002–1045.
- [19] Sougata Bose, Rasmus Ibsen-Jensen, and Patrick Totzke. “Bounded-Memory Strategies in Partial-Information Games.” In: *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science*. LICS ’24. New York, NY, USA: Association for Computing Machinery, 2024. ISBN: 9798400706608. DOI: 10.1145/3661814.3662096. URL: <https://doi.org/10.1145/3661814.3662096>.
- [20] Léonard Brice, Thomas A. Henzinger, and K. S. Thejaswini. “Finding Equilibria: Simpler for Pessimists, Simplest for Optimists.” In: *50th International Symposium on Mathematical Foundations of Computer Science (MFCS 2025)*. Vol. 345. 2025, 30:1–30:18. DOI: 10.4230/LIPIcs.MFCS.2025.30.
- [21] Krishnendu Chatterjee. *Stochastic ω -regular games*. University of California, Berkeley, 2007.
- [22] Krishnendu Chatterjee, Luca de Alfaro, and Thomas A Henzinger. “The complexity of quantitative concurrent parity games.” In: *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*. 2006, pp. 678–687.
- [23] Krishnendu Chatterjee, Rupak Majumdar, and Thomas A Henzinger. “Stochastic limit-average games are in EXPTIME.” In: *International Journal of Game Theory* 37.2 (2008), pp. 219–234.

- [24] Krishnendu Chatterjee, Rupak Majumdar, and Marcin Jurdzinski. “On Nash Equilibria in Stochastic Games.” In: *Computer Science Logic, 18th International Workshop, CSL 2004*. Vol. 3210. Lecture Notes in Computer Science. Springer, 2004, pp. 26–40. doi: 10.1007/978-3-540-30124-0_6.
- [25] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. “The Complexity of Computing a Nash Equilibrium.” In: *SIAM Journal on Computing* 39.1 (2009), pp. 195–259. doi: 10.1137/070699652.
- [26] Argyrios Deligkas, John Fearnley, Themistoklis Melissourgos, and Paul G. Spirakis. “Approximating the Existential Theory of the Reals.” In: *Web and Internet Economics - 14th International Conference, WINE 2018, Oxford, UK, December 15-17, 2018, Proceedings*. Ed. by George Christodoulou and Tobias Harks. Vol. 11316. Lecture Notes in Computer Science. Springer, 2018, pp. 126–139. doi: 10.1007/978-3-030-04612-5_9. URL: https://doi.org/10.1007/978-3-030-04612-5_9.
- [27] Kousha Etessami and Mihalis Yannakakis. “Recursive concurrent stochastic games.” In: *Logical Methods in Computer Science* 4 (2008).
- [28] Hugh Everett. “Recursive games.” In: *Contributions to the Theory of Games* 3.39 (1957), pp. 47–78.
- [29] Jerzy Filar and Koos Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, 1997.
- [30] Vojtěch Forejt, Marta Kwiatkowska, Gethin Norman, and David Parker. “Automated Verification Techniques for Probabilistic Systems.” In: *Formal Methods for Eternal Networked Software Systems: 11th International School on Formal Methods for the Design of Computer, Communication and Software Systems, SFM 2011, Bertinoro, Italy, June 13-18, 2011. Advanced Lectures*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 53–113. ISBN: 978-3-642-21455-4. doi: 10.1007/978-3-642-21455-4_3. URL: https://doi.org/10.1007/978-3-642-21455-4_3.
- [31] Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen. “Approximating the value of a concurrent reachability game in the polynomial time hierarchy.” In: *International Symposium on Algorithms and Computation*. Springer. 2013, pp. 457–467.
- [32] Semyon Aranovich Gershgorin. “Über die Abgrenzung der Eigenwerte einer Matrix.” In: *Izvestija Akademii Nauk SSSR, Serija Matematika* 7.3 (1931), pp. 749–754.
- [33] Hugo Gimbert and Wiesław Zielonka. “Blackwell optimal strategies in priority mean-payoff games.” In: *International Journal of Foundations of Computer Science* 23.03 (2012), pp. 687–711.
- [34] Hugo Gimbert and Wiesław Zielonka. “Discounting infinite games but how and why?” In: *Electronic Notes in Theoretical Computer Science* 119.1 (2005), pp. 3–9.
- [35] Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Peter Bro Miltersen. “The complexity of solving reachability games using value and strategy iteration.” In: *International Computer Science Symposium in Russia*. Springer. 2011, pp. 77–90.

- [36] Kristoffer Arnsfelt Hansen, Michal Koucký, Niels Lauritzen, Peter Bro Miltersen, and Elias P Tsigaridas. “Exact algorithms for solving stochastic games.” In: *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 2011, pp. 205–214.
- [37] Kristoffer Arnsfelt Hansen, Michal Koucký, and Peter Bro Miltersen. “Winning Concurrent Reachability Games Requires Doubly-Exponential Patience.” In: *LICS*. IEEE Computer Society, 2009, pp. 332–341.
- [38] Kristoffer Arnsfelt Hansen and Steffan Christ Sølvsten. “ $\exists\mathbb{R}$ -Completeness of Stationary Nash Equilibria in Perfect Information Stochastic Games.” In: *45th International Symposium on Mathematical Foundations of Computer Science, MFCS 2020*. Vol. 170. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, 45:1–45:15. URL: <https://doi.org/10.4230/LIPIcs.MFCS.2020.45>.
- [39] Narendra Karmarkar. “A new polynomial-time algorithm for linear programming.” In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. 1984, pp. 302–311.
- [40] Claude Lefèvre. “Optimal Control of a Birth and Death Epidemic Process.” In: *Operations Research* 29.5 (1981), pp. 971–982. ISSN: 0030364X, 15265463. URL: <http://www.jstor.org/stable/170234> (visited on 01/06/2025).
- [41] Richard J. Lipton, Evangelos Markakis, and Aranyak Mehta. “Playing large games using simple strategies.” In: *Proceedings 4th ACM Conference on Electronic Commerce (EC-2003), San Diego, California, USA, June 9-12, 2003*. Ed. by Daniel A. Menascé and Noam Nisan. ACM, 2003, pp. 36–41. DOI: 10.1145/779928.779933. URL: <https://doi.org/10.1145/779928.779933>.
- [42] Donald A Martin. “The determinacy of Blackwell games.” In: *The Journal of Symbolic Logic* 63.4 (1998), pp. 1565–1581.
- [43] Jean-François Mertens and Abraham Neyman. “Stochastic games.” In: *International Journal of Game Theory* 10 (1981), pp. 53–66.
- [44] John F. Nash. “Equilibrium points in n -person games.” In: *Proceedings of the National Academy of Sciences* 36.1 (1950), pp. 48–49. DOI: 10.1073/pnas.36.1.48.
- [45] Miquel Oliu-Barton. “New algorithms for solving zero-sum stochastic games.” In: *Mathematics of Operations Research* 46.1 (2021), pp. 255–267.
- [46] Marcus Schaefer and Daniel Štefankovič. “Fixed Points, Nash Equilibria, and the Existential Theory of the Reals.” In: *Theory of Computing Systems* 60.2 (Feb. 2017), pp. 172–193. DOI: 10.1007/s00224-015-9662-0.
- [47] Lloyd Stowell Shapley. “Stochastic games.” In: *Proceedings of the national academy of sciences* 39.10 (1953), pp. 1095–1100.
- [48] Lloyd Stowell Shapley and Roger Snow. “Basic solutions of discrete games.” In: *Contributions to the Theory of Games* 1 (1952), pp. 27–35.
- [49] Dawei Shi, Robert J Elliott, and Tongwen Chen. “On finite-state stochastic modeling and secure estimation of cyber-physical systems.” In: *IEEE Transactions on Automatic Control* 62.1 (2016), pp. 65–80.

- [50] Eilon Solan. “Continuity of the value of competitive Markov decision processes.” In: *Journal of Theoretical Probability* 16 (2003), pp. 831–845.
- [51] Florent Teichteil-Königsbuch, Ugur Kuter, and Guillaume Infantes. “Incremental plan aggregation for generating policies in MDPs.” In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*. AAMAS ’10. International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 1231–1238. ISBN: 9780982657119.
- [52] Wolfgang Thomas. “Languages, automata, and logic.” In: *Handbook of Formal Languages: Volume 3 Beyond Words*. Springer, 1997, pp. 389–455.
- [53] Michael Ummels. “Stochastic multiplayer games: theory and algorithms.” PhD thesis. RWTH Aachen, Germany, Jan. 2010.
- [54] Michael Ummels and Dominik Wojtczak. “The Complexity of Nash Equilibria in Stochastic Multiplayer Games.” In: *Logical Methods in Computer Science* Volume 7, Issue 3 (Sept. 2011). DOI: 10.2168/LMCS-7(3:20)2011. URL: <https://lmcs.episciences.org/1209>.

Appendix A

Declaration of the Use of Generative AI and AI Tools

Generative AI was used in the preparation of this thesis for one purpose only: the typesetting and formatting of the manuscript. The tool used was Claude Code (Anthropic).

Generative AI was not used for the research reported in this thesis. The models, the theorems, the proofs, and the scientific writing are the work of the author and the co-authors of the two publications on which this thesis is based, and were produced without the involvement of such tools. No content of the thesis was generated by an AI system.

The author takes full responsibility and ownership of all content in this thesis, has verified its accuracy, and confirms that it meets established scientific and academic standards.

