



Department of Mathematics
Applied and Numerical Analysis and Optimization and Data Analysis

Evolutionary Games for Global Function Minimization

Master's Thesis by Nathanael Bosch

Examiner: Prof. Dr. Massimo Fornasier

Advisor: Prof. Dr. Massimo Fornasier

Submission Date: October 15, 2018

I hereby confirm that this is my own work, and that I used only the cited sources and materials.

Garching, October 15, 2018

Nathanael Bosch

Abstract

Global optimization is a very important research area due to its large range of applications in many real-world problems in science and engineering. In recent years, evolutionary and swarm principles have been widely researched for intelligent optimization algorithms. In this thesis we propose a new global optimization method based on spatially inhomogeneous evolutionary games.

We review the corresponding theoretical results on well-posedness of the underlying differential equations and present the required measure theoretic preliminaries. In order to improve the application of the model to the problem of global optimization, we propose to adapt the dynamics to reflect the objective value of each player. The resulting ordinary differential equation is well-posed and we show the existence and uniqueness of its Lagrangian solution under compactness assumptions on the initial population.

Finally, we present the numerical results. Our optimization algorithm successfully minimized different non-convex functions in a range of scenarios and even shows exponential convergence. We discuss the observed behavior and analyze specific situations. Finally, we compare our algorithm to another consensus-based optimization method and explain the respective limits and strengths.

Zusammenfassung

Globale Optimierung ist ein überaus wichtiges Forschungsgebiet mit unzähligen Anwendungsmöglichkeiten. In den letzten Jahren wurden in diesem Feld unter anderem auch evolutionäre Prinzipien und Schwarmsimulationen genauer untersucht. Aufbauend auf räumlich inhomogener Spieltheorie stellen wir in dieser Arbeit eine neue globale Optimierungsmethode vor.

Wir wiederholen die theoretischen Ergebnisse zu den zugrundeliegenden Differentialgleichungen und erklären die notwendigen Grundlagen aus der Maßtheorie. Um das Modell besser auf globale Optimierungsprobleme anwenden zu können, schlagen wir eine Änderung der Replikatorgleichungen vor. Das daraus hervorgehende Problem ist gut gestellt. Mit Hilfe einer Kompaktheitsannahme an die Anfangsbevölkerung beweisen wir daraufhin die Existenz und Eindeutigkeit der zugehörigen Langrange-Lösungen.

Anschließend zeigen wir unsere numerische Ergebnisse. Unser Optimierungsalgorithmus minimiert erfolgreich unterschiedliche nicht-konvexe Funktionen mit verschiedenen Anfangswerten und zeigt eine exponentielle Konvergenz. Wir erörtern das beobachtete Verhalten und analysieren spezifische Situationen. Zuletzt vergleichen wir unseren Algorithmus mit einer anderen Optimierungsmethode, welche auch auf einer Konsensusbildung aufbaut, und wir bewerten die jeweiligen Stärken und Einschränkungen.

Contents

Basic Terminology and Notation	ix
1 Introduction	1
2 Spatially Inhomogeneous Evolutionary Games	3
2.1 A Brief Introduction to Game Theory	3
2.2 Evolutionary Game Theory	4
2.2.1 Individuals with pure strategies	4
2.2.2 Infinite or continuous strategies	5
2.2.3 Individuals with mixed strategies	6
2.3 Spatially Inhomogeneous Replicator Dynamics	7
2.4 Measure Theoretic Preliminaries	9
2.4.1 Notation and distances in the space of measures	9
2.4.2 Differentiable curves in the space of measures	14
2.4.3 Ordinary differential equations in the space of measures	15
2.5 Existence and Uniqueness of Lagrangian Solutions	16
2.5.1 Problem statement	16
2.5.2 Structural properties of the interaction term f	18
2.5.3 Stability estimates on b_t	19
2.5.4 Stability estimates on \mathbf{Y}_Λ	19
2.5.5 Proof by contractivity	21
3 Global Function Minimization	23
3.1 Problem Statement	23
3.2 The Optimization Method by Carrillo et al.	24
3.3 Global Optimization with Evolutionary Games	25
3.3.1 Motivation	25
3.3.2 Convergence and solution	26
3.4 Weighted Replicator Dynamics	26
4 Spatially Compact Populations and Weighted Replicator Dynamics	29
4.1 Spatially compact Initial Distributions	29
4.2 Local Bounds on the Payoff Function J	31
4.3 Local Bounds on the Interaction Term f	31
4.4 Properties of the Weight Function w	33

4.5	Existence and Uniqueness of Lagrangian Solutions	37
4.5.1	General problem and solution	37
4.5.2	Stability estimates	39
4.5.3	Proof by contractivity	43
5	Implementation	45
5.1	Discrete Time Solutions	45
5.2	Numerical Adaptations	46
5.2.1	Finite populations and strategies	46
5.2.2	Numerically stable weight computation	47
5.2.3	Different step sizes for strategy and location	48
5.2.4	Parametrized adaptive step size	49
5.3	The Payoff Function J	50
5.4	Global Minimization Algorithm	53
6	Numerical Results	57
6.1	Minimizing a Non-Convex Function	57
6.1.1	Symmetrically distributed starting locations	57
6.1.2	One-sided starting locations	62
6.2	Ackley Benchmark	66
6.3	Modified Ackley Function	70
6.4	The Global Optimization Method by Carrillo et al.	72
6.4.1	Results	72
6.4.2	Comparison with our method	72
6.4.3	Functions with infinitely many minima	75
7	Conclusion	79
List of Figures		81
Bibliography		83

Basic Terminology and Notation

Measure theory: Let (X, d_X) be a metric space.

$C(X)$	space of real continuous functions on X
$\mathcal{M}(X)$	Space of signed Borel measures on X with finite total variation
$\mathcal{M}_+(X)$	Space of nonnegative measures on X
$\mathcal{P}(X)$	Space of probability measures on X
$\mathcal{M}_0(X)$	Space of measures on X with zero mean
$\text{Lip}(f)$	Lipschitz constant of a function f
$\text{Lip}_b(X)$	Space of bounded Lipschitz functions on X
$f_\# \mu$	Push-forward of μ through f , see Definition 2.1
B_r	Open ball of radius r around 0 in (X, d_X)
$W_1(\mu, \nu)$	1-Wasserstein distance between $\mu, \nu \in \mathcal{P}(X)$, see Definition 2.3 and Theorem 2.4
$\ \cdot\ _{\text{TV}}$	Total variation norm, see Definition 2.5
$\ \cdot\ _{\text{Lip}}$	Lipschitz norm, see Definition 2.7
$\ \cdot\ _{\text{BL}}$	Bounded Lipschitz norm, see Definition 2.8

Spatially inhomogeneous evolutionary games:

U	Compact metric space of available strategies
$S := \mathbb{R}^d \times \mathcal{P}(U)$	Playerspace in spatially inhomogeneous evolutionary games
$J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$	Payoff function
$F(U)$	Aarens-Eells space, see Eq. (2.4.8)
$Y := \mathbb{R}^d \times F(U)$	Image space of the interaction term f
$f : S \times S \rightarrow Y$	Interaction term, see Eq. (2.5.4)
$b_\Sigma(t, y)$	Time dependent vector field used for the ODE, see Eq. (2.5.3)
$\mathbf{Y}_\Sigma(t, s, y)$	Induced flow map given the ODE, see Eq. (2.5.5)
\mathcal{A}	Space of possible evolutions of the initial population $\bar{\Sigma}$, see Eq. (2.5.22)

Spatially compact populations and weighted replicator dynamics:

S_0, S_t	Spatially compact playerspace, see Eqs. (4.1.5) and (4.1.6)
$\omega_g^\alpha : \mathbb{R}^d \rightarrow (0, 1]$	Weight function, see Eq. (3.4.1)
$w_\Sigma : S \rightarrow (0, 1]$	Normalized weight function, see Eq. (3.4.2)
$\tilde{b}_\Sigma(t, y)$	Modified time dependent vector field, see Eq. (4.5.4)
$\tilde{\mathbf{Y}}_\Sigma(t, s, y)$	Modified flow map, induced by the ODE Eq. (4.5.5)
$\tilde{\mathcal{A}}$	Space of possible populations, see Eq. (4.1.7)

Global optimization algorithm parameters: See also Section 5.4 for more details.

h	Step size
γ	Parameter for the adaptive step size on the strategy update
M, s_{\min}, s_{\max}	Parameters for the available set of strategies U
ϵ	Parameter for the payoff function J_ϵ , see Section 5.3
x_{\min}, x_{\max}	Define the interval $[x_{\min}, x_{\max}]$ from which we sample the initial locations
I	Maximum number of iterations
α	Parameter used in the weight function w_Σ , see Eq. (3.4.2)

Chapter 1

Introduction

Optimization problems arise in almost every field of science, engineering, and business. The standard approach begins by designing an objective function that can model the problem's objectives while incorporating all the constraints. In most cases, the objective function defines the optimization problem as a minimization task of the form

$$\min_{x \in \Omega} g(x), \quad \Omega \subset \mathbb{R}^d \text{ a domain.} \quad (1.0.1)$$

The function g is then also called the *cost function*, and can be very complicated and highly non-convex. Because of this complexity it is often impossible to find a minimum analytically, and consequently, these problems have to be addressed by numerical algorithms.

Unfortunately, there is often no guarantee to reliably find a global minimum, as the algorithm may be trapped in the local minima of the cost function. A notable example with this behavior is the method of steepest descent, and gradient-based methods in general are often prone to converge to a local minimum. But this issue is well-known, and gradient descent is still commonly used in many tasks where the gradient is fast to compute, as the speed of computation is sometimes more important than the quality of the result. Still, finding the global minimum is always the preferred outcome. This is the main focus in the field of global optimization.

There are many different approaches to the problem, leading to both deterministic and stochastic global optimization methods. In recent years, *metaheuristics* play an increasing role in this field. In general, they may be considered as high level concepts for exploring search spaces by using different strategies, often designed in a way to ensure a balance between exploration of the search space and exploitation of the accumulated search experience [BR03]. Their name is due to the fact that a majority of these methods lack a proper mathematical justification, as it is often not proven that a given method is capable of finding an optimal solution when provided with sufficient information. In many cases, metaheuristics exploit evolutionary principles and swarm intelligence. Notable examples are Evolutionary Algorithms [BFM97; Ree16], the Ant Colony Optimization, Genetic Algorithms, Particle Swarm Optimization and Simulated Annealing [HS88; HKS89].

In this thesis we propose a novel global optimization algorithm. Spatially inhomogeneous evolutionary games [Amb+18] describe the evolution of a population of players, which compete with each other and change their location according to their own current strategy. By modifying the underlying replicator dynamics and appropriately choosing the payoff function, we were able to use this dynamic process to globally minimize functions.

Chapter 2 begins with an introduction to classical evolutionary game theory and presents variations of increasing complexity. By adding a spatial component we arrive at a mathematical formulation of spatially inhomogeneous evolutionary games [Amb+18]. This model was recently proposed by Ambrosio et al. and the authors provide thorough theoretical results on its well-posedness. We present parts of their results and provide the necessary measure theoretic notions.

In Chapter 3 we discuss the application of this method to the problem of global optimization. However, before we develop our method we formally define the mathematical problem and present another consensus-based optimization method by Carrillo et al. [Car+16]. We explain the underlying mechanism of their method and their notion of solution, which provides a useful inspiration. We proceed and motivate the use of evolutionary games for global optimization, and formulate the notion of convergence to a global minimum. Finally, we argue that adaptations to the dynamics of our evolutionary model could be very beneficial in order to successfully minimize a function and we propose a modified replicator dynamics.

We then provide theoretical results on the well-posedness of this adapted model in Chapter 4. We introduce some additional assumptions relating to compactness of the support of the initial population, which leads to a compact support throughout the finite time evolution. We infer boundedness and Lipschitz continuity on the interaction term, the payoff function and the introduced weights. Finally, we retrace the proof as done in Chapter 2 and formally prove the modified parts.

In order to apply the method numerically, we present the construction of discrete time solutions for the continuous differential equation in Chapter 5. We provide a numerically stable way to compute the weights and discuss further adaptations for the practical use of the method. We explain our concrete choice of payoff function, and present the final algorithm as it was used together with an explanation of all its parameters.

Finally, in Chapter 6 we present numerical results. We evaluated our optimization method on multiple non-convex functions with different properties and difficulties. Additionally, we investigate the outcome for varying starting distributions and report their influence. As a comparison, we show results on the performance of the optimization method of Carrillo et al. We discuss the different properties of each algorithm and conclude the thesis.

Chapter 2

Spatially Inhomogeneous Evolutionary Games

In [Amb+18] Ambrosio et al. introduce spatially inhomogeneous evolutionary games. The authors provide general notions of evolutionary game theory and introduce a model, in which each player consists not only of a strategy but also of a location. Using tools from measure theory and calculus in Banach spaces, Ambrosio et al show results on the well-posedness of the occurring ordinary differential equations and prove existence and uniqueness for both Lagrangian and Eulerian solutions.

We start this chapter with a brief introduction to game theory and describe a simple symmetric game, together with the occurring notions and aspects. We proceed and introduce evolutionary games, gradually moving from individuals with pure strategies towards players with locations and continuous mixed strategies. We then provide necessary preliminary notions, in order to conclude this chapter with the formal proof on existence and uniqueness of Lagrangian solutions.

2.1 A Brief Introduction to Game Theory

In classical, non-evolutionary game theory, we often consider symmetrical games. Each player, I and II, has to choose one of n strategies, denoted by u_1, \dots, u_n . If I chooses u_i and II chooses u_j , then player I obtains a *payoff* $J_{i,j}$ and player II obtains $J_{j,i}$. The goal for each player in this game is to obtain the maximal payoff.

One example is the well-known game of Rock-Paper-Scissors. Each player chooses between three available strategies u_1, u_2, u_3 , and the payoff function is described by the matrix

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}. \quad (2.1.1)$$

If a player happens to know the strategy of the other player before playing the game, he will adapt his strategy in order to win. Therefore, if both players knew the chosen strategy of the other player it would create a vicious cycle of mutual outguessing.

In [Neu28], John von Neumann found a way to avoid this dead-end: By randomly selecting one of the strategies, each with equal probability, it becomes impossible for the co-player to adapt towards a winning strategy. If both players use this randomized strategy, then no player has an incentive to deviate. We call these randomized strategies *mixed strategies*. The stable situation in which no player deviates, even with full knowledge of the strategy of the other player, is called a *mixed Nash equilibrium*, first introduced in [VM47] (see also [Rou16] for an introduction to game theory, including many notions of equilibria).

2.2 Evolutionary Game Theory

In *evolutionary game theory* [Sig10] we consider *populations* of players. These players interact according to the classical notion of game theory, evaluating a payoff function after choosing a strategy each. Instead of relating this payoff to a utility of each individual, we relate it to the reproductive success of this player in the population.

Let us again consider the Rock-Paper-Scissors example: If one half of the population uses the strategy “rock” and the other half plays “scissors”, then the fraction playing “scissors” should decrease over time as the payoff of these players is generally lower than the payoff of players playing “rock”.

With this Darwinian evolution concept we create a dynamic population in which the occurrence of strategies changes over time, depending on their current success. In the following sections we formalize the described idea. We introduce different examples and models, starting with populations on a discrete set of strategies and then progressing towards individuals with continuous mixed strategies.

2.2.1 Individuals with pure strategies

Let us consider a finite set of N available strategies $u_1, \dots, u_N \in U$. We denote the fraction of the population using strategy u_i by σ_i . As players only differ by their strategy, we can describe the state of the whole population using only the vector $(\sigma_1, \dots, \sigma_N)$. The payoff of playing strategy u_i against u_j is described by the payoff function $J(u_i, u_j)$, with $J : U \times U \rightarrow \mathbb{R}$.

Consider a player using strategy u_i . In the next encounter the probability of meeting a co-player using strategy u_j is σ_j . The expected payoff for a random encounter is therefore given by the term

$$\sum_{j=1}^N J(u_i, u_j) \sigma_j.$$

To calculate the relative success of this player we need to compare his payoff to the expected payoff of all players in the population. The average payoff in the population

can be similarly described by

$$\sum_{i=1}^N \sigma_i \left(\sum_{j=1}^N J(u_i, u_j) \sigma_j \right),$$

which leads to the relative success of strategy u_i :

$$\Delta_N(u_i) := \sum_{j=1}^N J(u_i, u_j) \sigma_j - \sum_{i=1}^N \sum_{j=1}^N \sigma_i J(u_i, u_j) \sigma_j. \quad (2.2.1)$$

In the final step we describe the evolution of the population. As stated above, we want to relate the relative success of u_i to the reproductive success of the individual playing u_i . We therefore consider the (per capita) growth rate $\dot{\sigma}_i/\sigma_i$ and relate it to the relative success $\Delta_N(u_i)$ in the following differential equation:

$$\frac{\dot{\sigma}_i}{\sigma_i} = \Delta_N(u_i)$$

The evolution of the entire population can now be described by the following system of ordinary differential equations:

$$\dot{\sigma}_i = \left(\sum_{j=1}^N J(u_i, u_j) \sigma_j - \sum_{i=1}^N \sum_{j=1}^N \sigma_i J(u_i, u_j) \sigma_j \right) \sigma_i, \quad i = 1, \dots, N. \quad (2.2.2)$$

This system is known as *replicator dynamics* in the literature of evolutionary games (see [Sig10; HS98; HS83]). It is considered as one of the most popular dynamic game models as its sets of accumulation points and its steady states are closely related to the Nash equilibria of the game described by the payoff matrix $A = (J(u_i, u_j))_{i,j}$ [HS98, Thm. 7.2.1].

2.2.2 Infinite or continuous strategies

In Section 2.2.1 we worked on a finite set of strategies U . We want to extend these notions to infinite or continuous strategies, which is by now largely addressed in literature [Bom90; Cre05; Des+08; Gal11; OHR07; Nor08; OR01; RR14]. It can be viewed as a natural limit for $N \rightarrow \infty$ of the system

$$\sigma_t^N := \sum_{j=1}^N \sigma_{j,t} \delta_{u_j} \in \mathcal{P}(U), \quad (2.2.3)$$

where $\mathcal{P}(U)$ is the space of probability measures over U and δ_{u_j} is the Dirac-delta, which can be described heuristically as a probability measure having all its mass in the point u_j .

For any σ we denote

$$\Delta_\sigma(u) := \left(\int_U J(u, u') d\sigma(u') - \int_U \int_U J(w, u') d\sigma(w) d\sigma(u') \right).$$

We then define the evolution of the probability measure σ_t^N as

$$\dot{\sigma}_t^N = \Delta_\sigma \sigma_t^N,$$

or in weak form

$$\frac{d}{dt} \int_U \varphi(u) d\sigma_t^N(u) = \int_U \varphi(u) \Delta_{\sigma_t^N}(u) d\sigma_t^N(u), \quad (2.2.4)$$

for any $\varphi \in C(U)$. Note that $\Delta_{(\sigma^N)}(u_i) = \Delta_N(u_i)$ for $\sigma^N = \sum_{j=1}^N \sigma_j \delta_{u_j}$, such that the above definition coincides with Eq. (2.2.1) for finite N .

In order to analyze well-posedness of Eq. (2.2.4) for $N \rightarrow \infty$, Ambrosio et al. [Amb+18] assume that the initial conditions $\sigma_0^N \rightharpoonup \bar{\sigma}$ for a given $\bar{\sigma} \in \mathcal{P}(U)$, and then deduce that $\sigma_t^N \rightharpoonup \sigma_t$ for any t , where σ is the solution to

$$\frac{d}{dt} \int_U \varphi(u) d\sigma_t(u) = \int_U \varphi(u) \Delta_{\sigma_t}(u) d\sigma_t(u), \quad \sigma(0) = \bar{\sigma}.$$

Thus, we can view mixed strategies over a continuous set of strategies as the limits of strategies of the form Eq. (2.2.3) for $N \rightarrow \infty$.

2.2.3 Individuals with mixed strategies

Building up on Section 2.2.1, we consider a population of individuals endowed with some *mixed strategy* each, behaving as follows: Players meet randomly to play the game, but instead of having a single pure strategy, players draw a strategy at random, according to their respective mixed strategy.

Each player in this population is fully described by its mixed strategy $\sigma \in \mathcal{P}(U)$ on the set of pure strategies U . The space U is not restricted to the finite case anymore (see Section 2.2.2), allowing for any compact metric space. To generalize further, we can describe the whole population of possibly infinite players by a probability measure on the space of mixed strategies, that is $\Sigma \in \mathcal{P}(\mathcal{P}(U))$. For instance in the simplest case, for a finite number N of players, this can be seen as $\Sigma = \frac{1}{N} \sum_{i=1}^N \delta_{\sigma_i}$, giving one N -th of its mass to each player. Note that this definition does not distinguish two players with the same mixed strategy, which is the desired behavior. The payoff function is again given by some $J : U \times U \rightarrow \mathbb{R}$.

In order to describe how the population should evolve we need to describe how each individual has to adapt its mixed strategy. We first adopt the point of view of a fixed

individual in the population with mixed strategy σ . If this player meets a co-player, who in turn uses the mixed strategy σ' , the expected payoff for choosing strategy u is

$$\mathbb{E}_{u' \sim \sigma'} [J(u, u')] = \int_U J(u, u') d\sigma'(u'). \quad (2.2.5)$$

Therefore, the expected payoff for playing u in any random encounter, given the current population Σ , is

$$\mathbb{E}_{\sigma' \sim \Sigma} [\mathbb{E}_{u' \sim \sigma'} [J(u, u')]] = \int_{\mathcal{P}(U)} \left(\int_U J(u, u') d\sigma'(u') \right) d\Sigma(\sigma'). \quad (2.2.6)$$

The average payoff for the mixed strategy σ of the player is the expected value of the term Eq. (2.2.6) with respect to the probability distribution σ , that is

$$\mathbb{E}_{w \sim \sigma} [\mathbb{E}_{\sigma' \sim \Sigma} [\mathbb{E}_{u' \sim \sigma'} [J(w, u')]]] = \int_U \int_{\mathcal{P}(U)} \int_U J(w, u') d\sigma'(u') d\Sigma(\sigma') d\sigma(w), \quad (2.2.7)$$

which enables us to describe the relative success of picking strategy u , in relation to the current mixed strategy σ , as

$$\Delta_{\Sigma, \sigma}(u) := \int_{\mathcal{P}(U)} \int_U J(u, u') d\sigma'(u') d\Sigma(\sigma') - \int_U \int_{\mathcal{P}(U)} \int_U J(w, u') d\sigma'(u') d\Sigma(\sigma') d\sigma(w). \quad (2.2.8)$$

In order to describe the evolution of the population, we need to describe how each individual has to adapt its mixed strategy. We follow the same principle as in Section 2.2.1: The probability of successful strategies should increase. This is described by

$$\dot{\sigma}(u) = \Delta_{\Sigma, \sigma}(u) \sigma(u), \quad u \in U, \quad (2.2.9)$$

but with $\Delta_{\Sigma, \sigma}$ as defined in Eq. (2.2.8).

2.3 Spatially Inhomogeneous Replicator Dynamics

The previous sections described *spatially homogeneous dynamical games*. In [Amb+18] the authors add a spatial inhomogeneity, by assuming that players do not only consist of a (mixed) strategy, but also of a position in space. The strategies are used in order to play against other players and their evolution follows similar principles as in Section 2.2.3, but additionally they are drawn randomly to evolve the positions of each player.

With these definitions, the individuals of this population can be described as $y = (x, \sigma)$, where $x \in \mathbb{R}^d$ denotes the location and $\sigma \in \mathcal{P}(U)$ denotes a mixed strategy for some compact metric space U . We denote the space of these individuals by $S := \mathbb{R}^d \times \mathcal{P}(U)$, and the whole population can be described as in Section 2.2.3 by a

probability measure over this set, that is $\Sigma \in \mathcal{P}(S)$. Again, we are interested in the evolution of each individual in the population. Let $y = (x, \sigma) \in S$ be such a player.

The player adjusts its location by randomly drawing a strategy u according to its mixed strategy σ , and then changing its position using a suitable function $e : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$. In the simplest case we could have $e(x, u) = u$ with the set of strategies $U \in \mathbb{R}^d$ describing possible directions of movement. As this process happens in continuous time, the velocity of the player is obtained as the expected outcome of this random process, namely

$$\dot{x} = \mathbb{E}_{u \sim \sigma}[e(x, u)] = \int_U e(x, u) d\sigma(u).$$

By defining the map $a : S \rightarrow \mathbb{R}^d$ as

$$a(y) = a(x, \sigma) := \int_U e(x, u) d\sigma(u) \quad (2.3.1)$$

we can now conveniently write $\dot{x} = a(y)$.

The strategy of the player evolves according to similar dynamics as in Section 2.2.3, but taking the spatial dimension into account. The payoff function describing the interactions of particles now depends on both the position and the chosen strategy. We therefore consider a Lipschitz-continuous payoff function $J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$, and then define the relative success of strategy u , compared to the current mixed strategy σ , given the population Σ , as

$$\begin{aligned} \Delta_{\Sigma, (x, \sigma)}(u) &:= \int_S \int_U J(x, u, x', u') d\sigma'(u') d\Sigma(x', \sigma') \\ &\quad - \int_U \int_S \int_U J(x, w, x', u') d\sigma'(u') d\Sigma(x', \sigma') d\sigma(w). \end{aligned}$$

Note that this definition corresponds directly to Eq. (2.2.8), adjusted by including the spatial component. The evolution law for the mixed strategy σ of the player at x is then given by a replicator dynamics, and can be written as

$$\dot{\sigma} = \Delta_{\Sigma, (x, \sigma)} \sigma. \quad (2.3.2)$$

Together with Eq. (2.3.1) we compactly write the evolution for y as

$$\dot{y} = (\dot{x}, \dot{\sigma}) = (a(x, \sigma), \Delta_{\Sigma, (x, \sigma)} \sigma) := b_{\Sigma}(y). \quad (2.3.3)$$

Using tools from measure theory and calculus in Banach spaces Ambrosio et al. [Amb+18] studied the well-posedness of Eq. (2.3.3), together with the existence, uniqueness, and stability of its solutions. Notably, their results cover two different notions of solution: *Lagrangian* and *Eulerian* solutions.

Our focus lies on the Lagrangian notion of solution, which we define in Definition 2.16. In Section 2.5 we briefly retrace the corresponding parts of the proof of existence and uniqueness of Lagrangian solutions, as done by Ambrosio et al., as we later follow their procedure in order to show the same results on a modified replicator dynamics. The proof requires multiple notions and definitions which we therefore introduce in the following section.

2.4 Measure Theoretic Preliminaries

We closely follow [Amb+18, Section 2] and we use the same notation. As we do not prove all of the results contained in [Amb+18] in detail we provide only the necessary definitions.

2.4.1 Notation and distances in the space of measures

In the following let (X, d_x) be a metric space. We denote by $\mathcal{M}(X)$ the space of signed Borel measures in X with finite total variation and by $\mathcal{P}(X)$ its convex subset of probability measures. Further we denote by $\mathcal{M}_+(X)$ and $\mathcal{M}_0(X)$ the subspaces of positive measures and of measures with 0 mean, respectively.

Let $f : X \rightarrow \mathbb{R}$ be a Lipschitz-continuous function. We denote its Lipschitz constant by

$$\text{Lip}(f) := \sup_{\substack{x,y \in X \\ x \neq y}} \frac{|f(x) - f(y)|}{d_X(x, y)}, \quad (2.4.1)$$

and the space of bounded Lipschitz functions on X by $\text{Lip}_b(X)$.

Let $\mu \in \mathcal{M}_+(X)$ a nonnegative Borel measure on X and $f : X \rightarrow Y$ a μ -measurable function. We define the push-forward measure as done in [AFP00, Def. 1.70].

Definition 2.1 (Push-forward measure)

Let (X, \mathcal{E}) and (Y, \mathcal{F}) be measure spaces, and let $f : X \rightarrow Y$ be so that $f^{-1}(F) \in \mathcal{E}$ for all $F \in \mathcal{F}$. For any positive or real measure μ on (X, \mathcal{E}) we define a measure $f_\# \mu$ in (Y, \mathcal{F}) by

$$f_\# \mu(F) := \mu(f^{-1}(F)) \quad \forall F \in \mathcal{F}.$$

Further, the following change of variables formula holds for every $f_\# \mu$ -integrable Borel function $g : Y \rightarrow \mathbb{R}$:

$$\int_Y g d f_\# \mu = \int_X g \circ f d \mu. \quad (2.4.2)$$

Definition 2.2 (Probability measures with finite p -th moment)

Let (X, d_X) be a complete separable metric space, and $\mathcal{P}(X)$ the set of probability measures on X . We define the set of probability measures with finite p -th moment as

the subset

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) : \int_X d_X(x, \bar{x})^p d\mu(x) < +\infty \text{ for some } \bar{x} \in X \right\}.$$

Given a complete and separable metric space (X, d_X) , the *1-Wasserstein distance* endows the space $\mathcal{P}_1(X)$ with a distance. We define it as in [AGS05, Sec. 7.1].

Definition 2.3 (Wasserstein distance)

Let (X, d_X) be a complete separable metric space. The *1-Wasserstein distance* between two probability measures $\mu, \nu \in \mathcal{P}_1(X)$ is defined by

$$W_1(\mu, \nu) := \min \left\{ \int_{X \times X} d_X(x, y) d\Pi(x, y) : \Pi(A \times X) = \mu(A), \Pi(X \times B) = \nu(B) \right\}. \quad (2.4.3)$$

Thanks to the Kantorovich duality, the Wasserstein distance has another equivalent representation which is often more practical to work with. This theorem and its proof can be found in [AGS05, Thm. 6.1.1].

Theorem 2.4 (Duality formula)

The 1-Wasserstein distance, as defined above in Definition 2.3, can equivalently be written in the following way:

$$W_1(\mu, \nu) = \sup \left\{ \int \varphi d(\mu - \nu) : \varphi \in \text{Lip}_b(X), \text{Lip}(\varphi) \leq 1 \right\}. \quad (2.4.4)$$

In [AGS05, Section 7.1] the authors further show that W_1 does indeed define a distance on $\mathcal{P}_1(X)$, such that $(\mathcal{P}_1(X), W_1)$ is a metric space.

Given a compact metric space (U, d) we will now endow the spaces $\mathcal{M}(U)$, $\text{Lip}(U)$ and $(\text{Lip}(U))'$ with appropriate norms.

Definition 2.5 (Total variation norm)

Let $\sigma \in \mathcal{M}(U)$ be a signed Borel measure on U . We define the *total variation norm* as

$$\|\sigma\|_{\text{TV}} := \sup \left\{ \int_U \varphi d\sigma : \varphi \in C(U), |\varphi| \leq 1 \right\}$$

Remark 2.6

$(\mathcal{M}(U), \|\cdot\|_{\text{TV}})$ is a Banach space.

Definition 2.7 (Lipschitz-norm)

For any $\varphi \in \text{Lip}(U)$ we define

$$\|\varphi\|_{\text{Lip}} := \sup_{u \in U} |\varphi(u)| + \text{Lip}(\varphi). \quad (2.4.5)$$

This induces the *bounded Lipschitz norm* (BL norm) on the space of measures. In [Amb+18, Sec. 2.1] Ambrosio et al. introduce it more thoroughly as a norm on the dual of $\text{Lip}(U)$, together with the corresponding discussion about the relation of $\mathcal{P}(U)$ and $(\text{Lip}(U))'$. We chose to not include this aspect as it is not of direct relevance for this thesis, but we refer to [Amb+18, Sec. 2.1] for the stronger definition, as well as to Riesz Theorem [AFP00, Thm. 1.54, Remark 1.57] for a thorough result on the isometrical isomorphy of $(\mathcal{M}(U), \|\cdot\|_{\text{TV}})$ and $(\text{C}(U))'$.

Definition 2.8 (Bounded Lipschitz norm)

For any $\nu \in \mathcal{M}(U)$ we define the bounded Lipschitz norm on the space of Borel measures

$$\|\nu\|_{\text{BL}} := \sup \left\{ \int \varphi d\nu : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\}. \quad (2.4.6)$$

Proposition 2.9

For any $\varphi, \psi \in \text{Lip}(U)$ it holds

$$\|\varphi\psi\|_{\text{Lip}} \leq \|\varphi\|_{\text{Lip}} \|\psi\|_{\text{Lip}}. \quad (2.4.7)$$

We provide the following proof, as it is not included in [Amb+18].

Proof. Using the fact that the Lipschitz constant is always positive, we can prove the proposition with the following computation.

$$\begin{aligned} \|\varphi\psi\|_{\text{Lip}} &= \sup_{u \in U} |(\varphi\psi)(u)| + \text{Lip}(\varphi\psi) \\ &= \sup_{u \in U} |\varphi(u)| |\psi(u)| + \sup_{\substack{u, v \in U \\ u \neq v}} \frac{\varphi(u)\psi(u) - \varphi(v)\psi(v)}{d_U(u, v)} \\ &= \sup_{u \in U} |\varphi(u)| |\psi(u)| + \sup_{\substack{u, v \in U \\ u \neq v}} \frac{\varphi(u)\psi(u) - \varphi(v)\psi(u) + \varphi(v)\psi(u) - \varphi(v)\psi(v)}{d_U(u, v)} \\ &\leq \sup_{u \in U} |\varphi(u)| |\psi(u)| + \sup_{\substack{u, v \in U \\ u \neq v}} \frac{\varphi(u)\psi(u) - \varphi(v)\psi(u)}{d_U(u, v)} + \sup_{\substack{u, v \in U \\ u \neq v}} \frac{\varphi(v)\psi(u) - \varphi(v)\psi(v)}{d_U(u, v)} \\ &\leq \sup_{u \in U} |\varphi(u)| \cdot \sup_{u \in U} |\psi(u)| + \text{Lip}(\varphi) \sup_{u \in U} |\psi(u)| + \text{Lip}(\psi) \sup_{u \in U} |\varphi(u)| \\ &\leq \left(\sup_{u \in U} |\varphi(u)| + \text{Lip}(\varphi) \right) \cdot \left(\sup_{u \in U} |\psi(u)| + \text{Lip}(\psi) \right) \\ &= \|\varphi\|_{\text{Lip}} \|\psi\|_{\text{Lip}}. \end{aligned}$$

□

In [Amb+18] Ambrosio et al. need a linear structure in the space of probability measures, which $\mathcal{P}(U)$ can not provide. They introduce a closed subspace of $(\text{Lip}(U))'$, also known as the *Arens-Eells* space in the literature:

$$F(U) := \overline{\text{span}(\mathcal{P}(U))}^{\|\cdot\|_{\text{BL}}} \subset (\text{Lip}(U))'. \quad (2.4.8)$$

Note that $\mathcal{M}(U) \subset F(U)$.

In the last part of this section we show that for any measures $\mu_1, \mu_2 \in \mathcal{P}(U)$ the distance induced by the bounded Lipschitz norm is equivalent to the 1-Wasserstein distance:

Proposition 2.10

Given two probability measures $\mu_1, \mu_2 \in \mathcal{P}(U)$ it holds

$$\|\mu_1 - \mu_2\|_{BL} \leq W_1(\mu_1, \mu_2) \leq (1 + D_U) \|\mu_1 - \mu_2\|_{BL}, \quad (2.4.9)$$

with

$$D_U := \min_{x_0 \in U} \max_{x_1 \in U} d(x_0, x_1) \leq \text{diam}(U),$$

where $\text{diam}(U) := \sup\{d_U(x, y) | x, y \in U\}$ is the diameter.

We will prove this result in multiple parts.

Lemma 2.11

For any measure $\nu \in \mathcal{M}_0(U)$ and every 1-Lipschitz function $\varphi : U \rightarrow \mathbb{R}$ we have

$$\int_U \varphi d\nu \leq \|\nu\|_{BL} \cdot (D_U + 1).$$

Proof (Lemma 2.11). It is easy to show the influence of shifting and scaling on the Lipschitz constant and the supremum, respectively. That is, for any $a, b \in \mathbb{R}$ it holds

$$\begin{aligned} \text{Lip}\left(\frac{\varphi(x) - a}{b}\right) &= \frac{\text{Lip}(\varphi)}{b}, \\ \sup_{x \in U} \frac{\varphi(x) - a}{b} &= \frac{\sup_{x \in U} \varphi(x) - a}{b}. \end{aligned}$$

Choosing some $x_0 \in U$, this leads to

$$\left\| \frac{\varphi(x) - \varphi(x_0)}{\sup_{x \in U} d(x, x_0) + 1} \right\|_{\text{Lip}} = \frac{\sup_{x \in U} \varphi(x) - \varphi(x_0)}{\sup_{x \in U} d(x, x_0) + 1} + \frac{\text{Lip}(\varphi)}{\sup_{x \in U} d(x, x_0) + 1}.$$

As φ is Lipschitz, we know that

$$\varphi(x) \leq \varphi(x_0) + \text{Lip}(\varphi) \cdot d(x, x_0) \quad \forall x \in U,$$

and therefore also

$$\sup_{x \in U} \varphi(x) \leq \varphi(x_0) + \text{Lip}(\varphi) \cdot \sup_{x \in U} d(x, x_0).$$

It follows

$$\begin{aligned}
 \left\| \frac{\varphi(x) - \varphi(x_0)}{\sup_{x \in U} d(x, x_0) + 1} \right\|_{\text{Lip}} &\leq \frac{\varphi(x_0) + \text{Lip}(\varphi) \cdot \sup_{x \in U} d(x, x_0) - \varphi(x_0)}{\sup_{x \in U} d(x, x_0) + 1} \\
 &\quad + \frac{\text{Lip}(\varphi)}{\sup_{x \in U} d(x, x_0) + 1} \\
 &= \frac{\text{Lip}(\varphi) \cdot (\sup_{x \in U} d(x, x_0) + 1)}{\sup_{x \in U} d(x, x_0) + 1} \\
 &= \text{Lip}(\varphi).
 \end{aligned}$$

As ν has mean 0, the previous statement gives

$$\begin{aligned}
 \int_U \varphi d\nu &= \int_U \varphi(x) d\nu(x) - \int_U \varphi(x_0) d\nu(x) \\
 &= \int_U (\varphi(x) - \varphi(x_0)) d\nu(x) \\
 &= \int_U \left(\frac{\varphi(x) - \varphi(x_0)}{\sup_{z \in U} d(z, x_0) + 1} \right) d\nu(x) \cdot \left(\sup_{x \in U} d(x, x_0) + 1 \right) \\
 &\leq \|\nu\|_{\text{BL}} \cdot \left(\sup_{x \in U} d(x, x_0) + 1 \right).
 \end{aligned}$$

We can conclude the proof by taking the minimum of x_0 , as the previous result holds for any choice $x_0 \in U$.

$$\int_U \varphi d\nu \leq \|\nu\|_{\text{BL}} \cdot (D_U + 1). \quad \square$$

Proposition 2.12

For any measure $\nu \in \mathcal{M}_0(U)$ it holds

$$\|\nu\|_{\text{BL}} \leq \sup \left\{ \int \varphi d\mu : \varphi \in \text{Lip}(U), \text{Lip } \varphi \leq 1 \right\} \leq \|\nu\|_{\text{BL}} \cdot (D_U + 1).$$

Proof (Proposition 2.12). As Lemma 2.11 holds for any 1-Lipschitz function $\varphi : U \rightarrow \mathbb{R}$ it follows

$$\sup \left\{ \int \varphi d\mu : \varphi \in \text{Lip}(U), \text{Lip } \varphi \leq 1 \right\} \leq \|\nu\|_{\text{BL}} \cdot (D_U + 1).$$

For the other inequality we remind of the definition of the bounded Lipschitz norm as

$$\|\nu\|_{\text{BL}} = \sup \left\{ \int \varphi d\nu : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\}.$$

Now as

$$\left\{ \varphi \in \text{Lip}(U) : \|\varphi\|_{\text{Lip}} = 1 \right\} \subset \left\{ \varphi \in \text{Lip}(U) : \text{Lip}(\varphi) = 1 \right\},$$

we conclude the proof with

$$\begin{aligned} \|\nu\|_{\text{BL}} &= \sup \left\{ \int \varphi d\mu : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\} \\ &\leq \sup \left\{ \int \varphi d\mu : \varphi \in \text{Lip}(U), \text{Lip } \varphi \leq 1 \right\}. \end{aligned} \quad \square$$

We can now prove Proposition 2.10.

Proof (Proposition 2.10). Applying Proposition 2.12 to $\nu = \mu_1 - \mu_2$, with $\mu_1, \mu_2 \in \mathcal{P}(U)$, leads to

$$\|\mu_1 - \mu_2\|_{\text{BL}} \leq \sup \left\{ \int \varphi d(\mu_1 - \mu_2) : \varphi \in \text{Lip}(U), \text{Lip } \varphi \leq 1 \right\} \leq \|\mu_1 - \mu_2\|_{\text{BL}} \cdot (D_U + 1).$$

In the final step we use the duality formula Theorem 2.4 and get

$$\|\mu_1 - \mu_2\|_{\text{BL}} \leq W_1(\mu_1, \mu_2) \leq \|\mu_1 - \mu_2\|_{\text{BL}} \cdot (D_U + 1). \quad \square$$

2.4.2 Differentiable curves in the space of measures

When we introduced spatially inhomogeneous evolutionary games we presented a differential equation, Eq. (2.3.2), describing the evolution of the population. As the mixed strategy σ is a probability measure this differential equation lies in the space of measures. Over the course of the following two sections we will discuss the well-posedness of such differential curves, with respect to different properties of σ and for varying norms.

In the following, we consider two curves $t \mapsto \sigma_t \in \mathcal{P}(U)$ and $t \mapsto \nu_t \in F(U)$, for $t \in [0, T]$ and the Arens-Eells space $F(U)$ (see Eq. (2.4.8)). We assume that both σ and ν are continuous with respect to the BL norm. We want to give a meaning to the differential equation

$$\frac{d}{dt} \sigma_t = \nu_t, \quad t \in [0, T]. \quad (2.4.10)$$

Lemma 2.13

The classical formulation of Eq. (2.4.10) as an ODE in the Banach space $F(U)$ is equivalent to the weak formulation of Eq. (2.4.10), written as

$$\int_U \varphi d\sigma_t - \int_U \varphi d\sigma_s = \int_s^t \langle \nu_\tau, \varphi \rangle d\tau, \quad \forall \varphi \in \text{Lip}(U), s, t \in [0, T]. \quad (2.4.11)$$

Remark 2.14

The notation $\langle \nu_\tau, \varphi \rangle$ used the embedding into the dual of $\text{Lip}(U)$:

$$\langle \nu_\tau, \varphi \rangle = \int_U \varphi d\nu_\tau. \quad (2.4.12)$$

Proof. Since ν is continuous, the map

$$t \mapsto N_t = \int_0^t \nu_\tau d\tau$$

is continuously differentiable, and its derivative exists in the classical sense of

$$\lim_{h \rightarrow \infty} \left\| \frac{N_{t+h} - N_t}{h} - \nu_t \right\|_{\text{BL}} = 0 \quad \forall t \in [0, T].$$

With this and Eq. (2.4.12) it follows

$$\int_U \varphi d\sigma_t = \int_U \varphi d\sigma_0 + \int_U \varphi d \left(\int_s^t \nu_\tau d\tau \right),$$

which implies

$$\sigma_t = \sigma_0 + \int_0^t \nu_\tau d\tau \quad \forall t \in [0, T].$$

This concludes the proof. \square

In [Amb+18, Section 2.3], Ambrosio et al. proceed and state similar results in the smaller space $\mathcal{M}(U) \subset F(U)$, assuming $\nu \in \mathcal{M}(U)$ with

$$\int_0^T \|\nu_t\|_{\text{TV}} dt < +\infty.$$

Notably, they show absolute continuity of the curve $t \mapsto \sigma_t \in \mathcal{M}(U)$ with respect to the TV norm. In particular, if $\|\nu_t\|_{\text{TV}} \in L^\infty(0, T)$, then σ_t is Lipschitz continuous. Further, if ν is continuous with respect to the TV norm, then $t \mapsto \sigma_t \in \mathcal{M}(U)$ is even continuously differentiable, namely

$$\lim_{h \rightarrow \infty} \left\| \frac{\sigma_{t+h} - \sigma_t}{h} - \nu_t \right\|_{\text{TV}} = 0 \quad \forall t \in [0, T].$$

2.4.3 Ordinary differential equations in the space of measures

We want to generalize the results of the previous discussion further: Instead of a curve ν_t as the right hand side in Eq. (2.4.10), let us now consider a time dependent family of operators $A : [0, T] \times \mathcal{P}(U) \rightarrow F(U)$. We are considering the following ODE on $F(U)$:

$$\frac{d}{dt} \sigma_t = A(t, \sigma_t), \quad t \in [0, T]. \quad (2.4.13)$$

We briefly restate results for this situation, corresponding to those in Section 2.4.2:

1. We assume that the map $A : [0, T] \times \mathcal{P}(U) \rightarrow F(U)$ is continuous, when both $\mathcal{P}(U)$ and $F(U)$ are endowed with the BL topologies. Then $t \mapsto \sigma_t \in F(U)$ is continuously differentiable, and therefore Eq. (2.4.13) holds in the classical sense, that is

$$\lim_{h \rightarrow \infty} \left\| \frac{\sigma_{t+h} - \sigma_t}{h} - A(t, \sigma_t) \right\|_{\text{BL}} = 0 \quad \forall t \in [0, T]. \quad (2.4.14)$$

2. If $A : [0, T] \times \mathcal{P}(U) \rightarrow \mathcal{M}(U)$ and

$$\sup_{t \in [0, T]} \|A(t, \sigma_t)\|_{\text{TV}} < +\infty,$$

then σ_t is a Lipschitz curve with respect to the total variation norm.

3. If $A : [0, T] \times \mathcal{P}(U) \rightarrow \mathcal{M}(U)$ is continuous with respect to the total variation norm, then Eq. (2.4.14) improves to

$$\lim_{h \rightarrow \infty} \left\| \frac{\sigma_{t+h} - \sigma_t}{h} - A(t, \sigma_t) \right\|_{\text{TV}} = 0 \quad \forall t \in [0, T]. \quad (2.4.15)$$

2.5 Existence and Uniqueness of Lagrangian Solutions

In this section we proceed as in the original approach of Ambrosio et al. in [Amb+18]. We start by restating the formal setting, together with the ODE describing the dynamics of the population. Then, we present the Lagrangian notion of solution and state the main theorem on its existence and uniqueness. In order to prove this theorem we include the multiple stability estimates on the different terms and functions which were introduced, and then use these to conclude with a contraction argument. For more details and proofs of the statements and results we refer to the original paper.

2.5.1 Problem statement

Ambrosio et al. introduce the formal problem and the different notions of solution for the general continuous case in [Amb+18, Section 3.3].

In spatially inhomogeneous games, each player lies in the space $S = \mathbb{R}^d \times \mathcal{P}(U)$, which can be endowed with the distance

$$d_C(y_1, y_2) := |x_1 - x_2| + \|\sigma_1 - \sigma_2\|_{\text{BL}}, \quad y_i = (x_i, \sigma_i), \quad (2.5.1)$$

where $\|\cdot\|_{\text{BL}}$ is the bounded Lipschitz norm as defined in Definition 2.8. Let $\bar{\Sigma} \in \mathcal{P}(S)$ be an arbitrary initial distribution of players. The first component of S , \mathbb{R}^d , is not compact. Therefore, we assume that the first moment of the first marginal of $\bar{\Sigma}$ is finite:

$$\int_S |x| d\bar{\Sigma}(x, \sigma) = M_0 < +\infty. \quad (2.5.2)$$

Since U is compact, Eq. (2.5.2) holds if and only if $\bar{\Sigma} \in \mathcal{P}_1(S)$, where $\mathcal{P}_1(S)$ is the set of probability measures with finite first moment (see also Definition 2.2). For every continuous curve $t \in [0, T] \mapsto \Sigma_t \in \mathcal{P}_1(S)$ we now define a time dependent vector field

$$b_\Sigma(t, y) = b_{\Sigma_t}(y) = \int_S f(y, y') d\Sigma_t(y'), \quad (2.5.3)$$

with $f = (f_x, f_\sigma)$ defined as

$$\begin{cases} f_x(y, y') = f_x(x, \sigma, x', \sigma') := a(x, \sigma) = \int_U e(x, u) d\sigma(u) \\ f_\sigma(y, y') = f_\sigma(x, \sigma, x', \sigma') \\ := \left(\int_{\tilde{U}} J(x, \cdot, x', u') d\sigma'(u') - \int_U \int_U J(x, w, x', u') d\sigma'(u') d\sigma(w) \right) \sigma. \end{cases} \quad (2.5.4)$$

With the Arens-Eells space $F(U) = \overline{\text{span}(\mathcal{P}(U))}^{\|\cdot\|_{BL}}$, the function $f(y, \cdot)$ defines a continuous function with $S \rightarrow \mathbb{R}^d \times F(U) =: Y$. Notice that $f(y, \cdot)$ has at most linear growth by linearity of the integrals in Eq. (2.5.4) and the Lipschitz continuity of J . As $\Sigma_t \in \mathcal{P}_1(S)$ we can interpret the integral in Eq. (2.5.3) as a Bochner integral.

Remark 2.15 (Bochner integration)

The Bochner integral [Boc33] extends the definition of the Lebesgue integral to functions that take values in a Banach space, as the limit of integrals of simple functions. The formal definition can also be found in [Amb+18, Appendix A.2], as well as the integrability criterion which states that f is Σ_t -Bochner integrable. This holds as with $\Sigma_t \in \mathcal{P}_1(S)$ we have

$$\int_S \|f(y, y')\|_Y d\Sigma_t(y') < +\infty.$$

Let us now consider the following ODE in Y :

$$\dot{y}_t = b_{\Sigma_t}(y_t), \quad y_s = y. \quad (2.5.5)$$

A solution to Eq. (2.5.5) satisfies

$$\begin{cases} \dot{x}_t = a(x_t, \sigma_t) = \int_U e(x_t, u) d\sigma_t(u), \\ \dot{\sigma}_t = \left(\int_S \left(\int_U J(x, \cdot, x', u') d\sigma'(u') \right. \right. \\ \left. \left. - \int_U \int_U J(x, w, x', u') d\sigma'(u') d\sigma(w) \right) d\Sigma_t(x', \sigma') \right) \sigma, \end{cases} \quad (2.5.6)$$

Given the ODE in Eq. (2.5.5), we denote by $\mathbf{Y}_\Sigma(t, s, y)$ the induced *flow map* describing the evolution of a player, that is $\mathbf{Y}_\Sigma(t, s, y) := y_t$. This can be interpreted as a map which, given the state $y \in S$ of a player at time $s \in [0, t]$, returns the new state at time $t \in [0, T]$. As this can be applied to every y satisfying the ODE we can transport the whole measure describing the population, such that

$$\hat{\Sigma}_t = \mathbf{Y}_\Sigma(t, 0, \cdot)_\# \bar{\Sigma}. \quad (2.5.7)$$

Ambrosio et al. introduce two different notions of solution for this problem, Lagrangian and Eulerian. In this work we will only consider the Lagrangian solution, which corresponds to the idea given above. The curve Σ , describing the population over the whole time interval $[0, T]$, should be self-transported according to the flow map induced by the ODE. The following definition is included in [Amb+18, Def. 3.1].

Definition 2.16 (Lagrangian solution)

Let $\Sigma \in C^0([0, T]; (\mathcal{P}_1(S), W_1))$ and $\bar{\Sigma} \in \mathcal{P}_1(S)$. We say that Σ is a *Lagrangian solution* starting from $\bar{\Sigma}$, if

$$\Sigma_t = \mathbf{Y}_\Sigma(t, 0, \cdot)_\# \bar{\Sigma} \quad \text{for every } t \in [0, T], \quad (2.5.8)$$

where $\mathbf{Y}_\Sigma(t, s, y)$ are the transition maps associated to the ODE Eq. (2.5.5).

The following theorem is contained in [Amb+18, Thm. 3.2], and states the main result regarding existence and uniqueness of Lagrangian solutions.

Theorem 2.17 (Existence and uniqueness of Lagrangian solutions)

Suppose that $J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$ and $e : \mathbb{R}^d \times U \rightarrow \mathbb{R}$ are Lipschitz maps and let $f : S \times S \rightarrow Y$ be defined as in Eq. (2.5.4). Then, for every initial distribution $\bar{\Sigma} \in \mathcal{P}_1(S)$, there exists a unique Lagrangian solution Σ with flow map $\mathbf{Y}_\Sigma(t, s, y)$.

This theorem contains the main statements about existence and uniqueness of Lagrangian solutions. In the following we retrace the proof by Ambrosio et al. in order to show that this theorem holds.

2.5.2 Structural properties of the interaction term f

The results contained in [Amb+18, Prop. 3.5] state that f is L -Lipschitz with L depending only on L_e , L_J and $\text{diam}(U)$. Additionally, f satisfies the following compatibility condition

$$\forall R > 0 \exists \theta > 0 : \quad y, y' \in S \cap B_R(0) \quad \Rightarrow \quad y + \theta f(y, y') \in S. \quad (2.5.9)$$

As f is L -Lipschitz, it also follows that for any arbitrary point $y_0 \in S$ it holds

$$\|f(y, y')\| \leq \|f(y_0, y_0)\| + L(\|y - y_0\| + \|y' - y_0\|). \quad (2.5.10)$$

2.5.3 Stability estimates on b_t

Let $\Lambda \in C([0, T]; (\mathcal{P}_1(S), W_1))$ be a curve of measures, not necessarily describing a valid evolution of a population. Using the properties from Section 2.5.2 we can show stability estimates on the time dependent vector field $b_\Lambda(t, y)$ as defined in Eq. (2.5.3). The following proposition corresponds to [Amb+18, Prop. 4.3].

Proposition 2.18 (Properties of b_Λ)

Let $\Lambda, \Lambda' \in C([0, T]; \mathcal{P}_1(S))$ and let $b_\Lambda, b_{\Lambda'}$ be defined as in Eq. (2.5.3), with $f : S \times S \rightarrow Y$ as in Section 2.5.2. Then the following properties hold:

- (i) $\|b_\Lambda(t, y)\| \leq \|f(y_0, y_0)\| + L\|y - y_0\| + L \int_S \|y' - y_0\| d\Lambda_t(y')$ for all $y_0 \in Y$;
- (ii) $\|b_\Lambda(t, y) - b_\Lambda(t, z)\| \leq L\|y - z\|$;
- (iii) $\|b_\Lambda(t, y) - b_\Lambda(s, y)\| \leq LW_1(\Lambda_t, \Lambda_s)$;
- (iv) $\|b_\Lambda(t, y) - b_{\Lambda'}(t, y)\| \leq LW_1(\Lambda_t, \Lambda'_t)$;
- (v) If there exists $\bar{R} > 0$ such that $\Lambda_t(S \setminus B_{\bar{R}}(0)) = 0$ for every $t \in [0, T]$, then

$$\forall R > 0 \exists \theta > 0 : y \in S, \|y\| \leq R, t \in [0, T] \Rightarrow y + \theta b_\Lambda(t, y) \in S. \quad (2.5.11)$$

2.5.4 Stability estimates on \mathbf{Y}_Λ

Using these properties we can then apply Brezis' Theorem [Bre73, Sec. I.3, Thm. 1.4, Cor. 1.1] on the well-posedness of ODEs in Banach spaces. The following version is also provided in [Amb+18, Appendix B] together with additional results.

Theorem 2.19 (Well-posedness of ODEs in Banach spaces)

Let $(E, \|\cdot\|_E)$ be a Banach space, C a closed convex subset of E and let $A(t, \cdot) : C \rightarrow E$, $t \in [0, T]$, be a family of operators satisfying the following properties:

- (i) There exists a constant $L \geq 0$ such that

$$\|A(t, c_1) - A(t, c_2)\|_E \leq L\|c_1, c_2\|_E \quad \text{for every } c_1, c_2 \in C \text{ and } t \in [0, T] \quad (2.5.12)$$

- (ii) For every $c \in C$ the map $i \mapsto A(t, c)$ is continuous in $[0, T]$;

- (iii) For every $R > 0$ there exists $\theta > 0$ such that

$$c \in C, \|c\|_E \leq R \Rightarrow c + \theta A(t, c) \in C. \quad (2.5.13)$$

Then for every $\bar{c} \in C$ there exists a unique curve $c : [0, T] \rightarrow C$ of class C^1 satisfying $c_t \in C$ for all $t \in [0, T]$ and

$$\frac{d}{dt} c_t = A(t, c_t), \quad c_0 = \bar{c}. \quad (2.5.14)$$

This theorem, together with the previously stated stability estimates on b_Λ (Proposition 2.18), guarantees the existence of a unique solution to the Cauchy problem

$$\dot{y}_r = b_\Lambda(r, y_r) \quad y_s = y, \quad (2.5.15)$$

inducing a flow map $\mathbf{Y}_\Lambda(t, s, \cdot) : S \rightarrow S$. Ambrosio et al. provide this result in [Amb+18, Cor. 4.4], including additional stability estimates on \mathbf{Y} .

Corollary 2.20 (Properties of the flow maps \mathbf{Y}_Λ)

Let $\Lambda, \Lambda^i \in C([0, T]; \mathcal{P}_1(S))$ and b_Λ as defined in Eq. (2.5.3), with $f : S \times S \rightarrow Y$ L -Lipschitz, and $y_0 \in Y$. Then

- (i) for every $y \in S$ and $s \in [0, T]$ there exists a unique solution $y_t = \mathbf{Y}_\Lambda(t, s, y)$ from $[s, T]$ to S of class C^1 of the Cauchy problem

$$\dot{y}_r = b_\Lambda(r, y_r) = b_{\Lambda_r}(y_r), \quad y_s = y; \quad (2.5.16)$$

- (ii) $\mathbf{Y}_\Lambda(t, 0, y)$ satisfies the estimate

$$\|\mathbf{Y}_\Lambda(t, 0, y) - y_0\| \leq (\|y - y_0\| + tB(\Lambda, t, y_0)) e^{Lt}, \quad (2.5.17)$$

where

$$B(\Lambda, t, y_0) := \|f(y_0, y_0)\| + L \max_{s \in [0, t]} \int_S \|y' - y_0\| d\Lambda_s(y'); \quad (2.5.18)$$

- (iii) $\mathbf{Y}_\Lambda(\cdot, 0, y)$ satisfies the estimate

$$\begin{aligned} \|\mathbf{Y}_\Lambda(t, 0, y) - \mathbf{Y}_\Lambda(t', 0, y)\| \leq \\ |t - t'| \left[B(\Lambda, T, y_0) + L (\|y - y_0\| + TB(\Lambda, T, y_0)) e^{LT} \right]; \end{aligned} \quad (2.5.19)$$

- (iv) $\mathbf{Y}_\Lambda(t, s, \cdot)$ satisfies the estimate

$$\|\mathbf{Y}_\Lambda(t, s, y) - \mathbf{Y}_\Lambda(t, s, y')\| \leq e^{L(t-s)} \|y - y'\|, \quad 0 \leq s \leq t \leq T; \quad (2.5.20)$$

- (v) more generally, $\mathbf{Y}_{\Lambda^1}, \mathbf{Y}_{\Lambda^2}$ satisfy the estimate for $0 \leq s \leq t \leq T$:

$$\begin{aligned} \|\mathbf{Y}_{\Lambda^1}(t, s, y^1) - \mathbf{Y}_{\Lambda^2}(t, s, y^2)\| \leq \\ e^{L(t-s)} \|y^1 - y^2\| + L \int_s^t e^{L(t-\tau)} W_1(\Lambda_\tau^1, \Lambda_\tau^2) d\tau. \end{aligned} \quad (2.5.21)$$

2.5.5 Proof by contractivity

Finally, in [Amb+18, Section 4.4], Ambrosio et al. conclude the proof using a contraction argument. They define the complete metric space

$$\mathcal{A} := \left\{ \Lambda \in C([0, T]; (\mathcal{P}_1(S), W_1)) : \Lambda_0 = \bar{\Sigma} \right\}, \quad (2.5.22)$$

and a map $\mathcal{T} : \mathcal{A} \rightarrow \mathcal{A}$ as

$$\mathcal{T}[\Lambda]_t := \mathbf{Y}_\Lambda(t, 0, \cdot)_\# \bar{\Sigma}, \quad (2.5.23)$$

where $\mathbf{Y}_\Lambda(t, 0, \cdot)$ is the flow map associated to b_Λ . Using the properties shown earlier, notably Corollary 2.20, the authors were able to show that \mathcal{T} is a contraction. The unique existence of a fixed point then follows using Banach's fixed-point theorem for metric spaces, and as this fixed point corresponds exactly to a Lagrangian solution of the original problem this concludes the proof of Theorem 2.17.

Chapter 3

Global Function Minimization

We first introduce and define the problem of global function minimization. In [Car+16], Carrillo et al. proposed a solution to this problem. We present their consensus-based global optimization method, giving a brief overview on the main components, and we summarize the results of their paper. We then proceed and discuss how to use evolutionary games in order to solve global function minimization.

Using the previously provided introduction to spatially inhomogeneous evolutionary games, we explain how the movement and evolution of players evolves over time and how this can be influenced. Finally, we claim that the existing method lacks an important aspect in order to be applicable to function minimization. In the original replicator dynamics each player is equivalent, but we argue that players with a low function value should be more influential in this dynamic process. Inspired by the optimization method of Carrillo et al., we introduce a weight function and adapt the replicator dynamics accordingly.

3.1 Problem Statement

We consider a multi-dimensional optimization problem of the form

$$\min_{x \in \Omega} g(x) \quad (3.1.1)$$

for a given cost function $g \in C(\Omega)$ on the compact domain $\Omega \subset \mathbb{R}^d$. Without loss of generality, we may assume g to be positive and defined on the whole space \mathbb{R}^d , by extending it outside Ω without changing its global minimum. We will use the notation

$$x_* := \arg \min g(x), \quad g_* := g(x_*). \quad (3.1.2)$$

Further, we assume g to be locally Lipschitz. For a formal definition of local Lipschitz continuity see Definition 4.6.

3.2 The Optimization Method by Carrillo et al.

In [Car+16] Carrillo et al. propose a method to solve the problem of global optimization. The authors use a particle- and consensus-based approach where the population is described by a stochastic process. It does not directly relate to evolutionary game theory, but we include it as a comparison. We present the main stochastic differential equations which govern the evolution of the population and explain the resulting algorithm. In Section 6.4 we also include numerical results for this method. Notably, one mechanism in the approach of Carrillo et al. is particularly interesting to us, and we will use it to improve our own optimization algorithm.

The authors consider a stochastic system of $N \in \mathbb{N}$ agents with position $X_t^i \in \mathbb{R}^d$, described by the stochastic differential equations

$$dX_t^i = -\lambda (X_t^i - m_t) dt + \sigma |X_t^i - m_t| dW_t^i, \quad (3.2.1)$$

$$m_t = \sum_{i=1}^N X_t^i \left(\frac{\omega_g^\alpha(X_t^i)}{\sum_{j=1}^N \omega_g^\alpha(X_t^j)} \right), \quad (3.2.2)$$

with $\lambda, \sigma > 0$ and the weight function $\omega_g^\alpha(x) = e^{-\alpha g(x)}$ for some appropriately chosen $\alpha > 0$.

This stochastic system can be interpreted in the following way. The term m_t constitutes a weighted mean of the positions of all agents, with exponentially increasing weight as the function value decreases. This implies that for α large enough this mean is located very close to the agent with minimum location. Agents then move in average towards this mean, with noise added proportional to the distance between an agent and the weighted mean. This enables agents with a position differing a lot from m_t to explore a larger portion of the graph of $g(x)$, while agents closer to m_t diffuse much less. Additionally, this allows for a stable consensus, as once all particles arrive at the same location there is no added noise anymore and the particles stop moving. The well-posedness of this system is discussed in [Car+16, Section 2].

Similarly to our procedure in Section 2.2.2 the authors move towards a continuous setting by defining the empirical measure

$$\rho_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}, \quad (3.2.3)$$

and then taking the limit $N \rightarrow \infty$. The authors formally reformulate the process and the dynamics, and show well-posedness in the continuous case in [Car+16, Section 3].

The main results are then shown in [Car+16, Section 4], proving the convergence to a uniform consensus under mild assumptions on the objective function g , that is

$$\rho_t \rightarrow \delta_{\hat{x}} \quad \text{as } t \rightarrow \infty, \quad (3.2.4)$$

for some $\hat{x} \in \mathbb{R}^d$ possibly depending on the initial density ρ_0 . By choosing $\alpha \gg 1$ sufficiently large this point \hat{x} may be made arbitrarily close to the global minimum x_* . Additionally, the convergence is shown to happen exponentially in time.

After further theoretical results on the pseudo-inverse distribution in [Car+16, Section 5], the authors present numerical results, minimizing the Ackley function [Ack87] (see also Eq. (6.2.1) and Fig. 6.9). They were able to confirm their results and show exponential convergence to the global minimum.

3.3 Global Optimization with Evolutionary Games

In the following we motivate the application of evolutionary games to the problem of global function minimization. We introduced the notion of players and populations in Chapter 2 and presented the general process and the dynamics in Section 2.3. Notably, in our method the population consists of players, each with its own strategy. This level of individuality does not appear in the previously presented method from Carrillo et al., where individuals are purely characterized by their location. We are therefore required to think in a different way when discussing a potential minimization method, always considering the individual players and the underlying game.

3.3.1 Motivation

In Section 2.3 we described the dynamics in the population, consisting of the location update by randomly sampling a strategy and the mixed strategy update depending on the relative success of each pure strategy. The ODEs in Eq. (2.5.6) then follow from this general description. However, there are three objects that are considered as given, which we could modify in order to influence the dynamics: The set of strategies U , the velocity field $e : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$, and the payoff function $J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$. From Theorem 2.17 it follows that for a specific U , e , and J , each initial population $\bar{\Sigma}$ has exactly one clear evolution, according to the dynamics Eq. (2.5.6).

In order to optimize a function, the idea is now to chose U , e and J in a way which guarantees the occurrence of an equilibrium, located exactly in the global minimum of the function. For example, a reasonable simplification which greatly increases the intuitive understanding of the process is the choice $U \subset \mathbb{R}^d$ and $e(x, u) = u$, such that U directly describes possible directions of movement. The exact choices as they were used in our optimization algorithm will be presented in Section 5.4.

3.3.2 Convergence and solution

Given the initial population $\bar{\Sigma} \in \mathcal{P}(S)$ we denote by $\Sigma : [0, T] \rightarrow \mathcal{P}(S)$ the Lagrangian solution as defined in Definition 2.16, which is the population that arises by following the dynamics described in Eq. (2.5.6).

We say that the population reached a steady state if there exists some $\hat{t} < T$ such that

$$\Sigma_t = \hat{\Sigma} \quad \text{for } \hat{t} \leq t \leq T, \quad (3.3.1)$$

where $\hat{\Sigma} \in \mathcal{P}(S)$. This does not yet imply convergence towards a single location or even convergence towards a global minimum, but it provides an important point of view: In order to successfully develop our algorithm we have to assure the existence of population states which do not change over time.

To successfully find the global minimum of a function we want to converge towards a stable state of the form

$$\hat{\Sigma} = \delta_{\hat{y}},$$

with the point of consensus $\hat{y} = (\hat{x}, \hat{\sigma})$ located very close to the global minimum $x_* = \arg \min g(x)$, that is

$$\|\hat{x} - x_*\| \leq \epsilon \quad \text{for some } \epsilon > 0. \quad (3.3.2)$$

3.4 Weighted Replicator Dynamics

The payoff function is the main component for the evolution of the mixed strategy of each player, which in turn decides on its movement. Therefore, it plays a central role in the dynamics of the population. At the same time, the payoff function defines a game between two players, thus it does not contain any information on the general state of the population but only describes 1-on-1 interactions. We argue that it is necessary to implement an additional mechanism, as the interactions with players $y = (x, \sigma)$ with a low function value $g(x)$ should have more influence on the population dynamics.

Motivated by the procedure of Carrillo et al. we give exponentially more weight to players with low function values. Let $\omega_g^\alpha(x) : \mathbb{R}^d \rightarrow (0, 1]$ be defined as in [Car+16], that is

$$\omega_g^\alpha(x) := \exp(-\alpha g(x)). \quad (3.4.1)$$

We introduce the weight function $w_{\Sigma_t} : S \rightarrow (0, 1]$, defined for any distribution $\Sigma_t \in \mathcal{P}_1(S)$ as

$$w_{\Sigma_t}(y) = w_{\Sigma_t}(x, \sigma) := \frac{\omega_g^\alpha(x)}{\int_S \omega_g^\alpha(x') d\Sigma_t(x', \sigma')}. \quad (3.4.2)$$

Note that with this choice, for any $\Sigma_t \in \mathcal{P}_1(S)$ and any $y \in S$ it directly follows $w_{\Sigma_t}(y) \in (0, 1]$ and

$$\int_S w_{\Sigma_t}(y) d\Sigma_t(y) = 1. \quad (3.4.3)$$

We then modify the ODE describing the strategy evolution, as given in Eq. (2.5.6), to the following ODE

$$\begin{aligned} \dot{\sigma}_t = & \left(\int_C w_{\Sigma_t}(x', \sigma') \left(\int_U J(x, \cdot, x', u') d\sigma'(u') \right. \right. \\ & \left. \left. - \int_U \int_U J(x, w, x', u') d\sigma'(u') d\sigma(w) \right) d\Sigma_t(x', \sigma') \right) \sigma. \end{aligned} \quad (3.4.4)$$

The evolution in location does not depend on interactions with other players, so that our alteration only influences the dynamics of how the mixed strategy of each individual gets adjusted.

Chapter 4

Spatially Compact Populations and Weighted Replicator Dynamics

We have by now introduced spatially inhomogeneous evolutionary games and we argued that it is possible to use this framework for global function minimization. In order to do so, we introduced a weighted version of the replicator dynamics.

In this chapter we will retrace the proof as done by Ambrosio et al. in [Amb+18] for our modified setting. In order to do so we introduce additional assumptions on the initial distribution, notably compact support in its first component, and use it in order to deduce bounds and Lipschitz properties of J , f , ω and w . Using these we can then reformulate the proof following the lines of Section 2.5.

4.1 Spatially compact Initial Distributions

In the following we will show that given an initial distribution $\bar{\Sigma}$ with compact support in its spatial component, the resulting population Σ_t will also always be compactly supported at any time $t \in [0, T]$, and additionally we can even calculate its boundary.

In order to simplify the discussion and notation we will start by making stronger assumptions on the set of strategies U and the function e , but they can actually be relaxed as we show in Remark 4.1 and an analogous discussion could still be done.

We consider the compact set of strategies $U \subset \mathbb{R}^d$ with

$$U \subset B_{R_U} \tag{4.1.1}$$

for some $R_U > 0$, where $B_r := \{x \in \mathbb{R}^d : \|x\| \leq r\}$ is the closed ball of radius $r > 0$ around 0. We also consider the velocity field $e(x, u) = u$. This leads to the following dynamic describing the change in location of some player $y = (x, \sigma)$

$$\dot{x} = a(y) = a(x, \sigma) = \int_U u d\sigma(u). \tag{4.1.2}$$

As $U \subset B_{R_U}$ it follows

$$|\dot{x}| \leq \int_U |u| d\sigma(u) \leq R_U. \tag{4.1.3}$$

Given some initial location $x_0 \in \mathbb{R}^d$ this leads to

$$|x_t| \leq |x_0| + t \cdot R_u. \quad (4.1.4)$$

Therefore, if the initial location of all particles is bounded, there exists a bound at any time $t > 0$.

Let us consider an initial population $\bar{\Sigma} \in \mathcal{P}(S)$ with compact support in its first marginal. This implies that there exists some $R_0 > 0$ so that

$$\text{supp}(\bar{\Sigma}) \subset B_{R_0} \times \mathcal{P}(U) =: S_0 \quad (4.1.5)$$

The flow map induced by a Lagrangian solution describing the evolution of a population, as defined in Definition 2.16, has to satisfy the ODE given in Eq. (4.1.2). Using Eq. (4.1.4) it follows that at any time $t \in [0, T]$ it holds

$$\text{supp}(\Sigma_t) \subset B_{R_t} \times \mathcal{P}(U) =: S_t, \quad (4.1.6)$$

where $R_t = R_0 + t \cdot R_U$. Note that $S_0 \subset S_t \subset S_T$ holds.

Motivated by the previous discussion, we now define the space $\tilde{\mathcal{A}} \subset \mathcal{A}$ as

$$\tilde{\mathcal{A}} := \left\{ \Lambda \in C([0, T]; (\mathcal{P}_1(S), W_1)) : \Lambda_0 = \bar{\Sigma} \text{ and } \text{supp}(\Lambda_t) \subset S_t \text{ for all } t \in [0, T] \right\}, \quad (4.1.7)$$

with \mathcal{A} as in Eq. (2.5.22). Any curve describing a valid population starting from $\bar{\Sigma}$, according to the dynamics in both location and strategy, has to be contained in $\tilde{\mathcal{A}}$.

Remark 4.1

We can relax the assumptions on the velocity field e and on the exact structure of U , as long as U is compact and e is Lipschitz continuous, as originally assumed in [Amb+18]. Indeed, we can still find a boundary on $|\dot{x}_t|$ given the original ODE $\dot{x} = \int_U e(x, u) d\sigma(u)$, by using the fact that e is L_e -Lipschitz:

$$\begin{aligned} |\dot{x}| &\leq \int_U |e(x, u)| d\sigma(u) \\ &\leq \sup_{u \in U} |e(x, u)| \\ &\leq \sup_{u \in U} |e(0, u)| + L_e \cdot |x|, \end{aligned}$$

leading to

$$|\dot{x}| \leq e^* + L_e \cdot |x|, \quad (4.1.8)$$

with $e^* := \sup_{u \in U} |e(0, u)|$. It follows

$$|x_t| \leq |x_0| \cdot \exp(t \cdot L_e) + t \cdot e^*. \quad (4.1.9)$$

Using this we can define

$$R_T := R_0 \cdot \exp(t \cdot L_e) + t \cdot e^*, \quad (4.1.10)$$

which also leads to a bound at time t . For any dynamic process following a spatial evolution as described in Eq. (2.5.6), the evolution of a spatially bounded initial distribution will always stay spatially bounded.

4.2 Local Bounds on the Payoff Function J

As both the spatial component of the population and the set of strategies U are compact, we can now deduce upper bounds on the payoff function J , using its Lipschitz property.

Lemma 4.2 (Bound on J)

Let $J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$ be L_J -Lipschitz. Then, for any $x, x' \in B_{R_T}$ and $u, u' \in B_{R_U}$ for some $R_U, R_T > 0$ the following bound holds

$$|J(x, u, x', u')| \leq |J(0, 0, 0, 0)| + 2L_J(R_T + R_U) =: \bar{J}. \quad (4.2.1)$$

Proof. As J is L_J -Lipschitz, it has to be L_J -Lipschitz in each variable. Thus for any $x, x' \in B_{R_T}$ and $u, u' \in B_{R_U}$ we can conclude

$$\begin{aligned} |J(x, u, x', u')| &\leq |J(0, 0, 0, 0)| + L_J(|x| + |u| + |x'| + |u'|) \\ &\leq |J(0, 0, 0, 0)| + 2L_J(R_T + R_U) =: \bar{J}. \end{aligned} \quad \square$$

4.3 Local Bounds on the Interaction Term f

Using the bound on the payoff function J we argue that there exist similar boundaries on the interaction term $f(y, y')$, as long as the population is contained in S_T .

Lemma 4.3 (Bound on f)

For any $y, y' \in S_T$, with S_T as defined in Eq. (4.1.6) given some $R_U, R_0 > 0$, it holds

$$\|f(y, y')\| \leq 2\bar{J} + L_J + R_U. \quad (4.3.1)$$

In order to prove this statement we will need the following estimate, which can also be found in [Amb+18, Lemma. 3.4].

Lemma 4.4

Let $\sigma \in \mathcal{P}(U)$ and $z \in \text{Lip}(U)$. Then the following estimate holds

$$\|z\sigma\|_{BL} \leq \|z\|_{Lip} \|\sigma\|_{BL}. \quad (4.3.2)$$

Remark 4.5

Here $z\sigma$ can be interpreted as a measure, in the sense of densities, according to $z\sigma(A) = \int_A z d\sigma$. For more information see the Radon-Nikodym theorem, presented e.g. in [AFP00, Thm. 1.28].

We start with the proof of Lemma 4.4.

Proof (Lemma 4.4). The following computation proves the statement.

$$\begin{aligned}
 \|z\sigma\|_{BL} &= \sup \left\{ \int_U \varphi d(z\sigma) : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\} \\
 &= \sup \left\{ \int_U \varphi(u) z(u) d\sigma(u) : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\} \\
 &\leq \left(\sup_{u \in U} z(u) \right) \cdot \sup \left\{ \int_U \varphi(u) d\sigma(u) : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\} \\
 &\leq \left(\sup_{u \in U} z(u) + \text{Lip}(z) \right) \cdot \sup \left\{ \int_U \varphi(u) d\sigma(u) : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \right\} \\
 &= \|z\|_{\text{Lip}} \cdot \|\sigma\|_{BL}.
 \end{aligned}$$

□

We can now proceed with the proof of Lemma 4.3

Proof (Lemma 4.3). Recall that we defined the norm on $Y = \mathbb{R}^d \times F(U)$ such that we have

$$\|f(y, y')\|_Y = \|f_x(y, y')\| + \|f_\sigma(y, y')\|_{BL},$$

with $f = (f_x, f_\sigma)$ defined as in Eq. (2.5.4). In order to find a boundary on $\|f_x(y, y')\|$, we use the fact that U is compact:

$$\|f_x(y, y')\| = \|f_x((x, \sigma), (x', \sigma'))\| = \left\| \int_U u d\sigma(u) \right\| \leq R_u.$$

Now for the second part, we use Lemma 4.4 as well as the L_J -Lipschitz property of

the payoff function J and the boundary shown in Lemma 4.2. It follows

$$\begin{aligned}
 \|f_\sigma(y, y')\| &= \|f_\sigma((x, \sigma), (x', \sigma'))\|_{BL} \\
 &= \left\| \left(\int_U J(x, \cdot, x', u') d\sigma'(u') - \int_{U \times U} J(x, w, x', u') d\sigma'(u') d\sigma(w) \right) \sigma \right\|_{BL} \\
 &\leq \left\| \int_U J(x, \cdot, x', u') d\sigma'(u') - \int_{U \times U} J(x, w, x', u') d\sigma'(u') d\sigma(w) \right\|_{Lip} \cdot \|\sigma\|_{BL} \\
 &= \left(\sup_{u \in U} \left| \int_U J(x, u, x', u') d\sigma'(u') - \int_{U \times U} J(x, w, x', u') d\sigma'(u') d\sigma(w) \right| \right. \\
 &\quad \left. + \text{Lip} \left(\int_U J(x, \cdot, x', u') d\sigma'(u') - \int_{U \times U} J(x, w, x', u') d\sigma'(u') d\sigma(w) \right) \right) \|\sigma\|_{BL} \\
 &\leq (2 \sup \{|J(x_1, u_1, x_2, u_2)| : x_1, x_2 \in B_{R_T}, u_1, u_2 \in B_{R_U}\} + L_J) \cdot \|\sigma\|_{BL} \\
 &\leq (2\bar{J} + L_J) \cdot \sup_{\substack{\varphi \in \text{Lip}(U) \\ \|\varphi\|_{Lip} \leq 1}} \int_U \varphi(u) d\sigma(u) \\
 &\leq (2\bar{J} + L_J) \cdot \sup_{\substack{\varphi \in \text{Lip}(U) \\ \|\varphi\|_{Lip} \leq 1}} \sup_{u \in U} \varphi(u) \\
 &\leq 2\bar{J} + L_J,
 \end{aligned}$$

as for any Lipschitz continuous function φ with $\|\varphi\|_{Lip} \leq 1$ we also have that $\sup_{u \in U} |\varphi(u)| \leq 1$. This concludes the proof and provides the following boundary on the interaction term:

$$\|f(y, y')\| \leq 2\bar{J} + L_J + R_U. \quad (4.3.3)$$

□

4.4 Properties of the Weight Function w

Before we can retrace the proof as done in Section 2.5, we still need additional properties on the weight function w , which we introduced in Eq. (3.4.2). In this section we will show boundedness and Lipschitz continuity of both ω_g^α and w_{Σ_t} on the space S_T , for any $\Sigma \in \widetilde{\mathcal{A}}$ and $t \in [0, T]$. Additionally we will show stability estimates of w_{Σ_t} with respect to the measure Σ_t .

In the following parts of this chapter we extensively use the local Lipschitz continuity. For completeness we include the following definition.

Definition 4.6 (Local Lipschitz continuity)

Given two metric spaces (X, d_X) and (Y, d_Y) , where d_X denotes the metric on the set X and d_Y is the metric on set Y , a function $f : X \rightarrow Y$ is called locally Lipschitz

continuous if for every $x \in X$ there exists a neighborhood U of x such that f restricted to U is Lipschitz continuous.

The following proposition makes the property more usable for our needs.

Proposition 4.7

If (X, d_X) is a locally compact metric space, then f is locally Lipschitz if and only if it is Lipschitz continuous on every compact subset of X .

Proof. The definition of a locally compact metric space implies that for every $x \in X$ there exists a compact neighborhood $U \ni x$. Therefore, if f is Lipschitz continuous on every compact subset of X it follows that f is Lipschitz on all compact neighborhoods of any point $x \in X$, which proves second part of the proposition.

In order to prove the first part, we suppose that there exists some compact subset $C \subset X$ such that f is not Lipschitz continuous on C . Then the function

$$S(x, y) := \frac{d_Y(f(x_n), f(y_n))}{d_X(x_n, y_n)}$$

has to be unbounded for $x, y \in C$, $x \neq y$, and there exist two sequences $x_n, y_n \in C$ such that $S(x_n, y_n) \rightarrow \infty$ as $n \rightarrow \infty$. Since C is compact there exist convergent subsequences $x_{n_k} \rightarrow \bar{x}$ and $y_{n_k} \rightarrow \bar{y}$ for $k \rightarrow \infty$. As f is continuous it has to be bounded on the compact set C , thus implying that $\bar{x} = \bar{y}$. Now by definition of f being locally Lipschitz, there has to exist a neighborhood U around \bar{x} such that $f|_U$ is Lipschitz continuous, leading to $S(x_{n_k}, y_{n_k}) \leq L$ for some finite $L > 0$. This contradiction concludes the proof. \square

Lemma 4.8 (ω_g^α is bounded and locally Lipschitz)

Let g be a positive locally Lipschitz continuous function and $\alpha > 0$. Then the function ω_g^α , as defined in Eq. (3.4.1) is locally Lipschitz and there exists $\omega_ > 0$ so that for any $t \in [0, T]$ and any $y = (x, \sigma) \in S_t$ it holds*

$$\omega_g^\alpha(x) \geq \omega_*. \quad (4.4.1)$$

Proof. We first prove the lower bound. For any $t \in [0, T]$ we have $S_t \subset S_T$. It is therefore sufficient to show that there exists $\omega_* > 0$ so that

$$\omega_* \leq \inf_{(x, \sigma) \in S_T} \omega_g^\alpha(x) = \inf_{x \in B_{R_T}} \omega_g^\alpha(x). \quad (4.4.2)$$

As g is continuous its supremum on a compact set has to be finite. Therefore there exists some $g^* \in \mathbb{R}$ so that for any $x \in B_{R_T}$ it holds $g^* \geq g(x)$. Using this we conclude

the proof of the lower bound with

$$\begin{aligned} \inf_{x \in B_{R_T}} \omega_g^\alpha(x) &= \inf_{x \in B_{R_T}} \exp(-\alpha g(x)) \\ &= \exp\left(-\alpha \left(\sup_{x \in B_{R_T}} g(x)\right)\right) \\ &\geq \exp(-\alpha(g^*)) \\ &=: \omega_*. \end{aligned}$$

In order to prove the local Lipschitz property of ω_g^α we use the fact that g is locally Lipschitz. As \mathbb{R}^d is a locally compact metric space, g is $L_{g,C}$ -Lipschitz continuous on any compact subspace $C \subset \mathbb{R}^d$. The exponential function is 1-Lipschitz on negative numbers, thus

$$\begin{aligned} |\omega_g^\alpha(x_1) - \omega_g^\alpha(x_2)| &= |\exp(-\alpha g(x_1)) - \exp(-\alpha g(x_2))| \\ &\leq |-\alpha(g(x_1) - g(x_2))| \\ &= \alpha |g(x_1) - g(x_2)| \\ &\leq \alpha L_{g,C} |x_1 - x_2|. \end{aligned}$$

This concludes that ω_g^α is Lipschitz on any compact space $C \subset \mathbb{R}^d$, and it is therefore locally Lipschitz. \square

Remark 4.9

The discussion in Section 4.1 was done in order to show, that it is enough to consider the space S_T , given some initial parameters $R_U, R_0 > 0$. As ω_g^α is locally Lipschitz it is notably Lipschitz on the space B_{R_T} , and we denote its Lipschitz constant there with L_ω .

We will now show similar results on w .

Lemma 4.10 (w bounded and Lipschitz)

Let g be a positive and locally Lipschitz function and $\alpha > 0$, let with S_t defined as in Eq. (4.1.6) for some $R_U, R_0 > 0$. For any $\Sigma \in \widetilde{\mathcal{A}}$ we define w_{Σ_t} as in Eq. (3.4.2) for any $t \in [0, T]$. Then w_{Σ_t} is Lipschitz on S_T and it holds

$$|w_{\Sigma_t}(y)| \leq \frac{1}{\omega_*}, \quad (4.4.3)$$

for any $y \in S$, ω_* as in Lemma 4.8.

Proof. We first prove the bound. As Σ_t is a probability distribution with $\text{supp}(\Sigma_t) \subset S_t \subset S_T$ and $0 < \omega_g^\alpha \leq 1$, it follows

$$\begin{aligned} |w_{\Sigma_t}(y)| &= \frac{\omega_g^\alpha(x)}{\int_S \omega_g^\alpha(x') d\Sigma_t(x', \sigma')} \\ &\leq \frac{1}{\int_S \omega_g^\alpha(x') d\Sigma_t(x', \sigma')} \\ &= \frac{1}{\int_{S_T} \omega_g^\alpha(x') d\Sigma_t(x', \sigma')} \\ &\leq \frac{1}{\inf_{(x', \sigma') \in S_T} \omega_g^\alpha(x')} \\ &\leq \frac{1}{\omega_*}. \end{aligned}$$

Next we show the Lipschitz property of w_{Σ_t} on S_T . As stated in Remark 4.9, the function ω_g^α is L_ω -Lipschitz on the space B_{R_T} . Using this together with the lower bound ω_* on ω_g^α , it follows that for any $y_1, y_2 \in S_T$ we have

$$\begin{aligned} |w_{\Sigma_t}(y_1) - w_{\Sigma_t}(y_2)| &= |w_{\Sigma_t}(x_1, \sigma_1) - w_{\Sigma_t}(x_2, \sigma_2)| \\ &= \left| \frac{\omega_g^\alpha(x_1)}{\int_S \omega_g^\alpha(x') d\Sigma_t(x', \sigma')} - \frac{\omega_g^\alpha(x_2)}{\int_S \omega_g^\alpha(x') d\Sigma_t(x', \sigma')} \right| \\ &\leq \frac{|\omega_g^\alpha(x_1) - \omega_g^\alpha(x_2)|}{\int_S \omega_g^\alpha(x') d\Sigma_t(x', \sigma')} \\ &\leq \frac{L_\omega |x_1 - x_2|}{\omega_*} \\ &\leq \frac{L_\omega}{\omega_*} (|x_1 - x_2| + \|\sigma_1 - \sigma_2\|_{BL}) \\ &= \frac{L_\omega}{\omega_*} d_S(y_1, y_2). \end{aligned}$$

This concludes the proof. \square

Remark 4.11

As $S_0 \subset S_t \subset S_T$ it directly follows that w_{Σ_t} is also Lipschitz on S_t for any $t \in [0, T]$.

Lemma 4.12 (Stability estimate on w)

Let $\Lambda, \Lambda' \in \mathcal{A}$ and $t, s \in [0, T]$. Let $w_{\Lambda_t}, w_{\Lambda'_s}$ be defined according to Eq. (3.4.2). Then for any $y \in S$ we have

$$\|w_{\Lambda_t}(y) - w_{\Lambda'_s}(y)\| \leq \frac{L_\omega}{\omega_*^2} \cdot W_1(\Lambda_t, \Lambda'_s). \quad (4.4.4)$$

Proof. As $\Lambda, \Lambda' \in \widetilde{\mathcal{A}}$ it follows that the support of both Λ and Λ' is bounded by S_T . Using this, Lemma 4.8, and the Lipschitz-property of ω_g^α on B_{R_T} , we can conclude the proof with the following computation:

$$\begin{aligned}
 \|w_{\Lambda_t}(y') - w_{\Lambda'_s}(y')\| &= \left\| \frac{\omega_g^\alpha(x')}{\int_S \omega_g^\alpha(z) d\Lambda_t(z, \sigma)} - \frac{\omega_g^\alpha(x')}{\int_S \omega_g^\alpha(z) d\Lambda'_s(z, \sigma)} \right\| \\
 &= \frac{\left\| \omega_g^\alpha(x') \cdot \left(\int_S \omega_g^\alpha(z) d\Lambda'_s(z, \sigma) - \int_S \omega_g^\alpha(z) d\Lambda_t(z, \sigma) \right) \right\|}{\left\| \int_S \omega_g^\alpha(z) d\Lambda_t(z, \sigma) \right\| \cdot \left\| \int_S \omega_g^\alpha(z) d\Lambda'_s(z, \sigma) \right\|} \\
 &\leq \frac{\left\| \omega_g^\alpha(x') \right\| \cdot \left\| \int_S \omega_g^\alpha(z) d(\Lambda'_s - \Lambda_t)(z, \sigma) \right\|}{\omega_*^2} \\
 &\leq \frac{1}{\omega_*^2} \cdot \left\| \int_{S_T} \omega_g^\alpha(z) d(\Lambda'_s - \Lambda_t)(z, \sigma) \right\| \\
 &\leq \frac{1}{\omega_*^2} L_\omega W_1(\Lambda_t, \Lambda'_s). \quad \square
 \end{aligned}$$

4.5 Existence and Uniqueness of Lagrangian Solutions

With these preliminary results we can now retrace the proof as done in Section 2.5. Most notably we will prove the stability estimates on the modified vector field \widetilde{b}_Σ , but most other parts can be proven exactly as done by Ambrosio et al. in [Amb+18] so that we often refer to the original proofs. For completeness we provide all results and statements, with the adapted notation where appropriate.

4.5.1 General problem and solution

We assume the initial distribution $\bar{\Sigma}$ to have compact support, that is

$$\text{supp}(\bar{\Sigma}) \subset S_0, \quad (4.5.1)$$

for some $R_0 > 0$. For any compact set of strategies U we can then define S_t, R_t and $\widetilde{\mathcal{A}}$ as done in Remark 4.1 and Section 4.1.

Following [Amb+18, Section 3.3] we now define the problem and state the main theorem regarding its solution. By endowing S with the distance

$$d_S(y_1, y_2) := |x_1 - x_2| + \|\sigma_1 - \sigma_2\|_{BL} \quad y_i = (x_i, \sigma_i). \quad (4.5.2)$$

it follows for the initial distribution that the first moment of the first marginal is finite:

$$\int_S |x| d\bar{\Sigma}(x, \sigma) = \int_{S_0} |x| d\bar{\Sigma}(x, \sigma) \leq \sup_{(x, \sigma) \in S_0} |x| = R_0 < \infty. \quad (4.5.3)$$

Therefore $\bar{\Sigma} \in \mathcal{P}_1(S)$. For every continuous curve $\Sigma \in C([0, T]; \mathcal{P}_1(S))$ we can then define a time dependent vector field

$$\tilde{b}_\Sigma(t, y) = \tilde{b}_{\Sigma_t}(y) = \int_S w_{\Sigma_t}(y') f(y, y') d\Sigma_t(y'), \quad (4.5.4)$$

with the unchanged interaction term $f = (f_x, f_\sigma)$ as defined in Eq. (2.5.4) and the weight function $w_{\Sigma_t} : S \rightarrow (0, 1]$ as defined in Eq. (3.4.2). The functions $w_{\Sigma_t}(\cdot)$ and $f(y, \cdot)$ are both continuous. Note also that as the support of Σ_t is bounded and both $w_{\Sigma_t}(\cdot)$ and $f(x, \cdot)$ are Lipschitz continuous, then their product on this bounded space is also Lipschitz continuous, thus providing a growth at most linear. Therefore $w_{\Sigma_t}(\cdot)f(y, \cdot)$ is Bochner integrable and we can therefore interpret the integral above as a Bochner integral (see Remark 2.15). We will now consider the following ODE in Y , using the modified vector field \tilde{b}_Σ :

$$\dot{y}_t = \tilde{b}_{\Sigma_t}(y_t), \quad y_s = y. \quad (4.5.5)$$

A solution to Eq. (4.5.5) then also satisfies

$$\begin{cases} \dot{x}_t &= a(x_t, \sigma_t) = \int_U e(x_t, u) d\sigma_t(u), \\ \dot{\sigma}_t &= \left(\int_C w_{\Sigma_t}(x', \sigma') \left(\int_U J(x, \cdot, x', u') d\sigma'(u') \right. \right. \\ &\quad \left. \left. - \int_U \int_U J(x, w, x', u') d\sigma'(u') d\sigma(w) \right) d\Sigma_t(x', \sigma') \right) \sigma, \end{cases} \quad (4.5.6)$$

which is an adjusted version of the unweighted case (see Eq. (2.5.6)). We then denote by $\tilde{\mathbf{Y}}_\Sigma(t, s, y)$ the flow map induced by Eq. (4.5.5).

With these adaptations we now want to show existence and uniqueness of a Lagrangian solution, following its definition in Definition 2.16, but this time regarding the ODE Eq. (4.5.5). This can be stated in the form of the following theorem.

Theorem 4.13 (Adapted: Existence and Uniqueness of Lagrangian solution) Suppose that $J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$ and $e : \mathbb{R}^d \times U \rightarrow \mathbb{R}$ are Lipschitz maps. Let $f : S \times S \rightarrow Y$ be defined as in Eq. (2.5.4), and $w_{\Sigma_t} : S \rightarrow (0, 1]$ be defined as in Eq. (3.4.2) for any $\Sigma_t \in \mathcal{P}_1(S)$. Then, for every initial distribution $\bar{\Sigma} \in \mathcal{P}_1(S)$ with support in S_0 there exists a unique Lagrangian solution Σ associated to the ODE Eq. (4.5.5). That is, for \tilde{b}_Σ as defined in Eq. (4.5.4) the associated flow map $\tilde{\mathbf{Y}}_\Sigma(t, s, \cdot)$ satisfies

$$\Sigma_t = \tilde{\mathbf{Y}}_\Sigma(t, 0, \cdot) \# \bar{\Sigma} \quad \text{for every } 0 \leq t \leq T. \quad (4.5.7)$$

Moreover, there exists $\tilde{L} \geq 0$ such that for every pair of initial data $\bar{\Sigma}^1, \bar{\Sigma}^2 \in \mathcal{P}_1(S)$, the corresponding solutions Σ_t^1, Σ_t^2 satisfy

$$W_1(\Sigma_t^1, \Sigma_t^2) \leq e^{\tilde{L}t} W_1(\bar{\Sigma}^1, \bar{\Sigma}^2) \quad \forall t \in [0, T]. \quad (4.5.8)$$

In order to prove this theorem we proceed as done in [Amb+18] (see also Section 2.5). As we did not change the interaction term f the original results as stated in Section 2.5.2 still hold, and we can directly proceed by showing the stability estimates on \tilde{b}_Σ and $\tilde{\mathbf{Y}}$.

4.5.2 Stability estimates

As we modified the time dependent vector field $\tilde{b}_\Sigma(t, \cdot) : S \rightarrow Y$, as defined in Eq. (4.5.4), it is necessary to formally prove stability estimates similar to [Amb+18, Prop. 4.3] in order to continue with the proof.

Proposition 4.14 (Properties of \tilde{b}_t)

Let $\Lambda, \Lambda' \in \mathcal{A}$, let $\tilde{b}_\Lambda, \tilde{b}_{\Lambda'}$ be defined as in Eq. (4.5.4), with f as in Eq. (2.5.4) L -Lipschitz and w as in Eq. (3.4.2). Then

- (i) $\|\tilde{b}_\Lambda(t, y)\| \leq \|f(y_0, y_0)\| + \tilde{L} \|y - y_0\| + \tilde{L} \int_S \|y' - y_0\| d\Lambda(y')$ for all $y_0 \in Y$;
- (ii) $\|\tilde{b}_\Lambda(t, y) - \tilde{b}_\Lambda(t, z)\| \leq \tilde{L} \|y - z\|$;
- (iii) $\|\tilde{b}_\Lambda(t, y) - \tilde{b}_\Lambda(s, y)\| \leq \tilde{L} W_1(\Lambda_t, \Lambda_s)$;
- (iv) $\|\tilde{b}_\Lambda(t, y) - \tilde{b}_{\Lambda'}(t, y)\| \leq \tilde{L} W_1(\Lambda_t, \Lambda'_t)$;

for some $\tilde{L} \geq L$, depending on the exact choice of weight function w , payoff function J and strategy space U .

Remark 4.15

While the constant \tilde{L} in this proposition is larger than in the original version, the actual dependence on the outer factors does not change by much. Section 2.5.2 states dependence of L on L_e, L_J and $\text{diam}(U)$, so that the only additional dependence we introduced is on the weight function w .

Proof. As $\sup |w_{\Lambda_t}(y')| = 1$, property (i) follows by applying Eq. (2.5.10):

$$\begin{aligned} \|\tilde{b}_\Lambda(t, y)\| &= \left\| \int_S w_{\Lambda_t}(y') f(y, y') d\Lambda_t(y') \right\| \\ &\leq \int_S |w_{\Lambda_t}(y')| \|f(y, y')\| d\Lambda_t(y') \\ &\leq 1 \cdot \int_S (\|f(y_0, y_0)\| + L(\|y - y_0\| + \|y' - y_0\|)) d\Lambda_t(y') \\ &= \|f(y_0, y_0)\| + L \|y - y_0\| + L \int_S \|y' - y_0\| d\Lambda_t(y') \\ &\leq \|f(y_0, y_0)\| + \tilde{L} \|y - y_0\| + \tilde{L} \int_S \|y' - y_0\| d\Lambda_t(y'). \end{aligned}$$

(ii) follows from the L-Lipschitz property of f in its first argument, together with Λ_t being a probability measure and $\int_S w_{\Lambda_t}(y) d\Lambda_t(y) = 1$

$$\begin{aligned}
 \|\tilde{b}_\Lambda(t, y) - \tilde{b}_\Lambda(t, z)\| &= \left\| \int_S w_{\Lambda_t}(y') (f(y, y') - f(z, y')) d\Lambda_t(y') \right\| \\
 &\leq \int_S |w_{\Lambda_t}(y')| \|f(y, y') - f(z, y')\| d\Lambda_t(y') \\
 &\leq \int_S |w_{\Lambda_t}(y')| (L \|y - z\|) d\Lambda_t(y') \\
 &\leq \int_S |w_{\Lambda_t}(y')| d\Lambda_t(y') \cdot L \|y - z\| \\
 &= L \|y - z\| \\
 &\leq \tilde{L} \|y - z\|.
 \end{aligned}$$

As the weight function depends on the used measure, the proof for (iii) can not be done in the same way as in the original paper. We first separate this term into two parts.

$$\begin{aligned}
 \|\tilde{b}_\Lambda(t, y) - \tilde{b}_\Lambda(s, y)\| &= \left\| \int_S w_{\Lambda_t}(y') f(y, y') d\Lambda_t(y') - \int_S w_{\Lambda_s}(y') f(y, y') d\Lambda_s(y') \right\| \\
 &\leq \left\| \int_S w_{\Lambda_t}(y') f(y, y') d\Lambda_t(y') - \int_S w_{\Lambda_t}(y') f(y, y') d\Lambda_s(y') \right\| \\
 &\quad + \left\| \int_S w_{\Lambda_t}(y') f(y, y') d\Lambda_s(y') - \int_S w_{\Lambda_s}(y') f(y, y') d\Lambda_s(y') \right\| \\
 &= \left\| \int_S w_{\Lambda_t}(y') f(y, y') d(\Lambda_t - \Lambda_s)(y') \right\| \\
 &\quad + \left\| \int_S (w_{\Lambda_t}(y') - w_{\Lambda_s}(y')) f(y, y') d\Lambda_s(y') \right\| \\
 &\leq \left\| \int_S w_{\Lambda_t}(y') f(y, y') d(\Lambda_t - \Lambda_s)(y') \right\| \\
 &\quad + \int_S \| (w_{\Lambda_t}(y') - w_{\Lambda_s}(y')) \| \|f(y, y')\| d\Lambda_s(y') \\
 &=: I + II.
 \end{aligned}$$

We will start with the first part I . Note that $\text{supp}(\Lambda_t - \Lambda_s) \subset S_T$. Using the Lipschitz properties of f and w_{Λ_t} on S_T , together with the bound shown in Lemma 4.3, we can

show that $(w_{\Lambda_t}(\cdot)f(y, \cdot))$ is Lipschitz on S_T by doing the following computation:

$$\begin{aligned}
 \|w_{\Lambda_t}(y'_1)f(y, y'_1) - w_{\Lambda_t}(y'_2)f(y, y'_2)\| &\leq \|(w_{\Lambda_t}(y'_1) - w_{\Lambda_t}(y'_2))f(y, y'_1)\| \\
 &\quad + \|w_{\Lambda_t}(y'_1)(f(y, y'_1) - f(y, y'_2))\| \\
 &\leq \sup_{y' \in S_T} f(y, y') \|w_{\Lambda_t}(y'_1) - w_{\Lambda_t}(y'_2)\| \\
 &\quad + \sup_{y' \in S_T} w_{\Lambda_t}(y') \|f(y, y'_1) - f(y, y'_2)\| \\
 &\leq (2\bar{J} + L_J + R_U) \cdot L_w \|y'_1 - y'_2\| + 1 \cdot L \|y'_1 - y'_2\| \\
 &= ((2\bar{J} + L_J + R_U)L_w + L) \|y'_1 - y'_2\| \\
 &= L_{f,w} \|y'_1 - y'_2\|,
 \end{aligned}$$

with $L_{f,w} := ((2\bar{J} + L_J + R_U)L_w + L)$.

Now, using the dual definition of the 1-Wasserstein norm (see Theorem 2.4) and the fact that $\frac{w_{\Lambda_t}(\cdot)f(y, \cdot)}{L_{f,w}}$ is 1-Lipschitz, we have

$$\begin{aligned}
 I &= \left\| \int_S w_{\Lambda_t}(y') f(y, y') d(\Lambda_t - \Lambda_s)(y') \right\| \\
 &= L_{f,w} \cdot \left\| \int_S \frac{w_{\Lambda_t}(y') f(y, y')}{L_{f,w}} d(\Lambda_t - \Lambda_s)(y') \right\| \\
 &\leq L_{f,w} \cdot \sup_{\substack{\varphi \in \text{Lip}_b(X) \\ \text{Lip}(\varphi) \leq 1}} \left\| \int_S \varphi(y') d(\Lambda_t - \Lambda_s)(y') \right\| \\
 &= L_{f,w} \cdot W_1(\Lambda_t, \Lambda_s).
 \end{aligned}$$

The bound on II follows from Lemma 4.12 and Lemma 4.3, as

$$\begin{aligned}
 II &= \int_S \|(w_{\Lambda_t}(y') - w_{\Lambda_s}(y'))\| \|f(y, y')\| d\Lambda_s(y') \\
 &\leq \sup_{y' \in S_T} \|w_{\Lambda_t}(y') - w_{\Lambda_s}(y')\| \cdot \sup_{y, y' \in S_T} \|f(y, y')\| \\
 &\leq \frac{1}{\omega_*^2} L_\omega W_1(\Lambda_t, \Lambda_s) \cdot (2\bar{J} + L_J + R_U).
 \end{aligned}$$

We can now conclude the proof of (iii):

$$\begin{aligned}
 \|\tilde{b}_\Lambda(t, y) - \tilde{b}_\Lambda(s, y)\| &\leq I + II \\
 &\leq L_{f,w} \cdot W_1(\Lambda_t, \Lambda_s) + \frac{1}{\omega_*^2} L_\omega W_1(\Lambda_t, \Lambda_s) \cdot (2\bar{J} + L_J + R_U) \\
 &= \left(L_{f,w} + \frac{1}{\omega_*^2} L_\omega \cdot (2\bar{J} + L_J + R_U) \right) W_1(\Lambda_t, \Lambda_s) \\
 &=: \tilde{L} W_1(\Lambda_t, \Lambda_s),
 \end{aligned}$$

with $\tilde{L} \geq L$. The proof of (iv) is analogous.

Finally, we will show the last statement (v). For any $R > 0$ we can find a $\theta > 0$ so that Eq. (2.5.9) holds, that is

$$y + \theta f(y, y') \in S \quad \forall y, y' \in S \cap B_R.$$

As for any $\Lambda_t \in \mathcal{P}(S)$ and any $y' \in S$ we have $w_{\Lambda_t}(y') \in (0, 1]$. Now, as S is convex, it holds

$$y + \theta w_{\Lambda_t}(y') f(y, y') \in S \quad \forall y, y' \in S \cap B_R, \forall \Lambda_t \in \mathcal{P}(S).$$

The proof then follows exactly as the original proof in [Amb+18]:

$$y + \theta \tilde{b}_\Lambda(t, y) = \int_S (y + \theta w_{\Lambda_t}(y') f(y, y')) d\Lambda_t(y') \in S,$$

as S is convex and closed and Λ_t is a probability measure. \square

We now effectively have similarly strong stability estimates on \tilde{b}_Λ as we had on b_Λ (see Proposition 2.18). We can now state the corresponding properties on $\tilde{\mathbf{Y}}_\Lambda$, but the actual content does not differ from Corollary 2.20. The proof can be done in the same way as done by Ambrosio et al. for [Amb+18, Cor. 4.4].

Corollary 4.16 (Properties of $\tilde{\mathbf{Y}}$)

Let $\Lambda, \Lambda^i \in \mathcal{A}$ and let $\tilde{b}_\Lambda, \tilde{b}_{\Lambda^i}$ be defined as in Eq. (4.5.4). Then

- (i) for every $y \in S$ and $s \in [0, T]$ there exists a unique solution $y_t = \tilde{\mathbf{Y}}_\Lambda(t, s, y)$ from $[s, T]$ to S of class C^1 of the Cauchy problem

$$\dot{y}_r = \tilde{b}_\Lambda(r, y_r) = \tilde{b}_{\Lambda_r}(y_r), \quad y_s = y; \quad (4.5.9)$$

- (ii) $\tilde{\mathbf{Y}}_\Lambda(t, 0, y)$ satisfies the estimate

$$\|\tilde{\mathbf{Y}}_\Lambda(t, 0, y) - y_0\| \leq (\|y - y_0\| + t \tilde{B}(\Lambda, t, y_0)) e^{\tilde{L}t}, \quad (4.5.10)$$

where

$$\tilde{B}(\Lambda, t, y_0) := \|f(y_0, y_0)\| + \tilde{L} \max_{s \in [0, t]} \int_S \|y' - y_0\| d\Lambda_s(y'); \quad (4.5.11)$$

(iii) $\tilde{\mathbf{Y}}_\Lambda(\cdot, 0, y)$ satisfies the estimate

$$\|\tilde{\mathbf{Y}}_\Lambda(t, 0, y) - \tilde{\mathbf{Y}}_\Lambda(t', 0, y)\| \leq |t - t'| \left[\tilde{B}(\Lambda, T, y_0) + \tilde{L} \left(\|y - y_0\| + T \tilde{B}(\Lambda, T, y_0) \right) e^{\tilde{L}T} \right]; \quad (4.5.12)$$

(iv) $\tilde{\mathbf{Y}}_\Lambda(t, s, \cdot)$ satisfies the estimate

$$\|\tilde{\mathbf{Y}}_\Lambda(t, s, y) - \tilde{\mathbf{Y}}_\Lambda(t, s, y')\| \leq e^{\tilde{L}(t-s)} \|y - y'\|, \quad 0 \leq s \leq t \leq T; \quad (4.5.13)$$

(v) more generally, $\tilde{\mathbf{Y}}_{\Lambda^1}, \tilde{\mathbf{Y}}_{\Lambda^2}$ satisfy the estimate for $0 \leq s \leq t \leq T$:

$$\|\tilde{\mathbf{Y}}_{\Lambda^1}(t, s, y^1) - \tilde{\mathbf{Y}}_{\Lambda^2}(t, s, y^2)\| \leq e^{\tilde{L}(t-s)} \|y^1 - y^2\| + \tilde{L} \int_s^t e^{\tilde{L}(t-\tau)} W_1(\Lambda_\tau^1, \Lambda_\tau^2) d\tau. \quad (4.5.14)$$

4.5.3 Proof by contractivity

We still consider the initial population $\bar{\Sigma} \in \mathcal{P}_1(S)$ with compact support in S_0 , with S_0 as in Section 4.1 for some $R_0 > 0$. Recall the definition of the space of possible valid populations (see Eq. (4.1.7)) as

$$\widetilde{\mathcal{A}} = \left\{ \Lambda \in C([0, T]; (\mathcal{P}_1(S), W_1)) : \Lambda_0 = \bar{\Sigma} \text{ and } \text{supp}(\Lambda_t) \subset S_t \text{ for any } t \in [0, T] \right\},$$

with S_t as introduced in Section 4.1, but we remind the reader of Remark 4.1 so that S_t and R_t do not depend on our specific assumptions on U or e , as long as U is compact and e is Lipschitz.

This metric space is complete when endowed with the distance

$$d_{\sup}(\Lambda^1, \Lambda^2) := \sup_{t \in [0, T]} W_1(\Lambda_t^1, \Lambda_t^2). \quad (4.5.15)$$

Indeed, as $(\mathcal{P}_1(S), W_1)$ is complete it follows that

$$(\mathcal{A}, d_{\sup}) \subset (C([0, T]; (\mathcal{P}_1(S), W_1)); d_{\sup})$$

is a complete metric space. Then, as $\widetilde{\mathcal{A}} \subset \mathcal{A}$ is a closed subspace it also has to be complete.

While $\widetilde{\mathcal{A}}$ differs from the space \mathcal{A} , as introduced in Eq. (2.5.22), we also define the map $\tilde{\mathcal{T}} : S \rightarrow S$ with the same procedure as in Eq. (2.5.23): given $\Lambda \in \widetilde{\mathcal{A}}$ we first compute the flow map $\tilde{\mathbf{Y}}_\Lambda(t, s, \cdot)$ associated to \tilde{b}_Λ and then we define the curve $\tilde{\mathcal{T}}[\Lambda] : [0, T] \rightarrow \mathcal{P}_1(S)$ by

$$\tilde{\mathcal{T}}[\Lambda]_t := \tilde{\mathbf{Y}}_\Lambda(t, 0, \cdot) \# \bar{\Sigma}. \quad (4.5.16)$$

In the original paper the authors state that $\tilde{\mathcal{T}}$ maps $\widetilde{\mathcal{A}}$ to $\widetilde{\mathcal{A}}$. While this still holds in our case, we still include the according proof.

Proposition 4.17

$\tilde{\mathcal{T}}$ maps $\widetilde{\mathcal{A}}$ to $\widetilde{\mathcal{A}}$.

Proof. For any $\Lambda \in \widetilde{\mathcal{A}}$ it is clear that $\tilde{\mathbf{Y}}_\Lambda(t, 0, \cdot)_{\#}\bar{\Sigma}$ is a probability measure. Additionally, it holds

$$\tilde{\mathbf{Y}}_\Lambda(0, 0, \cdot)_{\#}\bar{\Sigma} = \bar{\Sigma}.$$

It is left to show, that $\text{supp}(\tilde{\mathbf{Y}}_\Lambda(t, 0, \cdot)_{\#}\bar{\Sigma}) \subset S_t$.

Indeed, as $\tilde{Y}_\Lambda(t, 0, \cdot)$ has to satisfy the ODEs Eq. (4.5.6) we can deduce from the discussion done in Section 4.1 that given some $y_0 = (x_0, \sigma_0) \in S_0$ we also have a bound on the spatial component of $y_t := \tilde{Y}_\Lambda(t, 0, y_0)$. Notably, it has to hold $Y_\Lambda(t, 0, y_0) \in S_t$, which concludes that $\text{supp}(\tilde{\mathbf{Y}}_\Lambda(t, 0, \cdot)_{\#}\bar{\Sigma}) \in S_t$. \square

Now we have the same properties on $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{T}}$ as Ambrosio et al. showed originally on \mathcal{A} and \mathcal{T} : $\widetilde{\mathcal{A}}$ is a complete metric space and $\widetilde{\mathcal{T}} : \widetilde{\mathcal{A}} \rightarrow \widetilde{\mathcal{A}}$. Thanks to Corollary 4.16 we also have the same stability estimates as in Corollary 2.20, while taking the modified constant $\tilde{L} \geq L$ into account. We can therefore proceed exactly as originally done by Ambrosio et al. All the results can be proven identically as they do not depend on the exact choice of L , but only require some constant $L > 0$ such that we can replace it with the modified constant $\tilde{L} > 0$. We therefore chose to omit the proofs of the two remaining results and refer to [Amb+18].

Lemma 4.18

For every $\Lambda, \Lambda^1, \Lambda^2 \in \widetilde{\mathcal{A}}$ we have

$$W_1(\tilde{\mathcal{T}}[\Lambda]_t, \tilde{\mathcal{T}}[\Lambda]_s) \leq |t - s| \left[\tilde{B}(\Lambda, T, y_0) + \tilde{L} \left(\int_S \|y - y_0\| d\bar{\Sigma}(y) + T \tilde{B}(\Lambda, T, y_0) \right) e^{LT} \right], \quad (4.5.17)$$

$$W_1(\tilde{\mathcal{T}}[\Lambda^1]_t, \tilde{\mathcal{T}}[\Lambda^2]_t) \leq \tilde{L} \int_0^t e^{\tilde{L}(t-\tau)} W_1(\Lambda_\tau^1, \Lambda_\tau^2) d\tau, \quad (4.5.18)$$

where the constant $\tilde{B}(\Lambda, T, y_0)$ is defined in Corollary 4.16(ii) for $t = T$.

Corollary 4.19

The map $\tilde{\mathcal{T}}$ admits a unique fixed point, which provides the unique solution Σ in Theorem 4.13.

Chapter 5

Implementation

In the previous chapter we concluded the proof of existence and uniqueness of Lagrangian solutions for the proposed weighted replicator dynamics. With these results on well-posedness we are now able use this modified model on the problem of global function minimization. Chapter 3 gave a first overview on the problem statement and we argued that by choosing the payoff function J carefully, together with the weighting function w , the population should be able to converge to the global minimizer.

In this chapter we will present how a population, following the dynamics of spatially inhomogeneous evolutionary game theory, can be simulated numerically. After providing discrete time solutions for the differential equations describing the population, we introduce further numerical adaptations which were made in order to make the resulting algorithm more usable. We present and explain our concrete choice of payoff function J , and conclude the chapter with the final minimization algorithm we used in our numerical results.

5.1 Discrete Time Solutions

In [Amb+18, Section 1.5] Ambrosio et al. introduce discrete time solutions to the dynamics we described in Section 2.3. The authors are able to show that the family of discrete solutions has limit points for decreasing step-sizes. These correspond to solutions of the continuous initial value problem. In order to numerically solve the continuous differential equation we will therefore construct a discrete time solution, which can be arbitrarily close to the continuous solution by choosing the step size small enough. In the following we present this construction.

Let $[0, T]$ be a time interval and $h > 0$ a positive step size. For each time step $t \in \{0, h, 2h, \dots, T\}$ we iteratively build a population, which follows the dynamics as induced by the ODEs Eq. (4.5.6), by calculating the subsequent location and strategy of each individual $y = (x, \sigma)$ in the following way:

$$\begin{aligned} \sigma_{t+h} &\leftarrow \left(1 + h\Delta_{\Sigma_t, (x_t, \sigma_t)}\right) \sigma_t, \\ x_{t+h} &\leftarrow x_t + h \cdot \tilde{u}, \quad \text{where} \quad \tilde{u} \sim \sigma_{t+h}, \end{aligned} \tag{5.1.1}$$

with

$$\begin{aligned} \Delta_{\Sigma_t, (x_t, \sigma_t)} := & \int_S w_{\Sigma_t}(x', \sigma') \left(\int_U J(x_t, \cdot, x', u') d\sigma'(u') \right. \\ & \left. - \int_U \int_U J(x_t, w, x', u') d\sigma'(u') d\sigma_t(w) \right) d\Sigma_t(x', \sigma'). \end{aligned} \quad (5.1.2)$$

We chose to denote the assignment with “ \leftarrow ”, as in this chapter we adopt a more computational and algorithmic point of view: Given the particle at time t , we *compute* its state for time $t + h$. By iterating this process T/h times we build a discrete solution on the time interval $[0, T]$. Thus the computations in Eq. (5.1.1) define the update-step in our optimization algorithm.

Remark 5.1

Note that the σ -update corresponds to the standard forward Euler method: Using the finite difference formula for the derivative, we have for any function y ,

$$y'(t) \approx \frac{y(t+h) - y(t)}{h} \quad \text{for } h > 0 \text{ small,}$$

which then leads to

$$y(t+h) \approx y(t) + h(y'(t)) \quad \text{for } h > 0 \text{ small.}$$

Now by applying this formula to the function σ we get

$$\sigma_{i,t+h} \approx \sigma_{i,t} + h\dot{\sigma}_{i,t} = \sigma_{i,t} + h\Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}\sigma_{i,t},$$

for some small step size $h > 0$ and $\dot{\sigma}_{i,t}$ as in Eq. (4.5.6).

5.2 Numerical Adaptations

In this section we consider different aspects of the algorithm from a more practical point of view. In order to efficiently compute the time discrete solutions it is necessary to work with finite populations and a finite set of strategies. Further, we make observations on stability and instability of different parts of the computation and we propose further adaptations in order to make the algorithm more usable.

5.2.1 Finite populations and strategies

In our implementation we consider a finite set of strategies $U = \{u_i\}_{i=1}^{N_U} \subset B_{R_U}$ together with the velocity field $e(x, u) = u$. We further assume the payoff function J to be independent of the strategy chosen by the co-player, that is $J(x, u, x', u')$ does not depend on u' , and we will therefore only write $J(x, u, x')$.

We consider discrete populations consisting of N players. We can then formally write the discrete initial distribution $\bar{\Sigma}$ in the following way:

$$\bar{\Sigma} = \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,0}} \in \mathcal{P}_1(S), \quad (5.2.1)$$

where we still assume that $\bar{\Sigma}$ has bounded support, that is for any $i \in \{1, \dots, N\}$ we have $y_{i,0} \in S_0$ with S_0 as defined in Eq. (4.1.5). The evolution of a discrete population has to be discrete, and at any point in time $t \in [0, T]$ for some $T > 0$ we have

$$\Sigma_t = \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,t}} \quad \text{with } y_{i,t} \in S_t. \quad (5.2.2)$$

In order to simplify the notation, we describe the population at time t with the set $\{y_{i,t}\}_{i=1}^N$.

Using the weighted version of the replicator dynamics as introduced in Chapter 4, we can then write the resulting ODEs with discrete populations and strategies describing the evolution of some player $y_i = (x_i, \sigma_i)$ as

$$\begin{aligned} \dot{x}_i &= \sum_{u \in U} u d\sigma_i(u) \\ \dot{\sigma}_i &= \left(\sum_{j=1}^N w_{\Sigma_t}(x_j, \sigma_j) \left(J(x_i, \cdot, x_j) - \sum_{w \in U} J(x_i, w, x_j) \sigma_i(w) \right) \right) \sigma_i, \end{aligned} \quad (5.2.3)$$

with the weight function w_{Σ_t} as defined in Eq. (3.4.2).

Following the lines of Section 5.1 this leads to the following discrete time solution for all $y_i = (x_i, \sigma_i)$ and all $t \in \{0, h, 2h, \dots, T\}$:

$$\begin{aligned} \sigma_{i,t+h} &\leftarrow \left(1 + h \Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})} \right) \sigma_{i,t}, \\ x_{i,t+h} &\leftarrow x_{i,t} + h \cdot \tilde{u}, \quad \text{where} \quad \tilde{u} \sim \sigma_{i,t+1}, \end{aligned} \quad (5.2.4)$$

with

$$\Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})} = \sum_{j=1}^N w_{\Sigma_t}(x_{j,t}, \sigma_{j,t}) \left(J(x_{i,t}, \cdot, x_{j,t}) - \sum_{w \in U} J(x_{i,t}, w, x_{j,t}) \sigma_{i,t}(w) \right). \quad (5.2.5)$$

5.2.2 Numerically stable weight computation

The weight function $w_{\Sigma_t}(x, \sigma)$, introduced in Eq. (3.4.2), leads to numerical problems in its computation. In order to get good results we need to chose $\alpha \gg 0$ large, but this leads to very small positive values $\omega_g^\alpha(x) \ll 1$. Therefore, the sum of those weights

is close to zero and in the computation of $w_{\Sigma_t}(x, \sigma) = \frac{\omega_g^\alpha(x)}{\sum_{i=1}^N \omega_g^\alpha(x_i)}$ the division is numerically unstable.

We circumvent this problem using the following numerical trick:

$$\begin{aligned} w_{\Sigma_t}(x, \sigma) &= w_{\Sigma_t}(x, \sigma) \cdot \frac{\exp(-\alpha g(x_{\min}))}{\exp(-\alpha g(x_{\min}))} \\ &= \frac{\exp(\alpha g(x))}{\sum_{i=1}^N \exp(\alpha g(x_i))} \cdot \frac{\exp(-\alpha g(x_{\min}))}{\exp(-\alpha g(x_{\min}))} \\ &= \frac{\exp(\alpha(g(x) - g(x_{\min})))}{\sum_{i=1}^N \exp(\alpha(g(x_i) - g(x_{\min})))} \end{aligned}$$

where

$$x_{\min} := \arg \min_{(x, \sigma) \in \{y_{i,t}\}_{i=1}^N} g(x)$$

is the location of the player with the minimal function value in the current population.

This ensures that for at least one player $y_j = (x_j, \sigma_j)$, with location $x_j = x_{\min}$, we have $g(x_j) - g(x_{\min}) = 0$ and therefore $\exp(\alpha(g(x_j) - g(x_{\min}))) = 1$. For the sum this leads to $\sum_{i=1}^N \exp(\alpha(g(x_i) - g(x_{\min}))) \geq 1$, such that the division does not induce a numerical problem.

In the numerical simulations we will always compute the weights in the presented manner, but as we did not change the actual value of $w_{\Sigma_t}(x, \sigma)$ all results from Chapter 4 still hold.

5.2.3 Different step sizes for strategy and location

In order to have a better control over the simulation it appears to be reasonable to use different step sizes for the location update and the strategy update, such that we are able to tune them independently. With this argumentation we might therefore introduce two independent step sizes $h_x, h_\sigma > 0$, one for the location update and one for the strategy update. On the other hand, when working with the algorithm it might not seem intuitive to have to modify two parameters instead of one in order to adapt the general step size.

Therefore, instead of introducing two independent step sizes h_x, h_σ , we chose to use a main step size h and a scaling factor θ , which effectively leads to $h_x := h$ and $h_\sigma := h \cdot \theta$.

With $\Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}$ as in Eq. (5.2.5) the resulting update step is given as

$$\begin{aligned} \sigma_{i,t+h} &\leftarrow \left(1 + h\theta\Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}\right)\sigma_{i,t}, \\ x_{i,t+h} &\leftarrow x_{i,t} + h \cdot \tilde{u}, \quad \text{where} \quad \tilde{u} \sim \sigma_{i,t+1}. \end{aligned} \tag{5.2.6}$$

By introducing the additional parameter θ we are able to tune the step size used in the strategy update, independently of the location update, while still being able to change the overall resolution of the simulation in the single parameter h .

Remark 5.2

This adaptation complies with the theory we introduced so far. Indeed, we have

$$\begin{aligned}\theta\Delta_{\Sigma_t,(x_{i,t},\sigma_{i,t})} &= \theta \left(\sum_{j=1}^N w_{\Sigma_t}(x_{j,t}, \sigma_{j,t}) \cdot \left(J(x_{i,t}, \cdot, x_{j,t}) - \sum_{w \in U} J(x_{i,t}, w, x_{j,t}) \sigma_{i,t}(w) \right) \right) \\ &= \sum_{j=1}^N w_{\Sigma_t}(x_{j,t}, \sigma_{j,t}) \cdot \left(\theta J(x_{i,t}, \cdot, x_{j,t}) - \sum_{w \in U} \theta J(x_{i,t}, w, x_{j,t}) \sigma_{i,t}(w) \right) \\ &= \sum_{j=1}^N w_{\Sigma_t}(x_{j,t}, \sigma_{j,t}) \cdot \left(\tilde{J}(x_{i,t}, \cdot, x_{j,t}) - \sum_{w \in U} \tilde{J}(x_{i,t}, w, x_{j,t}) \sigma_{i,t}(w) \right),\end{aligned}$$

with $\tilde{J} := \theta J$. Now, as J is L_J -Lipschitz then \tilde{J} is (θL_J) -Lipschitz and all requirements for the theory, both in [Amb+18] and Chapter 4, are still fulfilled.

5.2.4 Parametrized adaptive step size

We begin this section with a discussion of the numerical stability and instability of the update step in regards to the chosen step size, and introduce an adaptive step size for the strategy update. The discrete time solutions (see Section 5.1) are shown to converge to the continuous solution for step sizes $h \rightarrow 0$. We are therefore always encouraged to chose the step size as small as possible. On the other hand, from a practical point of view there is always an incentive in choosing a large step size in order to obtain results in a short time. In the following we will discuss an upper bound for the step size and introduce a parametrized adaptive step size.

Recall the strategy update

$$\sigma_{i,t+h} \leftarrow \left(1 + h\theta\Delta_{\Sigma_t,(x_{i,t},\sigma_{i,t})} \right) \sigma_{i,t}.$$

For player i , at time step t , $\sigma_{i,t}$ is a probability distribution. Therefore, for all $u \in U$ it holds $\sigma_{i,t}(u) \geq 0$ and $\sum_{u \in U} \sigma_{i,t}(u) = 1$. At time $t + h$ the second property surely still holds, regardless of the step size h , as $\Delta_{\Sigma_t,(x_{i,t},\sigma_{i,t})}$ has mean 0. However, the first property might not be conserved in the update step. Indeed, for any $u \in U$ with negative $\Delta_{\Sigma_t,(x_{i,t},\sigma_{i,t})}(u)$ we can find some $h, \theta > 0$ with $\left(1 + h\theta\Delta_{\Sigma_t,(x_{i,t},\sigma_{i,t})}(u) \right) < 0$, and therefore

$$\sigma_{i,t+h}(u) = \left(1 + h\theta\Delta_{\Sigma_t,(x_{i,t},\sigma_{i,t})}(u) \right) \sigma_{i,t}(u) < 0.$$

Thus, $\sigma_{i,t+h}$ would not be a probability distribution anymore. In order to prevent this, h and θ have to be chosen such that the following holds:

$$\left(1 + h\theta \min_{u \in U} (\Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}(u))\right) > 0.$$

Noting that $\min_{u \in U} \Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}(u) < 0$ we deduce the upper bound

$$h\theta < \frac{1}{-\min_{u \in U} \Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}(u)}. \quad (5.2.7)$$

By introducing a new parameter $\gamma < \frac{1}{h}$ we can then define

$$\theta := \gamma \frac{1}{-\min_{u \in U} \Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}(u)},$$

such that Eq. (5.2.7) always holds and $\sigma_{i,t+h}$ is non-negative. While we could have replaced both h and θ at once, we chose this procedure in order to keep control over the resulting step size, while also keeping the advantages of having different step sizes in location and strategy.

The strategy update step, using a parametrized adaptive step size $\gamma < \frac{1}{h}$ for the strategy update, is then for all $(x_{i,t}, \sigma_{i,t}) \in \{y_{i,t}\}_{i=1}^N$:

$$\begin{aligned} \sigma_{i,t+h} &\leftarrow \left(1 + h\gamma \frac{\Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}}{-\min_{w \in U} \Delta_{\Sigma_t, (x_{i,t}, \sigma_{i,t})}(w)}\right) \sigma_{i,t}, \\ x_{i,t+h} &\leftarrow x_i + h \cdot \tilde{u}, \quad \text{where} \quad \tilde{u} \sim \sigma_{i,t+h}. \end{aligned} \quad (5.2.8)$$

5.3 The Payoff Function J

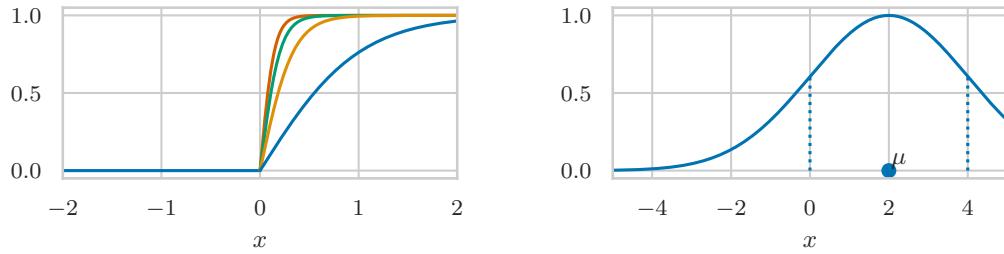
Before concluding this chapter with a presentation of the resulting algorithm, we explain our choice of payoff function and discuss its properties. For any $\epsilon > 0$ we chose the payoff function $J_\epsilon : \mathbb{R}^3 \rightarrow [0, 1]$ defined as follows:

$$J_\epsilon(x, u, x') = \exp\left(-\frac{(u - \tanh_+(3(g(x) - g(x')))) \cdot (x' - x)^2}{2|x' - x|^2 + \epsilon}\right), \quad (5.3.1)$$

where $\tanh_+(z) := \max\{0, \tanh(z)\}$. Note that by choosing $\epsilon > 0$ we bound the derivative of the payoff function, therefore it is Lipschitz continuous.

In order to understand this function we analyze it part by part. Figure 5.1a shows the function graph of $x \mapsto \tanh_+(bx)$. With increasing b it effectively constitutes a b -Lipschitz-continuous approximation of the function

$$x \mapsto \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{else,} \end{cases}$$



(a) Graph of $\tanh_+(bx)$ for $b \in \{1, 3, 5, 7\}$. (b) Modified Gaussian distribution $f(x|\mu, \sigma^2) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ with $\mu = 2$ and $\sigma = 2$. The dotted lines indicate the $[\mu - \sigma, \mu + \sigma]$ interval, containing 68% of its mass.

Figure 5.1: Illustrations of the different components of the proposed payoff function J .

which is basically an if-else-statement. Thus, we reinterpret J_ϵ as:

$$J_\epsilon(x, u, x') \approx \begin{cases} \exp\left(-\frac{(u-(x'-x))^2}{2|x'-x|^2+\epsilon}\right) & \text{if } g(x') < g(x) \\ \exp\left(-\frac{u^2}{2|x'-x|^2+\epsilon}\right) & \text{if } g(x') \geq g(x). \end{cases} \quad (5.3.2)$$

Next, we consider J_ϵ as a function $J_\epsilon(x, \cdot, x') : U \rightarrow [0, 1]$, and compare it to the function

$$f(x|\mu, \sigma^2) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

This function f corresponds almost to a Gaussian distribution, but it lacks the rescaling factor $\frac{1}{\sqrt{2\pi\sigma^2}}$ and always attains its maximum function value 1, at $x = \mu$. Figure 5.1b shows an example of its function graph with $\mu = 2$ and $\sigma = 1$. Disregarding the term $\epsilon > 0$, it becomes clear that the payoff function J_ϵ relates to this function f in the following way:

$$J_\epsilon(x, u, x') \approx \begin{cases} f(u|x' - x, |x' - x|^2) & \text{if } g(x') < g(x) \\ f(u|0, |x' - x|^2) & \text{if } g(x') \geq g(x). \end{cases} \quad (5.3.3)$$

Using this representation we understand J_ϵ even better: If the co-player x' has a lower function value than the player x , then the highest payoff is given for the strategy $\bar{u} := (x' - x)$, which lets x move towards x' , and the payoff decreases as strategies are farther away from \bar{u} . On the other hand, if the co-player x' has a higher function value the best strategy is to stay at the current location.

In order to interpret the meaning of the variance $|x' - x|^2$ we refer again to Fig. 5.1b. Assume $g(x') \ll g(x)$. We introduced the set of strategies as a subset of some interval

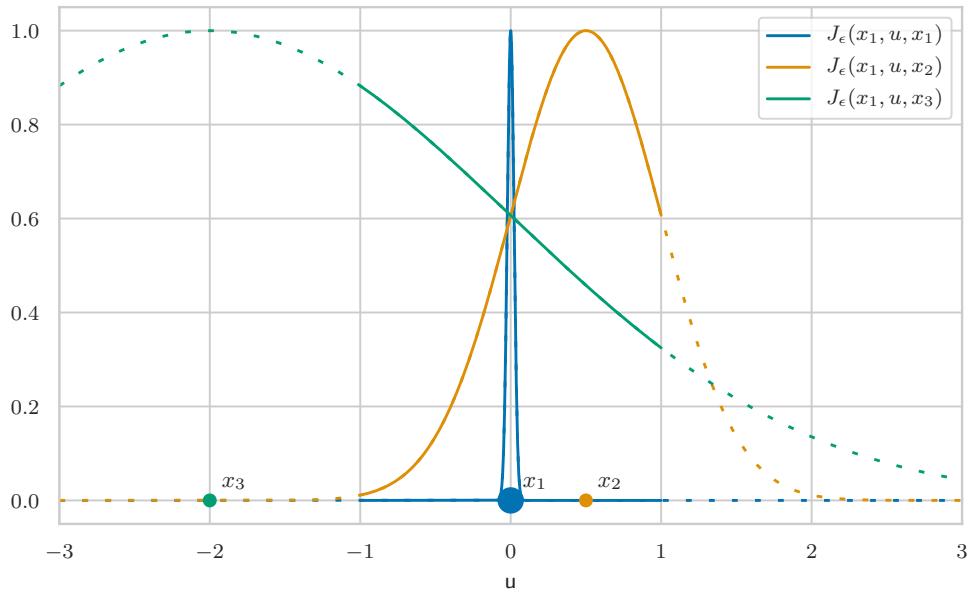


Figure 5.2: Visualization of the payoff function $J_\epsilon(x_1, \cdot, x')$ with $\epsilon = 10^{-3}$ and $U \subset [-1, 1]$. We present the point of view of some player x_1 , evaluating the payoff function against himself and two other co-players x_2, x_3 . We assume here that the co-players x_2, x_3 have a significantly lower function value than x_1 , leading to $\tanh_+(3(g(x_1) - g(x_2))) = \tanh_+(3(g(x_1) - g(x_3))) = 1$. The dashed continuation is added in order to understand the payoff function; With $U \subset [-1, 1]$ only the solid line is actually computed.

$[-s_{\max}, s_{\max}] \supset U$. Now, if the strategy with the highest payoff $\bar{u} = (x' - x)$ would be too far outside this interval, i.e. $|\bar{u}| \gg s_{\max}$, then most of the mass of the function could also be outside of $[-s_{\max}, s_{\max}]$ and all available strategies $u \in U$ might get a payoff very close to 0. By choosing the variance $\sigma^2 = |x' - x|^2$ we make sure that the boundary of the 68% interval (see Fig. 5.1b) lies exactly in 0, as

$$|\mu| - \sigma = |x' - x| - |x' - x| = 0.$$

Therefore there is always enough mass reaching the set of strategies U . We now consider the payoff function J_ϵ in its entirety, as visualized in Fig. 5.2.

After our previous discussion, the shape of the Gaussian distribution becomes clear. It is always broad enough to reach the interval U , which is chosen here as $U \subset [-1, 1]$. Additionally, we included the function graph of $J_\epsilon(x_1, u, x_1)$ in blue, which is relevant notably when the player with location x_1 has the lowest function value in the population.

5.4 Global Minimization Algorithm

With the proposed adaptations and the resulting update step (see Eq. (5.2.8)) we construct an algorithm, described in Algorithm 1, which iteratively computes these new locations and strategies for each player.

Algorithm 1 Global Minimization Algorithm

Input: Parameters $h, \gamma, \alpha \in \mathbb{R}_+$, $N, M \in \mathbb{N}$; Set of strategies U ; Payoff function J ;
Initial distribution $\{(x_i, \sigma_i)\}_{i=1}^N$.

```

1: for  $k = 1, \dots, M$  do
    Compute Delta:
    2:   for  $i = 1, \dots, N$  do
        3:         for  $u \in U$  do
        4:              $\Delta_i(u) \leftarrow \left( \sum_{j=1}^N w_{\Sigma_i}(x_j, \sigma_j) \cdot (J(x_i, u, x_j) - \sum_{w \in U} J(x_i, w, x_j) \sigma_i(w)) \right) \sigma_i(u)$ 
        5:         end for
        6:     end for
        Update strategy:
        7:     for  $i = 1, \dots, N$  do
        8:          $\Delta_{i\min} \leftarrow \min_{w \in U} \Delta_i(w)$ 
        9:         for  $u \in U$  do
        10:             $\sigma_i(u) \leftarrow \left( 1 + h\gamma \frac{\Delta_i(u)}{-\Delta_{i\min}} \right) \cdot \sigma_i(u)$ 
        11:        end for
        12:    end for
        Update location:
        13:    for  $i = 1, \dots, N$  do
        14:        Sample  $\tilde{u} \sim \sigma_i$ 
        15:         $x_i \leftarrow x_i + h\tilde{u}$ 
        16:    end for
        17:    if converged (see also Remark 5.4) then
        18:        BREAK
        19:    end if
20: end for
```

Remark 5.3 (Efficient computation of the algorithm)

The multiple sequential loops on i and u in Algorithm 1 might seem inefficient at first. These loops do not directly appear in our implementation, as we carefully represent Δ , σ and x as matrices or tensors, containing all information about all players. With linear algebra libraries such as NumPy¹ we use vectorized operations in order to perform the computations of these objects in a highly efficient and optimized manner.

¹<https://www.numpy.org/>

We chose to represent the steps in Algorithm 1 in this way as it visualizes the three-step procedure of sequentially computing Δ , σ and x instead of iterating over each player. Our complete code can be found in [Bos18].

Remark 5.4 (Early stopping)

In Section 3.3.2 we discussed the general notion of convergence, and described it as convergence of the population towards some single state y^* . For practical use in numerical simulations we therefore propose the following stopping criterion: If every player has a mixed strategy very close to δ_0 we interrupt the simulation. By choosing some $\lambda > 0$, e.g. $\lambda = 10^{-10}$, we then stop the computations once $\sigma_i(0) > 1 - \lambda$ for all $i \in \{1, \dots, N\}$.

In the following we clarify the meaning of each parameter, while specifying our exact choices where appropriate:

- h : The step size, introduced in Section 5.1.
- γ : The parameter for the adaptive step size on the strategy update, see Section 5.2.4
- J : The payoff function. We chose $J = J_\epsilon$ as specified in Section 5.3 with $\epsilon = 10^{-3}$, but other valid choices might also be possible.
- U : The set of available strategies. Given a pair number of strategies M and two parameters s_{\min}, s_{\max} we first create a linear grid in $V \in [\sqrt{s_{\min}}, \sqrt{s_{\max}}]$ with $M/2$ elements:

$$V := \left\{ \sqrt{s_{\min}}, \sqrt{s_{\min}} + \frac{\sqrt{s_{\max} - s_{\min}}}{M/2}, \sqrt{s_{\min}} + 2 \frac{\sqrt{s_{\max} - s_{\min}}}{M/2}, \dots, \sqrt{s_{\max}} \right\}. \quad (5.4.1)$$

Then we chose U as follows:

$$U := \left\{ \pm v^2 : v \in V \right\} \cup \{0\}. \quad (5.4.2)$$

With U chosen in this way we have $U \subset [-s_{\max}, s_{\max}]$ symmetric around 0 with a higher resolution close to 0 and a lower resolution for strategies with large values. Additionally it holds $\min_{u \in U} |u| = s_{\min}$. In Remark 5.5 we provide further explanations for this choice. Common choices for the parameters were $M = 40$, $s_{\min} = 5 \cdot 10^{-3}$ and $s_{\max} \in \{1, 100\}$.

- $(x_i, \sigma_i)_{i=1}^N$: The initial distribution. We chose the locations in the following way: After deciding on an interval $[x_{\min}, x_{\max}]$ we uniformly sample N locations $(x_i)_{i=1}^N$ of this interval. The initial mixed strategies of each player are then chosen as

$$\sigma_i(u) := \frac{e^{-\frac{1}{1-u^2}}}{\sum_{w \in U} e^{-\frac{1}{1-w^2}}} \quad \forall i \in \{1, \dots, N\}. \quad (5.4.3)$$

- I : The maximum number of iterations. Choosing this together with the step size h is equivalent to fixing a time interval $[0, T]$ and a time step h . We generally chose some I very large, e.g. $I = 1000000$, and stop the simulation early once it converged.
- α : Parameter used in the weight function w_Σ (see Eq. (3.4.2)). We use $\alpha = 1000$, unless specified differently.

Remark 5.5 (About the set of strategies U)

Our initial choice for the set of strategies was an equi-spaced grid

$$U = [-s_{\max}, s_{\max}] \cap \kappa \mathbb{Z}, \quad (5.4.4)$$

with $s_{\max} > 0$ and $\kappa > 0$ chosen such that the resulting number of strategies is reasonable. This choice already provided positive results in our early evaluations. But during the process of working with the algorithm and testing it on more difficult functions we made some realizations:

- i) $u^* := \arg \max_{u \in U} |u|$ large is beneficial for the initial exploration of the function graph;
- ii) $u_* := \arg \min_{u \in U} |u|$ has to be small for the population to converge, as it presents the minimal spatial step a player can make;
- iii) On the other hand, u_* too small leads to a very slow convergence as it takes longer for particles to stop moving;
- iv) The total number of strategies has a significant impact on the computational cost.

This leads to the following general idea: By squaring the equi-spaced grid (Eq. (5.4.4)) we get a higher resolution of small strategies and a sparser selection of large strategies, thus increasing the range of movement without increasing the number of strategies. The additional parameter s_{\min} was introduced in order to address the issues regarding small values of u_* and to improve the usability of the algorithm.

Chapter 6

Numerical Results

In this chapter we present numerical results of our global minimization algorithm applied to multiple different functions with varying initial distributions. In Section 5.4 we explained our minimization algorithm, together with all available parameters. While some of the parameters were already specified, the choices for the remaining parameters will always clearly be stated.

We visualize the minimization by presenting snapshots of the simulation. For each of the chosen iterations we show the function graph in blue and the players as circles positioned on this function graph. We can then see the movement of the players and the evolution of the population. We also visualized a selection of the mixed strategies of players, as well as the rate of convergence in function value.

6.1 Minimizing a Non-Convex Function

The first function on which we test our minimization algorithm is the following:

$$g_1(x) := x^2 - 5 \cos(10x) + 5. \quad (6.1.1)$$

It is a very simple example of a non-convex function, with a clear global minimum at $x = 0$. Note that its asymptotic behavior resembles x^2 , leading to a clear slope on a larger scale and even to monotonicity in $[c, \infty)$ for some c large enough as the function becomes very steep. Its graph, both on a smaller and a larger scale, can be seen in Fig. 6.1.

6.1.1 Symmetrically distributed starting locations

We simulated 20 particles with starting locations uniformly distributed in the interval $[-10, 10]$. Further, we use the payoff function J_ϵ described in Section 5.3 with $\epsilon = 10^{-4}$ and the set of strategies $U \subset [-1, 1]$ as in Eq. (5.4.2) with $M = 50$, $s_{\min} = 5 \cdot 10^{-3}$ and $s_{\max} = 1$. Finally, we chose the step sizes $h = 0.05$ and $\gamma = 0.05$. For additional explanation about parameters we refer to Section 5.4.

Figure 6.2 shows some of the iterations of the simulation. While a picture can not describe all details of the dynamic process it still gives a good impression of the general

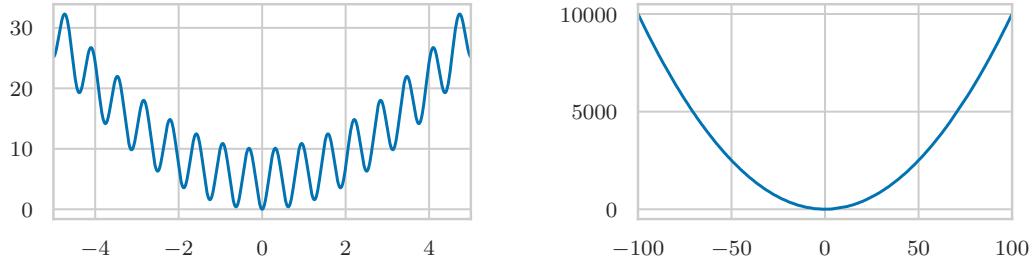


Figure 6.1: Graph of $g_1(x)$ (see Eq. (6.1.1)). On the left the non-convexity of the function can clearly be seen, while the larger scale on the right clearly reveals the asymptotic behavior of $x \mapsto x^2$.

progress and evolution in the population. Each circle corresponds to the location of one player. The colored circles represent a selection of 5 of these 20 players for which we also included plots of their respective mixed strategy at each of the represented iterations, shown in Fig. 6.3.

Observations

Overall, the simulation behaves as expected. Players start moving around and quickly adapt their mixed strategy in order to favor movement towards the player with the lowest function value. On the other hand, the player with the most favorable position adapts a strategy which strongly favors to stay at the current location.

To better understand the dynamics we look at one of the players in more detail, namely player y_2 in yellow. The player starts far on the left of the interval, such that its mixed strategy adapts in order to favor positive movement, as can be seen in Fig. 6.3, row 2 column 2. The player therefore gains momentum towards the right. In iteration 1500 we can see that it moved past the global minimum. During the remaining simulation it oscillates around the global minimum, and we see that the mixed strategy adapts depending on the side the player is on, until it joins the majority of the population in the global minimum. One notable observation is the emergence of a sharp peak in its mixed strategy, which can first be seen in iteration 1500. This peak occurred precisely when the player passed the global minimum as during this time it was the player with the lowest function value, which strongly promotes the strategy $u = 0$ due to our choice of J_ϵ . Therefore, once a player gets sufficiently close to being the player with the lowest function value of all players, it loses a significant amount of momentum. Over the course of many tests and simulations, we made the observation that this behavior is very important for the population to converge in reasonable time. On the other hand, a too extreme loss of momentum prevents the

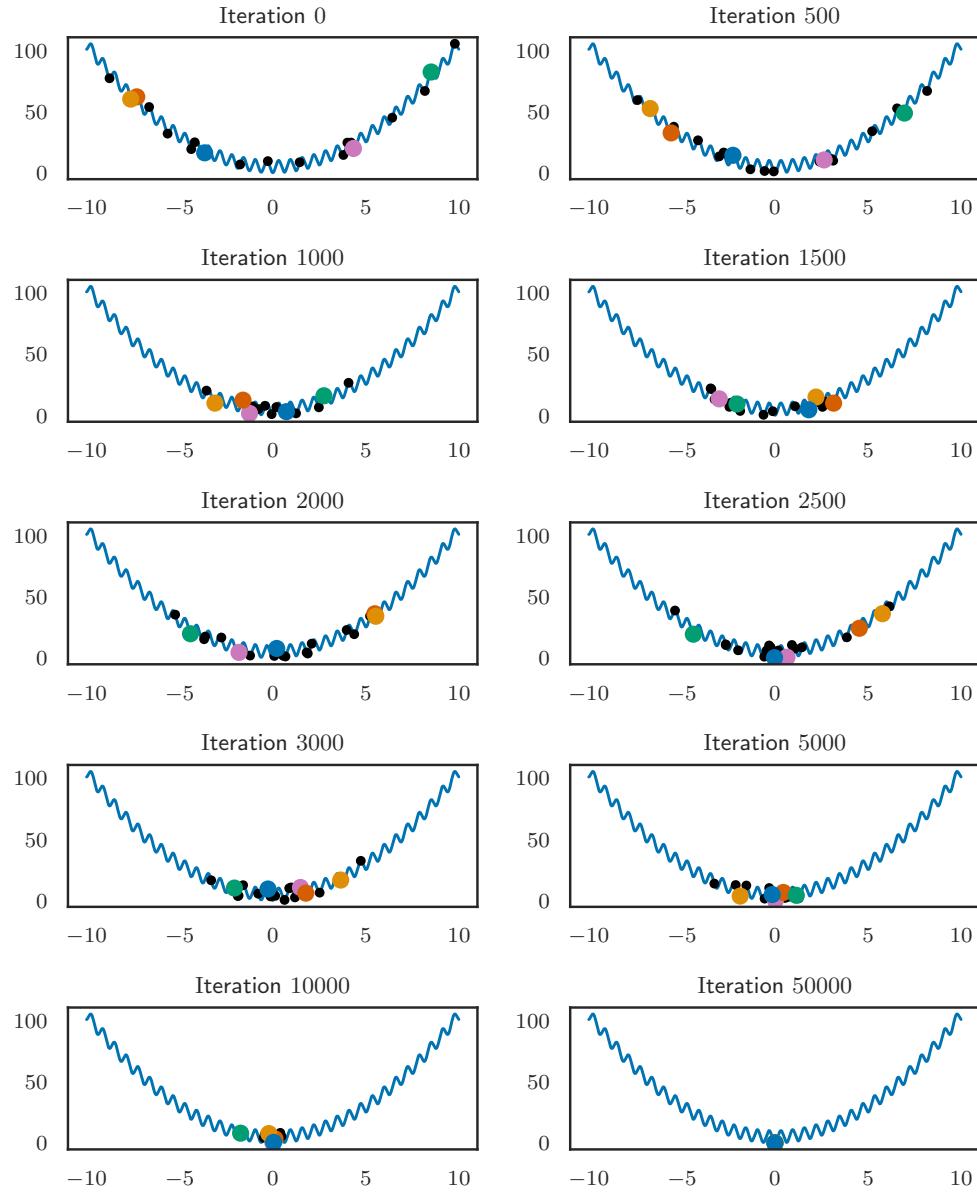


Figure 6.2: Example animation with 20 players. All points represent a player, both the larger colored circles and the smaller black circles. The colors match those in Fig. 6.3 in order to show the mixed strategy of each of those 5 players at each pictured iteration.

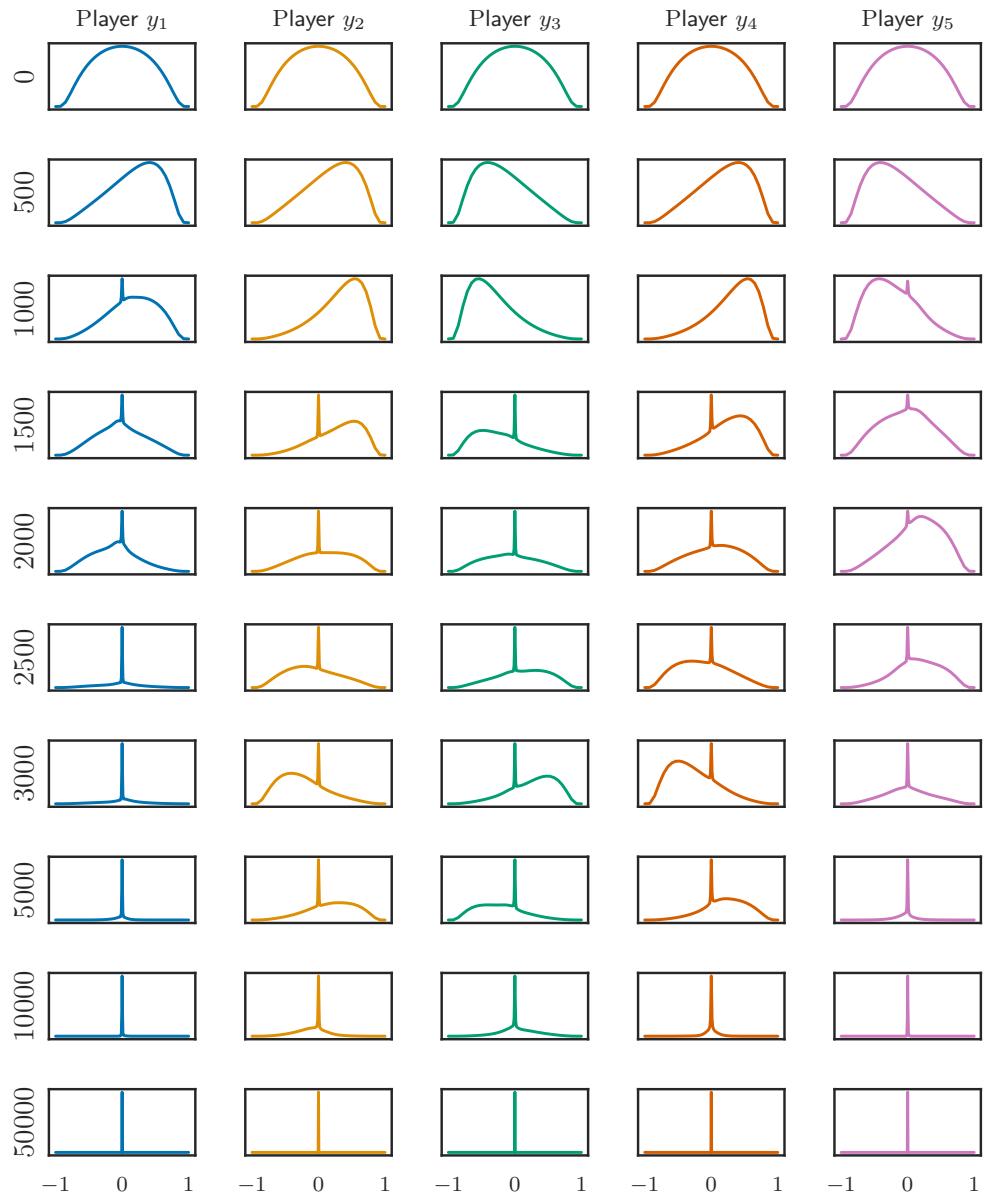


Figure 6.3: Mixed strategies corresponding to Fig. 6.2. Each row corresponds to one of the shown iterations, and each column to one of the colored players.

players from *overshooting* the player with lowest function value, which is crucial in order not to get stuck in a local minimum. We will discuss this property further in Section 6.1.2.

Rate of convergence

Finally, Fig. 6.4 shows the mean function value of the population in each iteration. Given the population $\{(x_{i,t}, \sigma_{i,t})\}_{t=1}^N$ at each time step $t \in \{0, \dots, I_{\max}\}$, with I_{\max} the maximum iteration of the simulation, this corresponds to the value of the function $F : \{0, \dots, I_{\max}\} \rightarrow \mathbb{R}_+$ defined as

$$F_1(t) = \frac{1}{N} \sum_{i=1}^N g_1(x_{i,t}). \quad (6.1.2)$$

In the presented plot the y -axis uses a logarithmic scale, which highlights the mostly exponential convergence. Note that between iteration 20000 and 40000 there appears an oscillation, due to the single last player to converge. While this player is constantly attracted by the rest of the population in the global minimum, its momentum decreases only slowly so that the player shoots past the global minimum multiple times. Once the momentum has sufficiently decreased the player joins the rest of the population and converges to the global minimum.

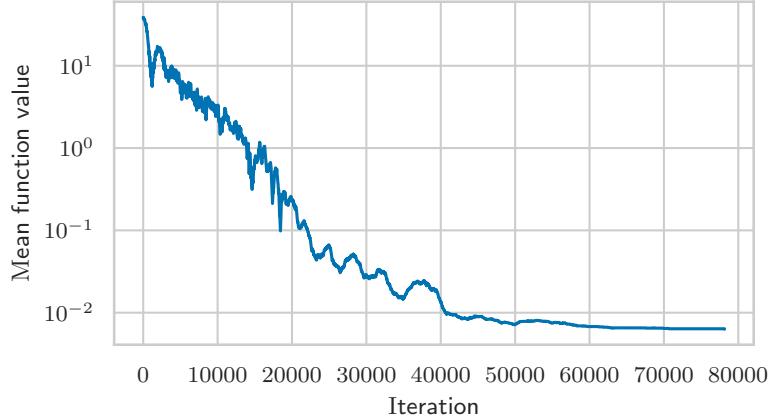


Figure 6.4: Mean function value of the simulation for each iteration, corresponding to the graph of the function F_1 as described in Eq. (6.1.2).

6.1.2 One-sided starting locations

The previous section showed very positive results, and with the chosen parameters we always achieved convergence to the global minimum. However, we always started with initial locations distributed over the interval $[-10, 10]$, which contains the global minimum at $x = 0$. The probability of having a player very close to the minimum is high, and as the other players then move towards this player in a continuous manner the global minimum will get discovered by someone, who then in turn attracts the rest of the population. We will therefore investigate the behavior when we start in an interval which does not contain the global minimum.

We worked again with 20 particles, but this time with starting locations uniformly in $[5, 10]$. All other parameters are chosen as before (see Section 6.1.1). The results can be seen in Figs. 6.5 and 6.6, where we again highlighted 5 particles in order to show the evolution of their mixed strategies. We tested this simulation multiple times and with the chosen parameters we always obtained positive results and successfully found the global minimum. The mean function-value can be seen in Fig. 6.7.

Observations

The simulation seems to behave similarly as the simulation before. However, there is one detail which is much more relevant in this example than in the symmetric situation of the previous section: If the players adapt their mixed strategy too quickly, the population can get stuck in a local minimum. Indeed, we can force this to happen by choosing γ too large, e.g. $\gamma = 10 = \frac{1}{2h}$, which can be seen in Fig. 6.8. Once a player $y_i = (x_i, \sigma_i)$ is the lowest in the population its mixed strategy adapts and $\sigma_i(0)$ gets very large, leading to very rare movement of the player. All other players move towards this player y_i , and in turn either become the lowest player and stop moving, or due to an upwards slope after a local minimum they change their direction as the lowest player is then to their right. Therefore, in order to successfully find the global minimum it is crucial that players do not lose too much momentum, which can be achieved by choosing a lower step size γ so that players do not adapt their strategy as quickly.

While this general property of γ does hold in any scenario, it would not necessarily lead to convergence to a local minimum in the case of initial locations uniformly in $[-10, 10]$, as players start on both sides of the global minimum. In order to converge to a single location players are forced to move through it and at some point in time the global minimum gets discovered by a player, who then might stop moving and wait until all other players converge towards it. Still, our results were generally better when choosing γ not too large.

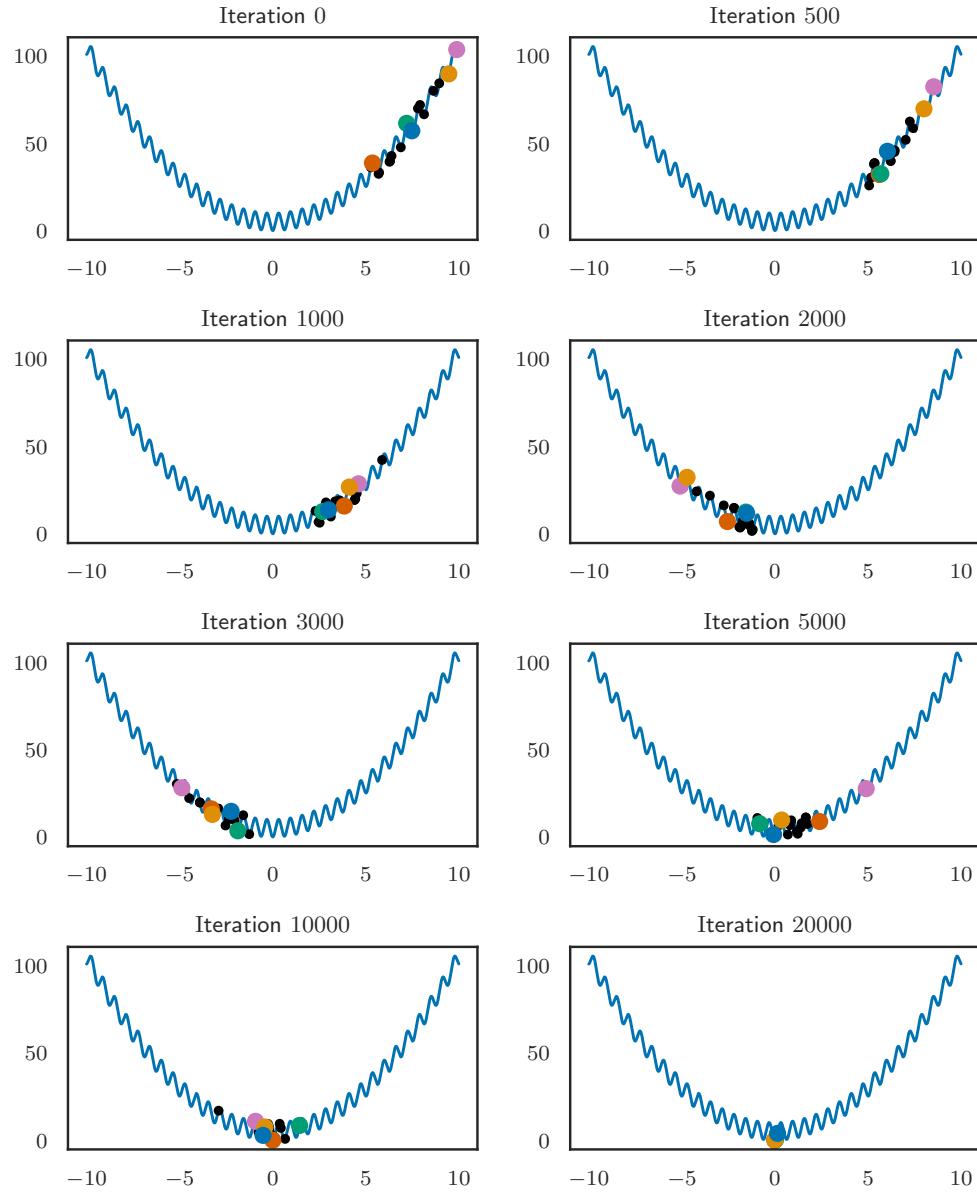


Figure 6.5: Example animation with 20 players, similar to the one shown in Fig. 6.2, where again each point represents a player in the population. The main difference is the one-sided initial distribution of the players in the interval $[5, 10]$.

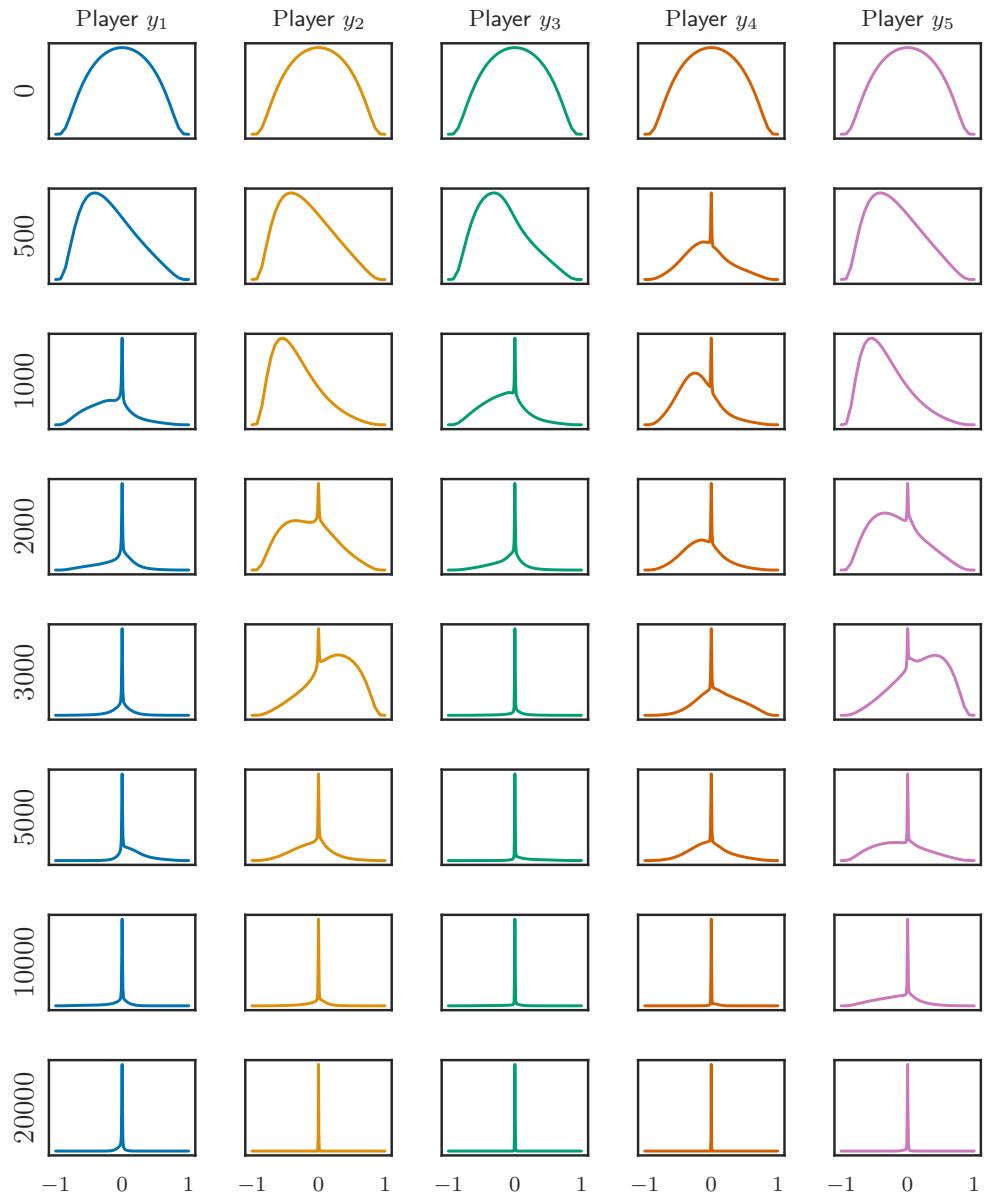


Figure 6.6: Mixed strategies corresponding to Fig. 6.5, with each row corresponding to one of the shown iterations and each column to one of the colored players.

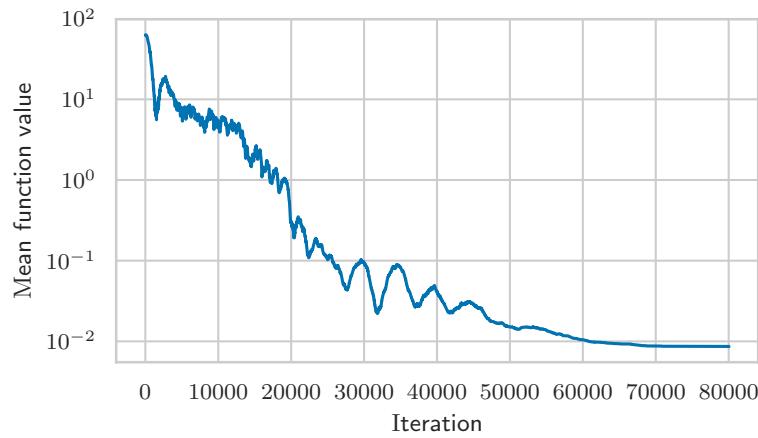


Figure 6.7: Mean function value of the simulation with one-sided initial locations.

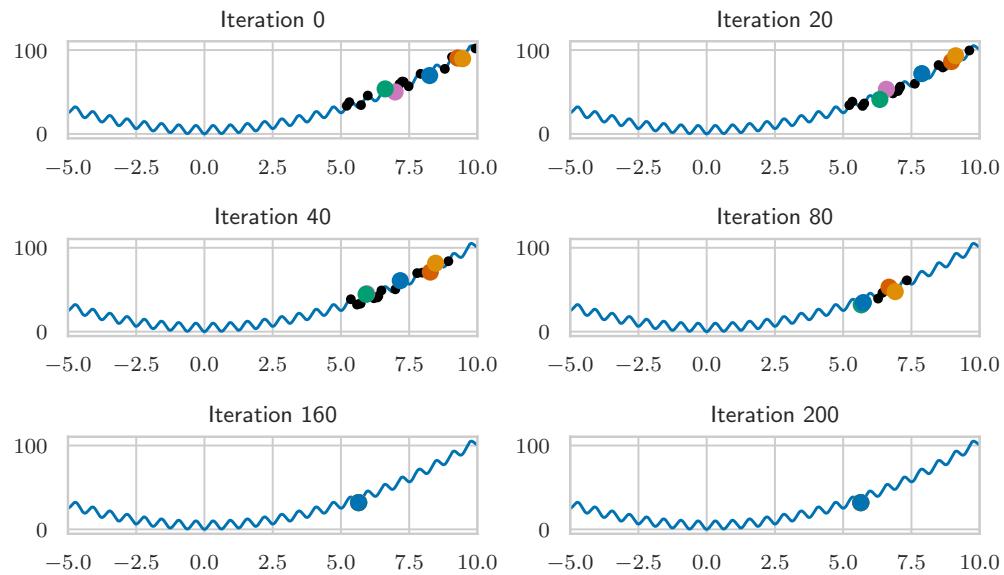


Figure 6.8: Unsuccessful minimization of g_1 due to γ being chosen too large.

6.2 Ackley Benchmark

We will now consider the Ackley function g_{Ackley} defined as

$$g_{\text{Ackley}}(x) := -20 \exp\left(-0.2 \cdot \sqrt{0.5x^2}\right) - \exp\left(\frac{\cos(2\pi x) + 1}{2}\right) + e + 20. \quad (6.2.1)$$

This is a well known benchmark for global optimization problems [AJ13], introduced by David Ackley in [Ack87] and also used by Carrillo et al. in [Car+16]. It is a highly non-convex function and has its global minimum at $x = 0$, with $g_{\text{Ackley}}(0) = 0$. Compared to g_1 as defined in Eq. (6.1.1), the main difference and difficulty lies in the asymptotic behavior, as can be seen in Fig. 6.9: There is no noticeable slope, therefore with starting locations far away from the well around $x = 0$ this task is even more difficult to solve. Additionally, for the function g_1 it holds $g_1(x) \geq x^2$ for all $x \in \mathbb{R}$. Conversely for g_{Ackley} the only convex function \tilde{g} with $g_{\text{Ackley}} \geq \tilde{g}$ is $\tilde{g}(x) = 0$.

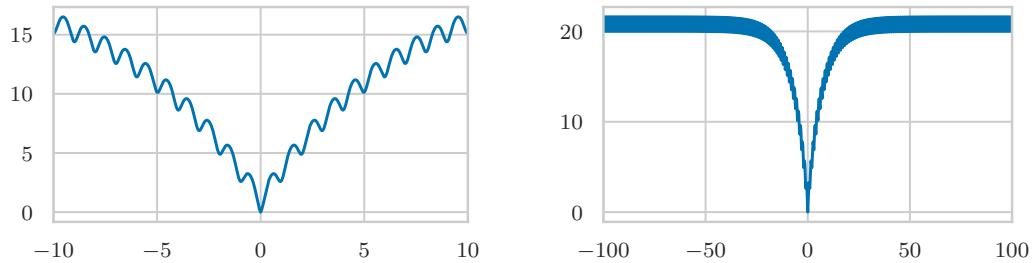


Figure 6.9: Graph of the Ackley function. It is highly non-convex, but a major difference to g_1 lies in its asymptotic behavior (see Fig. 6.1). As $|x|$ becomes large the Ackley function just oscillates between 20 and $20 + 2e$, with no noticeable general slope indicating some direction of descend.

We again include both a symmetric distribution around the global minimum at $x = 0$ and a one-sided distribution. In order to make the different asymptotic behavior more relevant, we also chose to work on a larger scale. The 50 locations for the symmetric initial distribution have been chosen uniformly in $[-100, 100]$, and the locations for the one-sided scenario uniformly in $[200, 500]$. In both scenarios, the other parameters were chosen as follows: We use the payoff function J_ϵ as in Section 5.3 with $\epsilon = 10^{-4}$, and the step sizes $h = 0.5$ and $\gamma = 0.05$. The set of strategies U is again chosen as in Eq. (5.4.2), but this time with $M = 40$, $s_{\min} = 5 \cdot 10^{-3}$ and $s_{\max} = 100$, such that $U \subset [-100, 100]$. We chose U in this way in order to enable farther movement.

Observations

With the chosen set of parameters, both simulations appeared very stable and always converged during our testing. The evolution of the population can be seen in Figs. 6.10 and 6.12, but we did not include a visualization of some individual mixed strategies, as they behave very similarly to those shown in Figs. 6.3 and 6.6. Again, the convergence of the mean function value seems exponential, as can be seen in Fig. 6.11.

In the first scenario, with the initial locations uniformly in $[-100, 100]$, the probability of one player having a lower function value than 20 is rather high. This player quickly attracts the rest of the population. In Fig. 6.10 we see for example the yellow player in iteration 0 with a very favorable position.

On the other hand, with starting locations in $[200, 500]$ this does not happen. The reason for success in this scenario lies in the fact that players during early iterations explore, as their initial mixed strategy is does not have a clear shape yet. With large movements $U \subset [-100, 100]$ they are then able to discover the well surrounding the global minimum, which then attracts all other players. While we only had positive results during testing for this situation, it is not difficult to create a situation in which the population might fail to find the global minimum. Indeed, with starting locations even further away, in $[1000, 4000]$ the population failed and converged to a local minimum, which can be seen in Fig. 6.13. Of course it might be possible to adapt some of the parameters, e.g. to decrease γ to enable further exploration, or to chose $s_{\max} >> 100$, but the main purpose of the example was merely to describe such a situation and to explain the reasons for this behavior.

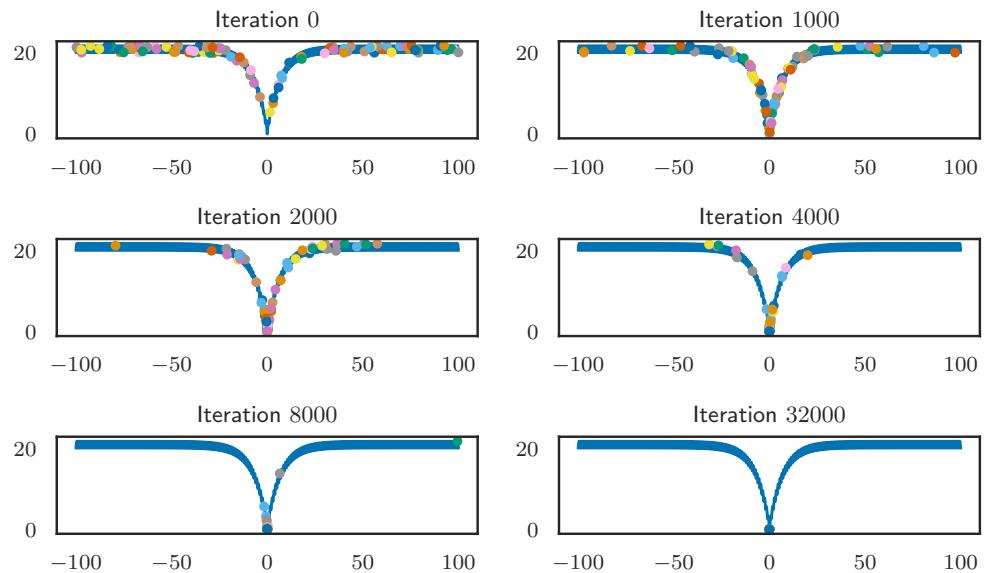


Figure 6.10: Ackley benchmark with a balanced starting distribution.

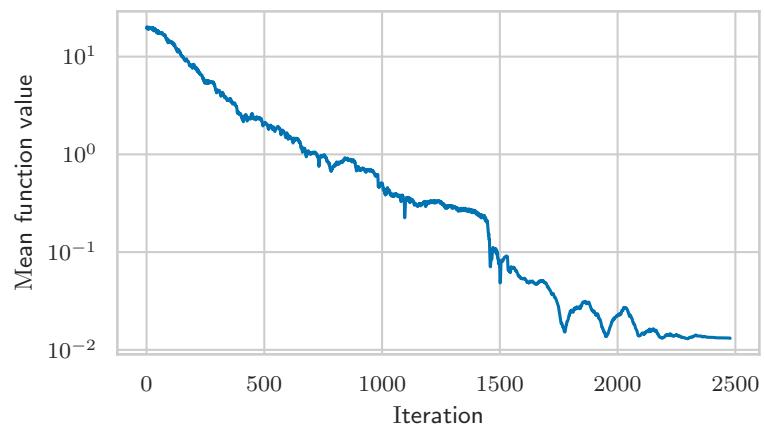


Figure 6.11: Mean function value during the minimization of the Ackley function with a balanced starting distribution.

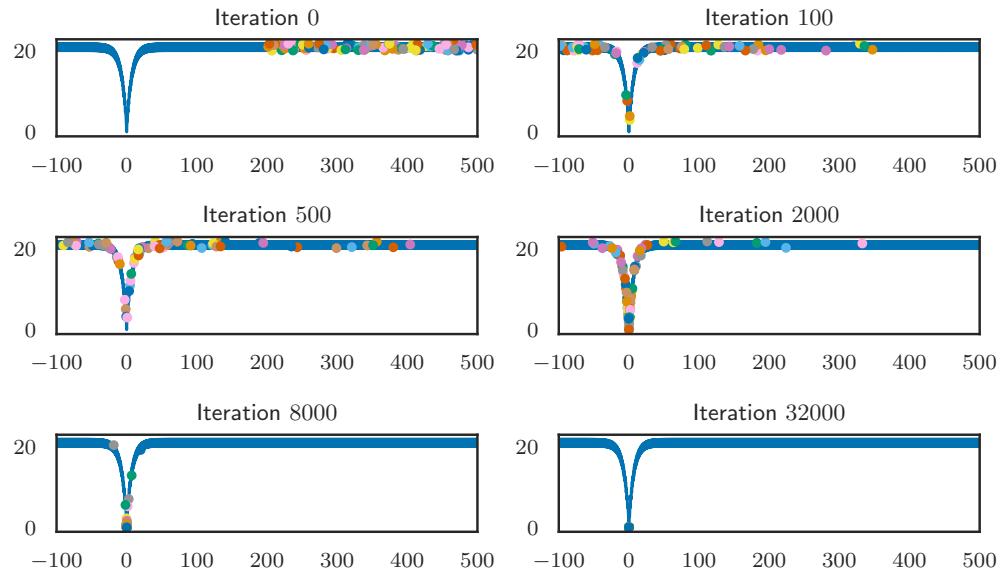


Figure 6.12: Successful minimization of the Ackley function given initial locations in $[200, 500]$.

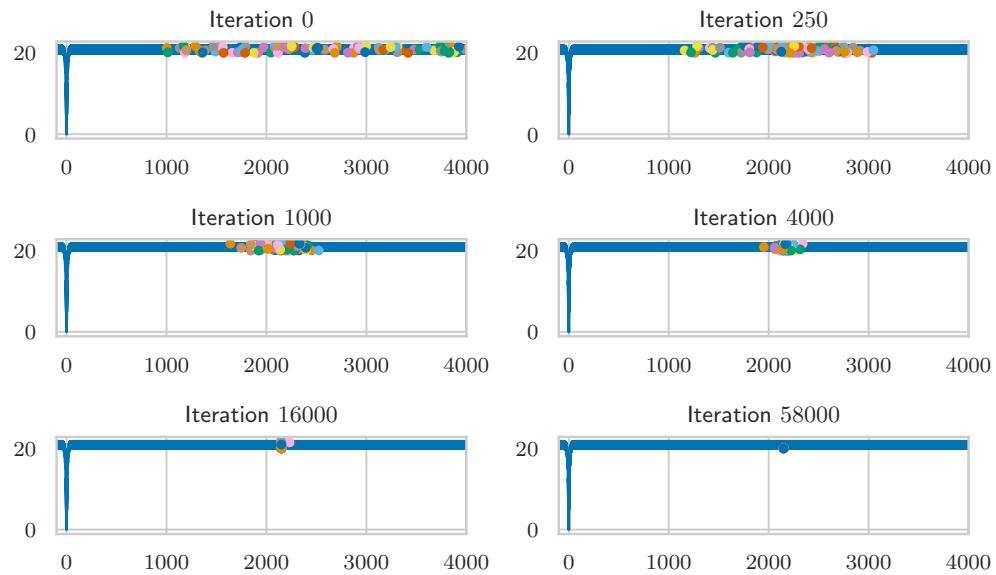


Figure 6.13: Unsuccessful minimization of the Ackley function g_{Ackley} with initial locations in $[1000, 4000]$.

6.3 Modified Ackley Function

As a final benchmark, we modified the Ackley function in the following way:

$$g_2(x) := \frac{2g_{\text{Ackley}}(x) + g_{\text{Ackley}}(x - 200) - g_{\text{Ackley}}(-200)}{2}. \quad (6.3.1)$$

While this function also has its global minimum at $x = 0$, together with the same asymptotic behavior as the original Ackley function, it has another additional region around $x = 200$ with lower function values. We wanted to test if the algorithm, given starting locations > 200 , correctly finds the global minimum or if it converges too quickly once a player found the second well. Figure 6.14 shows the corresponding function graph.

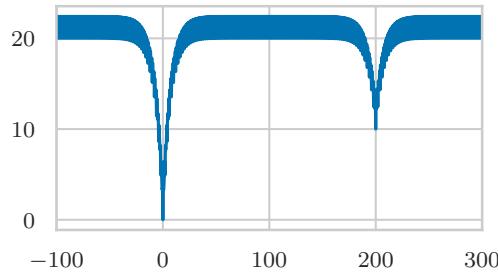


Figure 6.14: Modified Ackley function with two wells, see Eq. (6.3.1).

Figure 6.15 shows results with initial locations in $[300, 500]$. Similarly to the results on the Ackley function with one-sided initial locations, the crucial part are again the initial iterations. As the parameters were chose as before in (see Section 6.2), the players in the population start by randomly exploring the function graph. While in iteration 50 it can be seen that there is a player inside the well around 200, the players to the left kept their momentum long enough to discover the global minimum. The population was then attracted towards this location and proceeded to converge towards the global minimum.

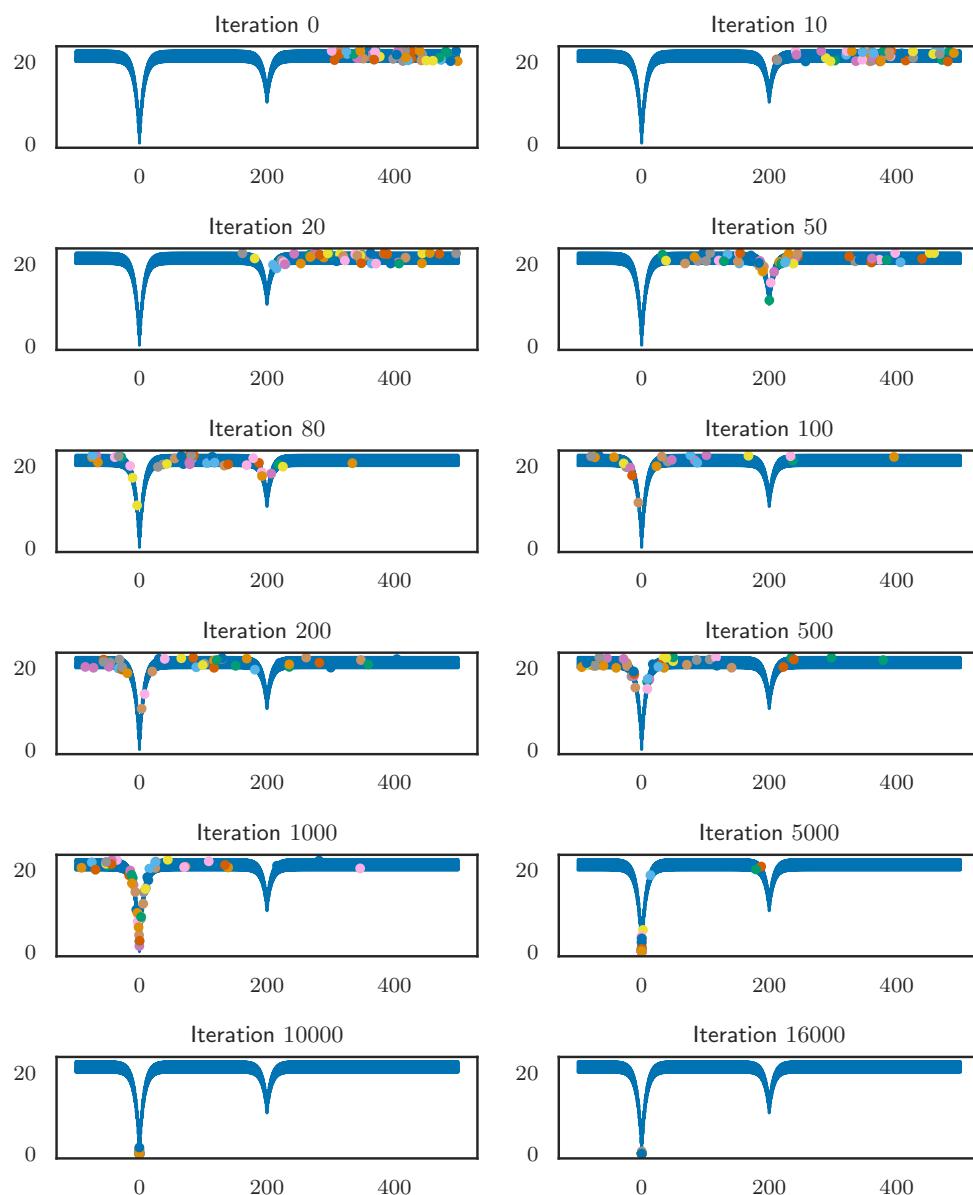


Figure 6.15: Successful minimization of the modified Ackley function (see Eq. (6.3.1)), given initial locations in $[300, 500]$.

6.4 The Global Optimization Method by Carrillo et al.

We will now compare both the results and the behavior of our algorithm with the consensus-based global optimization method proposed by Carrillo et al. in [Car+16]. In Section 3.2 we already presented the method briefly and explained the general functioning of the algorithm.

6.4.1 Results

We used the same parameters as described in [Car+16, Section 5.5]. Overall, the minimization algorithm had very convincing results on all of the test functions we chose and achieved fast convergence. It successfully minimized every function, even with starting points chosen very far away from the global minimum. Figure 6.16 shows one such example. The initial locations of the 5000 particles were all in the small interval [999, 1000]. Due to the functioning of the algorithm particles still started to explore the function graph, quickly reaching locations far away from the group. This behavior can be seen in iterations 100, 200 and 300. Then, once a particle found the region around the global minimum the weighted mean-location m_t , as defined in Eq. (3.2.2), lies inside this region, thus attracting all of the particles. Iteration 400 first shows this attraction of the population, and over the remaining iterations all particles converge towards the global minimum at $x = 0$.

In Fig. 6.17 we included another result. The starting locations were again all in the small interval [999, 1000], but this time we applied the optimization method to the modified Ackley function introduced in Eq. (6.3.1). Note how in iteration 1500 the population found the second minimum at 200, which directly attracted the whole population. Then, at some point between iteration 1500 and 2000, a particle found an even lower location around 0, and from iteration 2000 onwards we see a similar convergence behavior as before, where the global minimum slowly pulls the whole population in.

6.4.2 Comparison with our method

Both our algorithm and the optimization method by Carrillo et al. use a particle-based approach, but the underlying theory differs vastly.

In the algorithm by Carrillo et al., the population consists of agents which are defined by a position. Then, during a single step, each agent moves towards the weighted mean location m_t , which is exponentially closer to agents with a low function value. Finally, a Gaussian noise proportional to the distance to m_t is added to the position of the player. Due to this stochastic process agents jump around a lot, and change their location on the function graph in a non-continuous manner, almost resembling a gas which moves over the graph. This behavior enables them to discover a large part of the

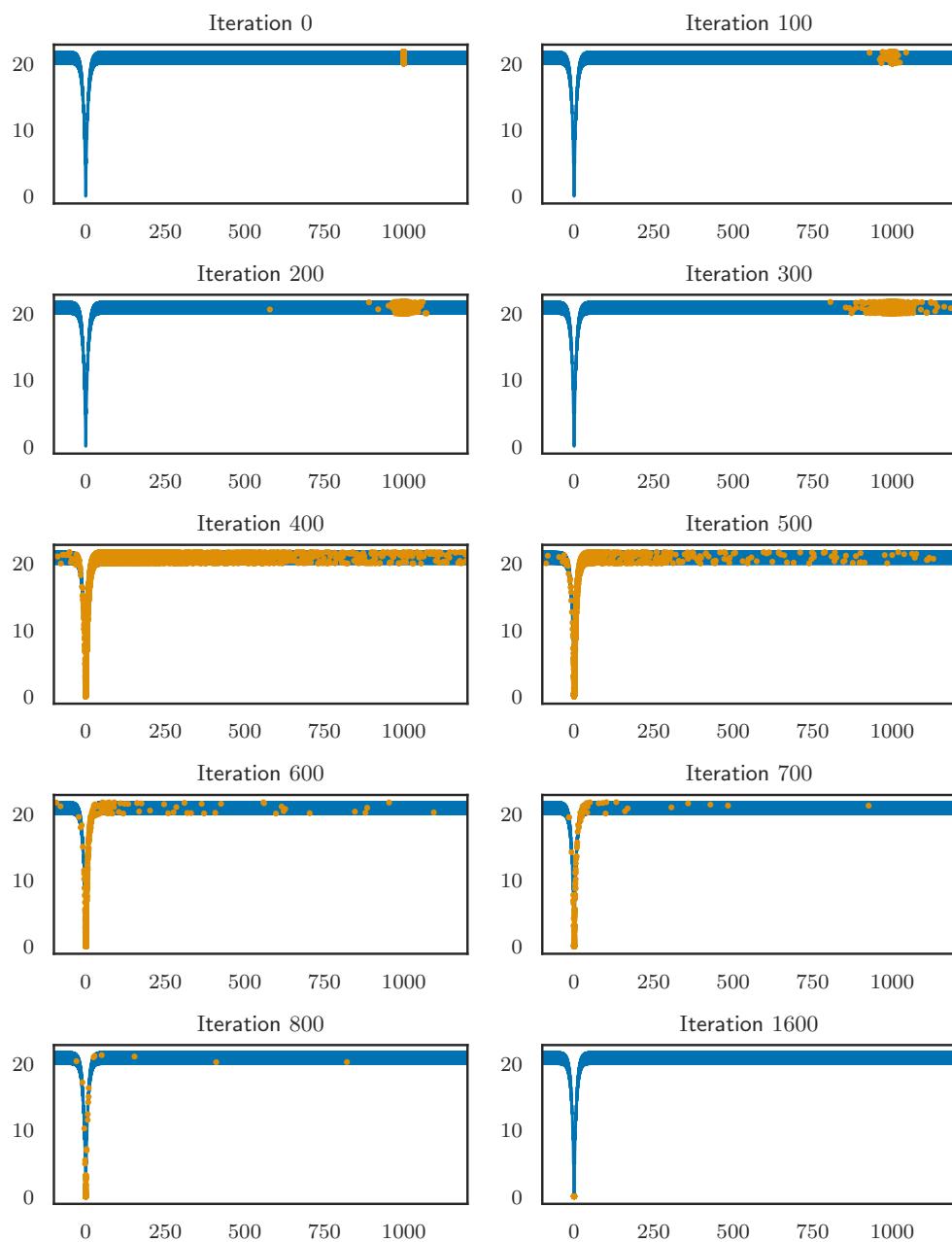


Figure 6.16: Results of the optimization method from Carrillo et al. on the Ackley benchmark, with 5000 initial locations in $[999, 1000]$.

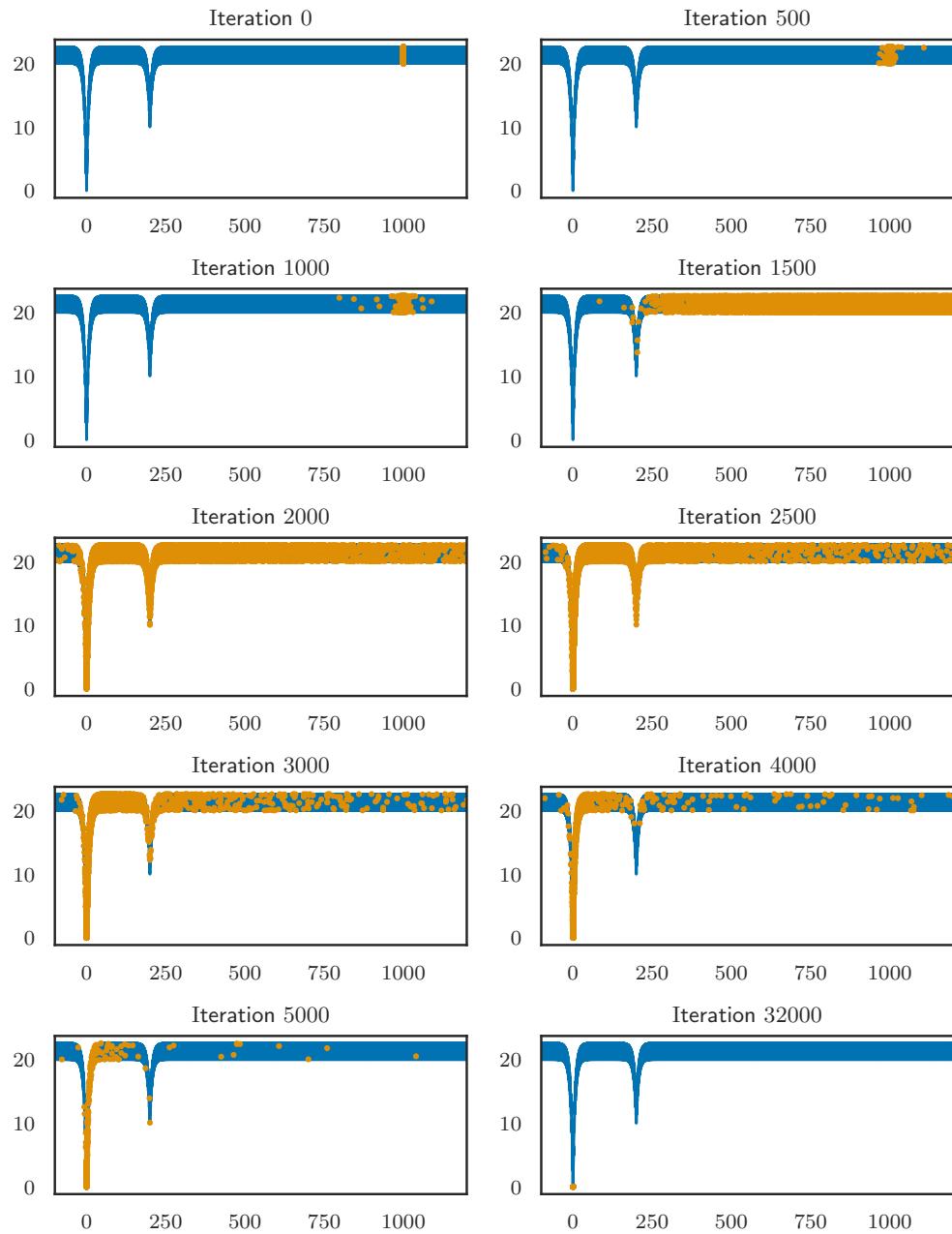


Figure 6.17: Results of the optimization method from Carrillo et al. on the modified Ackley function, with 10000 initial locations in [999, 1000].

graph very quickly. Further, the exponentially weighted mean location is comparatively fast to compute. With a population of N players it requires $O(N)$ evaluations of the objective function, and the new location of each player can then directly be calculated in a single step.

In our approach based on evolutionary game theory the population consists of players, each with location and strategy. The mechanism to update the locations of each player are, except for the reweighting, solely based on one-on-one interactions of the players. The computational costs are higher as these interactions have to be computed for each combination of player, strategy and co-player. With N players and M strategies this induces $O(MN^2)$ evaluations of the payoff function J for each update of the population, on top of the $O(N)$ evaluations of the objective function to calculate the weights. Further, the available options of movement for each player are limited as U is chosen before the minimization starts as a finite and bounded set. The velocity of the players is therefore bounded, and their capabilities of exploration and convergence are limited.

In this direct comparison, the optimization method of Carrillo et al. shows faster convergence at a lower computational complexity, due to the very different nature of both algorithms.

6.4.3 Functions with infinitely many minima

As a final result we applied both algorithms to the function

$$g(x) = \cos(2\pi x) + 1. \quad (6.4.1)$$

This example might seem unusual as the cosine does not have a unique global minimum, but the different behavior of each algorithm becomes very clear.

In Fig. 6.18 we show the locations of actors in the optimization method from Carrillo et al. We started with a population of 10000 points in the interval $[-5, 5]$. In the iterations 200 and 400 most mass still seems to be concentrated around the initial locations, but the number of particles which move away in order to explore the function graph increases over time. Iterations 600, 800 and 1000 clearly show the gas-like behavior of the population, spreading over the function graph. This behavior carries on indefinitely.

In comparison, our optimization method behaved very differently, as can be seen in Fig. 6.19. We started with a population of 50 particles with locations uniformly distributed in $[-20, 20]$, and the set of strategies U chosen with $s_{\max} = 1$. The other parameters were chosen as in Section 6.2. Quickly after the start of the simulation the players seem to group together. Between iteration 250 and 1000 the whole group has moved towards the right, without any clear point of attraction, so we assume this is purely due to randomness. Then, over the course of the remaining simulation, the population converges to a stable state. This is already very different to the

consensus-based optimization method by Carrillo et al., which never converged. Even in this ill-posed situation the population evolved in order to reach an equilibrium. Additionally, this example shows a behavior which did not appear in previous examples: The population does not converge towards a single location, but instead finds an equilibrium with players in two different global minima. In practice, the likelihood of the occurrence of such situations can be decreased by modifying parameters, such as α or the number of particles N . However, even with different parameters, a state consisting of a split population with each part lying in a different global minimum, with mixed strategies sufficiently close to $\sigma = \delta_0$, is a stable state. This is due to the behavior of the payoff function $J_\epsilon(x, \cdot; x')$ when $g(x) = g(x')$, as the highest payoff in this situation is to the strategy $u = 0$. Each player then adapts its mixed strategy and increase the preference for $u = 0$, therefore further decreasing the movement in the population. Thus, the concentration in multiple global minima represents a state of equilibrium.

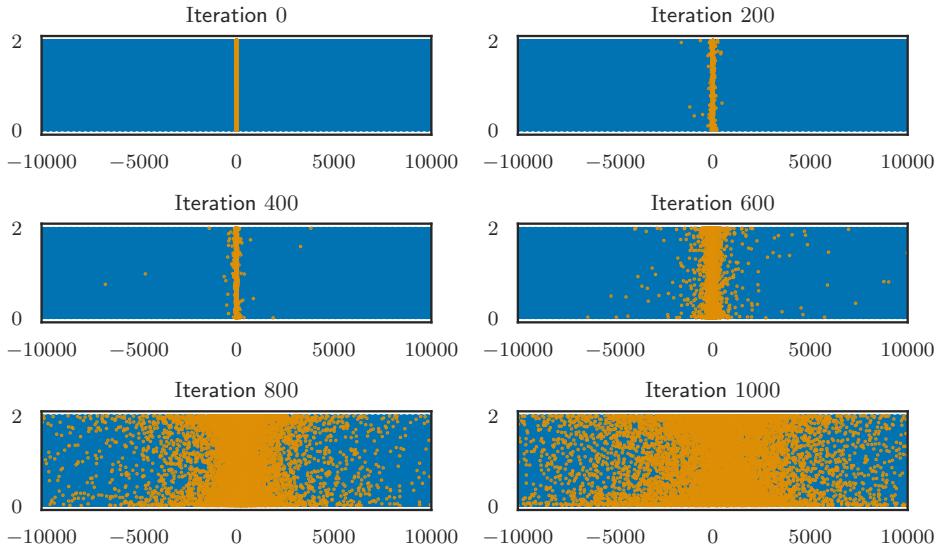


Figure 6.18: Visualization of the consensus-based optimization method of Carrillo et al., applied to the function $g(x) = \cos(2\pi x) + 1$. The agents quickly start exploring and gradually distribute over the whole visible function graph, without ever converging.

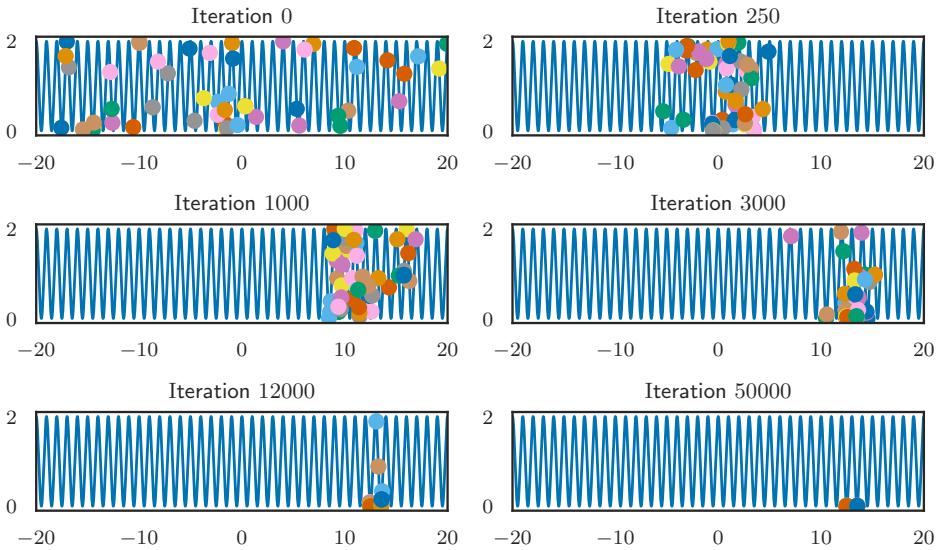


Figure 6.19: Visualization of our EGT-based optimization method, applied to the function $g(x) = \cos(2\pi x) + 1$. Note how the population converged towards two different global minima.

Chapter 7

Conclusion

The presented numerical results show a successful application of our minimization algorithm to the problem of global optimization. Our method was able to minimize highly non-convex functions with different starting positions, and we observed exponential convergence. However, we learned that players tend to lose momentum over time, which can be a disadvantage. Initial populations located very far away from the global minimum showed unsuccessful minimizations of the Ackley function as the population converged towards a local minimum. Still, from the point of view of evolutionary games the tendency to converge is a positive result by itself. The population successfully adapts and evolves until an equilibrium is reached.

During the development of the minimization method, we modified the underlying replicator dynamics of spatially inhomogeneous evolutionary games, in order to include considerations on each player's function value. We provided results on the well-posedness of the weighted replicator dynamics and proved the existence and uniqueness of Lagrangian solutions. After further numerical adjustments we proposed a suitable payoff function. The results show that by careful design of the game we are able to influence the behavior of each individual in the desired way. In our tests we generally converge towards a state of equilibrium, in which players are located at the global minimum. Nevertheless, various aspects of our global optimization algorithm could be further investigated. We made multiple simplifications in order to develop our algorithm, but their relaxation might be beneficial. Most notably, the proposed payoff function does not take the chosen strategy of the co-player into account. With its inclusion it might be possible to anticipate the planned direction of movement of the co-player and use it improve the results or to accelerate convergence. Finally, the results on higher-dimensional problems are still unclear and further research could be done in order to asses the performance of our algorithm.

We included the consensus-based optimization method of Carrillo et al. [Car+16], both as an inspiration for the development of our method and as a comparison for the numerical results. The presented behavior of this algorithm differs from ours, which we showed clearly in the attempted minimization of the cosine. The optimization method of Carrillo et al. explored the function graph indefinitely. On the other hand, in our algorithm the population reaches an equilibrium and finds two of the global minima.

This is both an advantage and a disadvantage: Our algorithm is robust to functions with infinitely many global minima and provides a result. However, during our tests on functions with a unique global minimum the tendency to explore showed more consistent convergences to the global minimum.

Finally, we investigated the complexity of our algorithm. For each iteration, our method requires more computations than the algorithm of Carrillo et al. But this computational disadvantage is directly tied to the very nature of evolutionary game theory: Each player is more than just a location. The population is made of individuals, which do not necessarily act in order to reach the best state of the population, but instead try to maximize their own personal payoff. The resulting dynamics are independent of the problem of global function minimization; spatially inhomogeneous evolutionary games were not developed with the goal of solving optimization problems. We presented a successful application of this model to global optimization, but the potential range of applications is much larger. Classical evolutionary games were often used to describe and explain the observed behavior of animals, as the dynamics capture their egoistic behavior and the notion of Darwinian fitness. With the inclusion of a spatial component the model could describe a much broader range of populations, and it might be well-suited in order to simulate groups of individuals in the real world. In future work, it would be interesting to see other fields of application for spatially inhomogeneous evolutionary games.

List of Figures

5.1	Illustrations of the different components of J	51
5.2	Visualization of the payoff function $J_\epsilon(x_1, \cdot, x')$	52
6.1	Graph of $g_1(x)$	58
6.2	Minimization of g_1	59
6.3	Minimization of g_1 - strategies.	60
6.4	Minimization of $g_1(x)$ - Mean function value.	61
6.5	Minimization of $g_1(x)$ with one-sided initial locations.	63
6.6	Minimization of $g_1(x)$ with one-sided initial locations - strategies.	64
6.7	Minimization of $g_1(x)$ with one-sided initial locations - convergence.	65
6.8	Unsuccessful minimization of g_1 due to γ being chosen too large.	65
6.9	Graph of the Ackley function.	66
6.10	Ackley benchmark with a balanced starting distribution.	68
6.11	Ackley benchmark - Mean function value.	68
6.12	Ackley benchmark with a one-sided starting distribution.	69
6.13	Unsuccessful minimization of the Ackley function.	69
6.14	Modified Ackley function with two wells.	70
6.15	Minimization of the modified Ackley function	71
6.16	Ackley benchmark of the optimization method from Carrillo et al.	73
6.17	Optimization method of Carrillo et al. on the modified Ackley function.	74
6.18	Consensus-based optimization method of Carrillo et al. on the cosine.	77
6.19	Our minimization algorithm on the cosine.	77

Bibliography

- [Ack87] D. H. Ackley. “A Connectionist Machine for Genetic Hillclimbing”. In: *The Kluwer International Series in Engineering and Computer Science* (1987). ISSN: 0893-3405.
- [AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Science Publications. Clarendon Press, 2000. ISBN: 9780198502456.
- [AGS05] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2005. ISBN: 9783764324285.
- [Amb+18] L. Ambrosio, M. Fornasier, M. Morandotti, and G. Savaré. *Spatially Inhomogeneous Evolutionary Games*. 2018.
- [AJ13] O. Askari-Sichani and M. Jalili. “Large-scale global optimization through consensus of opinions over complex networks”. In: *Complex Adaptive Systems Modeling* 1.1 (2013), p. 11. ISSN: 2194-3206.
- [BFM97] T. Back, D. B. Fogel, and Z. Michalewicz, eds. *Handbook of Evolutionary Computation*. 1st. Bristol, UK, UK: IOP Publishing Ltd., 1997. ISBN: 0750303921.
- [BR03] C. Blum and A. Roli. “Metaheuristics in combinatorial optimization”. In: *ACM Computing Surveys* 35.3 (Sept. 2003), pp. 268–308. ISSN: 0360-0300. DOI: 10.1145/937503.937505. URL: <http://dx.doi.org/10.1145/937503.937505>.
- [Boc33] S. Bochner. “Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind”. In: *Fundamenta Mathematicae* 20 (1933), pp. 262–176. ISSN: 1730-6329.
- [Bom90] I. M. Bomze. “Dynamical aspects of evolutionary stability”. In: *Monatshefte für Mathematik* 110.3-4 (Sept. 1990), pp. 189–206. ISSN: 1436-5081.
- [Bos18] N. Bosch. *Source Code of my Master’s Thesis*. Oct. 2018. DOI: 10.5281/zenodo.1458784. URL: <https://doi.org/10.5281/zenodo.1458784>.
- [Bre73] H. Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Mathematics Studies. Elsevier Science, 1973. ISBN: 9780080871165.

Bibliography

- [Car+16] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. “An analytical framework for a consensus-based global optimization method”. In: *ArXiv e-prints* (Jan. 2016). arXiv: 1602.00220 [math.AP].
- [Cre05] R. Cressman. “Stability of the replicator equation with continuous strategy space”. In: *Mathematical Social Sciences* 50.2 (Sept. 2005), pp. 127–147. ISSN: 0165-4896.
- [Des+08] L. Desvillettes, P. E. Jabin, S. Mischler, and G. Raoul. “On selection dynamics for continuous structured populations”. In: *Communications in Mathematical Sciences* 6.3 (2008), pp. 729–747. ISSN: 1945-0796.
- [Gal11] A. Galstyan. “Continuous strategy replicator dynamics for multi-agent Q-learning”. In: *Autonomous Agents and Multi-Agent Systems* 26.1 (Sept. 2011), pp. 37–53. ISSN: 1573-7454.
- [HS98] J. Hofbauer and K. Sigmund. “Evolutionary Games and Population Dynamics”. In: (1998).
- [HS83] J. Hofbauer and K. Sigmund. “Replicator dynamics”. In: *Evolutionary Games and Population Dynamics* (1983), pp. 67–85.
- [HS88] R. Holley and D. Stroock. “Simulated annealing via Sobolev inequalities”. In: *Communications in Mathematical Physics* 115.4 (Dec. 1988), pp. 553–569. ISSN: 1432-0916. DOI: 10.1007/bf01224127. URL: <http://dx.doi.org/10.1007/bf01224127>.
- [HKS89] R. A. Holley, S. Kusuoka, and D. W. Stroock. “Asymptotics of the spectral gap with applications to the theory of simulated annealing”. In: *Journal of Functional Analysis* 83.2 (Apr. 1989), pp. 333–347. ISSN: 0022-1236. DOI: 10.1016/0022-1236(89)90023-2. URL: [http://dx.doi.org/10.1016/0022-1236\(89\)90023-2](http://dx.doi.org/10.1016/0022-1236(89)90023-2).
- [Neu28] J. v. Neumann. “Zur Theorie der Gesellschaftsspiele”. In: *Mathematische Annalen* 100.1 (Dec. 1928), pp. 295–320. ISSN: 1432-1807.
- [Nor08] T. W. Norman. “Dynamically stable sets in infinite strategy spaces”. In: *Games and Economic Behavior* 62.2 (Mar. 2008), pp. 610–627. ISSN: 0899-8256.
- [OHR07] J. Oechssler, J. Hofbauer, and F. Riedel. “Brown-Von Neumann-Nash Dynamics: The Continuous Strategy Case”. In: *SSRN Electronic Journal* (2007). ISSN: 1556-5068.
- [OR01] J. Oechssler and F. Riedel. “Evolutionary dynamics on infinite strategy spaces”. In: *Economic Theory* 17.1 (Jan. 2001), pp. 141–162. ISSN: 0938-2259.

-
- [Ree16] C. R. Reeves. “Genetic Algorithms”. In: *Encyclopedia of Database Systems* (2016), pp. 1–5. DOI: 10.1007/978-1-4899-7993-3_562-2. URL: http://dx.doi.org/10.1007/978-1-4899-7993-3_562-2.
- [Rou16] T. Roughgarden. “Twenty Lectures on Algorithmic Game Theory”. In: (2016).
- [RR14] M. Ruijgrok and T. W. Ruijgrok. “An Effective Replicator Equation for Games with a Continuous Strategy Set”. In: *Dynamic Games and Applications* 5.2 (Aug. 2014), pp. 157–179. ISSN: 2153-0793.
- [Sig10] K. Sigmund. “Introduction to Evolutionary Game Theory”. In: *of Proceedings of Symposia in Applied Mathematics, American Mathematical Society*. 2010, pp. 1–26.
- [VM47] J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1947.