A Note on Dynamic Processes

by

Andrew McMillan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2022

Doctoral Committee:

Professor Silas Alben, Co-Chair
Professor Shaowei Ke, Co-Chair
Professor Joseph Conlon
Professor Zaher Hani

Andrew McMillan

andrewnm@umich.edu

ORCID iD: 0000-0001-8032-9831

ACKNOWLEDGEMENTS

This dissertation is the culmination of years of support and inspiration from teachers, friends, and family; there are far too many to enumerate, but I would like to extend particular gratitude to a few.

I am exceptionally lucky and grateful to have had such great advisors. I would like to thank my advisor Shaowei Ke, whose patience, kindness, and encouragement I never earned but will remain eternally grateful for. I thank my late advisor Charlie Doering (Go Blue) who showed me how to approach mathematics with the joy and curiosity of a child again, and perhaps just as significantly, the importance of a good bound. I also thank my advisor Silas Alben for stepping in during a tumultuous time.

I would like to thank Joseph Conlon and Zaher Hani for taking the time to serve on my dissertation committee as well as their instruction while at the University of Michigan. I would also like to thank Yin Lu Young for her mentorship and guidance, which made navigating my final year incredibly facile.

I am immensely grateful for having such exceptional friends. To Blazin' Squad, it's early days, innit? I thank them for their friendship and their continued reminders to never take myself too seriously. I thank Noah Gomez for his support, wisdom, and our continued, shared adventures that got me through the past five years and likely will through life. I'd also like to thank Patrick Nave for his continued friendship, thought-provoking discussions, and for being someone I can always call either about interesting problems or otherwise.

Finally, I would like to thank my family. I thank my grandparents for their continued encouragement and relentless positivity. I thank my parents for their encouragement, patience, and support even when my path was not so clear; I would not be anything close to the man I am today or hope to be without their guidance. I would like to thank my brothers, Alan and Alex, for their love and willingness to always bring a smile to my face.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

ABSTRACT

This dissertation is written in two parts. In the first part, we overview and extend a novel, robust and computationally efficient method come to be known as an auxiliary function method for long-time averages - which is capable of computing sharp upper or lower bounds on time averaged quantities in underlying dynamical variables via convex optimization and semidefinite programming techniques. We then turn to studying the validity of asymptotic methods for computing long-time statistics in nonlinear or nonautonomous dynamical systems. Asymptotic methods, such as Fourier expansions, the Galerkin method, and harmonic balance methods, are ubiquitous in the literature when studying dynamical systems, but as these methods produce only approximate solutions, it is natural to ask how well these approximate solutions agree with a system's true solutions. We show for the Duffing equation and the nonlinear, damped driven pendulum that the mean squared amplitude as produced by the harmonic balance method agrees quite well with the system's true solution. However, asymptotic methods fail to accurately predict the regions of stability for a parametrically driven, coupled oscillator system. We show that the regions of stability are particularly sensitive to the coupling effects across a broad range of modulation frequencies, and hence show the auxiliary function method as a more robust means of determining stability regions.

In the second part of this work, we first overview dynamic choice in the presence of uncertainty while discussing the classical paradigms of Von Neumann-Morgenstern expected utility [3] and discounted expected utility. We then discuss the ethical theory of utilitarianism from the perspective of Jeremy Bentham [4] and discuss its connections to decision theory - in particular, social choice theory. We briefly overview the social choice literature by reviewing the seminal work of Kenneth Arrow [5] and John Harsanyi [6] and subsequent results. Then we present a novel extension of Harsanyi's theorem to an infinite time horizon, multi-generation setting. Under some additional assumptions, a Pareto condition is equivalent to utilitarian aggregation and the utilitarian weights are unique. We analyze the properties of utilitarian weights,

such as the limiting behavior of utilitarian weights for distant future generations, and the comparative statics of utilitarian weights as the social discount factor or the social risk attitude changes. Among other findings, we show that a higher social discount rate is associated with a more unequal assignment of utilitarian weights across generations.

CHAPTER I

Key Quantities in Dynamical Processes

1.1 Introduction

In this chapter, we motivate the study of dynamically varying environments from both an applied and theoretical mathematics perspective as well as an economics perspective. In particular, we present two generic problems from the mathematics and economics literature and discuss their challenges and potential directions for research.

In the mathematics literature, the study of dynamical systems governed by ordinary differential equations (ODEs) is ubiquitous for their wide range of modeling capabilities [7, 8, 9] as well as their rich mathematical structure [10, 11]. From a modeling perspective, dynamical systems are employed in modeling fluid flow through various geometries, the firing and synchronization of neurons, the interaction of species via predator-prey dynamics, etc.[12]; the applications are far too extensive to enumerate. From a theoretical perspective, the analysis of dynamical systems can be prohibitively difficult due to the presence of nonlinear or non-autonomous terms. Therefore, modern research concerns itself with developing new methods for studying the essential characteristics of the underlying dynamics. In particular, time averages of functions of dynamical variables are often of greater interest than knowing the value of the function across a collection of times. Consequently, robust and efficient, computational methods are in need of development to compute these time averaged quantities.

In the economics literature, the study of dynamic decision makers is ubiquitous for its applications to decision theoretic machine learning, dynamic portfolio optimization or the trading of financial instruments, the admittance or recurrence of psychological behavior, etc. [13, 14, 15]. Modern research focuses on characterizing the choices of a dynamic decision maker in various environments from two perspectives—normative versus descriptive analysis. In normative analysis, decision theorists concern themselves with the outcomes of decisions or determining optimal decisions given constraints and assumptions. In descriptive analysis, decision theorists attempt to understand how a decision maker came to make the chosen decision. Modern decision theoretic research is frequently concerned with the axiomatic framework for which a decision maker's preferences can be represented with an analytically tractable function. In particular, functional representations in dynamic, social choice are of interest, and therefore, a rigorous mathematical framework is needed to analyze choice in this setting.

In both the mathematical and economic literature, the study of these dynamical processes can frequently be reduced or turns out to be equivalent to studying and characterizing crucial, key quantities. These key quantities are discussed further in §1.2 and §1.3.

1.2 Key quantities in dynamical systems

The solutions to dynamical systems are often complicated, analytically intractable, or chaotic. Therefore, modern research often focuses on statistics, e.g., long-time averages of key quantities. Time averages are generally sensitive to initial conditions, so it is natural to seek the largest or smallest averages across all trajectories and their respective initial conditions, as well as the extremal trajectories that realize them. Hence, the optimization problem to solve becomes

$$\begin{aligned} &\max_{\mathbf{x}_0 \in \mathrm{B}} \overline{\Phi(\mathbf{x}_0)} \\ &\text{s.t. } \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \\ \overline{\Phi}(\mathbf{x}_0) := &\limsup_{T \to +\infty} \frac{1}{T} \int_0^T \Phi(\mathbf{x}(t)) dt, \end{aligned} \tag{1.1}$$

where $\mathbf{x}_0 = \mathbf{x}(0)$, B is a suitably chosen set such that extrema exist, $\Phi(\cdot) : \mathbb{R}^n \to \mathbb{R}$, and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable vector field, and we define $\dot{\mathbf{x}} := \frac{d}{dt}\mathbf{x}$.

In view of (1.1), a natural question is: how does one recover these extremal averages? The naive, brute-force approach is to construct a large number of candidate trajectories. However, this foregoes the potential existence of pathological trajectories due to un-realized initial conditions in the large set of candidate trajectories. This also has the drawback of being quite computationally expensive and operationally limited to trajectories that are suitably stable. A second, more tractable and widely implemented approach is to make limiting assumptions about the system's solutions via perturbative asymptotic methods [16, 17]. However, this approach may fail to characterize the true behavior of solutions, especially across all initial conditions, and asymptotic analysis may dangerously fail to predict important characteristics of a system, such as regions of stability and instability [18]. An alternative approach that is broadly applicable – and often more tractable – is to construct sharp a priori bounds on long-time averages via convex optimization.

A modern approach, which has come to be known as *the auxiliary function method for long-time averages*, permits the computation of sharp bounds of long-time averages in autonomous dynamical systems whose terms are polynomial in the underlying dynamical variables. However, there are two key concerns. The first is that for various applications a dynamical system may include nonautonmous or nonpolynomial terms [19], and therefore, a robust framework to handle these situations is needed. Secondly, given the ability to compute sharp bounds of long-time averages, it is not clear how these results may differ from those established via the standard approach of perturbative asymptotic analysis. In this dissertation, we focus on dynamical systems that model oscillators or coupled oscillator systems whose parameters govern real world phenomena, and hence, the essential characteristics of the underlying system also depend on these parameters. Moreover, as unstable dynamics can be both costly and dangerous for either experimental or real world implementation [20, 21], we are concerned with the predictions of stability via perturbative asymptotic methods and juxtapose them with the predictions of the auxiliary function method for long-time averages. We concern ourselves with asking when do these two methods agree and if there are dynamic regimes for which agreement should be expected. We also investigate when each method has differing predictions and the consequences of using one method over the other.

1.3 Key quantities in dynamic decision processes

The study of dynamic decision processes analyzes the temporally interdependent decisions that occur due to external, environmental changes or due to the previous actions of the decision maker [22]. In practice, an experimentalist or analyst only observes the decision maker's choices, and hence, a rigorous mathematical framework is needed to analyze choice and their resulting consequences or implications.

For the purposes of terminology, we fix some set X, and we refer to the elements of X as prospects—prospects are the choice objects for our decision maker, and we refer to the set X as the choice domain. Hence, it is natural to model the decision maker's choices or preference as a binary relation $\succeq$ over the set X, where $\succeq$ is a subset of $X \times X$. This relation induces an ordering on X, which for two elements $x, y \in X$, the statement $x \succeq y$ is read as "x is preferred to y". This relation will be hereafter referred to as a *primitive*, and it is the key quantity to study. However, from a mathematical point of view, a primitive is quite abstract, and hence, it may be difficult to analyze and study its essential characteristics. Therefore, modern research is concerned with what properties of a primitive allow it to be completely characterized by an analytically tractable function. The problem to solve becomes:

$$\begin{gathered} \text{Find } U(\cdot) \\ \text{s.t. } U(x) \geq U(y) \iff x \succeq y \; \forall \, x, y \in X \\ \succeq \in \mathbb{M}, \end{gathered} \tag{1.2}$$

where $U(\cdot) : X \to \mathbb{R}$ is frequently called a utility function, $\succeq$ is the primitive, and $\mathbb{M}$ is a non-empty collection of binary relations over X that obey pre-specified properties. For example, if X is a topological space, then a pre-specified property could be imposing continuity on the elements of $\mathbb{M}$. Note that in contrast to (1.2), one could also begin with the desired function $U(\cdot)$ and ask which set $\mathbb{M}$ would allow a binary relation $\succeq$ to be represented by the function $U(\cdot)$. Once the analyst has recovered the function $U(\cdot)$, the study of the decision maker and their preferences becomes equivalent to studying the properties of $U(\cdot)$.

In this dissertation, we consider a dynamic decision maker representing a governing body or social planner within the framework of social choice theory—in particular, utilitarianism. This social planner must make a collection of decisions, and we concern ourselves with how this social planner makes choices in the presence of a population consisting of decision makers that express their own choices and values under the premise that the governing body altruistically cares about the members of the population.

The seminal work of Harsanyi [6] placed axiomatic utilitarianism on a firm mathematical footing providing conditions for which a social planner has a utilitarian representation. That is, Harsanyi was able to establish necessary and sufficient conditions for which a social planner's utility function takes on the form of a weighted summation of individual utility functions. However, Harsanyi's result is limited to dynamic choice with a finite time horizon. We extend Harsanyi's result to a natural, infinite time horizon setting, and with the extension in hand, we study the asymptotic properties of the summation's weights and the interdependent relationship between weights across time.

1.4 Outline of the dissertation

The focus of this dissertation is two fold. In Chapters II and III, we focus on key, long-time statistics in dynamical systems. Chapter II outlines a recently developed technique—the auxiliary function method for long-time averages, which allows one to obtain bounds on time-averaged quantities in dynamical variables. We first review the theoretical formalism of this method, and we then discuss how, in light of this auxiliary function method, obtaining bounds on time-averaged quantities can be viewed as a semi-definite programming problem, whose solution can be obtained via sum of squares programming. In Chapter III, we begin by providing a novel extension of the technique outlined in Chapter II to obtain bounds on time-averaged quantities for dynamical systems with non-autonomous or trigonometric dependence in the dynamical variables. We then study the validity of asymptotic expansions as solution approximations in three, classical dynamical systems. We show that asymptotic methods can perform quite well in recovering long-time statistics, and we also show that asymptotic methods can be dangerously unconservative with respect to a system's stability.

In Chapters IV and V, we focus on dynamic decision processes. In Chapter IV, we first outline the mathematical formalism and historical development of axiomatic decision theory with a discussion on descriptive decision theory, the classical Von–Neumann Morgenstern Theorem, and intertemporal choice in the presence of future uncertainty. We then give a brief description of the ethical theory utilitarianism from the perspective of Bentham, and we subsequently discuss the connections between Bentham's theory and decision theory—in particular, social choice theory. We discuss the seminal work of Kenneth Arrow and outline the relevant subsequent literature. We then overview the mathematical formalism of utilitarianism due to Harsanyi, and we also overview extensions of Harsanyi's work. In Chapter V, we provide a novel extension of Harsanyi's work to an infinite time horizon, multi-generation setting, and we give two proofs for our main theorem of varying abstraction. We conclude Chapter V by studying the asymptotic and intergenerational properties of the obtained extension.

Chapter 6 concludes with a synopsis and a discussion for potential future work.

CHAPTER II

Auxiliary Functions and Sum of Squares

2.1 Introduction

In this chapter, we first overview the theoretical formalism of the auxiliary function method for long-time averages, which allows one to determine upper or lower bounds on time averaged quantities for ordinary differential equations, and moreover, these bounds are global in the sense that they hold across all initial conditions for a specified domain. This method involves the choice of an auxiliary function for which optimal bounds are achieved. The use of auxiliary functions to prove bounds on long-time averages is similar in spirit to the use of Lyapunov functions to determine global stability of trajectories in ODEs. In a similar fashion to Lyapunov functions, auxiliary functions need only be defined over the dynamical variables, but they do not require actual knowledge of trajectories. We also discuss the historical development of researching time averaged quantities in the ergodic optimization literature. This is followed by a short, demonstrative example to provide intuition for the aforementioned ideas.

Next, we overview the polynomial optimization literature concerned with optimization problems that are subject to nonnegativity constraints. We discuss the natural strengthening of these constraints from nonnegativity to being a sum of squares and the resultant computational implications. We discuss several key theorems that provide intuition and enlighten why this strengthening is natural.

Finally, we discuss the connections of polynomial optimization with that of semidefinite programming. In particular, we see that sum of squares optimization problems can be convex optimization problems realized in a semidefinite program for which efficient numerical algorithms are available to solve problems of this type. We also discuss the connections between these computational approaches and the auxiliary function method for long-time averages.

2.2 Extremal time averages for autonomous dynamical systems

In order to introduce the auxiliary function method for long-time averages, we focus on determining upper bounds for time averages of functions of the dynamical variable for autonomous ODEs; lower bounds can be determined in a completely analagous fashion. Consider $\mathbf{x}(t) \in \mathbb{R}^d$ satisfying

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{2.1}$$

for continuously differentiable vector fields $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$. When there is no confusion, we denote the vector components of $\mathbf{x}(t)$ and $\mathbf{f}(\mathbf{x})$ as $x_i(t)$ and $f_i(\mathbf{x})$, respectively.

Given a quantity of interest $\Phi(\mathbf{x})$, define its long-time average along the trajectory $\mathbf{x}(t)$ with $\mathbf{x}(0) = \mathbf{x}_0$ by

$$\overline{\Phi}(\mathbf{x}_0) := \limsup_{T\to\infty} \frac{1}{T} \int_0^T \Phi(\mathbf{x}(t))dt. \tag{2.2}$$

The choice of $\Phi(\mathbf{x})$ is subject to the particular application in mind. For example, $\Phi(\mathbf{x})$ could be the power in a dynamical system modeling a circuit, the amplitude in a dynamical system modeling wave propagation, etc..

Let $\mathrm{B} \subset \mathbb{R}^d$ be compact in the usual topology on $\mathbb{R}^d$ and an invariant region of the phase space with respect to the dynamical system in (2.1) in the sense that trajectories that begin in B, remain in B for all time. In a dissipative system, B could be an absorbing compact set, or in a conservative system, B could be defined by constraints on dynamical invariants.

We are interested in the maximal long-time average among all trajectories (eventually) remaining in B, i.e.,

$$\overline{\Phi}^* = \max_{\mathbf{x}_0 \in \mathrm{B}} \overline{\Phi}(\mathbf{x}_0). \tag{2.3}$$

The fundamental questions are: what is the value of $\overline{\Phi}^*$ and what trajectories attain it?

Upper bounds on averages can be deduced using the fact that time derivatives of bounded functions average to zero. This elementary observation follows from the fact that for every $\mathrm{V}(\mathbf{x}) \in C^1(\mathrm{B})$—the set of continuously differentiable functions on

B—we have

$$
\begin{aligned}
0 &= \limsup_{T\to+\infty} \frac{\mathrm{V}(\mathbf{x}(T)) - \mathrm{V}(\mathbf{x}(0))}{T} \\
&= \limsup_{T\to+\infty} \frac{1}{T} \int\limits_0^T \frac{d}{dt} V(\mathbf{x}(t)) dt \\
&= \overline{\frac{d}{dt}\mathrm{V}(\mathbf{x}(t))} = \overline{\mathbf{f}(\mathbf{x}(t)) \cdot \nabla \mathrm{V}(\mathbf{x}(t))}.
\end{aligned}
\tag{2.4}
$$

We hereafter refer to any such $\mathrm{V}(\mathbf{x}) \in C^1(\mathrm{B})$ as an "auxiliary" function. Note that (2.4) holds for any auxiliary function, so there is an infinite family of functions with the same time average as $\Phi(\mathbf{x})$. In particular,

$$
\overline{\Phi}(\mathbf{x}_0) = \overline{[\Phi + \mathbf{f} \cdot \nabla \mathrm{V}]}(\mathbf{x}_0) \tag{2.5}
$$

for all $\mathrm{V} \in C^1(\mathrm{B})$. For any auxiliary function, one obtains an upper-bound on $\overline{\Phi}^*$ by bounding the right hand-side point-wise on B and subsequently maximizing the left hand side over initial data $\mathbf{x}_0$

$$
\overline{\Phi}^* \leq \max_{\mathbf{x}\in\mathrm{B}} [\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x})]. \tag{2.6}
$$

The bound on $\overline{\Phi}^*$ in (2.6) is useful because it does not require knowledge of trajectories. Moreover, since the bound in (2.6) holds for all trajectories and any auxiliary function, the best such *a priori* upper bound on $\overline{\Phi}^*$ is then

$$
\overline{\Phi}^* \leq \inf_{\mathrm{V}\in C^1(\mathrm{B})} \max_{\mathbf{x}\in\mathrm{B}} [\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x})]. \tag{2.7}
$$

The minimization over the right hand side of (2.7) is a *convex* optimization in the auxiliary function V. Indeed, define the functional $\mathcal{F} : C^1(\mathrm{B}) \to \mathbb{R}$ as

$$
\mathcal{F}(\mathrm{V}) = \max_{\mathbf{x}\in\mathrm{B}} [\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x}))], \tag{2.8}
$$

and insert a convex combination of auxiliary functions V_1, V_2 and apply the triangle inequality to deduce

$$
\begin{aligned}
&\mathcal{F}(\lambda \mathrm{V}_1 + (1-\lambda)\mathrm{V}_2) = \max_{\mathbf{x}\in \mathrm{B}}[\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla(\lambda \mathrm{V}_1(\mathbf{x}) + (1-\lambda)\mathrm{V}_2(\mathbf{x}))] \\
&= \max_{\mathbf{x}\in \mathrm{B}} [\lambda\{\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}_1(\mathbf{x})\} + (1-\lambda))\{\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}_2(\mathbf{x})\}] \\
&\leq \lambda \max_{\mathbf{x}\in \mathrm{B}} [\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}_1(\mathbf{x})] + (1-\lambda) \max_{\mathbf{x}\in \mathrm{B}}[\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}_2(\mathbf{x})] \\
&= \lambda \mathcal{F}(\mathrm{V}_1) + (1-\lambda)\mathcal{F}(\mathrm{V}_2).
\end{aligned}
$$

The remarkable fact is that the inequality in (2.7) is actually an *equality*, which is surmised in the following theorem due to [23].

Theorem II.1. *(Tobasco, Goluskin, and Doering; 2018, [23]) Suppose* $\boldsymbol{x} \in \mathbb{R}^d$ *satisfies (2.1) and* B *is a compact, invariant subset of* $\mathbb{R}^d$*. Then the following equality holds:*

$$\overline{\Phi}^* = \inf_{\mathrm{V}\in C^1(\mathrm{B})} \max_{\mathbf{x}\in \mathrm{B}}[\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x})]. \tag{2.9}$$

The proof of Theorem II.1 is established via a sequence of equalities that follows a minmax template from convex analysis. If $\phi_t(\mathbf{x})$ denotes the flow map for (2.1) and $\mathcal{P}(B)$ denotes the space of all Borel probability measures over the set B, a measure $\mu \in \mathcal{P}(B)$ is said to be ϕ_t-invariant if $\mu(\phi_t^{-1}(A)) = \mu(A)$ for all Borel sets A and for all t. An invariant measure is said to be ergodic if it assigns measure 0 or 1 to all ϕ_t-invariant Borel sets. In other words, for an ergodic measure, there are no ϕ_t-invariant Borel sets up to measure zero—intuitively, this is a type of mixing condition. A classical result is that the set of invariant probability measures on B is nonempty, convex, and weak-* compact, and its extreme points are ergodic [24].

The key observations for the proof are (i) that time averages can be realized as phase space averages against invariant measures, (ii) maximizing over invariant probability measures can be realized as a Lagrange multiplier problem where $\int f \cdot \nabla \mathrm{V}\, d\mu = 0$ for all $\mathrm{V} \in C^1(\mathrm{B})$ ensures μ is invariant, (iii) swapping the order of supremum and infimum can be performed due to standard abstract min-max theorems, and (iv) the $\sup_{\mu \in \mathrm{Pr(B)}} \int (\Phi + f \cdot \nabla \mathrm{V}) d\mu$ is realized by a delta-mass located where $\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x})$ assumes its maximum.

Thus arbitrarily sharp bounds on the maximal or minimal long-time average are available via convex optimization over auxiliary functions; optimal or sequences of near-optimal auxiliary functions produce optimal or sequences of increasingly near-optimal bounds.

Moreover, if $\mathrm{V} \in C^1(\mathrm{B})$ is an optimal auxiliary function, then it's straightforward to see that the corresponding optimal trajectory or trajectories reside in the subset of B where the continuous function $\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x}) = \overline{\Phi}^*$. In a similar fashion, if V is just near-optimal, then corresponding near-optimal trajectories spend a significant fraction of time in high altitude level sets of $\Phi(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x})$. Either way the auxiliary function approach can be used to localize extremal trajectories in the phase space.

On the surface, the minimization over auxiliary functions in (2.7) seems computationally intractable, as the optimization must be performed over an infinite dimensional function space—C^1. For ease of notation, we define the function $S(\mathbf{x})$ as

$$S(\mathbf{x}) := U - \Phi(\mathbf{x}) - \nabla \mathrm{V}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}), \tag{2.10}$$

so that an upper bound on Φ is implied by the nonnegativity of S, and (2.3) can be reduced to

$$\begin{aligned} &\min_{\mathrm{V} \in C^1(\mathrm{B})} U \\ &\text{s.t. } S(\mathbf{x}) \geq 0 \, \forall \mathbf{x} \in \mathrm{B}, \end{aligned} \tag{2.11}$$

as a pointwise constraint for all $\mathbf{x} \in \mathrm{B}$ is sufficient to obtain a global average constraint.

As stated, the problem reduces to determining the non-negativity of a given multivariate function. Unfortunately determining the non-negativity of multivariate functions is computationally indiscernible, but we will formulate the problem as a semidefinite program and perform suitable and a natural strengthening to make the problem computationally accessible in §2.4 and §2.5.

Meanwhile it is important to recognize that interest in extremal time averages is not new. In the abstract dynamical systems community, it goes under the name "ergodic optimization" [25, 26]. The ergodic optimization literature originated in the 1990s, with a large volume of the early work studying the dependence of a maximizing measure on the underlying flow map and a given function of the dynamical variables; these problems were born of earlier work on problems related to physics - in particular, problems concerning Lagrangian dynamics and the limiting zero temperature formalism of thermodynamics [27]. The literature would subsequently explode, and in part, it was motivated by conjectures that many quantities of interest in applications for chaotic dynamical systems are optimized, in a time averaged sense, on relatively simple unstable periodic orbits [28]. Those conjectures, in turn, underly "control of chaos" notions [29] that emerged earlier.

Rather than developing theoretical or quantitative computational tools to evaluate extremal time averages, however, the ergodic optimization field focused on more conceptual questions resulting in theorems such as that every ergodic measure is the unique maximizing measure for some continuous function. In our setting, this is precisely the statement that for every initial condition $\mathbf{x}_0 \in \mathrm{B}$, $\overline{\Phi}(\mathbf{x}_0) = \overline{\Phi}^*$ for some continuous function Φ. The ergodic optimization community recognized the variational structure reflected in (2.9) and given complete knowledge of the flow map, proposed a strategy to produce a sequence of increasingly near-optimal auxiliary functions [30]. We next provide a simple example to give intuition for the aforementioned ideas.

2.3 Example: sequences of near-optimal auxiliary functions

Consider the following one dimensional polynomial dynamical system and accompanying quantity of interest:

$$\begin{aligned} \dot{x} = x - x^3 = f(x) \\ \Phi(x) = x^2 \end{aligned} \tag{2.12}$$

The system in (2.12) possesses three classes of solutions corresponding to three classes of initial data:

$$\begin{aligned} x(t) &\to -1 \quad \text{for} \quad -\infty < x_0 < 0, \\ x(t) &\to 0 \quad \text{for} \quad x_0 = 0, \quad \text{and} \\ x(t) &\to +1 \quad \text{for} \quad 0 > x_0 > \infty. \end{aligned} \tag{2.13}$$

Therefore $\overline{\Phi}(0) = 0$ and $\overline{\Phi}(x_0) = 1$ for all $x_0 \neq 0$ so that $\overline{\Phi}^* = 1$. However, these observations are independent of any of the auxiliary function method formalism, so how might one discern this within the auxiliary function formulation?

In this example, it is easy to construct an optimal polynomial auxiliary function—$\mathrm{V}(x) = \frac{1}{2}x^2$. To see this, we note that

$$\begin{aligned} \Phi(x) + f(x)\mathrm{V}'(x) &= x^2 + (x - x^3)x \\ &= 2x^2 - x^4 \\ &= 1 - (x+1)^2(x-1)^2. \end{aligned} \tag{2.14}$$

That is, for this optimal auxiliary function $\Phi(x) + f(x)\mathrm{V}'(x) = \overline{\Phi}^* - S(x)$, where $S(x) = (x+1)^2(x-1)^2$ is a sum of polynomials where each term is squared.

Moreover, for this particular $\Phi(x)$ and optimal auxiliary function $\mathrm{V}(x)$, the quantity $\Phi(x) + f(x)\mathrm{V}'(x)$ achieves its pointwise maximum $\overline{\Phi}^*$ only when $x = \pm 1$. The points $x = \pm 1$ are therefore both optimal initial conditions, but it is also worth noting that any initial condition such that $x_0 \neq 0$ is optimal. Also, optimal trajectories such that every neighborhood thereof hosts *every* optimal trajectory 100% of the time over the infinite time interval of averaging.

Given our quantitative analytical knowledge of the flow map for this simple example, however, by relaxing the polynomial restriction we can also conceive a sequence of near-optimal auxiliary functions $\mathrm{V}_\epsilon \in C^1(\mathbb{R})$ so that that $\lim_{\epsilon\to 0}\{\Phi(x)+f(x)\mathrm{V}_\epsilon'(x)\} = \overline{\Phi}^*$ for *every* $x \neq 0$. Indeed, for every $\epsilon > 0$ define

$$\mathrm{V}_\epsilon(x) = \frac{1}{2}\ln\left(x^2+\epsilon\right) \tag{2.15}$$

so that

$$\begin{aligned} \Phi(x) + f(x)\mathrm{V}_\epsilon'(x) &= x^2 + \frac{(x-x^3)x}{x^2+\epsilon} \\ &= \frac{(1+\epsilon)x^2}{x^2+\epsilon}. \end{aligned} \tag{2.16}$$

It is evident that V_ϵ is an increasingly near-optimal sequence of auxiliary functions in the sense that

$$\inf_{\epsilon>0}\sup_{x}\{\Phi(x) + f(x)\mathrm{V}_\epsilon'(x)\} = \overline{\Phi}^*. \tag{2.17}$$

Furthermore, $\lim\limits_{\epsilon\to 0}\{\Phi(x) + f(x)\mathrm{V}_\epsilon'(x)\} = \overline{\Phi}^*$ for every $x \neq 0$ despite the limit of the sequence V_ϵ is not C^1, but even *more* is true about this sequence of auxiliary functions. For every $x \in \mathbb{R}$, we have

$$\lim_{\epsilon\to 0}\{\Phi(x) + f(x)\mathrm{V}_\epsilon'(x)\} = \overline{\Phi}(x). \tag{2.18}$$

That is, $\Phi(x)+f(x)\mathrm{V}_\epsilon'(x)$ is a sequence of functions such that its limit at each point in the phase space yields the infinite time average of $\Phi(\cdot)$ along the trajectories passing through that point.

While this impressive feature of the sequence $\mathrm{V}_\epsilon(x)$ is apparent in this simple example, such sequences of increasingly near-optimal auxiliary functions $\mathrm{V}_\epsilon(\mathbf{x})$ also exist more generally for well behaved $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ defined by sufficiently smooth vector fields $\mathbf{f}$. If we could deduce these sequences, then we could bypass the dynamics altogether to estimate and evaluate long time averages $\overline{\Phi}(\mathbf{x})$ along all trajectories. But, alas, as of now construction of such sequences requires explicit knowledge of the flow map—complete access to all information about all trajectories [30, 26]—so this approach is essentially tautological in an operational sense. At the present time, we are limited to the variational methods described in §2.2 to effectively compute sequences of increasingly optimal auxiliary functions.

2.4 Polynomial dynamical systems and sum of squares programs

The convex optimization problem in (2.11) is over the infinite dimensional vector space C^1 and is therefore intractable in general. However, for various applications, many differential equations are purely polynomial in their arguments. That is, one is frequently interested in $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ with $f_i(\mathbf{x})$ polynomial in its argument. If $f(\mathbf{x})$ is a polynomial vector field, then (2.11) can be made tractable by also restricting the class of auxiliary functions. If $\mathrm{V}(\mathbf{x})$ is restricted to being a polynomial of degree no larger than d, the resulting optimization problem is now performed over $\mathbb{P}_{n,d}$—the finite dimensional vector space of d-degree polynomials in n variables, and the constraint in (2.11) simplifies to determining whether a multi-variate polynomial is non-negative.

Unfortunately, determining the non-negativity of a multivariate polynomial is generally NP hard, however, except for an extremely limited set of examples such as uni-variate or quadratic polynomials [31]. However a natural strengthening, which has become standard in the polynomial optimization literature [32], is to insist that S is a sum of squares (SOS) polynomial.

Definition II.2. A polynomial $\mathrm{p}(\mathbf{x}) \in \mathbb{P}_{n,d}$ is a *SOS polynomial* if there is a finite collection of polynomials $p_i(\mathbf{x}) \in \mathbb{P}_{n,d}$ such that $\mathrm{p}(\mathbf{x}) = \sum_{i=1}^{N} [p_i(\mathbf{x})]^2$.

Let $\mathbb{P}^+_{n,d}$ be the positive cone of the space $\mathbb{P}_{n,d}$ and $\Sigma_{n,d}$ be the subset of $\mathbb{P}^+_{n,d}$ with SOS representations; it's obvious that d must be even for either set to be nonempty. Restricting one's attention to optimization problems with SOS constraints is useful because deciding whether a polynomial is in $\Sigma_{n,d}$ can be deduce in polynomial time in both n and d [33, 34]. Efficient algorithms have been developed for this purpose [35] based on the theoretical work of Shor [36].

Therefore, if S is assumed to be a SOS polynomial, a tractable optimization problem is of the form:

$$\begin{aligned} &\min_{V\in\mathbb{P}_{n,d}} U \\ &\text{s.t} \quad S \in \Sigma_{n,d}. \end{aligned} \tag{2.19}$$

However as written, (2.19) is a global SOS condition insofar as it insists upon a SOS representation for all $\mathbf{x} \in \mathbb{R}^n$, but in practice, one is frequently satisfied with the local positivity of a polynomial, and this was the original formulation of the problem in (2.11).

Due to the robustness of characterizing regions in phase space with polynomials, we can naturally restrict our attention to locality constraints defined in terms of only polynomials.

Definition II.3. A set K is called *semi-algebraic* if K is defined by finitely many polynomial equalities or inequalities. A semi-algebraic set $K \subset \mathbb{R}^n$ is generally of the form

$$K := \{\mathbf{x} \in \mathbb{R}^n \,|\, g_i(\mathbf{x}) \leq 0\,, h_j(\mathbf{x}) = 0 \text{ for } i = 1, \dots, m \text{ and } j = 1, \dots, k\}, \tag{2.20}$$

where $g_i(\mathbf{x}), h_j(\mathbf{x}) \in \mathbb{P}_{n,d}$ for each $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, k\}$.

The introduction of semi-algebraic sets is useful because they can be characterized via inner products with other polynomials [37]. Put plainly, one way of viewing the localized constraint of being within K is to say that for all $\mathbf{x} \in K$

$$\begin{gathered} \sum_{j=1}^{k} h_j(\mathbf{x}) r_j(\mathbf{x}) = 0\,,\ \forall r_j(\mathbf{x}) \in \mathbb{P}_{n,d} \quad \text{and} \\ \sum_{i=1}^{m} g_i(\mathbf{x}) s_i(\mathbf{x}) \leq 0\,,\ \forall s_i(\mathbf{x}) \in \mathbb{P}^+_{n,d}. \end{gathered} \tag{2.21}$$

Therefore if one wants to determine the positivity of a polynomial $p(\mathbf{x}) \in \mathbb{P}_{n,d}$ on a semi-algebraic set K, the problem can be realized by writing

$$\begin{gathered} \text{Find } r_1, .., r_k \text{ and } s_1, .., s_m \\ \text{s.t. } p(\mathbf{x}) + \sum_{i=1}^{k} h_i(\mathbf{x}) r_i(\mathbf{x}) + \sum_{j=1}^{m} g_j(\mathbf{x}) s_j(\mathbf{x}) \geq 0 \\ s_1, .., s_m \in \Sigma_{n,d}, \end{gathered} \tag{2.22}$$

where we've replaced the constraints that $s_i(\mathbf{x}) \in \mathbb{P}^+_{n,d}$ with the stronger constraint of having a SOS representation. Augmenting the problem with SOS constrained polynomials on the set K is frequently called the *S-procedure*, where the "S" comes from the SOS constraints on the s_i polynomials [37].

In returning to (2.11), we see that the nonnegativity of S can be replaced with the SOS condition and the insistence of local positivity can be replaced with the

S-procedure, so that (2.11) can be relaxed to

$$\begin{aligned} &\min \text{ U} \\ &\text{s.t. } S + \sum_{i=1}^{k} h_i(\mathbf{x}) r_i(\mathbf{x}) + \sum_{j=1}^{m} g_j(\mathbf{x}) s_j(\mathbf{x}) \in \Sigma_{n,d} \text{ and} \\ &s_1(\mathbf{x}), \ldots, s_n(\mathbf{x}) \in \Sigma_{n,\tilde{d}}, \end{aligned} \tag{2.23}$$

where d and $\tilde{d}$ are determined by the degree of the auxiliary function V and the degree of the polynomials defining the semi-algebraic set K.

A few key remarks are required here. The polynomials $f_i(\mathbf{x})$ are exogenously given as part of the dynamical system in question but there are choices to be made for polynomials $\Phi(\mathbf{x})$ and $\mathrm{V}(\mathbf{x})$. $\Phi(\mathbf{x})$ is chosen according to the particular application in mind. However, from a computational perspective, it turns out that the resulting U can generically be quite sensitive to the choice of $\mathrm{V}(\mathbf{x})$. In particular the degree of $\mathrm{V}(\mathbf{x})$ is pertinent, and the reason is two fold. Firstly, if the degree of $\mathrm{V}(\mathbf{x})$ is too small then the SOS constraint may fail to be feasible even within a reasonable tolerance for numerical error. Secondly, the resulting U may fail to be a sharp upper bound for $\overline{\Phi}$. The restriction that $\mathrm{V}(\mathbf{x})$ is polynomial is completely absent in (2.11), and moreover, the set $\mathbb{P}^+_{n,d}$ is strictly larger than $\Sigma_{n,d}$ except when the the polynomials in question are univariate, quadratic, or bivariate and quartic [38], but in general being SOS is not equivalent to non-negativity.

Therefore, it may seem unreasonable to expect sharp bounds can be achieved by restricting to a subset of the space of polynomials. However, there are two wonderful results by [39] and [40].

Theorem II.4. *(Lassere; 2007, [39]) Let $p \in \mathbb{P}^+_{n,d}$ with some global minimum p^*. Then for every $\epsilon > 0$ there exists a $N(p, \epsilon)$ such that*

$$p_\epsilon := p + \epsilon \sum_{k=0}^{N(p,\epsilon)} \sum_{j=1}^{n} \frac{x^{2j}}{k!} \in \Sigma_{n,d} \tag{2.24}$$

Hence, $\lim_{\epsilon \to 0} ||p - p_\epsilon||_{\ell^1} = 0$.

Theorem II.4 states that SOS polynomials are dense in the set of non-negative, real polynomials of arbitrary degree and of arbitrary dimension in the ℓ^1 norm of the polynomial's coefficients. Additionally, we also have the following result:

Theorem II.5. *(Lakshmi, Fantuzzi, Fernandez-Caballero, Hwang, and Chernyshenko; 2020, [40]) Suppose K is a compact, semi-algebraic set defined in terms of $\{g_i\}_{i=1}^m$.*

Let $s = \max_i deg(g_i)$, $r = deg(U - \Phi - \boldsymbol{f} \cdot \nabla V)$ *and* Γ_d *denote the set of polynomials that are a weighted sum of the* g_i*'s, where the weights are SOS polynomials of degree no more than* $r - s$*. If there exists* L *such that* $L - ||\boldsymbol{x}||^2 \in \Gamma_d$ *for some* d*, then*

$$\overline{\Phi}^* = \lim_{d \to \infty} \inf_{\substack{U \in \mathbb{R} \\ V \in \mathbb{P}_{n,d}}} \{U | \, U - \Phi - \boldsymbol{f} \cdot \nabla V \in \Gamma_d\}.$$

Therefore, by taking the polynomial degree of our auxiliary function to infinity and provided that the S-procedure is enforcing non-negativity only on a compact set, we are guaranteed to achieve the desired sharp bounds from the previously discussed theoretical formalism. In theory, we have lost nothing in restricting V to being polynomial.

In practice, one incrementally increases the allowed degree of V and the bounds are declared sharp if increasing the degree only yields small (near numerical precision) improvements in the bounds U. Often sharp bounds may be achieved for auxiliary functions of relatively small degree—say, around degree 8 or 10—that are computationally accessible on a standard laptop for systems with relatively low degrees of freedom.

There is quite a rich history in determining whether a polynomial, or more generally a rational function, can be written as a SOS or a sum of rational functions with square numerators and denominators dating back to Hilbert's 17th problem; see [41] for a historical review.

For computational applications, SOS-constrained optimization problems of the form (2.23) are frequently reformulated as a semidefinite program (SDP), which is a type of conic optimization problem. The key to this reformulation is the following theorem:

Theorem II.6. *Given a multi-variate polynomial* $p(\boldsymbol{x})$ *in* n *variables and of degree* $2d$*,* $p(\boldsymbol{x})$ *is representable as a sum of squares if and only if there exists a positive semi-definite and symmetric matrix* Q *such that*

$$p(\boldsymbol{x}) = z(\boldsymbol{x})^T Q z(\boldsymbol{x}),$$

where $z(\boldsymbol{x}) = [1, x_1, x_2, .., x_n, x_1x_2, .., x_n^d]$.

Therefore, determining whether an even degree, non-negative polynomial is a SOS is equivalent to finding a positive semi-definite and symmetric matrix, Q, such that

$$p(\mathbf{x}) = z(\mathbf{x})^T Q z(\mathbf{x}) \geq 0, \tag{2.25}$$

where $z(\mathbf{x})$ is a vector of suitably chosen polynomial basis functions. However, we remark that this representation is not unique as it depends on one's choice of basis. We compute an example for demonstration purposes: suppose we wish to represent $f(x,y) = 2x^4 + 5y^4 + x^2y^2$, a SOS polynomial, in the form of (2.25). Write

$$\begin{aligned} f(x,y) = 2x^4 + 5y^4 + x^2y^2 &= [x^2\, y^2\, xy]^T \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} [x^2\, y^2\, xy] \\ &= q_{11}x^4 + q_{22}y^4 + (q_{33} + 2q_{12})x^2y^2 + 2q_{13}x^3y + 2q_{23}xy^3. \end{aligned} \tag{2.26}$$

Equating coefficients we find

$$q_{11} = 2\,, q_{22} = 5,\; q_{33} + 2q_{12} = 1,\; q_{23} = 0\,,\; q_{13} = 0 \tag{2.27}$$

so that the matrix is positive semi-definite for $-\sqrt{10} \leq q_{12} \leq \frac{1}{2}$ with $q_{33} = 1 - 2q_{12}$.

Theorem II.6 allows one to reduce the task of determining whether a polynomial is SOS to a convex optimization problem subject to a matrix constraint for which there are efficient algorithms as already discussed.

2.5 Convex optimization via semidefinite programming

As discussed previously, computing upper and lower bounds on the quantity of interest, $\overline{\Phi}$, can be simplified to a convex optimization problem over a finite dimensional vector space of polynomials in a SDP under a natural strengthening. In general, a SDP takes $C, A_i \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^m$ for $i \in \{1, 2, .., m\}$ as inputs with the goal of determining

$$\begin{aligned} &\min \langle C, Q\rangle \\ \text{s.t. } &\langle A_i, Q\rangle = b_i \\ &Q \succeq 0, \end{aligned} \tag{2.28}$$

where for two matrices $B, D \in \mathbb{R}^{n\times n}$, $\langle B, D\rangle = \sum_{i,j} B_{i,j}D_{i,j}$ and $Q \succeq 0$ means Q is positive semi-definite.

The problem in (2.28) has a dual problem of the form

$$\begin{aligned} &\max_{y\in\mathbb{R}^n} b^T y \\ \text{s.t. } &\sum_{i=1}^{m} y_i A_i \succeq C \end{aligned} \tag{2.29}$$

Both (2.28) and (2.29) have regions of feasibility that are given by the set of all decision variables satisfying the constraints. In practice, the regions of feasibility are frequently nonempty and thus the SDP has the *strong duality property*—(2.28)— and (2.29) have the same optimal solution [34]. There are a number of primal-dual algorithms taking advantage of this property by solving both primal and dual problems simultaneously. Convergence to the optimal solution can be obtained by solving the problem iteratively, and the algorithm terminates when the "duality gap" between the two solutions falls within some pre-defined numerical tolerance. Sum-of-squares optimization problems can be expressed in the form (2.28) with the MATLAB package YALMIP [42]. See [43] for a review of SDPs and their applications.

CHAPTER III

Extremal Averages in Non-autonomous Dynamical Systems

3.1 Introduction

In this chapter, we study extensions of the auxiliary function method for long-time averages and its utility for studying the validity of asymptotic methods.

We first study non-autonomous and non-polynomial dynamical systems with trigonometric dependence and show how they can be reformulated as equivalent autonomous and polynomial dynamical systems. These systems are subject to algebraic constraints on the transformed dynamical variables, but we explain how these algebraic constraints can be efficiently, computationally realized in the SDPs previously explicated in §2.5.

We then turn to three classic dynamical systems—the Duffing equation, the damped, driven pendulum, and a system of two coupled, parametric oscillators—and show how the extended auxiliary function method for long-time averages can shed light on the study of long-term statistics. In particular, we concern ourselves with the following question: For a dynamical system subject to varying system parameters, how valid are the predictions of finitely truncated, asymptotic expansions for regions of stability and long time amplitudes? This question is natural and of relevance across many applied sciences as the reliance on invalid approximation methods may have costly and dangerous consequences for experimental or real-world implementation.

We remark that the auxiliary function method for long-time averages has already appeared in the literature. In particular, the method has proven to be useful in bounding heat transport for truncated models of Rayleigh-Bérnard convection [44], bounding the mean squared amplitude over the Van der Pol limit cycle and bounding stationary ensembles in stochastic dynamical systems [45], and finding unstable periodic orbits for the purposes of dynamic control [46]. We also note for the interested reader that there are extensions of this auxiliary function method to partial differential equations applications; see the work of [47, 48, 49] for explication. All of the aforementioned literature is part of an expanding community concerned with computational, polynomial optimization [50, 51, 52, 53]. However, to the author's knowledge, this work is the first time the auxiliary function method for long-time averages has been employed in studying nonlinear, driven oscillators, coupled, parametric oscillators, and used to investigate the legitimacy of asymptotic stability analysis; this is our primary, novel contribution.

3.2 Extensions to non-autonomous & non-polynomial dynamics

The theoretical formalism and computational implementation via SOS programming described in §2.2 depend very much on, respectively, the autonomous nature of the dynamics and the polynomial nature of the equations of motion.

However, models in many applications involve non-autonomous, i.e., "driven" systems, and non-polynomial vector fields. Therefore, it is useful to consider how broader classes of ODEs might be recast as autonomous polynomial systems. Periodically forced dynamics of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \cos(\omega t), \sin(\omega t)) \tag{3.1}$$

with $\mathbf{x} = (x_1, x_2, ..., x_d)$ are particularly interesting and ubiquitous. The traditional way of making such systems autonomous is to introduce a new coordinate $x_{d+1} = t$ and extend the system dimension from d to $d + 1$. However, this method has the drawback of introducing an unbounded dependent variable while retaining non-polynomial dependence on it.

For our purposes, these problems can be circumvented by introducing *two* new dynamical variables satisfying the polynomial sub-system

$$\begin{aligned} \dot{x}_{d+1} &= -\omega x_{d+2} \\ \dot{x}_{d+2} &= \omega x_{d+1} \\ \text{Subject to:} \quad x_{d+1}^2 + x_{d+2}^2 &= 1. \end{aligned} \tag{3.2}$$

One should note that equations of the form in (3.2) are frequently referred to as differential-algebraic systems—that is, a differential equation subject to an algebraic constraint. In order to write (3.2) as a standard dynamical system, the form seen in (2.1), one needs to omit the origin as a potential solution either via assumption, the S-procedure or other techniques.

This approach can also be used to formulate equivalent autonomous polynomial dynamics for both quasiperiodic and substantially more complex $\frac{2\pi}{\omega}$-periodic time dependences in the vector field. Employing a new pair of dynamical variables like those in (3.2) for each independent frequency allows for quasiperiodic time dependence, at least for quasiperiodicity involving only a *finite* number of independent frequencies. Other $\frac{2\pi}{\omega}$-periodic time functions can be expressed as finite linear combinations of $\cos(n\omega t)$ and $\sin(n\omega t)$, each of which in turn is a finite polynomial combination of $\cos(\omega t)$ and $\sin(\omega t)$. The overall order of the dynamical system necessarily increases but autonomous polynomial dynamics are still sufficient to capture the systems' dynamics.

A broad class of autonomous vector fields with trigonometric variable dependence can be handled similarly [54]. Consider, for example, vector fields $\mathbf{f}(\mathbf{x})$ where the components $f_1, \ldots, f_d$ depend polynomially on x_j for $j \neq i$ and on x_i via $\cos x_i$ or $\sin x_i$ but *not* on x_i itself, i.e.,

$$f_j = f_j(x_1, \ldots, x_{i-1}, \cos x_i, \sin x_i, x_{i+1}, \ldots, x_d) \text{ for each } j = 1, \ldots, d. \tag{3.3}$$

For ease of notation, let us denote the "angular" variable $x_i(t) = \theta(t)$ and the corresponding component of the vector field

$$f_i = \Omega(x_1, \ldots, \ldots, x_{i-1}, \cos\theta, \sin\theta, x_{i-1}, \ldots, x_d). \tag{3.4}$$

Then augment the system with two new variables evolving according to

$$\begin{aligned} \dot{x}_{d+1} &= -\Omega\, x_{d+2} \\ \dot{x}_{d+2} &= \Omega\, x_{d+1} \\ \text{Subject to:} \quad x_{d+1}^2 &+ x_{d+1}^2 = 1. \end{aligned} \tag{3.5}$$

Just as in (3.2), the differential-algebraic system (3.5) may be written as a standard dynamic system with various techniques omitting the origin or via assumption. The claim now is that solutions of the original d-dimensional system

$$\dot{x}_k = f_k(x_1, \ldots, x_{i-1}, \cos x_i, \sin x_i, x_{i+1}, \ldots, x_d) \text{ for } k = 1, \ldots, d \tag{3.6}$$

are in 1-to-1 correspondence with solutions of the $(d+1)$-dimensional system consisting of (3.5) and the remaining $d-1$ differential equations

$$\begin{gathered}\dot{x}_j = f_j(x_1, \ldots, x_{i-1}, x_{d+1}, x_{d+2}, x_{i+1}, \ldots, x_d) \\ \text{for } j = 1, \ldots, i-1, i+1, \ldots, d.\end{gathered} \tag{3.7}$$

In the following subsections, we illustrate these approaches and their robustness by studying the periodically forced Duffing equation, the damped, driven pendulum, and a system of coupled, parametric oscillators. We concern ourselves with the performance of asymptotic methods and their consequences for predicting dynamic stability.

3.3 The periodically driven Duffing equation

The periodically driven Duffing system is the non-autonomous second order, nonlinear ODE

$$\ddot{x} + \delta\dot{x} + \alpha x + \beta x^3 = F\cos(\omega t), \tag{3.8}$$

where δ is the damping term, α is the linear stiffness term, β is the strength of the nonlinear restoring force, F is the driving force's amplitude, and ω is the driving force's frequency.

The Duffing equation has received widespread attention in the engineering and applied physics literature for its various engineering applications as well as attention from the mathematical literature as a prototypical nonlinear model to investigate for the development of further analytic and theoretical methods [55].

The equation describes the motion of a damped harmonic oscillator with a more complicated potential than in comparison to the simple harmonic oscillator, and in particular, the Duffing equation has been used to model the harmonically excited pendulum, nonlinear isolators used to isolate vibrating sources from their surroundings, and large deflections of beams with nonlinear stiffness [55]. It is a simple paradigmatic model that without the sinusoidal forcing is asymptotically stable at zero—which can be shown in an elementary way by considering the Lyapunov functional:

$$V(x) = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\alpha x^2 + \frac{1}{4}\beta x^4, \tag{3.9}$$

However, by including the sinusoidal forcing, the Duffing equation exhibits dynamical hysteresis for variations in the frequency as well as chaotic behavior for a collection of chosen parameters [56, 57, 58, 59]. As the Duffing equation is highly non-linear with a cubic term, various solution approximation methods have appeared in the literature, such as the harmonic balance method, the Galerkin method, the Lindstedt–Poincare method, and homotopy methods [55]. As the sinusoidal forcing encourages periodicity or quasi-periodicity of solutions, the true solution should admit a Fourier expansion. From this observation, one can use the harmonic balance method, which produces $\frac{2\pi}{\omega}$-periodic approximate solutions via the ansatz

$$x(t) = A\cos(\omega t) + B\sin(\omega t), \tag{3.10}$$

as the true solution's frequency should intuitively match that of the driving frequency. Upon plugging the (3.10) ansatz into (3.8) and projecting onto $\cos(\omega t)$ and $\sin(\omega t)$, the harmonic balance method yields an implicit prediction for the frequency response curve in the form

$$\left[\left(\omega^2 - \alpha - \frac{3}{4}\beta R^2\right)^2 + (\delta\omega)^2\right] R^2 - F^2 = 0, \tag{3.11}$$

where $R = \sqrt{A^2 + B^2}$. For fixed parameters α, β, F, and δ, one can solve for the roots of (3.11) to deduce the oscillation amplitude R; the computations for Figure (3.1) were performed using MATLAB's fimplicit function.

When $\alpha > 0$ and $\beta > 0$ or $\beta < 0$ we say that the nonlinearly perturbed oscillator has been "stiffened" or "softened" and the frequency response curve tilts to the right or to the left, respectively; see Figure (3.1). In Figure (3.1), one can see the hysteresis phenomena. That is, if one traverses the curve from left to right, the branch of solutions that one comes across is different then traversing from right to left.

A natural question is to ask how well the harmonic balance method approximates true solutions of (3.8). In particular, we can compare its predictions with independent approaches to recover the frequency response curves like those in Figure (3.1). In the following, we employ the auxiliary function method for bounding long-time averages implemented in a SOS program and juxtapose the predictions of this auxiliary function method with the predictions of the harmonic balance method.

The Duffing equation (3.8) is not of the form (2.1), so we proceed by augmenting it with two additional variables to make the system autonomous. It is then realized

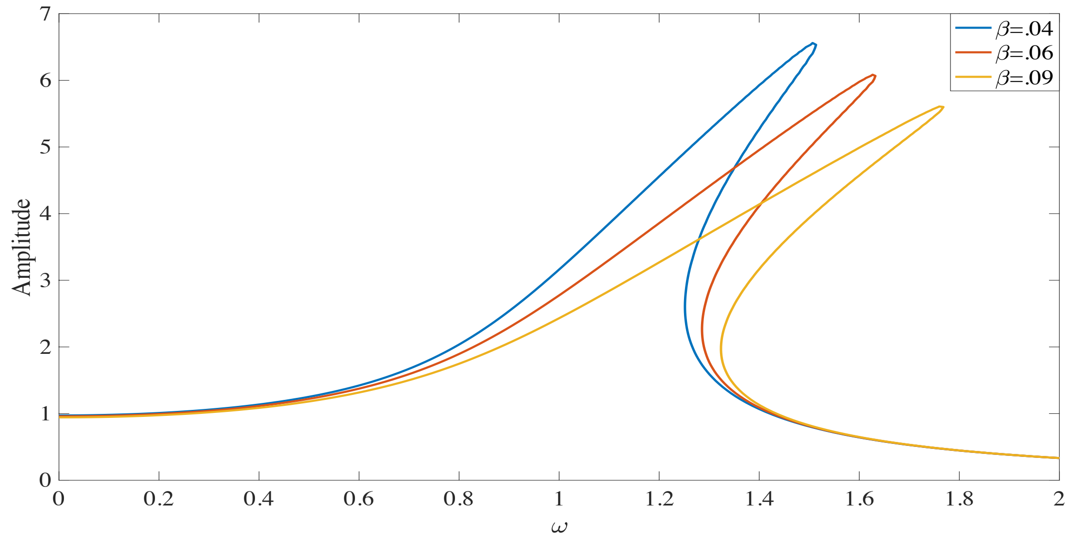

Figure 3.1: Harmonic balance approximate mean amplitude $R = \sqrt{A^2 + B^2}$ vs. driving frequency ω with $\delta = .1$, $\alpha = 1$, $F = 1$, and $\beta = .04, .06, .09$.

as the 4-dimensional first order system with one algebraic constraint

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= z_2 - \delta y - \alpha x - \beta x^3 \\ \dot{z_1} &= \omega z_2 \\ \dot{z_2} &= -\omega z_1 \\ \text{Subject to:} \quad z_1^2 + z_2^2 &= F^2, \end{aligned} \tag{3.12}$$

where the amplitudes of z_1 and z_2 will be enforced by the S-procedure so that $z_1 = F\sin(\omega t + \phi)$ and $z_2 = F\cos(\omega t + \phi)$ with phase ϕ determined by initial conditions but which is irrelevant for long-time averages.

When the function to be maximized is

$$\Phi(x, y, z_1, z_2) = x^2, \tag{3.13}$$

the relevant SOS program is

$$\begin{gathered} \min \mathrm{U} \\ \text{s.t. } \mathrm{U} - x^2 - \mathbf{f}(x, y, z_1, z_2) \cdot \nabla \mathrm{V} + \ldots \\ \cdots + \mathrm{S}(x, y, z_1, z_2)(F^2 - {z_1}^2 - {z_2}^2) \in \Sigma_{4,d} \\ \mathrm{S}(x, y, z_1, z_2) \in \Sigma_{4,\tilde{d}}, \end{gathered} \tag{3.14}$$

where d and $\tilde{d}$ are hyper-parameters determined by the degree of the auxiliary function as well as the S-procedure polynomial S.

We now systematically increase the polynomial degrees of both V and S until sharp bounds are achieved; lower bounds on $\overline{x^2}$ can be computed by negating Φ, performing the SOS program, and taking the absolute value of the resulting U.

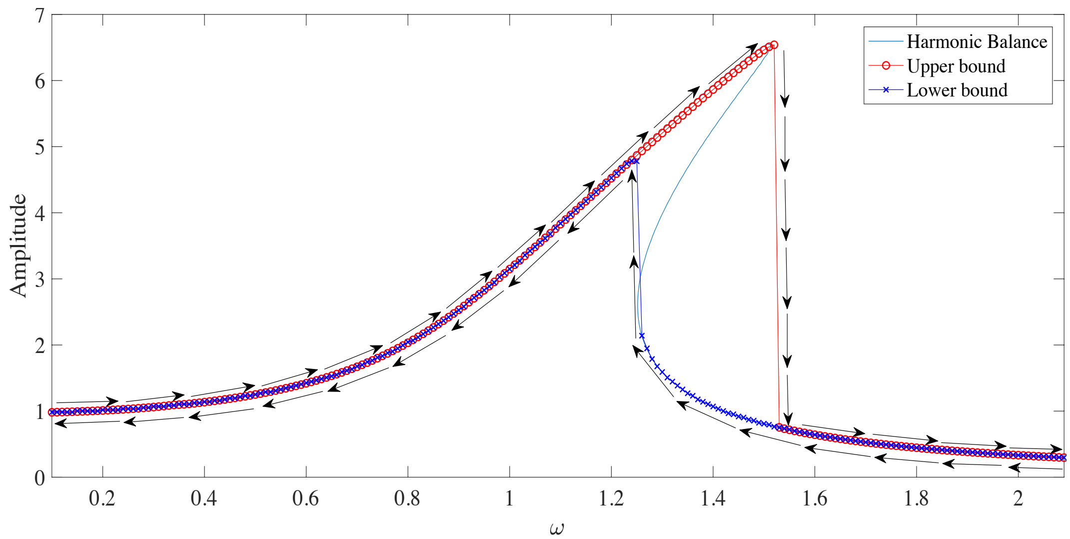

Figure 3.2: (McMillan and Doering; 2021, [1]) Harmonic balance approximate mean amplitude, $\sqrt{2\overline{x^2}}$, and upper and lower bounds on the solutions' mean amplitude vs. driving frequency ω with $\delta = .1$, $\alpha = 1$, $\beta = .04$, and $F = 1$ for a degree 10 polynomial auxiliary function.

The harmonic balance approximation of the mean amplitude agrees remarkably well with the upper and lower bounds on the true solution's mean amplitude; see Figure (3.2).

The differences between the upper and lower bounds plotted in Figure (3.3) suggest that they agree (to computational precision) for points on the frequency response curve that are single valued when the degree of the auxiliary function is sufficiently high. Not unexpectedly, there is an order 1 difference between the bounds when the curve is multi-valued. We conclude that for this sort of small amplitude forcing and weak nonlinearity, the harmonic balance approximation does exceptionally well quantitatively approximating the true solution's mean amplitude—even though the forcing and nonlinearity are strong enough to induce multi-stability and hysteresis.

It is worthwhile remarking that if we consider the degree of the auxiliary functions as a parameter then there seem to be ω-dependent thresholds in the parameter space for which the degree 6 bounds become sharp. In this example a transition occurs at

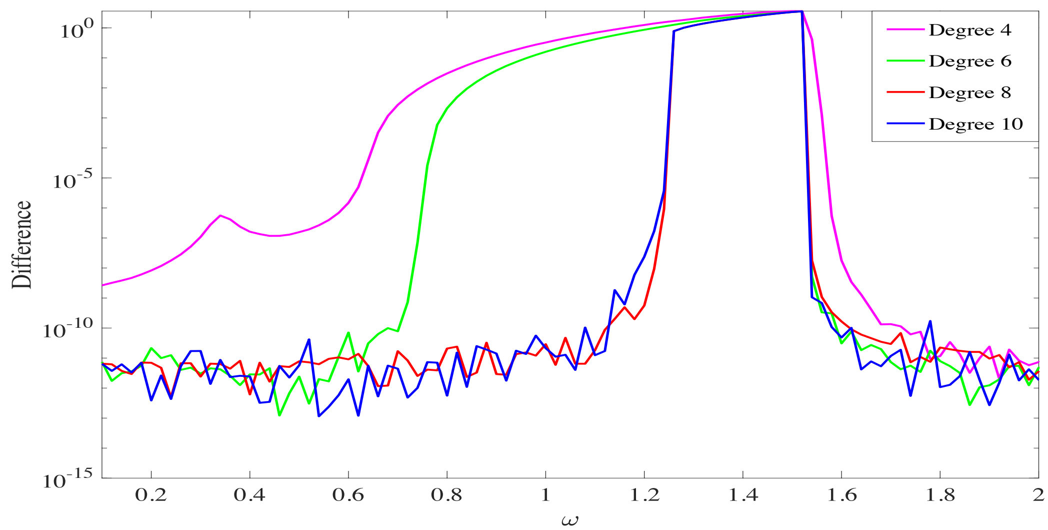

Figure 3.3: (McMillan and Doering; 2021, [1]) Difference between the upper and lower bounds on the solution's mean amplitude vs. driving frequency for degree 4, 6, 8 and 10 polynomial auxiliary functions.

$\omega \approx .7$ which, to our knowledge, is no particularly special frequency value. Under what conditions we might expect such transitions to occur, especially with a smoothly varying parameter such as ω, is an question for research in its own right.

The asymptotic accuracy of the harmonic balance approximation is remarkable but perhaps a bit unsurprising due to the system being asymptotically stable at zero without the sinusoidal forcing term. We next study the full pendulum system in a similar fashion.

3.4 The damped, periodically driven pendulum

Consider the damped and periodically driven pendulum dynamics defined by the non-polynomial and non-autonomous 2^{nd} order ODE

$$\ddot{\theta} + \gamma\dot{\theta} + \sin(\theta) = F\cos(\omega t), \tag{3.15}$$

where γ is the damping strength, F is the driving amplitude, and ω is the driving frequency. The equation in (3.15) models the swing of an object at a fixed pivot, where θ represents the angle of the object with respect to that pivot. The equation is made non-dimensional by choosing appropriate length and time scales such that θ is a function of dimensionless time.

For weak forcing, the sinusoidal non-linearity may be modeled by expanding the

$\sin(\theta)$ term in a Taylor series, and we employ a procedure similar to that of the Duffing example here to test the validity of the harmonic balance approximation.

We expect the two term expansion of the $\sin(\theta)$ term, that results in a Duffing equation, to perform poorly, however, for moderately large forcing amplitude. Hence, we expand the $\sin(\theta)$ term in a Taylor Series to 7th order and insert the (3.10) ansatz to obtain an approximate frequency response curve. The result is

$$\begin{aligned} &\frac{R^2(R^6+1152R^2-48R^4-9216)^2}{84934656}+R^2\omega^4 \\ &+\frac{R^2(R^6+4608(\gamma^2-2)+1152R^2-48R^4)\omega^2}{4608}-F^2=0, \end{aligned} \tag{3.16}$$

where $R=\sqrt{A^2+B^2}$.

The calculation is tedious and purely algebraic, but plots of R vs. ω for several forcing amplitudes are shown in Figure (3.4).

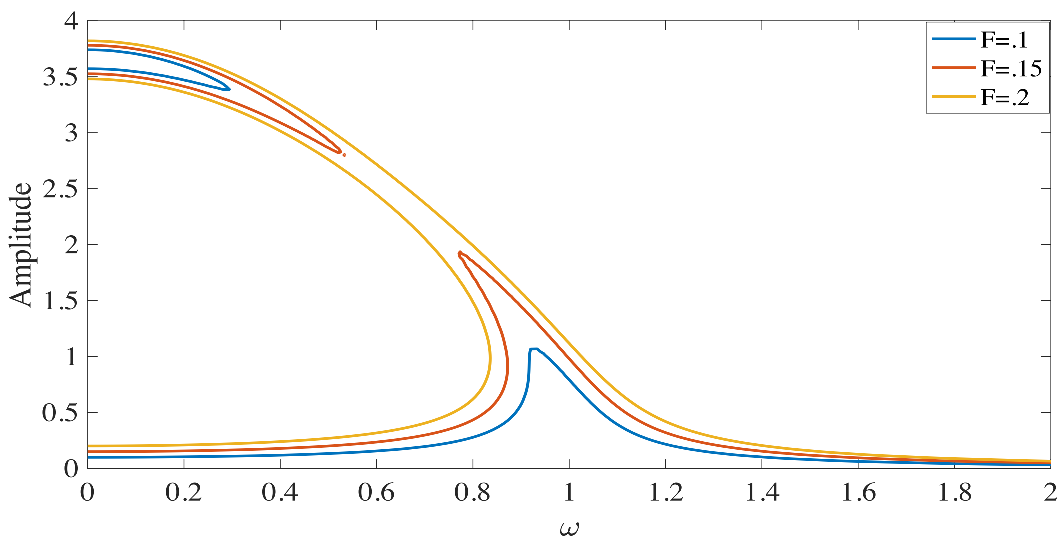

Figure 3.4: Plot of the harmonic balance approximate mean amplitude vs. ω with $\gamma = .1$ and $F = 0.10, 0.15$ and 0.20.

In this example, we seek to compare the mean mechanical energy $E=\frac{1}{2}(\dot{\theta})^2-\cos(\theta)$ from the harmonic balance approximation with auxiliary function bounds on solutions to (3.15). But as written, (3.15) is neither polynomial nor autonomous which prevents immediate implementation of the polynomial optimization via a SOS program.

Augmenting the system with four additional variables, however, we may re-write (3.15) as the 5-dimensional first order polynomial system with two algebraic constraints

$$\begin{aligned}
\dot{\phi} &= z_1 - \gamma\phi - \psi_1 \\
\dot{\psi}_1 &= \phi\psi_2 \\
\dot{\psi}_2 &= -\phi\psi_1 \\
\dot{z}_1 &= \omega z_2 \\
\dot{z}_2 &= -\omega z_1 \\
\text{Subject to:} \quad \psi_1^2 + \psi_2^2 &= 1 \\
z_1^2 + z_2^2 &= F^2.
\end{aligned} \tag{3.17}$$

The quantity of interest to extremize is the total energy plus z_1^2 given by

$$E + z_1^2 = \frac{1}{2}(\dot{\theta})^2 - \cos(\theta) + z_1^2 = \frac{1}{2}(\phi)^2 - \psi_2 + (z_1)^2 = \Phi. \tag{3.18}$$

The z_1^2 improves the computational conditioning of the SOS program. Admittedly, the authors do not know why this should be true. However, as $\overline{z_1^2} = \frac{1}{2}$, we can interpret the upper and lower bounds on the mean energy as a $\frac{1}{2}$ shift down and up, respectively.

Letting $\mathbf{x} = [\phi, \psi_1, \psi_2, z_1, z_2]$, the SOS program for upper bounds becomes

$$\begin{gathered}
\min \mathrm{U} \\
\text{s.t. } \mathrm{U} - \Phi(\mathbf{x}) - \mathbf{f}(\mathbf{x}) \cdot \nabla \mathrm{V}(\mathbf{x}) + C_1(\mathbf{x}) + C_2(\mathbf{x}) \in \Sigma_{5,d} \\
S_1, S_2 \in \Sigma_{5,\tilde{d}},
\end{gathered} \tag{3.19}$$

where $C_1(\mathbf{x}) = S_1(\mathbf{x})(F^2 - {z_1}^2 - {z_2}^2)$ and $C_2(\mathbf{x}) = S_2(\mathbf{x})(1^2 - {\psi_1}^2 - {\psi_2}^2)$, and both d and $\tilde{d}$ are determined by the degree of the S-procedure polynomials $S_1(\mathbf{x})$ and $S_2(\mathbf{x})$ as well as the auxiliary function $\mathrm{V}(\mathbf{x})$. Lower bounds on $\overline{\Phi}$ are computed just as in the Duffing setting.

Performing the SOS program in (3.19), we find that this auxiliary function method's lower bound on the mean energy and the harmonic balance approximation to the mean energy can agree quite nicely—for sufficiently weak forcing; see Figure 3.5. As the forcing amplitude increases, however, the harmonic balance approximation increasingly fails. On the other hand the upper bound in Figure 3.5 clearly does not correspond to the harmonic balance approximations we found. Indeed, the upper bound with $\overline{\Phi}^* \approx 1.5$ in Figure 3.5 suggests that there is a solution that spends most of the time oscillating weakly around $\theta = \pi$ as illustrated in Figure 3.6. Due to

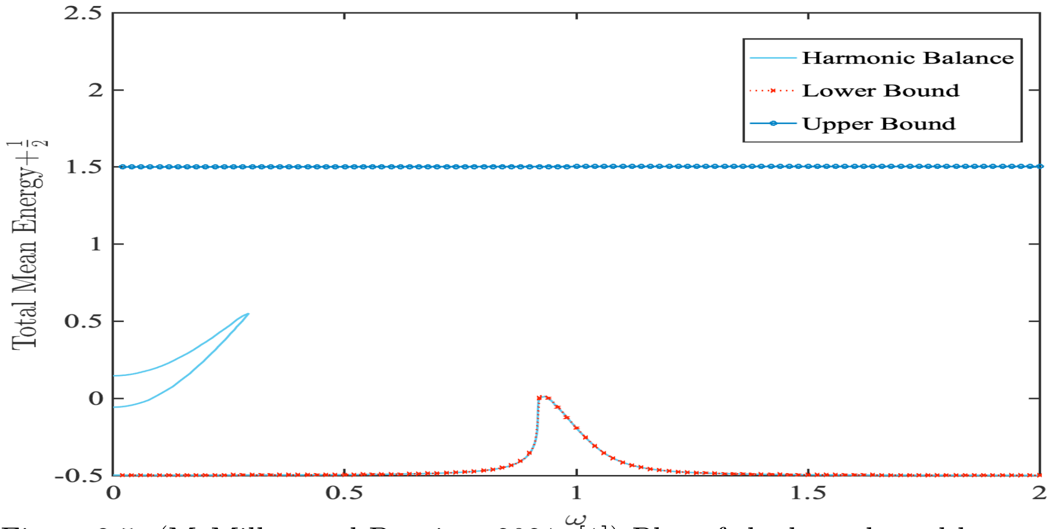

Figure 3.5: (McMillan and Doering; 2021, [1]) Plot of the bounds and harmonic balance approximate total mean energy vs. driving frequency, ω, with $\gamma = .1$, $F = 0.10$ and degree 6 polynomial auxiliary functions.

its dynamical instability, however, one would never expect to discover it via direct numerical simulation.

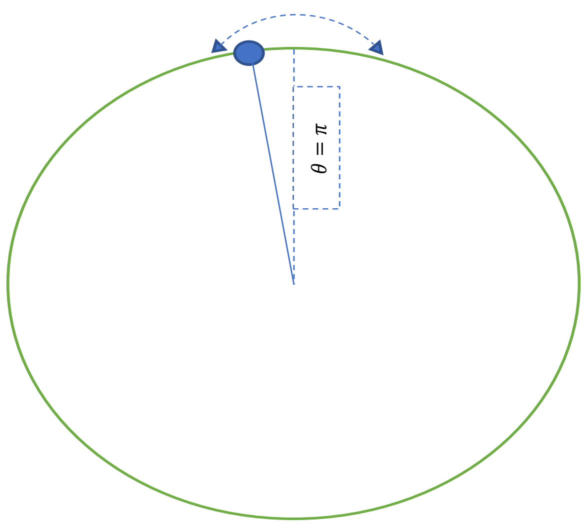

Figure 3.6: A potentially unstable solution oscillating about a neighborhood of $\theta = \pi$.

With this interpretation in mind, we can make the linear change of variables such that $\theta' = \pi - \theta$. Then when θ has low potential energy θ' has high potential energy and vice versa. Figure 3.7 is the harmonic balance approximation with the Taylor expansion performed about $\theta = \pi$—the analog of Figure 3.4.

Meanwhile Figure 3.8 shows that the harmonic balance approximation of the high

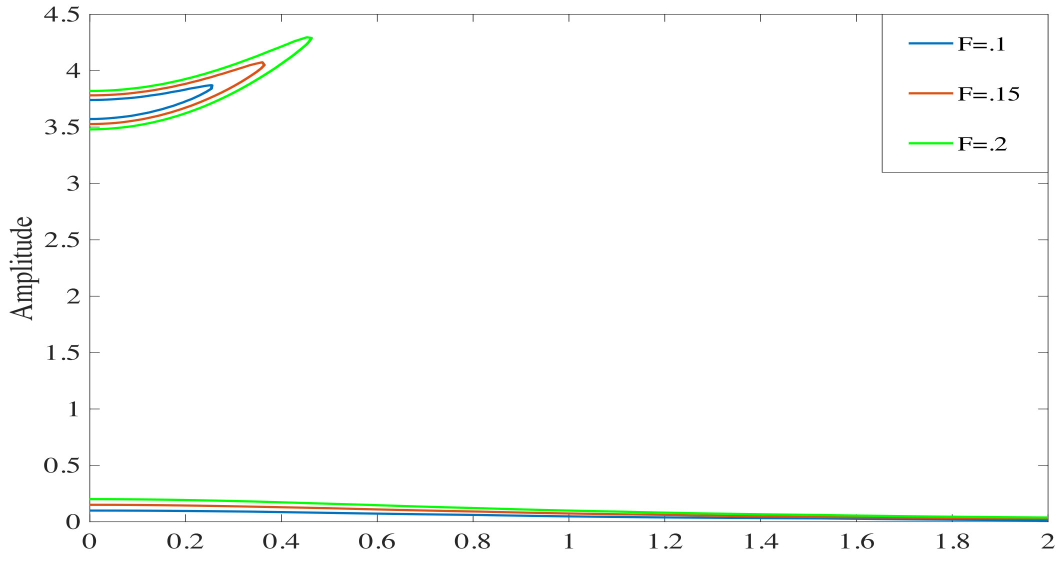

Figure 3.7: (McMillan and Doering; 2021, [1]) Harmonic balance approximate mean amplitude about $\theta = \pi$ vs. driving frequency ω with $\gamma = .1$ and $F = 0.10$, 0.15 and 0.20.

potential solution's total mean energy agrees quite well with the auxiliary function upper bound on the true solution's total mean energy.

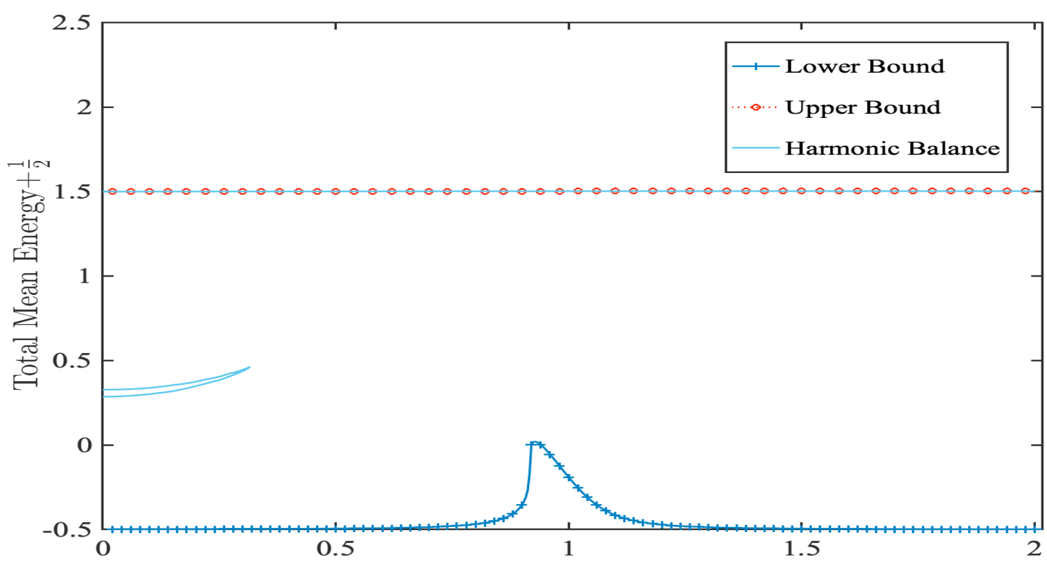

Figure 3.8: (McMillan and Doering; 2021, [1]) Plot of the bounds and harmonic balance approximation (about $\theta = \pi$) of $\overline{\Phi}$ vs. driving frequency ω with $\gamma = .1$, $F = 0.20$ and degree 6 polynomial auxiliary functions.

This example illustrates one of the operational "quirks" of this auxiliary function method: it produces upper bounds or lower bounds on the chosen Φ across all potential initial conditions including those that breed not readily observed unstable solutions.

Of course the knowledge of the existence of such unstable solutions is frequently a concern—it is certainly the central concern for control-of-chaos applications—but if one is interested in estimates of long-time averages of Φ on particular solutions (or branches of solutions) there is currently no supplementary procedure that one can employ to ensure that the bounds computed correspond with specific trajectories unless one is only interested in a particular, compact region of phase space, for which the S-procedure is readily available.

3.5 A coupled, parametric oscillator system

The next general model of interest is a parametrically driven, coupled oscillator system of the form

$$\begin{aligned} \ddot{x}_1 + {\omega_0}^2[1 + h\sin(\gamma t + \frac{\phi}{2})]x_1 + \omega_0 g\dot{x}_1 - \omega_0 r\dot{x}_2 = 0 \\ \ddot{x}_2 + {\omega_0}^2[1 + h\sin(\gamma t - \frac{\phi}{2})]x_2 + \omega_0 g\dot{x}_2 + \omega_0 r\dot{x}_1 = 0. \end{aligned} \tag{3.20}$$

In (3.20), ω_0 denotes the proper frequency of the oscillators and g is the intrinsic loss term, which is taken to be equal for both oscillators for the sake of simplicity. The h and γ terms are the intensity and frequency of the parametric stiffness terms, respectively. We focus on $h \in [0, 1]$ and $\gamma \in [-3, 3]$, as this is the regime which encapsulates most physically relevant phenomena. The coupling strength r describes an energy preserving coupling between the oscillators, which corresponds to rotations in the (x_1, x_2) plane and preserves the system's total energy. We note that in the limit as $r \to 0$ that the system in (3.20) becomes equivalent to two, decoupled parametric oscillators described by two, damped Mathieu equations. Finally, ϕ denotes the phase difference between the oscillators. Just as in [18], we focus on the cases $\phi = 0, \frac{\pi}{2}$, and π, and we remark that the system in (3.20) for the aforementioned ϕ values exhibits three potential resonance frequencies at $\gamma = 2\Omega_r$ and $\gamma = 2\Omega_r \pm \omega_0 r$, where $\Omega_r = \omega_0\sqrt{1 + (r^2 - g^2)/4}$. Throughout the remainder of this section, frequencies such as γ are given in units of ω_0.

Frequently, a primary interest is in studying the effects of the oscillators coupling strength, the type of coupling, the phase differences between the oscillators, and the magnitude, frequency and phase of the driving force on the long term statistics of the system's solutions [60, 61, 62]. Although most real world applications involve oscillators with differing phases or differing frequencies, most of the analysis that appears in the literature has ignored both of these essential features while also ignoring the effects of coupling strength, especially at frequencies away from parametric resonance [18].

Ignoring frequencies away from parametric resonance frequencies is natural if one is interested in studying phase-locking or if one has energy harvesting applications in mind [63, 64, 65, 66]. However, real-world systems typically operate over a broad range of parametric, modulation frequencies, and many engineering and biological applications including coupled oscillator systems are designed to operate away from parametric resonance to avoid excessive vibration, noise, and accelerate fatigue [67, 68, 69]. The reason these essential features have primarily been ignored is almost exclusively due to the problem's difficulty and the limitations of perturbative asymptotic methods [70].

From a computational perspective, direct numerical simulation (DNS) is incredibly expensive. Moreover, one is limited to the study of trajectories that are numerically stable, which crucially depends on one's choice of initial conditions for parametric oscillators. Frequently, one is interested in unstable solutions in particular for the "control of chaos" [29]. These concerns naturally pose the problems of which initial conditions to choose and a robust method to determine whether one's findings are generically sensitive to initial data.

From a theoretical perspective, past and on-going research often investigates stability regions within the parameter phase space [61, 60] [1], and without the knowledge of exact solutions, approximate solutions are frequently derived via perturbative, asymptotic methods in the form of finitely truncated, solution expansions, such as Fourier or eigenfunction expansions [71, 72, 73, 74, 75]. The principal idea behind these truncated expansions is that a sufficiently small number of terms in the full expansion should serve as a viable solution approximation, and hence, all other terms may be disregarded; from hereafter, we refer to these disregarded terms as higher order terms. Deciding which terms to consider or disregard is often nontrivial, especially in coupled systems, and even given a fixed choice, the truncated solution approximation may still fail to be a viable approximation [76]. Moreover, these asymptotic methods fail to capture the full range of the parameters in question by restricting attention to regions of phase space for which the modulation frequency is the same or an integer multiple of the resonance frequencies [76, 18].

[1] Throughout the remainder of this dissertation, we write stability region(s) to mean the linear stability of the underlying model within the parameter state space.

In addition, as we demonstrate and explain throughout this section, the problems concerning the validity of asymptotic, approximate solutions can be exacerbated for coupled, oscillator systems. In particular, the effects of the oscillator's coupling term can make the higher order terms non-negligible. As higher order terms are frequently neglected, asymptotic methods often forego when or how these higher order terms become relevant [77].

We demonstrate that the auxiliary function method for long-time averages, is incredibly robust in addressing the difficulties previously explicated. On computational grounds, this method is computationally efficient and foregoes the need to manually check or randomly select various initial conditions, such as in [78], since this auxiliary function method determines extremal time averages across *all* initial conditions. On theoretical grounds, our method is robust even in the face of highly parametric, coupled dynamics and does not require approximate expansions or restrictions to only small parameter values. Therefore, we can recover regions of stability and long term statistics without needing to decide on an asymptotic, approximate solution method, decide which higher order terms can be neglected, or attempt to determine if the higher order terms become non-negligible due to coupling effects.

We make use of the auxiliary function method for long-time averages in order to investigate the effects of higher order terms that are ignored in perturbative asymptotic methods. We also compare this auxiliary function method's results against the results predicted by perturbative asymptotic analysis. In particular, we investigate the effects of coupling terms on the system's region of stability and the legitimacy of asymptotic methods in the presence of these coupling effects across a broad range of physically relevant modulation frequencies at and away from parametric resonance.

Indeed to make the system in (3.20) accessible via this auxiliary function method, (3.20) can be written as a first order, coupled system with exclusively polynomial terms:

$$\begin{aligned}
\dot{x}_1 &= y_1 \\
\dot{x}_2 &= y_2 \\
\dot{y}_1 &= -\omega_0^2[1 + hx_3]x_1 - \omega_0 g y_1 + \omega_0 r y_2 \\
\dot{y}_2 &= -\omega_0^2[1 + hx_3]x_2 - \omega_0 g y_2 - \omega_0 r y_1 \\
\dot{x}_3 &= \gamma x_4 \\
\dot{x}_4 &= -\gamma x_3 \\
&\text{Subject to: } x_3^2 + x_4^2 = 1,
\end{aligned} \tag{3.21}$$

where $x_3 = \sin(\gamma t)$, $x_4 = \cos(\gamma t)$, and $\phi = 0$. The system in (3.21) can be written similarly for additional values of ϕ via elementary trigonometric identities.

The constraint $x_3^2+x_4^2 = 1$ is enforced via the S-procedure and forces the variables x_3 and x_4 to be uniquely determined. However, we remark that the S-procedure is not enforced when $\gamma = \phi = 0$, as (3.21) becomes independent of trigonometric terms. We employ the auxiliary function method for long-time averages on the quantity of interest

$$\Phi = x_1^2 + x_2^2. \tag{3.22}$$

Hence, computing $\overline{\Phi}^*$ gives the maximal long-time average of the summed, mean squared amplitudes, which allows us to distinguish between small and large amplitude, parameter dependent solutions. The intuition being that small or large amplitude solutions should generically correlate with asymptotically stable or unstable solutions, respectively. Since the maximal long-time averages will be either zero or infinite in the regions of stability and instability, respectively, the auxiliary function method will allow us to trace out the threshold between stability and instability. The fineness of the threshold, moreover, only depends on one's mesh size for the system parameters; the mesh size is 80×80 for every figure in this section.

With Φ as defined in (3.22), the sum of squares program to solve becomes:

$$\begin{gathered} \min\ \mathrm{U} \\ \text{s.t. } \mathrm{U} - \Phi - \mathbf{f}(x_1, x_2, y_1, y_2, x_3, x_4) \cdot \nabla \mathrm{V} + S(x_1, x_2, y_1, y_2, x_3, x_4)(1 - x_3^2 - x_4^2) \in \Sigma_{6,d} \\ S(x_1, x_2, y_1, y_2, x_3, x_4) \in \Sigma_{6,\tilde{d}} \end{gathered} \tag{3.23}$$

where $\mathbf{f}(x_1, x_2, y_1, y_2, x_3, x_4)$ is the polynomial vector field as defined in (3.21), both d and $\tilde{d}$ are determined by the degree of the S-procedure polynomials as well as the auxiliary function $\mathrm{V}(\mathbf{x})$, and we have enforced the S-procedure at all γ values other than zero. The convex optimization problems in this paper were solved using Mosek [79] paired with Yalmip [80].

We remark that similar computations using the auxiliary function method have already appeared in [1] and were previously discussed in §3.3, where the model of interest was the Duffing oscillator. In particular, it was shown that the auxiliary function method accurately reproduces the Duffing equation's frequency response curve and parameter dependent hysteresis phenomena as derived by harmonic balance down to numerical precision. This further validates the auxiliary function method as a tool to study oscillator dynamics.

To the author's knowledge, this is the first time that the auxiliary function method has been applied to studying parametric, coupled oscillators or has been used to explicitly study dynamical stability, especially in the context of linear stability analysis. Moreover, as we suspect and as we show, the dynamics of (3.20) are substantially more complicated than that which can be described by asymptotic methods. The reason is that in contrast to the Duffing equation, our model has two *coupled*, parametric oscillators and the non-autonomous forcing appears parametrically instead of externally. Hence, the synchronization of the two oscillators plays a crucial role in their stability and that synchronization pivotally depends on one's choice of system parameters.

In a similar fashion to [1], the parameter dependent computations of $\overline{\Phi}^*$ will be used to recover stability region boundaries and the validity of these results will be corroborated via asymptotic analysis. We corroborate the findings of the auxiliary function method for long-time averages via comparison with the approximate solutions established by Floquet theory. This is a standard method which is discussed by many sources on asymptotic analysis; the specific reference used here is [81].

Consider $x(t) \in \mathbb{R}$ satisfying

$$\ddot{x}(t) + F(t)x(t) = 0, \tag{3.24}$$

where $F(t)$ is a periodic function with period T. Equation (3.24) is a second order differential equation and thus has two linearly independent solutions $x_1(t)$ and $x_2(t)$.

Since $x_1(t)$ and $x_2(t)$ are linearly independent, they span the solution space of (3.24). In particular, there must exist $\mathbf{M} \in \mathbb{R}^{2\times 2}$ such that

$$\begin{bmatrix} x_1(t+T) \\ x_2(t+T) \end{bmatrix} = \mathbf{M} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \tag{3.25}$$

where $\mathbf{M}$ is called the *monodromy matrix* of $x_1(t)$ and $x_2(t)$. A monodromy matrix can be constructed for any pair of linearly independent solutions of (3.24). One can show that the determinant of the monodromy matrix is 1 and eigenvalues of the monodromy matrix are independent of the pair of solutions that one chooses to start with.

Suppose λ_1, λ_2 are eigenvalues of a monodromy matrix $\mathbf{M}$ corresponding to solutions $x_1(t)$, $x_2(t)$. For each λ_j there exists a linear combination $x_{\lambda_j}(t)$ of $x_1(t)$ and $x_2(t)$ such that

$$x_{\lambda_j}(t+T) = \lambda_j x_{\lambda_j}(t). \tag{3.26}$$

Floquet's theorem states that we may write

$$x_{\lambda_j} = e^{-i\mu_j t} u_{\lambda_j}(t), \tag{3.27}$$

where $u_{\lambda_j}(t)$ is periodic with period T and $\mu_j \in \mathbb{C}$ is such that $\lambda_j = e^{-i\mu_j T}$. Since $u_{\lambda_j}(t)$ is periodic, one may expand $u_{\lambda_j}(t)$ as a Fourier series and recover the analytic solution $x_{\lambda_j}(t+T)$ to within arbitrary accuracy by recursively solving for the Fourier amplitudes.

For (3.20) with $\phi = 0$, we generally follow the method outlined in [18], where the study focused on frequencies at or near parametric resonance due to an interest in phase-locking, but our novel contribution is that we do not limit our attention to resonant frequencies.

The equations in (3.20) can be decoupled by performing a change of basis and defining

$$x_\pm(t) = x_1 \pm i x_2, \tag{3.28}$$

Adding the first line of (3.20) to $\pm i$ times the second line yields

$$\ddot{x}_\pm + \omega_0^2 [1 + h \sin(\gamma t)] x_\pm + \omega_0 \left(g \pm i r\right) \dot{x}_\pm = 0, \tag{3.29}$$

where the solutions in the new basis $x_\pm(t)$ exhibit real and imaginary loss terms. We then let

$$x_\pm(t) = e^{-\frac{(g \pm i r)\omega_0 t}{2}} y_\pm(t) \tag{3.30}$$

for $y_\pm(t)$ to be determined. Upon substituting (3.30) into the decoupled system of (3.29), we arrive at:

$$\ddot{y}_\pm + \omega_0^2 \left[1 - \frac{(g \pm i r)^2}{4} + h \sin(\gamma t)\right] y_\pm(t) = 0. \tag{3.31}$$

Since the coefficient of $y_\pm$ in (3.31) is periodic with period $T = \frac{2\pi}{\gamma}$, then according to Floquet Theory, we may write

$$y_\pm(t) = e^{-i\mu t} f_\pm(t), \tag{3.32}$$

where $f_\pm(t)$ are some periodic functions with period $T = \frac{2\pi}{\gamma}$. Wherever $\mathrm{Im}(\mu) > 0$, $y_\pm(t)$ will be unstable as $t \to \infty$. Since $f_\pm(t)$ are periodic, we may express them as

Fourier series

$$f_{\pm}(t) = \sum_{n \in \mathbb{N}} A_n^{\pm} e^{in\gamma t}, \tag{3.33}$$

where $A_n^{\pm}$ denotes the amplitude of the nth Fourier coefficient for $f_{\pm}(t)$, respectively. If we substitute the expression in (3.33) into (3.31) and collect the coefficients of $e^{in\gamma t}$, we get the recursion relation for the Fourier coefficients $A_n^{\pm}$ as derived by [18]:

$$D_{\pm,n}(\mu) A_n^{\pm} + i\frac{\omega_0^2 h}{2}(A_{n+1}^{\pm} - A_{n-1}^{\pm}) = 0, \tag{3.34}$$

where $D_n(\mu) = \omega_0^2 - (n\gamma - \mu^2) + i\omega_0(g + ir)(n\gamma - \mu)$. For a visual, we can conveniently write (3.34) as the infinite dimensional matrix:

$$\begin{pmatrix} \ddots & & & \vdots & & & \iddots \\ -i\frac{\omega_0^2 h}{2} & D_{-1}(\mu) & i\frac{\omega_0^2 h}{2} & 0 & 0 & & \\ \cdots & 0 & -i\frac{\omega_0^2 h}{2} & D_0(\mu) & i\frac{\omega_0^2 h}{2} & 0 & \cdots \\ & 0 & 0 & -i\frac{\omega_0^2 h}{2} & D_1(\mu) & -i\frac{\omega_0^2 h}{2} & \\ \iddots & & & \vdots & & & \ddots \end{pmatrix} \tag{3.35}$$

We note that if the driving intensity h is less than one, $A_n^{\pm}$ is coupled most strongly to $A_{n\pm1}^{\pm}$, with coupling proportional to h. Without loss of generality, we consider $n = 0$. Therefore, one recovers a matrix equation of the form:

$$\operatorname{Det}\begin{pmatrix} \mathbf{D}_{-1}(\mu) & \mathbf{M}(h) & 0 \\ \mathbf{M}^*(h) & \mathbf{D}_0(\mu) & \mathbf{M}(h) \\ 0 & -\mathbf{M}^*(h) & \mathbf{D}_{+1}(\mu) \end{pmatrix} = 0, \tag{3.36}$$

where one defines

$$\begin{aligned} \mathbf{D}_n(\mu) &= \begin{pmatrix} D_{+,n}(\mu) & 0 \\ 0 & \mathbf{D}_{-,n}(\mu) \end{pmatrix} \\ \mathbf{M}(h) &= \begin{pmatrix} \frac{ih\omega_0^2}{2} & 0 \\ 0 & \frac{ih\omega_0^2}{2} \end{pmatrix}. \end{aligned} \tag{3.37}$$

Note in (3.36) that enforcing the determinant to vanish ensures that there are non-trivial solutions to (3.31). In contrast, it was also shown in [18] that the simplifying assumptions of $r = g = 0$ and ignoring the coupling effects of D_{+1} yield the matrix equation:

$$\operatorname{Det}\begin{pmatrix} \omega_0^2-(\gamma-\mu)^2 & \frac{ih\omega_0^2}{2} \\ -\frac{ih\omega_0^2}{2} & \omega_0^2-\mu^2 \end{pmatrix}=0, \tag{3.38}$$

The matrix equation in (3.38) is a consequence of assuming the coupling effects as well as the higher order effects are negligible in the asymptotic analysis. Hence, we hereafter refer to the solutions of (3.38), which is the same as presented in [18], as the simplified, uncoupled solution.

Similar to [18], we focus on the solutions of Equation (3.36) that display parametric resonance at $\gamma = 2\Omega_r, 2\Omega_r \pm \omega_0 r$ and the solutions of Equation (3.38) that display parametric resonance at $\gamma = 2\Omega_r$; we present the solutions below. In particular, we show that the simplifying assumptions have highly non-trivial and significant effects on the system's regions of stability at and away from parametric resonance frequencies. We first compare the stability boundary of the results established via (3.36) with the stability boundary of the simplified, uncoupled solution of (3.38). We then compare the stability boundaries as predicted by asymptotic analysis with the stability boundary as predicted by the auxiliary function method for long time averages.

Indeed, we employ MATLAB's solve function to find the roots in μ of (3.36), whose solutions are sufficiently complicated and long to not be included in the text, and this yields six solutions. In Figure 3.9, we plot the contours of the imaginary part of one of the combined boundaries of the six solutions of our more generalized result in comparison to the simplified, uncoupled solution to (3.38) for a wide range of parameter values: We remark that implicitly solving for the roots of μ is computationally sensitive to very small numerical error. Hence, the stability boundary seen in Figure 3.9 had miniscule variations depending on the software used. However, the general shape of the stability boundary is consistent across software.

In Figure 3.9, we see that the more general, higher order asymptotic analysis solution is a dramatic and surprising improvement on the simplified, uncoupled solution, as the region of instability for the simplified, uncoupled asymptotic result is a subset of the region of instability for the more general asymptotic result. While there is good agreement between the two solutions for $|\gamma| > 1$, there is a large region within the parameter space, $|\gamma| \leq 1$, for which the simplified, uncoupled asymptotic analysis solution predicts stability, while the more general, higher order solution reveals instability.

The fact that the simplified, uncoupled asymptotic expansion stability solution fails so drastically for $|\gamma| \leq 1$ is quite a surprising finding. This means the simplified

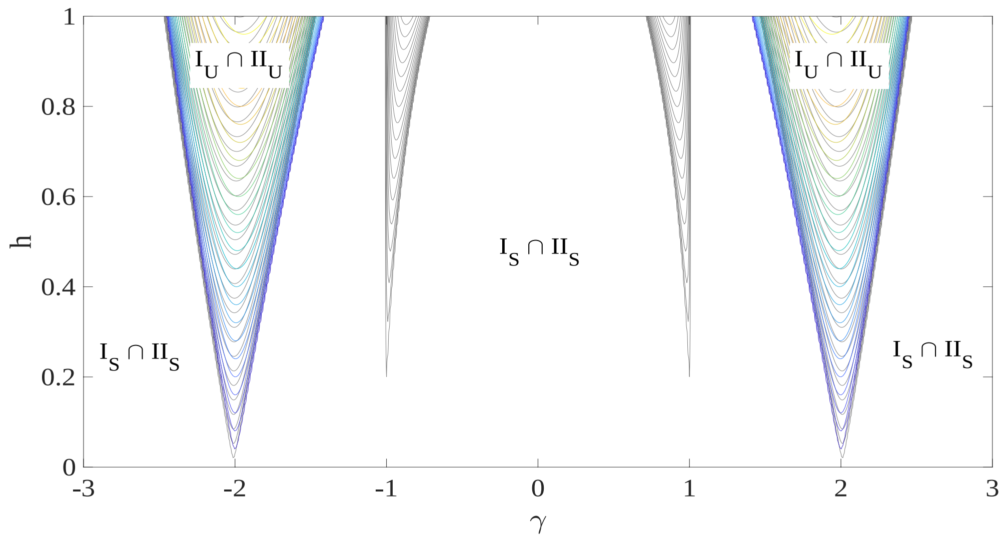

Figure 3.9: (McMillan and Young; 2022, [2]) The simplified, uncoupled asymptotic solution to Equation (3.38) (shown with colored, solid contour lines) and the more general asymptotic solution to Equation (3.36) (shown with dashed, gray contour lines). The markers $\mathrm{I_S}$, $\mathrm{I_U}$, $\mathrm{II_S}$, and $\mathrm{II_U}$ denote the stable (S) or unstable (U) regions of the simplified (I) and general (II) solutions, respectively. The parameter values are $r = .2$, $g = .01$, and $\phi = 0$.

solutions are dangerously un-conservative in predicting regions of stability. However, it then becomes natural to ask if the more general, higher order solution also fails to fully describe the stability boundary of the solutions to (3.20). That is, we may ask if including successively higher order terms in the asymptotic expansion would lead to such drastic changes in the stability boundary as seen between the two solutions in Figure 3.9. In order to address this question, we compare the stability regions as predicted by the general asymptotic analysis with the stability regions predicted by the auxiliary function method for long time averages.

Upon computing the stability boundary with the auxiliary function method for long time averages via the SDP in (3.23), we arrive at the solid black line seen Figure 3.10:

In Figure 3.10 [2], we see that the more general asymptotic analysis result agrees quite well with the results of the auxiliary function method. In particular and most noticeably, both methods capture a pair of narrow, protruding tongues that occur at $\gamma \in \{-1, 1\}$. However, there is still a quite large range of γ values for which the auxiliary function method predicts potential instability, but the asymptotic analysis

[2]In Figure (3.10), the solution is independent of h and there is not a singularity at $\gamma = 0$; the appearance is simply a consequence of line thickness.

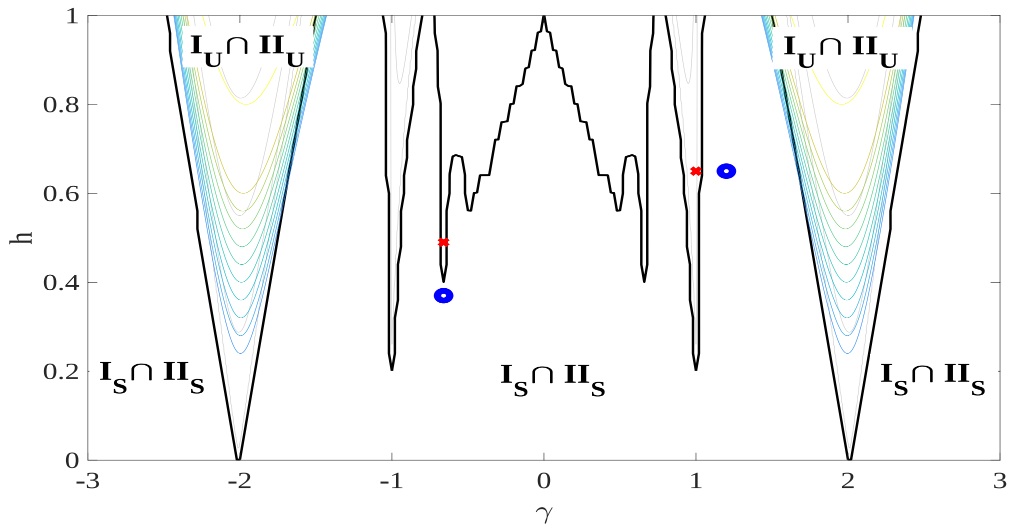

Figure 3.10: (McMillan and Young; 2022, [2]) There is strong agreement between our auxiliary function method (shown with a thick, solid black line) with the theoretical predictions of the simplified, uncoupled asymptotic analysis (shown with colored, solid contour lines) and the more general, higher order asymptotic analysis (shown with dashed, gray contour lines) for $r = .2$, $g = .01$, and $\phi = 0$. The open blue circles and red crosses indicate where direct numerical simulation results are shown in Figure 3.11 for stable and unstable points, respectively.

solution to Equation (3.36) predicts stability.

In order to validate the extended region of instability as predicted by the auxiliary function method for long time averages results, we choose four points within the parameter space to perform direct numerical simulation using ode45 in MATLAB. The results show that the auxiliary function method for long time averages is corroborated via direct numerical simulation.

That is, for points marked by red crosses inside the region of instability, there is a blow up of the solutions, with the time histories shown on the right hand side of Figure 3.11. However, for points marked by open, blue circles—just outside the instability region—solutions experience decay and hence are stable; the corresponding time histories are shown on the left hand side of Figure 3.11.

Additionally, Figure 3.11 displays three trajectories for three randomly chosen initial conditions for each of the four chosen points in parameter space. We also note that the above figure only plots the trajectory of $x_1(t)$ for our model, but when we plotted the trajectories for $x_2(t)$, a similar pattern of blow-up or stability still held.

This is perhaps a case study example of how the auxiliary function method can lend itself to finding regions of instability where perturbative methods fail. Strik-

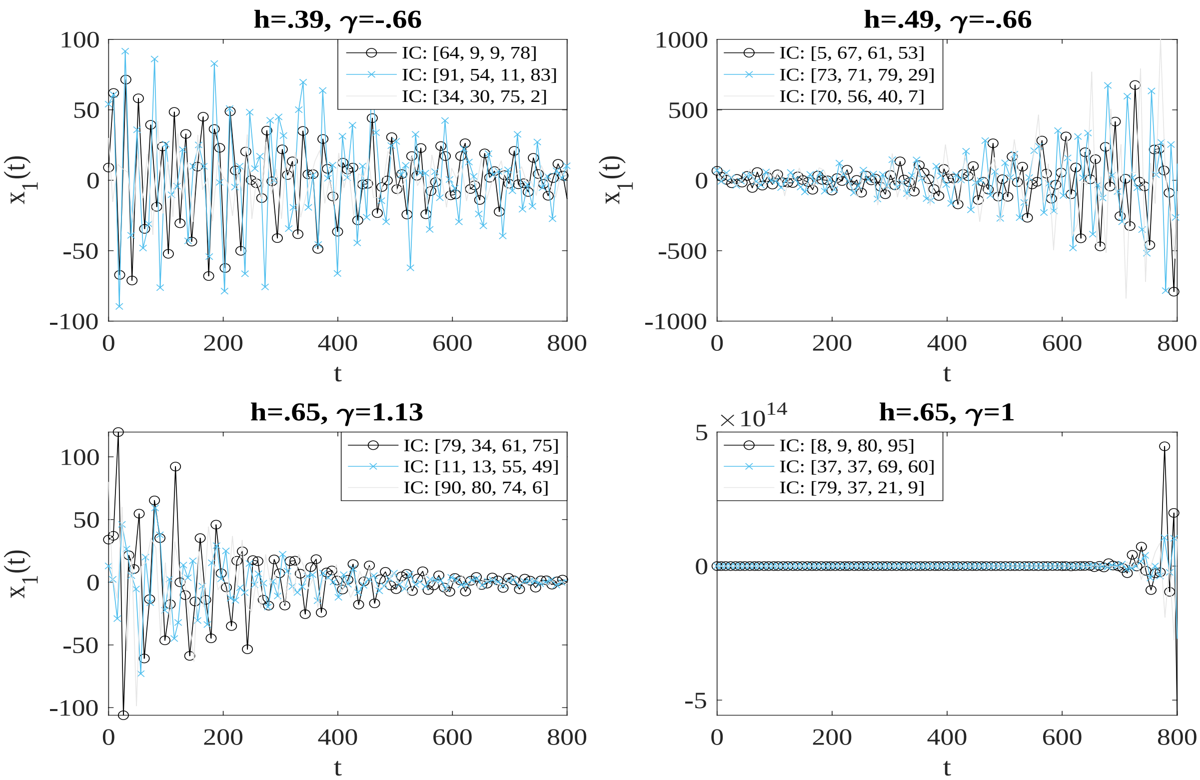

Figure 3.11: The time histories predicted by DNS for the four points marked in Figure 3.10. The left two plots correspond to the two blue, open circles just outside the region of instability for Figure 3.10, and the two right plots correspond to the two red crosses just inside the region of instability for Figure 3.10. In the above plot, an initial condition vector IC: [a,b,c,d] corresponds to $x_1(0) = a$, $x_2(0) = b$, $y_1(0) = c$, and $y_2(0) = d$ for Equation (17) with $r = 0.2$, $g = 0.01$, and $\phi = 0$.

ingly, the perturbative method for this system seems to fail quite drastically with a large portion of the stability diagram not being captured by the perturbative method but instead by the auxiliary function method. Therefore, the auxiliary function method has the advantage of being able to recover the true stability boundary both at and away from parametric resonance, while doing so in a computationally efficient way. We remark that the auxiliary function method is quite computationally efficient despite the polynomial representation, seen in (3.21) , containing five degrees of freedom. In particular, the CPU times corresponding to the computation of $\overline{\Phi^*}$ at the points $(\gamma, h) = (-1.13, .650), (-1.00, .650), (-.660, .390)$, and $(-.660, .490)$ are 3.4844, 2.6744, 2.8574, and 2.5737 seconds for a degree 8 auxiliary function and a degree 6 S-procedure enforcement on nothing more than a standard laptop using a single core, 2.2 GHZ processor.

In view of the above results, it then becomes natural to analyze how varying the system parameters changes the above regions of stability. That is, we study the effects of varying the coupling strength in the next section.

In order to study the affects of varying the coupling strength, we consider (3.20) for $r = .4$ and compare the stability boundary with the prior results of the both auxiliary function method as well as the asymptotic analysis for $r = .2$ and $g = .01$. We do not consider negative values for r because the coupling term appears negatively in the first equation and positively in the second equation of (3.20); hence, a sign change of r just swaps the role of $x_1(t)$ and $x_2(t)$, respectively. Also, we focus on $\gamma \in [-1.5, 1.5]$, as this is where the discrepancies between the auxiliary function method and asymptotic analysis were previously established.

Performing the same procedure with the auxiliary function method as in the previous section as well as incorporating the predictions of the asymptotic analysis, we find in Figure 3.12 that the region of stability for the equations with $r = .2$ is a subset for the region of the stability for the equations with $r = .4$. Intuitively, one would expect this to be true because as the r-value increases, the rate of energy transfer from one oscillator to another increases and hence initial transients are far less likely to become unstable. Figure 3.12 shows that the discrepancies between the auxiliary function method and the asymptotic analysis solution for $r = .4$ persist. Moreover, the discrepancies seem to grow as the value of r is increased. Therefore, as the coupling strength between the oscillators increases, the asymptotic analysis solution to (3.36) increasingly fails to capture the true stability boundary of the system, because the coupling terms amplify the relative importance of the higher order terms, particularly away from parametric resonance. We also note that Figure 3.12 is for $g = .01$, but if one is interested in the affects of g, we can similarly compute regions of stability for fixed r and different g values. The expected result would simply be a shift up of the region of stability appearing in Figure 3.9. The reason is that g appears on the damping term and hence increasing the damping coefficient should increase the size of the region of stability.

Finally, we study the effects of varying the phase value ϕ. In a fashion similar to [18], we focus on $\phi = 0, \frac{\pi}{2}, \pi$. However, it is important to remark that [18] was unable to determine a closed form, asymptotic solution for varying ϕ. We also could not find a closed form solution to the analog of (3.36) with $\phi \neq 0$. The ability to find the stability boundary for general system parameters proves to be another benefit of the auxiliary function method. Indeed, performing the analogous computations

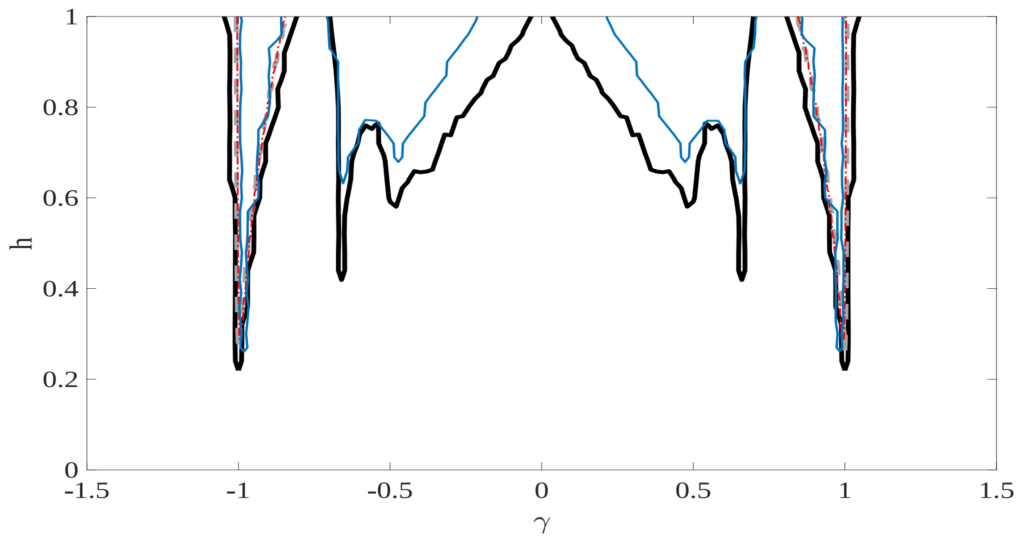

Figure 3.12: (McMillan and Young; 2022, [2]) The stability boundary as predicted by the auxiliary function method is shown with a thick, solid black line and a thick, solid blue line for $r = .2$ and $r = .4$, respectively, while the asymptotic analysis boundary is shown with a thin, red dashed line and a thin, gray dotted line for $r = .2$ and $r = .4$, respectively, with $g = .01$ and $\phi = 0$.

via the auxiliary function method and choosing the same Φ as before, we arrive at Figure 3.13. Figure 3.13 shows that the auxiliary function method reproduces the

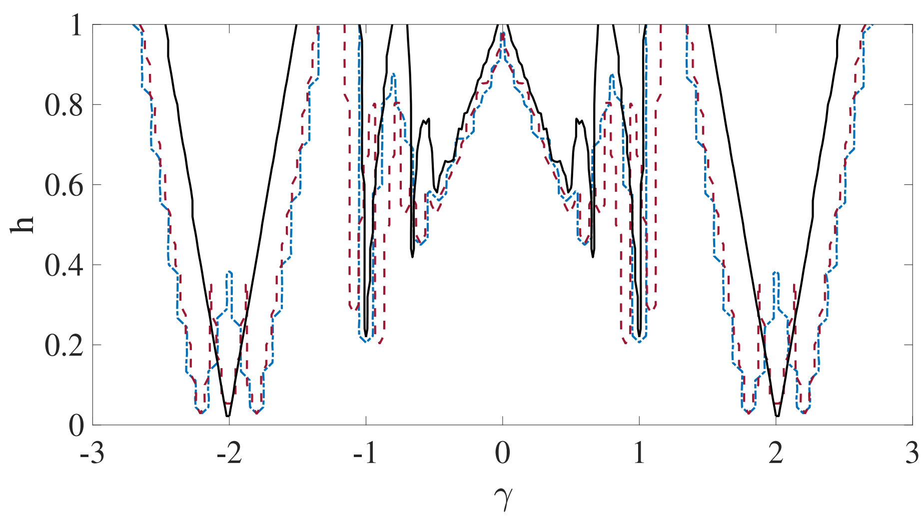

Figure 3.13: (McMillan and Young; 2022, [2]) The stability boundary as predicted by the auxiliary function method for $\phi = 0$ (shown with a thick, solid black line), $\phi = \frac{\pi}{2}$ (shown with a thin, dashed blue line), and $\phi = \pi$ (shown with a thin, dashed red line).

protruding tongues around $\gamma = \pm 2\Omega_r$ and $\gamma = \pm 2\Omega_r \pm \omega_0 r$ with $\phi = 0$, $\phi = \pi$, and $\phi = \frac{\pi}{2}$ corresponding to one, two, and three tongues respectively. Moreover, note that varying ϕ has non-trivial effects throughout the entirety of parameter space, and not only about $\gamma = \pm 2\Omega_r, \pm 2\Omega_r \pm \omega_0 r$, as varying the phase of the parametric oscillator term can lead to more localized tongue formation.

Since the phase of a parametric oscillator can often be random, it is important to identify the instability boundary when varying phases are taken into consideration, especially when there is non-trivial disagreement between the true stability boundary and the boundary as predicted by asymptotic analysis.

CHAPTER IV

Dynamic Choice and Utilitarianism

4.1 Introduction

In this chapter, we first overview some classic results in decision theory. In particular, we discuss the seminal Von Neumann–Morgenstern theorem [3], concerning the axiomatization of a preference relation satisfying Von Neumann–Morgenstern rationality, as a special case of dynamic choice occurring in only one time period—static choice. We then move on to discuss the Von Neumann–Morgenstern theorem's connections with choice occurring in finitely or infinitely many periods. In particular, we overview the exponentially discounted, expected utility framework initially proposed by Samuelson [82]. In both frameworks, we present and discuss the relevant axioms and their intuition as well as the underlying preference relation's functional representation; we also provide a brief discussion of the subsequent literature relating to the descriptive limitations of the underlying models.

Second, we provide a brief overview of the ethical theory of utilitarianism from the perspective of Jeremy Bentham [4]. We overview the major tenants of utilitarianism, which are the maximization of pleasure and the minimization of pain, as well as discuss Bentham's framework for achieving these extrema. We then overview alternative utilitarian frameworks and explicate the connections between utilitarianism and decision theory—in particular, utilitarianism's connections with social choice theory. We discuss the historical development of what is to be considered the "ultimate good" of utilitarianism from an economics perspective, which ultimately falls on the criteria established by John Harsanyi [83]—the satisfaction of preferences within the Von Neumann–Morgenstern expected utility paradigm.

Finally, we overview the developments of social choice theory by beginning with the seminal work of Kenneth Arrow and discussing the subsequent reaction within the economics literature. We conclude the chapter by discussing the work of John

Harsanyi on decision theoretic utilitarianism and the extensions of Harsanyi's work that appear in the literature. In addition, we discuss the results of Zhou [84] that will be used in the new results of Chapter V and explain the limitations of Zhou's result as well as motivate the need for this work.

4.2 Dynamic choice under uncertainty

It is natural to analyze decisions that are made dynamically while also in the presence of an uncertain future. The study of dynamic choice is thus concerned with choices such that different actions lead to different outcomes that are realized at various stages over time. Moreover, as the future is uncertain, we think of these future outcomes as also being uncertain.

In order to fix ideas, we begin with the static case, which can be thought of as choice occurring in only one period. For any topological space X, we define $\Delta(X)$ to be the set of all Borel probability measures over X. Let $\succeq$ be a binary relation over $\Delta(X)$, where $\succeq$ is the preference relation of some specified individual and we refer to elements of $\Delta(X)$ as prospects. We sometimes refer to $\succeq$ as a *weak preference*, and it induces two other binary relations on $\Delta(X)$. For any two prospects $\mathbf{p}, \mathbf{q} \in \Delta(X)$, we define $\succ$, the *strict preference*, as $\mathbf{p} \succ \mathbf{q} \iff \mathbf{p} \succeq \mathbf{q}$ and $\mathbf{q} \not\succeq \mathbf{p}$[1], and we define $\sim$, *indifference*, as $\mathbf{p} \succ \mathbf{q} \iff \mathbf{p} \succeq \mathbf{q}$ and $\mathbf{q} \succeq \mathbf{p}$.

With the preference relations in hand, we concern ourselves with characterizing the decision maker's preferences under the assumption that they behave subject to consistent rules or axioms. There are of course various axioms that one could invoke on the preference relations to describe different behavioral aspects, but a very simple criteria is that of rationality. We assume that our decision maker behaves with Von Neumann–Morgenstern rationality [3], which is surmised with the following four axioms:

Axiom 1. *(Completeness) For any $\boldsymbol{p}, \boldsymbol{q} \in \Delta(X)$, one of the following hold: $\boldsymbol{p} \succ \boldsymbol{q}$, $\boldsymbol{q} \succ \boldsymbol{p}$, or $\boldsymbol{q} \sim \boldsymbol{p}$.*

Axiom 2. *(Transitivity) For any $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in \Delta(X)$, if $\boldsymbol{p} \succ \boldsymbol{q}$ and $\boldsymbol{q} \succ \boldsymbol{r}$, then $\boldsymbol{p} \succ \boldsymbol{r}$; similarly for $\sim$.*

Axiom 3. *(Continuity) For any $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in \Delta(X)$, if $\boldsymbol{p} \succeq \boldsymbol{q} \succeq \boldsymbol{r}$, then there exists $\lambda \in [0, 1]$ such that $\lambda \boldsymbol{p} + (1 - \lambda)\boldsymbol{r} \sim \boldsymbol{q}$.*

[1]The notation $\mathbf{p} \not\succeq \mathbf{q}$ means the logical negation of the relation statement.

Axiom 4. *(Independence) For any* $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in \Delta(X)$ *and* $\lambda \in (0,1]$, $\boldsymbol{p} \succeq \boldsymbol{r} \iff \lambda\boldsymbol{p} + (1-\lambda)\boldsymbol{q} \succeq \lambda\boldsymbol{r} + (1-\lambda)\boldsymbol{q}$.

The completeness axiom says that for any two prospects the decision maker is decisive and can say which prospect they prefer over the other or if they are indifferent between the two. The transitivity axiom says that the decision maker respects their own orderings; there are not prospects for which order reversals are possible. The continuity axiom says that there is a threshold for which weakly better and weakly worse prospects become equally preferred to something in between; there are no prospects that are infinitely preferred over others. The independence axiom says that if one prospect is preferred over another then adding extraneous prospects to both in an identical fashion should not alter the preference.

All four axioms are intuitively appealing while also decently seem to capture a notion of rationality. Moreover, the above four axioms on the primitive $\succeq$ lead to the following theorem:

Theorem IV.1. *(Von Neumann–Morgenstern; 1947, [3]) Suppose* $\succeq$ *over* $\Delta(X)$ *satisfies axioms 1-4. Then there exists a function* $u : X \to \mathbb{R}$ *such that*

$$\boldsymbol{p} \succeq \boldsymbol{q} \iff \mathbb{E}_{\boldsymbol{p}}[u] \geq \mathbb{E}_{\boldsymbol{q}}[u], \tag{4.1}$$

where $\mathbb{E}_{\boldsymbol{p}}[u] := \int_X u(x)d\boldsymbol{p}$. *Moreover, the function* $u(\cdot)$ *is unique up to a positive linear transformation.*

The interpretation of Theorem IV.1 is that under axioms 1–4 a decision maker's preferences are essentially represented by an expected value operator, where this expectation is taken over an endogenously revealed utility function $u(\cdot)$. Therefore, a decision maker behaves in such a way as to maximize their expected utility despite the decision maker potentially not knowing their own utility function $u(\cdot)$. Theorem IV.1 has been met with criticism in the literature. In particular, it has been shown that decision makers frequently violate the theorem's criteria which have been studied under the Allais paradox [85] and the framing effect [86].

However, Theorem IV.1 is concerned with static choice, and we would like to introduce a framework for which consumption occurs in more than one period. For the dynamic setting, we again take as primitive a binary relation $\succeq$, but the choice domain is now $\mathbb{P} := \Delta\big(\prod_{t\in\mathcal{T}} X_t\big)$ [2], where $\mathcal{T}$ may be a finite or countably infinite index set. In this framework, the index set $\mathcal{T}$ represents time, and we think of an element

[2] In §5.2, we choose $\prod_{t\in\mathcal{T}} \Delta(X)_t \subset \mathbb{P}$ as the choice domain.

$\mathbf{p} \in \mathbb{P}$ as an uncertain, sequence of future consumption. For the sake of simplicity in this section, we take $X_t = X$ for all t—that is, the same set of consumption goods is available to the decision maker in each period.

Analogously to the static case, preferences are defined over a space of probability distributions in order to capture uncertainty about future consumption. However, in the dynamic setting, there are further behavioral aspects that should be captured. In particular, ideas such as being consistent over time or displaying impatience are commonplace, and they should be considered within any axiomatic framework of dynamic choice.

A common dynamic choice framework is the *exponentially discounted, expected utility* framework [87]. The central assumption to the exponential, discounted expected utility framework is that all of the diverse motives and behavioral characterizations can be captured with a single parameter—the discount rate. Intuitively, a decision maker's discount rate is meant to capture the relative weighting of present versus future consumption. For example, a highly impatient individual should have a relatively high discount rate, and the converse should be true for a patient individual. In [88], Koopman was the first to axiomatize an expected discount utility maximizer for the infinite time horizon setting. However, Koopman's original work was noted to have fundamental flaws regarding rigor, so instead, we make use of the work appearing in [87], and for simplicity, we assume that $|\mathcal{T}| < \infty$—that is, there are a finite number of time periods.

In order to state the axioms, we introduce the following definitions:

Definition IV.2. For any $\mathbf{p} \in \mathbb{P}$, $\mathbf{p}_i$ is *the marginal measure of* $\boldsymbol{p}$ *on the i-th component of* $X^{|\mathcal{T}|}$. That is, $\mathbf{p}_i(B) = \mathbf{p}(X^{i-1} \times B \times X^{|\mathcal{T}|-i})$ for any Borel set B.

Definition IV.3. For any $\mathbf{p} \in \mathbb{P}$, $\mathbf{p}_i^c$ is the *marginal measure of* $\boldsymbol{p}$ *on all but the i-th component of* $X^{|\mathcal{T}|}$. That is, $\mathbf{p}_i^c(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \sum_{x \in X} \mathbf{p}(x_1, \dots, x_{i-1}, x, \dots, x_n)$. Note that if X is infinite, then the summation is replaced with integration.

Definition IV.4. For the preference $\succeq$ on $\mathbb{P}$, we define the *marginal preference* $\succeq_M$ on $\Delta(X)$ by $\mathbf{p}^* \succeq_M \mathbf{q}^* \iff \mathbf{p} \succeq \mathbf{q}$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{P}$ such that $\mathbf{p}_i = \mathbf{p}^*$ and $\mathbf{q}_i = \mathbf{q}^*$ for all $i \in \{1, \dots, |\mathcal{T}|\}$.

With the above definitions, we are now ready to state the axioms:

Axiom 5. *(Persistence)* $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}$ *and* $\forall \boldsymbol{p}^*, \boldsymbol{q}^* \in \Delta(X)$ *if there exists* $i \in \{1, \dots, |\mathcal{T}|\}$ *such that* $\boldsymbol{p}_i = \boldsymbol{p}^*$, $\boldsymbol{q}_i = \boldsymbol{q}^*$, $\boldsymbol{p}_i^c = \boldsymbol{q}_i^c$, *then* $\boldsymbol{p} \succ \boldsymbol{q} \iff \boldsymbol{p}^* \succ \boldsymbol{q}^*$.

Axiom 6. *(Consistency)* $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}$ *if* $\boldsymbol{p}_i = \boldsymbol{q}_i$ *for all* $i \in \{1, \dots, |\mathcal{T}|\}$, *then* $\boldsymbol{p} \sim \boldsymbol{q}$.

Axiom 7. *(Stationarity) There exists some* $\boldsymbol{x} \in \Delta(X)$ *such that if* $\boldsymbol{p}, \boldsymbol{q}, \tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}} \in \mathbb{P}$ *with* $\boldsymbol{p}_n = \boldsymbol{q}_n = \tilde{\boldsymbol{p}}_1 = \tilde{\boldsymbol{q}}_1 = \boldsymbol{x}$, $\boldsymbol{p}_n^c = \boldsymbol{q}_n^c$, *and* $\tilde{\boldsymbol{p}}_1^c = \tilde{\boldsymbol{q}}_1^c$, *then* $\boldsymbol{p} \succ \boldsymbol{q} \iff \tilde{\boldsymbol{p}} \succ \tilde{\boldsymbol{q}}$.

The persistence axiom says that the marginal preference $\succeq_M$ is a "faithful" weak order; it preserves preferences. The consistency axiom says that the behavior of the marginal preference remains unchanged across all $|\mathcal{T}|$ factors; it does not matter which factor we consider at which time. The stationarity axiom says that swapping the first and final marginal measure does not change preferences.

Under the aforementioned axioms, the natural analog of Theorem IV.1 in the dynamic setting is thus the following:

Theorem IV.5. *Let* $\succeq$ *be a binary relation over* $\mathbb{P}$. *Suppose* $\succeq$ *satisfies the Von Neumann–Morgenstern axioms in addition to axioms 5–7. Then there exists a function* $v : X \to \mathbb{R}$ *and a unique number* $\delta > 0$ *such that*

$$\boldsymbol{p} \succeq \boldsymbol{q} \iff \sum_{i=1}^{|\mathcal{T}|} \delta^{i-1} \mathbb{E}_{\boldsymbol{p}_i}[v] \geq \sum_{i=1}^{|\mathcal{T}|} \delta^{i-1} \mathbb{E}_{\boldsymbol{q}_i}[v]$$

In Theorem IV.5, the parameter δ is called the discount factor, which precisely measures the decision maker's relative weighting of present versus future consumption, and the function $v(\cdot)$ is a Von Neumann–Morgenstern expected utility function. Hence, the decision maker determines the overall utility of a consumption element by taking into consideration the expected utility of the consumption good in each period determined by the marginal distribution, discounting that expected utility in proportion to which period the consumption occurs, and then simply summing those discounted values. In the literature, the parameter δ is frequently taken to be less than one, which can be axiomatized via some axiom that is meant to represent impatience [87].

However, we remark that the exponential, expected discounted utility framework has been shown to have descriptive limitations such as the common difference effect for which individuals behave inconsistently over time [89], and this lead to numerous generalizations such as quasihyperbolic discounting [90], generalized hyperbolic discounting [91], liminal discounting [92], and rank-dependent discounted utility [93].

With the formalism of both a static and dynamic decision maker in hand, we now turn to utilitarianism and discuss its connections to decision theoretic analysis in both static and dynamic choice settings.

4.3 Utilitarianism

Utilitarianism from the perspective of Bentham is a normative, ethical theory concerning the maximization of happiness and well-being for a collection of individuals. Indeed, in the seminal work, "An Introduction to the Principles of Morals and Legislation" [4], Bentham proposed a hedonistic moral philosophy according to which an action is right if it tends to promote pleasure or happiness, and an action is wrong if it tends to promote unhappiness or pain. Bentham's theory has been coined utilitarianism and is considered a type of consequentialism, where the consequences of happiness or pain determine the moral rightness or wrongness of an action [4].

Of course levels of pain and happiness associated with a specified action are subjective to an individual, so from the onset, it was not clear how an individual could properly make choices in this hedonistic framework. To overcome this moral ambiguity, Bentham suggested individuals perform a hedonic calculus, which is an algorithmic calculation to determine whether an action is right or wrong based on seven criteria: intensity—the strength of the pleasure, duration—the longevity of the pleasure, certainty or uncertainty—the likelihood or unlikelihood of the pleasure occurring, propinquity—the temporal resolution of the pleasure, fecundity—the likelihood that the pleasure will be followed by other similar pleasures, and extent—the number of people impacted.

Furthermore, in order to connect his ethical theory to political philosophy, Bentham attempted to establish rules for a legislator or governing body on when to invoke a particular piece of legislation and to establish criteria for when and how this governing body should act. We hereafter refer to any legislator or governing body as a *social planner*. In the words of Bentham,

> "Pleasures then, and the avoidance of pains, are the ends that the legislator has in view . . . The business of government is to promote the happiness of the society, by punishing and rewarding . . . In proportion as an act tends to disturb that happiness, in proportion as the tendency of it is pernicious, will be the demand it creates for punishment. [4]"

Therefore, in Bentham's framework, a social planner's moral imperative is to promote pleasure and demote pain for the constituents of the underlying social planner.

Bentham's theory very quickly received widespread attention in both philosophical as well as economic studies. In the economic setting, many researchers set out to determine, concretely, the ultimate good of utilitarianism and a proper framework for social utility. The hedonistic framework that Bentham proposed would quickly fall out of favor as it could not account for people being motivated by, for example, social status, knowledge, success, altruism, etc [83]. In [94], Moore would go on to propose an ideal utilitarianism, which measured social utility by the total amount of "mental states of intrinsic worth". However, Moore's theory also had serious descriptive difficulties, as it appealed to what he called "nonnatural qualities", which would prove to be an unsatisfactory metaphysical theory. Ultimately, Harsanyi's theory of preference utilitarianism would survive [83]. Harsanyi argued that the difficulties of the aforementioned theories could be escaped by following common economic practice—that is, defining social utility in terms of preferences. This practice can be justified by both the biblical and Kantian principle of "*treat others in the way that we want to be treated*", and this justification allows for a simple, essentially tautological solution in deciding *how* people would like to be treated—namely, to be treated in accordance with our own preferences. This framework provided the flexibility of allowing individuals to choose for themselves what actions or outcomes have intrinsic value and then insisting that a social planner respect their choices [5].

This economic criteria of preference satisfaction was eventually linked to the Von Neumann–Morgenstern paradigm of expected utility. In particular, if operating within a social utility framework for which the social planner should altruistically care about individuals and take into account the preferences of individuals, past and current research has been concerned with determining which criteria a social planner's preferences should satisfy as well as the functional representation of this social preference. Bentham suggested that a social planner's ultimate good should be a weighted summation of individual's pleasures minus their pain [4]; in this way, a governing body would then non-trivially take into account the preferences and valuations of each individual—an idealization of democracy. Therefore, in the Von Neumann–Morgenstern paradigm, it then became natural to ask what axioms on the social planner's preference will make the social preference equivalent to a social utilitarian representation—a weighted summation of individual expected utilities.

4.4 Harsanyi's theorem and extensions

As Bentham suggested that a utilitarian decision maker's utility should consist of a weighted summation of individual utilitys, it became natural to seek under what conditions the decision maker's primitive would yield such a weighted summation representation. However by 1950, Kenneth Arrow had published his seminal work "A Difficulty in the Concept of Social Welfare" [95]. In [95], Arrow concerned himself with combining the preferences of individuals into an aggregate social preference and the intuitively appealing conditions that such an aggregation should satisfy. The problem, in the words of Arrow, is the following:

> "Certain properties which every reasonable social choice function should posses are set forth. The possibility of fulfilling these conditions is then examined. If we are lucky, there will be exactly one social choice function that will satisfy them. If we are less fortunate, there can be several choice functions satisfying the conditions or axioms. Finally, it will be the height of bad luck if there exists no function fulfilling the desired conditions."

Arrow's Impossibility Theorem would prove to be the " height of bad luck ". Indeed, the theorem states that a planner's aggregating function could not satisfy the four intuitively appealing criterion of unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives [95]—let alone a social aggregation representation that was additively separable such as a utilitarian weighted summation.

Arrow's work would subsequently send wide-ranging, ripples throughout the social choice community and literature. Perhaps the only matter equally as wide-ranging as the volume was the form of the responses. Various responses ranged from attempted refutations, alternate solutions, relaxations or compromises, pessimistic resignations, or denials concerning the applicability of Arrow's analysis to various problems [96]. On the side of pessimism, many economists believed Arrow's result suggested a coherent or satisfactory conception of social welfare was unattainable. A wide array of more optimistic economists would go on to seek natural relaxations of Arrow's hypotheses in order to reconcile a suitable notion of welfare. By 1955, Harsanyi would adopt the Pareto efficiency hypothesis and address the problem of social welfare from a utilitarian point of view [6].

To fix terminology, let $\succeq_i$ with $i \in I$ for some set I and $\succeq$ be a collection of binary relations over $\Delta(X)$. For each $i \in I$, $\succeq_i$ will be called an individual preference and $\succeq$ will be called the social planner's preference. We think of the planner as altruistically caring about the preferences of individuals in a democratic sense. In order to make this precise, we have the following definitions:

Definition IV.6. Suppose $\succeq_i$ for $i \in I$ and $\succeq$ are binary relations over some set Y. The relation $\succeq$ is said to satisfy Pareto indifference, weak Pareto, and strong Pareto if the following, respective conditions are met:

- Pareto indifference: $\mathbf{p} \sim_i \mathbf{q} \Rightarrow \mathbf{p} \sim \mathbf{q} \quad \forall \mathbf{p}, \mathbf{q} \in Y$

- Weak Pareto: $\mathbf{p} \succeq_i \mathbf{q} \Rightarrow \mathbf{p} \succeq \mathbf{q} \quad \forall \mathbf{p}, \mathbf{q} \in Y$
- Strong Pareto: $\mathbf{p} \succ_i \mathbf{q} \Rightarrow \mathbf{p} \succ \mathbf{q} \quad \forall \mathbf{p}, \mathbf{q} \in Y$

The interpretation of the Pareto conditions is clear. They state that upon a unanimous preference between two prospects that the social planner must also respect those preferences. By utilizing the Pareto conditions and placing preferences within the expected utility framework of Von Neumann–Morgenstern, Harsanyi's 1955 work was the first to place utilitarian aggregation on a firm mathematical foundation. Indeed, he proved the following result:

Theorem IV.7. *(Harsanyi; 1955, [6]) Suppose $\succeq_i$ with $i \in I$ with $|I| < \infty$ and $\succeq$ satisfy the Von Neumann–Morgenstern axioms and suppose that Pareto Indifference is satisfied. Let U_i be an expectational representation of $\succeq_i$, and let U be an expectational representation of $\succeq$. Then there exist numbers w_i and b such that for all $\boldsymbol{p} \in \Delta(X)$:*

$$U(\boldsymbol{p}) = \sum_{i=1}^{n} w_i U_i(\boldsymbol{p}) + b$$

(a) Suppose weak Pareto is satisfied. Then the w_i are non-negative.
(b) Suppose Strong Pareto is satisfied. Then the w_i are positive.
(c) The w_i are unique if and only if Independent Prospects is satisfied.

It is quite beautiful that the intuitively appealing and simply stated Pareto axioms give the weighted summation that Bentham had described so many years earlier. Harsanyi's result would subsequently ignite a wide ranging academic discussion. Several authors went on to debate the ethical relevance and consequences of his results as they pertain to utilitarianism [97, 98], while others have attempted to make Harsanyi's original analysis more rigorous, provide alternative proofs, or extend the generality of the results [99, 100, 101, 102, 103].

However, all of the aforementioned analyses were limited to a crucial assumption of Harsanyi—the number of individuals appearing in the aggregation is finite. In particular and in the context of dynamic choice, one is frequently working with an infinite time horizon, and hence, the number of individuals in the utilitarian aggregation is also infinite. In view of this, the work by [84] was the first to attempt a rigorous formulation of Harsanyi's original result with infinitely many individuals.

Theorem IV.8. *(Zhou; 1997, [84]) Suppose individual preferences $\succeq_i$ with $i \in I$ and the social planner's preference $\succeq$ satisfy the Von Neumann–Morgenstern axioms* *with expectational representations U_i and U, respectively. The social planner and individual preferences satisfy weak Pareto if and only if there exists a nonnegative, linear functional β on span$\{U_i\}$ such that*

$$U(\boldsymbol{p}) = \beta(U(\boldsymbol{p}; \cdot)) \quad \forall \boldsymbol{p} \in \Delta(X). \tag{4.2}$$

We note that the set I in Theorem IV.8 is completely general, and the proof of Theorem IV.8 is purely constructive. That is, one can construct a positive linear functional on the vector space spanned by individual utility functions and show that the planner's utility function is equal to this functional using standard mixture space arguments. However, the linear functional β can not be extended to a positive linear functional on the space of all continuous functions over I, and hence, it fails to be easily characterized. However, if one additionally assumes that the individual and social planner's utility representations separate points and that I is a compact metric space, then Theorem IV.8 holds with the added structure that β can now be represented as integration against a countably additive measure; this is Theorem 2' in [84].

Unfortunately, Zhou's result has a critical pitfall. The pitfall is that Zhou's result does not yield utilitarianism—that is, a weighted summation of individual utility, so the extension is not complete. In the subsequent chapter, we provide an intuitive, dynamic setting for which Harsanyi's result holds with countably infinitely many individuals.

CHAPTER V

Utilitarianism in the Infinite Time Horizon Setting

5.1 Introduction

In this chapter, we study utilitarian aggregation in the infinite time horizon setting. We consider a natural, dynamic setting for which time is discrete and infinite. Therefore, any Pareto condition involves countably infinitely many individuals, as there are infinitely many future generations of individuals.

In order to keep our setting as general as possible, the population and the set of consumption goods may differ across generations in order to capture the special cases of time dependent resources and technology. We show that under some additional, mild assumptions, Harsanyi's utilitarianism theorem [6] can be extended to our setting.

In [6], Harsanyi requires that individual preferences and the social preference have expected utility representations. In our setting, a natural analogue will be that individual preferences and the social planner's preference can be represented by discounted, expected utility functions. However, to again preserve the generality of our results, we do not require that the discounted utility functions have exponential discount functions, and we do not require that future-generation individuals' discounted utility functions be related to past generations'.

We first show that Harsanyi's [6] utilitarianism theorem can be extended to our setting. In particular, we provide two proofs of our main theorem regarding the extension of Harsanyi's result. In the first proof, we make use of Theorem IV.8 and measure theoretic results concerning the additivity of measures. Put concretely, when there are countably infinitely many individuals, the Pareto condition still implies that the planner's utility function is a linear functional of individuals' utility functions.

This linear functional is represented by the integration of individuals' utility with respect to a finitely additive measure. To have utilitarian aggregation, the measure must be countably additive. In the second proof, we make use of the function analytic result of the generalized Farka's lemma and the theory concerning positive operators on ordered, Banach vector spaces. The generalized Farka's lemma allows us to treat the determination of utilitarian weights as a linear program, and the Pareto condition will imply that this linear program as a solution.

Both proofs rely on mainly two assumptions. First, the social planner and individuals have discounted utility functions, which is a quite general assumption for which most utility functions in dynamic, economic models would satisfy. Second, for each consumption sequence, individuals' discounted utility functions are uniformly bounded. This is almost a necessary condition, because if there exists some consumption sequence that leads to infinite utility for infinitely distant future generations, the utilitarian aggregation may not be well defined. Additionally, we also provide a new condition that allows the utilitarian weights to be uniquely determined. This is done by taking a classic affine independence assumption from static settings—that ensures that the utilitarian weights are uniquely determined in those settings—and extending it to our setting.

After extending the result of Harsanyi [6] using the normatively appealing Pareto condition and establishing that the utilitarian weights are uniquely determined, we then naturally analyze the intergenerational and asymptotic properties of the utilitarian weights. These analyses might be particularly useful in understanding, for example, inequality or fairness based on the utilitarian weights implied by a social discounted utility function. We show that asymptotically, future-generation individuals' utilitarian weights diminish exponentially at a rate equal to the social discount rate. Therefore, roughly speaking, a higher social discount rate is associated with a more unequal assignment of utilitarian weights across generations.

In addition, we illustrate some counter-intuitive properties of the utilitarian weights in a simple setting. We show that when the social discount factor converges to the discount factor of one family of individuals (who share the same discount function and same instantaneous utility function), but the social risk attitude converges to the risk attitude of a different family of individuals, only the utilitarian weights of the former family converge to zero, regardless of the relative speed of convergence.

We remark that several papers have extended [6] to a setting with infinitely many individuals, and this work will largely build on the aggregation results of [84]. In [84], Zhou allows for an arbitrary number of individuals and shows that the Pareto condition holds if and only if the planner's expected utility function is a linear functional of individuals' expected utility functions. The linear functional, however, does not always take the form of utilitarian aggregation, in which case we may not be able to analyze the properties of utilitarian weights. Similar types of results are established by [104], although they further allow individuals' and the planner's preferences to violate completeness and continuity. We will explain below how our approach avoids this issue. In [105], the set of individuals is a probability space with the number of individuals being countably infinite but the measure being nonatomic and finitely additive—defined on the power set of the set of individuals. Therefore, this is also different from the utilitarian aggregation in our case.

The static version of our assumption that ensures the uniqueness of utilitarian weights would assume that individuals' expected utility functions are affinely independent. This affine independence assumption is used in [106] and shown in [107] to be equivalent to the independent prospects condition introduced by [108] and later used by [109] and [110].

The work discussed in this chapter was done in collaboration with Shaowei Ke and Tangren Feng and subsequently published in the Journal of Mathematical Economics [111].

5.2 Preferences

In view of utilitarianism, it is natural to postulate a social planning, decision maker who must make dynamic decisions, and at each time step, they consider the individuals in their society at both current and future times or generations.

In order to fix mathematical ideas, let $\mathbf{T} := \{1, \ldots, T\}$ denote the set of generations/periods, in which $1 \leq T \leq +\infty$. In each generation $t \in \mathbf{T}$, $1 \leq N_t < +\infty$ individuals live for one period and consume a public consumption good; we remark that in similar dynamic settings, economists typically assume that each N_t is either finite or compact. We assume that N_t is finite so that our results can be directly compared with [6]. Let $\mathbf{N}_t := \{1, \ldots, N_t\}$. The set of period-$t$ public consumption goods is $\Delta(X_t)$, in which $\Delta(X_t)$ is the set of Borel probability measures on a compact metric space X_t. This assumption covers the case in which each individual has his own consumption. We only need to view each period-t public consumption good as an N_t-tuple of individual consumption, and let each individual care only about his own component. A typical consumption sequence is denoted by $\mathbf{p} = (p_1, \ldots, p_T) \in \mathbf{P} := \prod_{t \in \mathbf{T}} \Delta(X_t)$.[1]

[1] See [112] for a discussion of an alternative recursive choice domain.

Each $\Delta(X_t)$ is endowed with the topology of weak convergence, and $\mathbf{P}$ is endowed with the product topology. We identify a degenerate probability measure in $\Delta(X_t)$ that assigns probability 1 to $x_t \in X_t$ with x_t. We use $\mathbf{x} = (x_1, \ldots, x_T)$ to denote a consumption sequence that only consists of such degenerate probability measures.

An individual lives for only one period, but he altruistically cares about future generations' consumption. Each generation-t individual i, $t \in \mathbf{T}$ and $i \in \mathbf{N}_t$, has a preference $\succsim_{i,t}$ over consumption sequences. Let $\mathbf{T}_0 := \{0, \ldots, T-1\}$. We assume that each $\succsim_{i,t}$ is represented by some continuous *discounted utility* function:

$$U_{i,t}(\mathbf{p}) = \sum_{\tau=t}^{T} \delta_{i,t}(\tau - t) u_{i,t}(p_\tau, \tau), \tag{5.1}$$

in which $\delta_{i,t} : \mathbf{T}_0 \to \mathbb{R}_+$ with $\delta_{i,t}(0) = 1$ is his *discount function*, and the continuous expected utility function $u_{i,t}(\cdot, \tau) : \Delta(X_\tau) \to \mathbb{R}$ is his *instantaneous utility function* for period-τ consumption.[2] When $T = +\infty$, we require $\{\delta_{i,t}(\tau)\}_{\tau=0}^{\infty} \in \ell^1$ and for any $\mathbf{p} \in \mathbf{P}$, $\{u_{i,t}(p_\tau, \tau)\}_{\tau=t}^{\infty} \in \ell^\infty$; that is, the former is an absolutely summable sequence and the latter is a bounded sequence.

In each period $t \in \mathbf{T}$, there is a social planner who has a preference $\succsim_t$ over consumption sequences. We assume that her preference $\succsim_t$ is also represented by some continuous discounted utility function:

$$U_t(\mathbf{p}) = \sum_{\tau=t}^{T} \delta_t(\tau - t) u_t(p_\tau, \tau), \tag{5.2}$$

in which $\delta_t : \mathbf{T}_0 \to \mathbb{R}_+$ with $\delta_t(0) = 1$ is her discount function, and the continuous expected utility function $u_t(\cdot, \tau) : \Delta(X_\tau) \to \mathbb{R}$ is her instantaneous utility function for period-τ consumption.

Similarly, when $T = +\infty$, we require that $\{\delta_t(\tau)\}_{\tau=0}^{\infty} \in \ell^1$ and for any $\mathbf{p} \in \mathbf{P}$, $\{u_t(p_\tau, \tau)\}_{\tau=t}^{\infty} \in \ell^\infty$. We assume that for any $\tau \in \mathbf{T}$, there exists some $x_\tau^h, x_\tau^l \in X_\tau$ such that

$$u_{i,t}(x_\tau^l, \tau) = u_t(x_\tau^l, \tau) = 0 = 1 - u_{i,t}(x_\tau^h, \tau) = 1 - u_t(x_\tau^h, \tau) \tag{5.3}$$

and $u_{i,t}(\cdot, \tau), u_t(\cdot, \tau) \geq 0$ for any $t \in \mathbf{T}$ and $i \in \mathbf{N}_t$.

Assumption (5.3) helps us rule out uninteresting cases (e.g., see the discussion

[2] We assume that $U_{i,t}(\mathbf{p})$ does not depend on past consumption. We cannot observe an individual choosing between consumption streams with different past consumption; that is, there is no revealed-preference foundation for utility over past consumption.

after Definition V.1) and simplify proofs. It says that there exist some public goods in each period whose utility would be 0 and 1 for all individuals and the planner, respectively. A similar but weaker assumption is used in [84].

Then, we further assume that x_τ^l is unanimously the worst period-τ public good, such as the extinction of the τth generation. However, we do not assume that x_τ^h is unanimously the best public good in each period. This allows for the possibility that, for example, current-generation individuals believe that the best public goods for future generations are strictly better than theirs due to technological advancements.

In [112], the authors introduce a Pareto condition to the setting with multiple generations: *intergenerational Pareto*. It is shown that intergenerational Pareto is not only useful in avoiding the impossibility results in [113] and [114], but also in understanding which social discount rates are reasonable. The idea of intergenerational Pareto is simple: Because the planner's decision affects both current- and future-generation individuals—and how the current generation thinks about the future may well differ from how future generations will think—the planner should take the current generation as well as the future generations into account when aggregating individual preferences. Intergenerational Pareto is defined as follows:

Definition V.1. The planner's preference $\{\succsim_t\}_{t\in\mathbf{T}}$ is intergenerationally Pareto if fixing any $\mathbf{p}, \mathbf{q} \in \mathbf{P}$ and $t \in \mathbf{T}$, $\mathbf{p} \succsim_{i,s} \mathbf{q}$ for every $s \in \{t, \dots, T\}$ and $i \in \mathbf{N}_s$ implies $\mathbf{p} \succsim_t \mathbf{q}$.

Intergenerational Pareto says that in any period t, if all current- and future-generation individuals prefer a consumption sequence $\mathbf{p}$ to another sequence $\mathbf{q}$, the planner should agree. Note that if the planner is always indifferent, intergenerational Pareto holds trivially, but due to (5.3), this uninteresting case is ruled out.

5.3 Unique utilitarianism with $T < +\infty$

In [6], Harsanyi points out that if individuals' and the planner's preferences have expected utility representations, the Pareto condition is equivalent to utilitarianism. This equivalence is established under the assumption that the number of individuals is finite. Since (5.1) and (5.2) are expected utility functions defined on $\mathbf{P}$, if T is finite, Harsanyi's finding applies to our setting [3].

[3] The mixture operation defined on $\mathbf{P}$ is as follows: For any $\mathbf{p}, \mathbf{q} \in \mathbf{P}$ and $\lambda \in [0,1]$, $\lambda\mathbf{p}+(1-\lambda)\mathbf{q} := (\lambda p_1 + (1-\lambda)q_1, \dots, \lambda p_T + (1-\lambda)q_T)$. Each $\lambda p_t + (1-\lambda)q_t$ is a standard mixture of probability measures in expected utility theory.

Theorem V.2. *(Harsanyi; 1955, [6]) Suppose $T < +\infty$. The planner's preference $\{\succsim_t\}_{t \in \mathbf{T}}$ is intergenerationally Pareto if and only if for any $t \in \mathbf{T}$, there exists a finite sequence of nonnegative numbers $\{\omega_t(i,s)\}_{s \in \{t,\dots,T\}, i \in \mathbf{N}_s}$ such that*

$$U_t = \sum_{s=t}^{T} \sum_{i=1}^{N_s} \omega_t(i,s) U_{i,s}. \tag{5.4}$$

Throughout the remainder of this work, if (5.4) holds, whether T is finite or not, we call the nonnegative numbers $\{\omega_t(i,s)\}_{s \in \{t,\dots,T\}, i \in \mathbf{N}_s}$ *utilitarian weights.*

Due to (5.3), (5.4) implies that the current period utilitarian weights satisfy recursive conditions between future period utilitarian weights and the discount function $\delta_t(\cdot)$. For example, if we consider the consumption sequence $(x_1^l \dots, x_{t-1}^l, x_t^h, x_{t+1}^l, x_{t+2}^l, \dots, x_T^l)$ for (5.4), we immediately have $\sum_{i=1}^{N_t} \omega_t(i,t) = 1$. Moreover, if we consider the consumption sequence $(x_1^l, \dots, x_t^l, x_{t+1}^h, x_{t+2}^l, x_{t+3}^l, \dots, x_T^l)$ for (5.4), we have $\sum_{i=1}^{N_t} \omega_t(i,t)\delta_{t,i}(1) + \sum_{i=1}^{N_{t+1}} \omega_t(i,t+1) = \delta_t(1)$; and so on. The same holds for our main theorem—Theorem V.5—as the argument is identical.

Due to the level of generality for our setting, Theorem V.2 covers many useful special cases, such as the case in which individual and social discount functions are exponential, the case in which future generations inherit the current generation's preferences, the case in which the population may grow or shrink, the case in which technology (the set of consumption goods) may change, etc.

However, Theorem V.2 is silent about the uniqueness of the utilitarian weights. If they are not unique, it may be difficult to analyze the asymptotic and intergenerational properties of the utilitarian aggregation. Although some assumptions have been introduced in the literature to allow the utilitarian weights to be determined uniquely, we introduce a new assumption below that is easy to understand and use in our dynamic setting.

Definition V.3. We say that an n-tuple of real-valued functions defined on the same domain, $(f_1, \dots, f_n)$, is *linearly independent* if for any $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, $\sum_{j=1}^{n} \lambda_j f_j = 0$ implies $\lambda_j = 0$, $j = 1, \dots, n$.

The following result shows that when some of individuals' instantaneous utility functions are linearly independent, utilitarian weights can be uniquely determined.

Proposition V.4 (Feng, Ke, and McMillan; 2022, [111])**.** *Suppose $T < +\infty$ and for any $t \in \mathbf{T}$, there exists a finite sequence of nonnegative numbers $\{\omega_t(i,s)\}_{s \in \{t,\dots,T\}, i \in \mathbf{N}_s}$*

such that

$$U_t = \sum_{s=t}^{T} \sum_{i=1}^{N_s} \omega_t(i,s) U_{i,s}.$$

Fixing any $t \in \mathbf{T}$*, if* $\{u_{i,s}(\cdot, s)\}_{i\in\mathbf{N}_s}$ *is linearly independent for every* $s \in \{t, \ldots, T\}$*,* $\{\omega_t(i,s)\}_{s\in\{t,\ldots,T\}, i\in\mathbf{N}_s}$ *is unique.*

Proof. Fix an arbitrary $t \in \mathbf{T}$. Since $U_t = \sum_{s=t}^{T} \sum_{i=1}^{N_s} \omega_t(i,s) U_{i,s}$, for any $\mathbf{p} \in \mathbf{P}$,

$$\begin{aligned} \sum_{\tau=t}^{T} \delta_t(\tau - t) u_t(p_\tau, \tau) &= \sum_{s=t}^{T} \sum_{i=1}^{N_s} \omega_t(i,s) \left(\sum_{\tau=s}^{T} \delta_{i,s}(\tau - s) u_{i,s}(p_\tau, \tau) \right) \qquad (5.5) \\ &= \sum_{\tau=t}^{T} \left(\sum_{s=t}^{\tau} \sum_{i=1}^{N_s} \omega_t(i,s) \delta_{i,s}(\tau - s) u_{i,s}(p_\tau, \tau) \right). \end{aligned}$$

We prove the uniqueness of $\{\omega_t(i,s)\}_{s\in\{t,\ldots,T\}, i\in\mathbf{N}_s}$ by induction. First, consider any $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq t$. Then, (5.5) becomes

$$u_t(p_t, t) = \sum_{i=1}^{N_t} \omega_t(i,t) u_{i,t}(p_t, t)$$

for any $p_t \in \Delta(X_t)$. Since $\{u_{i,t}(\cdot, t)\}_{i\in\mathbf{N}_t}$ is linearly independent, $\{\omega_t(i,t)\}_{i\in\mathbf{N}_t}$ is unique.

Next, suppose for some fixed $s \geq t+1$, $\{\omega_t(i,\hat{s})\}_{\hat{s}\in\{t,\ldots,s-1\}, i\in\mathbf{N}_{\hat{s}}}$ is uniquely determined. Consider another $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq s$. Then, (5.5) implies

$$\begin{aligned} \delta_t(s-t) u_t(p_s, s) &= \sum_{\hat{s}=t}^{s} \sum_{i=1}^{N_{\hat{s}}} \omega_t(i,\hat{s}) \delta_{i,\hat{s}}(s - \hat{s}) u_{i,\hat{s}}(p_s, s) \\ \sum_{i=1}^{N_s} \omega_t(i,s) u_{i,s}(p_s, s) &= \delta_t(s-t) u_t(p_s, s) - \sum_{\hat{s}=t}^{s-1} \sum_{i=1}^{N_{\hat{s}}} \omega_t(i,\hat{s}) \delta_{i,\hat{s}}(s - \hat{s}) u_{i,\hat{s}}(p_s, s). \end{aligned}$$

Since $\{\omega_t(i,\hat{s})\}_{\hat{s}\in\{t,\ldots,s-1\}, i\in\mathbf{N}_{\hat{s}}}$ is uniquely determined, the right-hand side of the equation above is uniquely determined. Hence, $\{\omega_t(i,s)\}_{i\in\mathbf{N}_s}$ is uniquely determined because $(u_{i,s}(\cdot, s))_{i\in\mathbf{N}_s}$ is linearly independent. Therefore, we know that $\{\omega_t(i,s)\}_{s\in\{t,\ldots,T\}, i\in\mathbf{N}_s}$ is unique. □

To have unique utilitarian weights in every period, Proposition V.4 only requires that $\{u_{i,t}(\cdot, t)\}_{i\in\mathbf{N}_t}$ be linearly independent for every $t \in \mathbf{T}$; that is, for each generation of individuals, their instantaneous utility functions for their own consumption are linearly independent. It does not require that, for example, $\{u_{i,t}(\cdot, \tau)\}_{i\in\mathbf{N}_t}$ be linearly independent when $\tau \neq t$.

Static versions of the linear independence assumption in Proposition V.4 have already appeared in the literature; see [108], [107], [106], [109], and [110]). In [106], Mongin assumes affine independence rather than linear independence, but due to (5.3), affine independence reduces to linear independence in our setting.

There are two ways to understand the linear independence assumption. First, similar to the intuition in [108] and [106], $\{u_{i,t}(\cdot, t)\}_{i\in\mathbf{N}_t}$ is linearly independent if and only if these functions can be separated from each other by some period-t consumption.[4] Second, if we think of $\{u_{i,t}(\cdot, t)\}_{i\in\mathbf{N}_t}$ as being drawn randomly from the space of continuous expected utility functions defined on $\Delta(X_t)$ such that the size of X_t is larger than N_t, generically $\{u_{i,t}(\cdot, t)\}_{i\in\mathbf{N}_t}$ should be linearly independent.

5.4 Unique utilitarianism with $T = \infty$

In many economic models, the time horizon is infinite. When $T = +\infty$, intergenerational Pareto requires that the planner aggregate countably infinitely many individuals' preferences, in which case Harsanyi's (1955) utilitarianism theorem no longer applies. The difficulty of generalizing Harsanyi's utilitarianism theorem to the case with countably infinitely many individuals (without assuming a nonatomic measure) can be seen from Zhou's Theorems 1 and 2 [84]. The Pareto condition along with a mild boundedness assumption only implies that the planner's expected utility function is a linear functional of individuals' expected utility functions, which is represented by an integration of individuals' utility with respect to a finitely additive measure. For the linear functional to take the form of utilitarian aggregation, which necessitates a countably additive measure, additional assumptions are needed.

Intuitively, one possibility to rule out linear functionals that cannot be written as utilitarian aggregations is to impose further restrictions on individuals' and the planner's utility functions. Our assumptions about their utility functions, which are discounted utility functions, already provide more structure than [84]. The result below shows that those assumptions are almost sufficient. Together with a boundedness assumption, they ensure that the measure is countably additive and hence Harsanyi's [6] utilitarianism theorem continues to hold in our setting.

[4] The independent prospects condition used in some of these papers is equivalent to assuming linear independence (see the proof in [108]), which states that for each $i \in \mathbf{N}_t$, we can find some $p_t^i, q_t^i \in \Delta(X_t)$ such that $u_{i,t}(p_t^i, t) \neq u_{i,t}(q_t^i, t)$ and $u_{i,t}\left(p_t^j, t\right) = u_{i,t}\left(q_t^j, t\right)$ for any $j \neq i$.

Theorem V.5 (Feng, Ke, and McMillan; 2022, [111]). *Suppose $T = +\infty$ and for each $\mathbf{p} \in \mathbf{P}$, $\sup_{i,s} U_{i,s}(\mathbf{p}) < +\infty$. The planner's preference $\{\succsim_t\}_{t\in\mathbf{T}}$ is intergenerationally Pareto if and only if for any $t \in \mathbf{T}$, there exists a sequence of nonnegative numbers $\{\omega_t(i,s)\}_{s\geq t, i\in\mathbf{N}_s} \in \ell^1$ such that*

$$U_t = \sum_{s=t}^{\infty} \sum_{i=1}^{N_s} \omega_t(i,s) U_{i,s}. \tag{5.6}$$

Fixing any $t \in \mathbf{T}$, if in addition, $\{u_{i,s}(\cdot, s)\}_{i\in\mathbf{N}_s}$ is linearly independent for every $s \geq t$, $\{\omega_t(i,s)\}_{s\geq t, i\in\mathbf{N}_s}$ is unique.

A new assumption introduced in Theorem V.5 requires that for each $\mathbf{p} \in \mathbf{P}$, $\sup_{i,s} U_{i,s}(\mathbf{p}) < +\infty$. This upper bound assumption is uniform across generations and individuals, but not across consumption sequences. It is a sufficient condition for (5.6) to be well defined, and it is almost necessary—if this assumption does not hold, the right-hand side of (5.6) may diverge. In fact, we have the following result that shows the pointwise bound assumption is tight in our setting.

Proposition V.6 (Feng, Ke, and McMillan; 2022, [111]). *Suppose the following assumptions hold: $X_t = X = [0,2]$, $N_t = N = \{1,2\}$, $\delta_t(\tau - t) = \delta^{\tau-t}$, $\delta_{i,t}(\tau - t) = \delta_i^{\tau-t}$, $u_{i,s}(p_\tau, t) = \int_{[0,2]} x^{i+s} dp_\tau(x)$, and $u_t(p_\tau, \tau) = \alpha_1 u_{1,t}(p_\tau, \tau) + \alpha_2 u_{2,t}(p_\tau, \tau)$ with $\alpha_1 + \alpha_2 = 1$. Then, Theorem V.5 fails.*

Proof. It is clear that the above $u_{i,s}(\cdot, \tau)$ functions are continuous expected utility functions on $\Delta(X)$, which are linearly independent in each period. Moreover, there also exist x_τ^l and x_τ^h such that $u_{i,s}(x_\tau^l, \tau) = u_t(x_\tau^l, \tau) = 0 = 1 - u_{i,s}(x_\tau^h, \tau) = 1 - u_t(x_\tau^h, \tau)$ by taking x_τ^l and x_τ^h to be Dirac masses centered at zero and one, respectively.

Clearly, the pointwise bound assumption is violated under the assumptions, but all other hypotheses of Theorem V.5 are satisfied. By way of contradiction, suppose that Theorem 1's conclusion holds. Then there exists a unique, nonnegative sequence $\{\omega_t(i,s)\}_{s\geq t, i\in\mathbf{N}} \in \ell^1$ such that

$$U_t(\mathbf{p}) = \sum_{s=t}^{\infty} \sum_{i=1}^{2} w_t(i,s) U_{i,s}(\mathbf{p}).$$

A lower bound on $U_t(\mathbf{p})$ can be established as follows:

$$\begin{aligned} U_t(\mathbf{p}) &= \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)U_{i,s}(\mathbf{p}) \\ &= \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)\sum_{\tau=s}^{\infty}\delta_i^{\tau-s}u_{i,s}(p_\tau,\tau) \\ &= \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)\delta_i^{s-s}u_{i,s}(p_s,s) + \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)\sum_{\tau=s+1}^{\infty}\delta_i^{\tau-s}u_{i,s}(p_\tau,\tau) \\ &\geq \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)u_{i,s}(p_s,s). \end{aligned}$$

Choosing $\mathbf{p} \in \mathbf{P}$ such that p_s is the Dirac mass centered at 2 for $s \geq t$ yields

$$\begin{aligned} U_t(\mathbf{p}) &\geq \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)2^{i+s} \\ &\geq \sum_{s=t}^{\infty}\sum_{i=1}^{2} w_t(i,s)2^{s}. \end{aligned}$$

Similar to Theorem V.15, we have that $w_t(i,s) = \alpha_i$ if $s = t$, and we can establish a summed recursive relationship for the $w_t(i,s)$ weights of the following form:

$$\sum_{i=1}^{2} w_t(i,s) = \sum_{i=1}^{2}\sum_{\tau=t}^{s-1}[\delta\delta_{i,\tau}(s-1-\tau) - \delta_{i,\tau}(s-\tau)]w_t(i,\tau) \quad \text{for } s > t.$$

Finally, we show that the above series diverges via the ratio test and hence $U_t(\mathbf{p})$ also diverges, which breeds a contradiction.

$$\begin{aligned} \liminf_{s\to\infty} \frac{2^{s+1}\sum_{i=1}^{2} w_t(i,s+1)}{2^s\sum_{i=1}^{2} w_t(i,s)} &= \liminf_{s\to\infty} \frac{2\sum_{i=1}^{2}\sum_{\tau=t}^{s}[\delta\delta_i^{s-\tau} - \delta_i^{s+1-\tau}]w_t(i,\tau)}{\sum_{i=1}^{2}\sum_{\tau=t}^{s-1}[\delta\delta_i^{s-1-\tau} - \delta_i^{s-\tau}]w_t(i,\tau)} \\ &\geq \liminf_{s\to\infty} \frac{2\sum_{i=1}^{2}\sum_{\tau=t}^{s-1}[\delta\delta_i^{s-\tau} - \delta_i^{s+1-\tau}]w_t(i,\tau)}{\sum_{i=1}^{2}\sum_{\tau=t}^{s-1}[\delta\delta_i^{s-1-\tau} - \delta_i^{s-\tau}]w_t(i,\tau)} \\ &= \liminf_{s\to\infty} \frac{2\sum_{i=1}^{2}\sum_{\tau=t}^{s-1}[\delta - \delta_i]\delta_i^{s-\tau}w_t(i,\tau)}{\sum_{i=1}^{2}\sum_{\tau=t}^{s-1}[\delta\delta_i^{-1} - 1]\delta_i^{s-\tau}w_t(i,\tau)}, \end{aligned}$$

in which the limit is strictly greater than 1 provided that $\delta - \delta_i > \frac{1}{2}(\frac{\delta}{\delta_i} - 1)$; this holds, for example, if $\delta = .9$ and $\delta_i \in (.5, .9)$. Hence, we have a contradiction. $\square$

We provide two proofs of Theorem V.5 of varying abstraction in §5.4.1 and §5.4.2

in order to uncover the underlying machinery.

5.4.1 Proof 1 of Theorem V.5

We first introduce some terminology and facts that will be useful in the proof. Let (X, Σ) be a 2-tuple consisting of a set X and an algebra Σ over X.

Definition V.7. A nonnegative, finitely additive measure μ on Σ is called *purely finitely additive* if for every countably additive measure ν on Σ that $0 \leq \nu \leq \mu$ implies $\nu = 0$.

Pure finite additivity is therefore the anti-thesis of countable additivity. Moreover, as the $\{0,1\}$-valued, finitely additive measures over Σ are in one-to-one correspondence with the set of ultrafilters over Σ, it follows that purely finitely additive measures are convex combinations of finitely additive measures with ultrafilter representations. An elementary, yet useful fact is that a purely finitely additive measure over $2^{\mathbb{N}}$ identically vanishes on the finite subsets of $\mathbb{N}$—the set of natural numbers [115].

Fix an arbitrary $t \in \mathbf{T}$. Let $\mathbf{J} := \{\mathbf{j} = (i, s) : i \in \mathbf{N}_s, s \geq t\}$. Note that the individuals' discounted utility functions, $U(\mathbf{j}; \mathbf{p}) := U_{i,s}(\mathbf{p})$, are continuous and bounded functions over $\mathbf{J}$ for every fixed $\mathbf{p} \in \mathbf{P}$, since $\mathbf{J}$ is discrete and there is a pointwise bound across the $U_{i,s}(\mathbf{p})$ functions. Due to Theorem IV.8, there exists a nonnegative linear functional β on $V := \text{span}(\{U_{i,s}\})$ such that

$$U_t = \beta(U_{i,s})$$

Under the assumptions of (5.3), the linear functional β can be extended to $C_b(\mathbf{J})$ by the Krein–Rutman theorem [84], which is a more refined version of the Hahn–Banach theorem. By the Riesz representation theorem, this functional can be represented as a nonnegative, finite, finitely additive measure, μ, on $\mathbf{J}$, so that the planner's preference takes on the form

$$U_t(\mathbf{p}) = \int_{\mathbf{J}} U(\mathbf{j}, \mathbf{p}) d\mu(\mathbf{j}).$$

Therefore, since μ is discretely supported, it suffices to show that μ is countably additive for the planner to have a countably infinite discounted utilitarian representation.

By [116], μ uniquely decomposes into a countably additive and purely finitely additive part, denoted by μ_c and μ_{pf}, respectively, both of which are nonnegative.

By the countable additivity and discrete support of μ_c, we obtain:

$$\begin{aligned}
U_t(\mathbf{p}) &= \int\limits_{\mathbf{J}} U(\mathbf{j},\mathbf{p})d\mu(\mathbf{j}) \\
&= \int\limits_{\mathbf{J}} U(\mathbf{j},\mathbf{p})d\mu_c(\mathbf{j}) + \int\limits_{\mathbf{J}} U(\mathbf{j},\mathbf{p})d\mu_{pf}(\mathbf{j}) \\
&= \sum_{s=t}^{\infty}\sum_{i=1}^{N_s} \mu_c(i,s)\sum_{\tau=s}^{\infty}\delta_{i,s}(\tau-s)u_{i,s}(p_\tau,\tau) + \int\limits_{\mathbf{J}} U(\mathbf{j},\mathbf{p})d\mu_{pf}(\mathbf{j}).
\end{aligned}$$

To see that the right-hand most integral vanishes, let $\mathbf{p} = (p_1, p_2, \ldots) \in \mathbf{P}$ and for any $\tilde{t} \in \mathbf{T}$, let $\mathbf{p}_{\tilde{t}} = \big(p_1, p_2, \ldots, p_{\tilde{t}-1}, p_{\tilde{t}}, x^l_{t+1}, x^l_{t+2}, \ldots\big)$. Note that for any finite $\tilde{t}$, the right-hand most integral vanishes in the expression for $U_t(\mathbf{p}_{\tilde{t}})$ since only finitely many individuals' utility is positive and μ_{pf} is purely finitely additive.

Let $\epsilon > 0$. Since $U(\mathbf{j},\mathbf{p})$s are nonnegative for all $\mathbf{j} \in \mathbf{J}$ and $\mathbf{p} \in \mathbf{P}$, $(\delta_t(\tau))_{\tau=0}^{\infty} \in \ell^1$, and for any $\mathbf{p} \in \mathbf{P}$, $(u_t(p_\tau,\tau))_{\tau=t}^{\infty} \in \ell^\infty$, we can find t^* such that

$$\begin{aligned}
\epsilon &> \big|U_t(\mathbf{p}) - U_t(\mathbf{p}_{t^*})\big| \\
&\geq \sum_{s=t}^{\infty}\sum_{i=1}^{N_s} \mu_c(i,s)\sum_{\tau=t^*+1}^{\infty}\delta_{i,s}(\tau-s)u_{i,s}(p_\tau,\tau) + \int\limits_{\mathbf{J}} U(\mathbf{j},\mathbf{p})d\mu_{pf}(\mathbf{j}) \\
&\geq \int\limits_{\mathbf{J}} U(\mathbf{j},\mathbf{p})d\mu_{pf}(\mathbf{j}) \\
&\geq 0.
\end{aligned}$$

Since ϵ was arbitrary, the result follows.

There are a couple of remarks in order in view of the first proof for Theorem V.5. The proof has the advantage of being concise and each step is mathematically simple. The key ideas are simply that the intergenerational Pareto condition allow the social planner's utility function to be represented as a linear functional on the individual utility functions due to [84], and since both the social planner's and individual's utility functions separate points, the underlying linear functional is actually integration against a measure. Moreover, due to the individual utility functions being additively separable, this measure must be countably additive, which is intuitive enough because otherwise the utility functions would be primarily supported on a finite number of temporal coordinates.

However, the proof is unfortunately somewhat opaque from an economics perspective. The reason is that our proof crucially depends on the characterization that finitely additive measures have a unique decomposition into a countably additive and purely finitely additive part due to [116], and purely finitely additive measures are rather pathological objects. That is, they are convex combinations of finitely additive measures with ultrafilter representations, which are highly non-constructive mathematical objects. From a mathematical perspective, if one is willing to swallow the axiom of choice, there is little to be criticized here. However, from an economics perspective, this non-constructive property is of limited practical interest. In particular, one is unable to write down an example and see *why* the purely finitely additive part should vanish for a utilitarian aggregator.

5.4.2 Proof 2 of Theorem V.5

Our second proof of Theorem V.5 has the advantage of being transparent, but it is unfortunately significantly more difficult and relies on deeper function analytic machinery. We make use of the generalization of Farka's lemma for dual pairs due to [117], and to state the lemma, we first introduce some definitions.

Definition V.8. A *dual pair* is a 3-tuple (A, A', ϕ) consisting of two vector spaces A and A' and a function $\phi : A \times A' \to \mathbb{R}$ such that (i) ϕ is bilinear, (ii) if $\phi(a, a') = 0$ for every $a \in A$, then $a' = 0$, and (iii) if $\phi(a, a') = 0$ for every $a' \in A'$, then $a = 0$.

Properties (ii) and (iii) are called the separation properties; see [117] and [115].

Definition V.9. A nonempty subset $S \subset A$ is a *convex cone* if $\alpha a + \beta b \in S$ for any $\alpha, \beta \geq 0$ and $a, b \in S$.

Definition V.10. The *anticone* S' of a convex cone S is defined by $S' := \{a' \in A' \mid \phi(a, a') \geq 0 \, \forall a \in S\}$.

Definition V.11. Suppose (A, A', ϕ) and (B, B', φ) are dual pairs. If both A and B are equipped with a norm, a map $\psi : A \to B$ is *strongly continuous* if it is a continuous map between the topologies generated by the norms on A and B. A map $\psi : A \to B$ is *weakly* continuous if it is continuous map between the weak topologies on A and B.

Definition V.12. The *adjoint* of a weakly continuous linear map $\psi : A \to B$ is defined as the map $\psi' : B' \to A'$ that satisfies

$$\phi(a, \psi'(b')) = \varphi(\psi(a), b')$$

for any $a \in A$ and $b' \in B'$.

We state Craven and Koliha's generalized Farkas' lemma below.

Theorem V.13. *(Craven and Koliha; 1977, [117]) Let (A, A', ϕ) and (B, B', φ) be dual pairs, let S be a convex cone in A, and let $\psi : A \to B$ be a weakly continuous linear map. If $\psi(S)$ is closed in the weak topology and $b \in B$, the following statements are equivalent:*

1. *The equation $\psi(a) = b$ has a solution $a \in S$.*

2. *$\psi'(b') \in S' \Rightarrow \varphi(b, b') \geq 0$.*

We now turn to the second proof of Theorem V.5. Indeed, the "if" part is obvious, and we only prove the "only-if" part.

Recall that the generation-t individual i's utility function is $U_{i,t}(\mathbf{p}) = \sum_{\tau=t}^{\infty} \delta_{i,t}(\tau - t)u_{i,t}(p_\tau, \tau)$ and the social planner's utility function in period t is $U_t(\mathbf{p}) = \sum_{\tau=t}^{\infty} \delta_t(\tau - t)u_t(p_\tau, \tau)$.

Since X_t is compact and metrizable for each $t \in \mathbf{T}$, $\mathbf{X} := \prod_{t=1}^{\infty} X_t$ is also compact and metrizable in the product topology. Let $A = \ell^1, A' = \ell^\infty, B = C_b(\mathbf{X})$, and $B' = ca(\mathbf{X})$, in which ℓ^∞ is the set of bounded sequences, $C_b(\mathbf{X})$ is the set of continuous and bounded functions on $\mathbf{X}$, and $ca(\mathbf{X})$ is the set of countably additive finite signed measures on $\mathbf{X}$. The norm of A' and B is the sup norm, denoted by $||\cdot||_{\ell^\infty}$ and $||\cdot||_{C_b(\mathbf{X})}$, respectively. The norm of B' is the total-variation norm, and for any $\{a_n\}_{n=1}^{\infty} \in A$, its norm is equal to $\sum_{n=1}^{\infty} |a_n|$ and denoted by $||\cdot||_{\ell^1}$.

It is not hard to show that by defining

$$\phi(a, a') = \sum_{n=1}^{\infty} a_n a'_n$$

and

$$\varphi(b, b') = \int_{\mathbf{X}} b db'$$

for any $a \in A, a' \in A', b \in B,$ and $b' \in B'$ that (A, A', ϕ) and (B, B', φ) are dual pairs.

Fix an arbitrary $t \in T$. For any sequence in ℓ^1 denoted by $\vec{\omega}_t = (\omega_t(1,t), \ldots, \omega_t(N_t, t), \omega_t(1, t+1), \ldots, \omega_t(N_{t+1}, t+1), \ldots)$, define an operator $\psi : A \to B$ such that for each $\mathbf{x} \in \mathbf{X}$,

$$\psi(\vec{\omega}_t)(\mathbf{x}) = \sum_{s=t}^{\infty} \sum_{i=1}^{N_s} \omega_t(i, s) U_{i,s}(\mathbf{x})$$

For this operator to be well defined, we need to verify two things. First, since for each fixed $\mathbf{x} \in \mathbf{X}$, we have $\{U_{i,s}(\mathbf{x})\}_{s \geq t, i \in \mathbf{N}_s} \in \ell^\infty$ and therefore $\psi(\vec{\omega_t})(\mathbf{x})$ is well defined.

Second, in order for ψ's codomain to indeed be B, we need to show that for $\vec{\omega_t} \in \ell^1$, $\psi(\vec{\omega_t})$ is a continuous function defined on $\mathbf{X}$. However, this is easily deduced as for each fixed $\mathbf{x} \in \mathbf{X}$, $\{U_{i,s}(\mathbf{x})\}_{s \geq t, i \in \mathbf{N}_s} \in \ell^\infty$, and therefore, finite truncations of $\psi(\vec{\omega_t})$ are continuous and converge monotonically pointwise to $\psi(\vec{\omega_t})$—since $\{U_{i,s}(\mathbf{x})\}_{s \geq t, i \in \mathbf{N}_s}$ is a nonnegative sequence due to assumption (5.3). By Dini's Theorem, the convergence is uniform. Hence, $\psi(\vec{\omega_t})$ is a continuous function defined on $\mathbf{X}$.

Next, we'd like to verify that ψ is weakly continuous. Note that ψ is a linear map between Banach spaces. Moreover, if a linear map between Banach spaces is strongly continuous, it is also weakly continuous [118]. Therefore, it suffices to show that ψ is strongly continuous. However, ψ is a positive operator[5] between Banach lattices[6] due to assumption (5.3), and hence, ψ is strongly continuous [119].

Next, we show that $\psi(S)$ is closed in the weak topology induced by $ca(\mathbf{X})$. Since S is convex and ψ is linear, $\psi(S) \subset C_b(\mathbf{X})$ is convex. By [118], $\psi(S)$ as a convex subset of a normed Banach space is closed in the weak topology if and only if it is closed in the norm topology. Therefore, we only need to show that $\psi(S)$ is closed in the norm topology.

Indeed, take a sequence $\{f_n\}_{n=1}^\infty$ of $\psi(S)$ such that f_n converges to $f \in C_b(\mathbf{X})$ in sup norm. We want to show that $f \in \psi(S)$; that is, there exists some $\vec{\omega_t} \in S$ such that $\psi(\vec{\omega_t}) = f$. Since $f_n \in \psi(S)$, there exists a sequence $\{\vec{\omega_t}^n\}_{n=1}^\infty \in S$ such that for any $n \in \mathbb{N}$ and $\mathbf{x} \in \mathbf{X}$,

$$f_n(\mathbf{x}) = \sum_{s=t}^{\infty} \sum_{i=1}^{N_s} \omega_t^n(i,s) U_{i,s}(\mathbf{x})$$

We first show that $\{\vec{\omega_t}^n\}_{n=1}^\infty \in S$ is uniformly bounded in the ℓ^1 norm. Indeed, since the sequence $\{f_n\}_{n=1}^\infty$ are continuous functions on a compact metric space, sup norm convergence implies that $\{f_n\}_{n=1}^\infty$ is an equicontinuous family. Moreover, define $\mathbf{x}^h$ by

$$\mathbf{x}^h = (x_1^h, x_2^h, \dots) \in \mathbf{X}.$$

[5]An operator $T : A \to B$ between two ordered vector spaces A and B is said to be positive if $a \geq 0 \Rightarrow T(a) \geq 0$.

[6]A Banach lattice X is a partially ordered Banach space with norm $||\cdot||$ and ordering $\geq$ such that for two elements $x, y \in X$, $|x| \leq |y| \Rightarrow ||x|| \leq ||y||$, where $|x| := x \vee -x$ under the usual lattice operations.

We have

$$f_n(\mathbf{x}^h) = \sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t^n(i,s)\sum_{\tau=0}^{\infty}\delta_{i,s}(\tau).$$

Since $f_n(\mathbf{x}^h)$ converges to $f(\mathbf{x}^h)$, we know that there exists some $\rho > 0$ such that $f_n(\mathbf{x}^h) \leq \rho$ for any $n \in \mathbb{N}$, and because $\sum_{\tau=0}^{\infty}\delta_{i,s}(\tau)$'s are greater than 1,

$$\sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t^n(i,s) \leq \rho$$

for any $n \in \mathbb{N}$. Thus, the sequence $\{\vec{\omega_t}^n\}_{n=1}^{\infty}$ is uniformly bounded.

We next wish to show that the sequence $\{\vec{\omega}_t^n\}_{n=1}^{\infty}$ is uniformly summable and admits a convergent subsequence with respect to the ℓ^1 topology. Let $\mathbf{x}^{h\times s} := (x_1^h,\ldots,x_s^h,x_{s+1}^l,x_{s+2}^l,\ldots)$. For any $\mathbf{x},\mathbf{y} \in \mathbf{X}$, the metric

$$d(\mathbf{x},\mathbf{y}) := \sup_{\tau\in\mathbf{T}}\left\{\frac{\min\{||x_\tau - y_\tau||_\tau, 1\}}{\tau}\right\}$$

induces the product topology on $\mathbf{X}$, where $||\cdot||_\tau$ denotes the metric on X_τ. Intuitively, when $\mathbf{x}$ and $\mathbf{y}$ are close, x_τ and y_τ are close when τ is small, but x_τ and y_τ can be far apart when τ is large. Therefore, when s is large, $\mathbf{x}^h$ and $\mathbf{x}^{h\times s}$ are close.

By the equicontinuity of the sequence $\{f_n\}_{n=1}^{\infty}$, for any $\epsilon' > 0$, there exists some $\kappa' > 0$ such that for any $\tilde{\kappa}' \geq \kappa'$ and $n' \in \mathbb{N}$,

$$\epsilon' > |f_{n'}(\mathbf{x}^h) - f_{n'}(\mathbf{x}^{h\times\tilde{\kappa}'})| = \Big|\sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t^{n'}(i,s)U_{i,s}(\mathbf{x}^h) - \sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t^{n'}(i,s)U_{i,s}(\mathbf{x}^{h\times\tilde{\kappa}'})\Big|$$

This shows that for any $\epsilon' > 0$, there exists $s' > 0$ such that for any $\tilde{s}' \geq s'$ and $n' \in \mathbb{N}$,

$$\sum_{s=\tilde{s}'}^{\infty}\sum_{i=1}^{N_s}\omega_t^{n'}(i,s) < \epsilon'$$

Therefore, the sequence $\{\omega_t^n\}_{n=1}^{\infty}$ is uniformly summable. As a consequence of the Frechet–Kolmogorov theorem, a sequence in ℓ^1 is relatively compact if and only if it is uniformly bounded and uniformly summable [120]. Thus, $\{\omega_t^n\}_{n=1}^{\infty}$ is relatively compact. By the Eberlein–Smulian theorem, relative compactness is equivalent to relative sequential compactness, which implies that there exists a convergent subsequence $\{\omega_t^{n_k}\}_{n_k=1}^{\infty}$ that converges to some $\omega_t^* \in \ell^1$.

In addition, since ℓ^1 is a Banach lattice, the positive cone, S, is closed in the ℓ^1 topology [119]. Then, because $\{\omega_t^{n_k}\}_{n_k=1}^{\infty}$ is in S, it follows that $\omega_t^* \in S$. Since f_n

converges to f, this implies that $\sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t^{n_k}U_{i,s}(\mathbf{x})$ converges to $f(\mathbf{x})$ in sup norm. Therefore, we have

$$\sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t^*(i,s)U_{i,s}(\mathbf{x}) = f(\mathbf{x}).$$

This shows that $\psi(S)$ is strongly closed and therefore weakly closed.

Finally, we want to show that the following equation

$$U_t = \sum_{s=t}^{\infty}\sum_{i=1}^{N_s}\omega_t(i,s)U_{i,s} = \psi(\vec{\omega_t}) \tag{5.7}$$

has a nonnegative solution; that is, there exists some $\vec{\omega}_t \in S$ that solves (5.7). For any $\mu \in B'$, $\psi'(\mu) \in S'$ if and only if $\phi(a,\psi'(\mu)) = \varphi(\psi(a),\mu) \geq 0$ for any $a \in S$. In turn, $\varphi(\psi(a),\mu) \geq 0$ for any $a \in S$ implies that

$$\int_{\mathbf{X}} U_{i,s}d\mu \geq 0 \tag{5.8}$$

for any $s \geq t$ and $i \in \mathbf{N}_s$. By Theorem V.13, we know we can find a nonnegative solution $\vec{\omega}_t \in S$ for (5.7) if we can show that for any $\mu \in B'$ such that (5.8) with $s \geq t$ and $i \in \mathbf{N}_s$ implies $\int_X U_t d\mu \geq 0$. A useful fact is that because each $U_{i,s}$ is additively separable in time, we have

$$\int_{\mathbf{X}} U_{i,s}(\mathbf{x})d\mu = U_{i,s}(\mathbf{p}^{\mu}), \tag{5.9}$$

where the components of $\mathbf{p}^{\mu}$ are p_{τ}^{μ}- the marginal distribution of μ on X_{τ}. More precisely, by applying the Fubini–Tonelli theorem, we have

$$\int_{\mathbf{X}} U_{i,s}(\mathbf{x})d\mu = \int_{\mathbf{X}} \Big[\sum_{\tau=s}^{\infty}\delta_{i,s}(\tau-s)u_{i,s}(x_{\tau},\tau)\Big]d\mu = \sum_{\tau=s}^{\infty}\delta_{i,s}(\tau-s)\Big[\int_{\mathbf{X}} u_{i,s}(x_{\tau},\tau)d\mu\Big].$$

By the Hahn–Jordan decomposition theorem, μ can be uniquely decomposed into $a\mu_+ - \beta\mu_-$ in which $\alpha,\beta \geq 0$ and μ_+ and μ_- are some probability measures on $\mathbf{X}$. Therefore, the expression in (5.8) becomes

$$\alpha\int_{\mathbf{X}} U_{i,s}d\mu_+ \geq \beta\int_{\mathbf{X}} U_{i,s}d\mu_-$$

for any $s \geq t$ and $i \in \mathbf{N}_s$. Moreover, just as in (5.9), the probability measures μ_+

and μ_- can be identified with some $\mathbf{p}, \mathbf{q} \in \mathbf{P}$, in which p_τ and q_τ are the marginal distributions of μ_+ and μ_- on X_τ, respectively. Hence, (5.8) becomes

$$\alpha U_{i,s}(\mathbf{p}) \geq \beta U_{i,s}(\mathbf{q})$$

for any $s \geq t$ and $i \in \mathbf{N}_s$. Without loss of generality, suppose that $\alpha \geq \beta$ and denote $\mathbf{x}^l$ as the sequence $(x_1^l, x_2^l, \dots)$. By assumption, we have $U_{i,s}(\mathbf{x}^l) = 0$ for any $s \geq t$ and $i \in \mathbf{N}_s$, and (5.8) becomes

$$U_{i,s}(\mathbf{p}) \geq \frac{\beta}{\alpha} U_{i,s}(\mathbf{q}) + \Big(1 - \frac{\beta}{\alpha}\Big) U_{i,s}(\mathbf{x}^l)$$

for any $s \geq t$ and $i \in \mathbf{N}_s$. Since the $U_{i,s}$'s are time-additively separable, the above equation implies that for every $s \geq t$ and $i \in \mathbf{N}_s$, the generation-s individual i prefers $\mathbf{p}$ to $\frac{\beta}{\alpha}\mathbf{q} + \Big(1 - \frac{\beta}{\alpha}\Big)\mathbf{x}^l$, in which $\frac{\beta}{\alpha}\mathbf{q} + \Big(1 - \frac{\beta}{\alpha}\Big)\mathbf{x}^l \in \mathbf{P}$ as a convex mixture between $\mathbf{q}$ and $\mathbf{x}^l$. By intergenerational Pareto, this implies that

$$\begin{aligned} U_t(\mathbf{p}) &\geq \frac{\beta}{\alpha} U_t(\mathbf{q}) + \Big(1 - \frac{\beta}{\alpha}\Big) U_t(\mathbf{x}^l) \\ \alpha U_t(\mathbf{p}) &\geq \beta U_t(\mathbf{q}) \\ \int_{\mathbf{X}} U_t d\mu &\geq 0. \end{aligned}$$

Therefore, (5.8) has a nonnegative solution. Finally, the uniqueness of the utilitarian weights follows from the same argument as in the proof of Lemma V.4.

The proof is demonstrative from two points of view. First, if we compare our approach with [84] and the first proof of Theorem V.5, we see that we have worked backwards. In [84], one constructs the planner's linear functional on the linear subspace generated by individual's utility functions, and extends the functional to a larger space. Theorem 2 in [84] then obtains the utilitarian representation of the planner's preference via the Riesz representation theorem, because the (larger) space is the space of continuous functions defined on a compact metrizable set of individuals.

This technique fails in our setting, because our corresponding space is the space of continuous functions over $\mathbb{N}$. Such an extension may have the desired utilitarian representation, but theoretical guarantees can only be achieved if the extended functional is continuous with respect to the strict topology in the sense of [121], and there is no such guarantee in Zhou's setting. Our result works backwards in the sense that instead of constructing and extending a functional and invoking the Riesz representation theorem to obtain the utilitarian representation, we start with the utilitarian representation, which can be thought of as an uncountable set of potential planners, and show that the generalized Farka's lemma, paired with the Pareto condition and the linear independence assumption, allows us to comb away all members of this set except for one.

One may be left curious as to why the utilitarian representation works in such a pathological space as ℓ^1. Without reflexivity, compactness is difficult to recover, which is often at the heart of representation results using separation arguments. This brings us to the second point. For a generic set of individuals' utility functions, there does not seem to be a way to perform this combing procedure, as the ℓ^1 space is too large, and reducing the number of potential planners seems impossible even via the generalized Farka's lemma, unless we have certain closedness and compactness conditions. It turns out that we can recover those conditions in our setting. As can be seen in the proof of the closedness of S, the space of discounted utility functions allows us to control the tail end of positive sequences in ℓ^1, which grants uniform control over such sequences. In particular, one is allowed to recover the desired compactness of a subset of ℓ^1.

5.5 Comparative static analyses of the utilitarian weights

Being able to determine utilitarian weights uniquely is crucial to the analysis of utilitarian weights' properties. In order to motivate and gain intuition for our subsequent results, consider the following example. Suppose for any $t \in \mathbf{T}$, $N_t = 2$, $X_t = [0, 1]$, and

$$U_{1,t}(\mathbf{p}) = \sum_{\tau=t}^{T} 0.8^{\tau-t} u_1(p_\tau) \text{ and } U_{2,t}(\mathbf{p}) = \sum_{\tau=t}^{T} 0.9^{\tau-t} u_2(p_\tau)$$

for some linearly independent continuous expected utility functions u_1 and u_2 defined on $\Delta([0, 1])$. Therefore, for any $s, t \in \mathbf{T}$ and $i \in \{1, 2\}$, the generation-t individual i shares the same discount factor and instantaneous utility function with the generation-s individual i. One may think of individuals with the same i as ancestors and descendants who share the same preference parameters. Suppose in period 1, the planner's exponentially discounted utility function is

$$U_1(\mathbf{p}) = \sum_{\tau=t}^{T} \delta^{\tau-t} u_\alpha(p_\tau),$$

in which $\delta \in (0.9, 1)$ and for some $\alpha \in (0, 1)$,

$$u_\alpha = \alpha u_1 + (1-\alpha) u_2. \tag{5.10}$$

According to Theorem V.2, Proposition V.4, and Proposition 3 of [112], U_1 is equal to some utilitarian aggregation of $U_{i,t}$'s, $i \in \{1, 2\}$ and $t \in \mathbf{T}$; that is, for some nonnegative numbers $\omega(i, s)$'s,

$$U_1 = \sum_{s=1}^{T} \sum_{i=1}^{2} \omega(i, s) U_{i,s}.$$

Suppose we want to understand what happens to the individuals' utilitarian weights as the social discount factor δ approaches 0.9 and the social risk attitude parameter αapproaches 1. As δ approaches 0.9, the planner's discount factor is arbitrarily close to individual 2's discount factor. However, as α approaches 1, the planner's instantaneous utility function is arbitrarily close to individual 1's instantaneous utility function. Hence, although it seems that either individual 1's or 2's utilitarian weights will likely converge to 0, it is not obvious which one's will. One may think that this will depend on which parameter converges faster. As will be shown below, the rate of convergence does not matter.

It is difficult to examine this question if the utilitarian weights cannot be uniquely determined. Here, because u_1 and u_2 are linearly independent, by Proposition V.4, we can determine the utilitarian weights uniquely. Consider any consumption sequence with $p_t = x_t^l$ for any $t > 1$. We immediately have for any $p \in \Delta([0, 1])$,

$$\omega(1, 1) u_1(p) + \omega(2, 1) u_2(p) = u_\alpha(p).$$

Because u_1 and u_2 are linearly independent, there is only one way to write u_α as a convex combination of u_1 and u_2. By (5.10),

$$\omega(1, 1) = \alpha \text{ and } \omega(2, 1) = 1 - \alpha.$$

Next, consider another consumption sequence with $p_t = x_t^l$ for any $t \neq 2$. We have for any $p \in \Delta([0, 1])$,

$$\begin{aligned} \delta u_\alpha(p) &= 0.8\omega(1, 1)u_1(p) + 0.9\omega(2, 1)u_2(p) + \omega(1, 2)u_1(p) + \omega(2, 2)u_2(p) \\ u_\alpha(p) &= \frac{1}{\delta}[0.8\omega(1, 1) + \omega(1, 2)]u_1(p) + \frac{1}{\delta}[0.9\omega(2, 1) + \omega(2, 2)]u_2(p), \end{aligned}$$

which implies that

$$\frac{1}{\delta}[0.8\omega(1,1)+\omega(1,2)]=\alpha \text{ and } \frac{1}{\delta}[0.9\omega(2,1)+\omega(2,2)]=1-\alpha.$$

Therefore, $\omega(1,2)=\alpha(\delta-0.8)$ and $\omega(2,2)=(1-\alpha)(\delta-0.9)$. A similar calculation shows that for any $s\geq 2$,

$$\omega(1,s)=\alpha\delta^{s-2}(\delta-0.8) \text{ and } \omega(2,s)=(1-\alpha)\delta^{s-2}(\delta-0.9).$$

It can be seen that first, in any period $s\geq 2$, $\omega(1,s)/\omega(2,s)$ is constant. Second, as α approaches 1 and δ approaches 0.9,

$$\omega(1,s)\not\to 0 \text{ and } \omega(2,s)\to 0 \text{ for any } s\in\mathbf{T},$$

regardless of the rate of convergence of α or δ. In fact, even if α is fixed and we only let the planner's discount factor δ go to 0.9—which is individual 2's discount factor—surprisingly, individual 2's utilitarian weight $\omega(2,s)$ still converges to 0 for any $s\in\{2,\ldots,T\}$. Using the intuition from the above example, we derive a more general result—Theorem V.15, but first, we introduce the following definition:

Definition V.14. For each discount function $\delta_{i,t}(\cdot)$, let $\beta_{i,t}(\tau):=\delta_{i,t}(\tau)/\delta_{i,t}(\tau-1)$, $\tau\in\{1,\ldots,T-1\}$, be its relative discount function. We say that the relative discount function $\beta_{i,t}(\cdot)$ is θ-regular if it is weakly increasing and bounded above by some $\theta\in(0,1)$. We say that a discount function is θ-regular if its relative discount function is θ-regular.

A generic $\beta_{i,t}(\tau)$ will be called a relative discount factor. The relative discount function captures the instantaneous discounting for consumption that is τ periods ahead relative to consumption that is $\tau-1$ periods ahead. Regular discount functions are closely related to (weak) present bias; see [112] for a more detailed discussion. For example, the relative discount function of an exponential discount function is constant, and the relative discount function of a quasi-hyperbolic discount function is strictly increasing between $\tau=1$ and $\tau=2$ and constant afterward. With the above definition and intuition from the example, we have the following theorem:

Theorem V.15 (Feng, Ke, and McMillan; 2022, [111])**.** *Suppose $T=+\infty$, and for some $\mathbf{N}=\{1,\ldots,N\}$ and compact metric space X, $N_t=N$ and $X_t=X$ for every $t\in\mathbf{T}$. Assume that for some $\theta\in(0,1)$, θ-regular discount functions $\delta_{i,t}$s with $t\in\mathbf{T}$ and $i\in\mathbf{N}$, and linearly independent instantaneous utility functions $\{u_i\}_{i\in\mathbf{N}}$,* *individuals' discounted utility functions take the following form:*

$$U_{i,t}(\mathbf{p}) = \sum_{\tau=t}^{T} \delta_{i,t}(\tau - t) u_i(p_\tau, \tau)$$

for any $t \in \mathbf{T}$ and $i \in \mathbf{N}$. Moreover, for some $\delta \in \big(\sup_{i\in\mathbf{N},t\in\mathbf{T},\tau\geq t} \beta_{i,t}(\tau), 1\big)$ and $\alpha_1, \ldots, \alpha_N \in [0,1]$ such that $\sum_{i\in\mathbf{N}} \alpha_i = 1$, we have

$$U_t(\mathbf{p}) = \sum_{\tau=t}^{T} \delta^{\tau-t} u(p_\tau, \tau)$$

for any $t \in \mathbf{T}$, in which $u = \sum_{i\in\mathbf{N}} \alpha_i u_i$. Then, the planner's preference $\{\succsim_t\}_{t\in\mathbf{T}}$ is intergenerationally Pareto if and only if for any $t \in \mathbf{T}$, $s \geq t$, and $i \in \mathbf{N}$, the utilitarian weight satisfies the following recursive formula

$$\omega_t(i,s) = \begin{cases} \alpha_i, & \text{if } s = t, \\ \sum\limits_{\tau=t}^{s-1} [\delta \cdot \delta_{i,\tau}(s-1-\tau) - \delta_{i,\tau}(s-\tau)]\omega_t(i,\tau), & \text{if } s > t. \end{cases} \tag{5.11}$$

Proof. Since $\beta_{i,t}$'s are bounded above by $\theta < 1$, $\sup_{i\in\mathbf{N},t\in\mathbf{T},\tau\geq t} \beta_{i,t}(\tau) \leq \theta$ and hence δ is well defined. The sufficiency part follows from Theorem $S2$ in [112]. To prove the necessity part, suppose the planner's preference $\{\succsim_t\}_{t\in\mathbf{T}}$ is intergenerationally Pareto. Fix any $t \in \mathbf{T}$. By Theorem V.5, there exists a unique sequence of nonnegative numbers $\{\omega_t(i,s)\}_{s\geq t, i\in\mathbf{N}} \in \ell^1$ such that

$$U_t = \sum_{s=t}^{\infty} \sum_{i=1}^{N} \omega_t(i,s) U_{i,s}. \tag{5.12}$$

We only need to verify (5.11).

First, consider any $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq t$. Equation (5.12) implies

$$\sum_{i=1}^{N} \alpha_i u_i(p_t, t) = u(p_t, t) = \sum_{i=1}^{N} \omega_t(i,t) u_i(p_t, t)$$

for any $p_t \in \Delta(X_t)$. Since $\{u_i\}_{i\in\mathbf{N}}$ is linearly independent, we have $\omega_t(i,t) = \alpha_i$. Next, consider another $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq t+1$. Equation (5.12) implies

$$\sum_{i=1}^{N} \delta\alpha_i u_i(p_{t+1}, t+1) = \delta u(p_{t+1}, t+1) = \sum_{i=1}^{N} [\omega_t(i,t)\delta_{i,t}(1) + \omega_t(i,t+1)] u_i(p_{t+1}, t+1)$$

for any $p_{t+1} \in \Delta(X_{t+1})$. Again, since $\{u_i\}_{i\in\mathbf{N}}$ is linearly independent, we have

$$\omega_t(i,t+1) = \delta\alpha_i - \omega_t(i,t)\delta_{i,t}(1) = [\delta - \delta_{i,t}(1)]\omega_t(i,t).$$

Since $\delta > \max_{i\in\mathbf{N}} \delta_{i,t}(1)$, $\omega_t(i,t+1)$ are nonnegative. Then, for any $s > t+1$, by considering any $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq s$, we can follow similar calculations to obtain

$$\omega_t(i,s) = \sum_{\tau=t}^{s-1} [\delta \cdot \delta_{i,\tau}(s-1-\tau) - \delta_{i,\tau}(s-\tau)]\omega_t(i,\tau).$$

Again, $\omega_t(i,s)$'s are nonnegative because $\delta > \sup_{i\in\mathbf{N}, t\in\mathbf{T}, \tau\geq t} \beta_{i,t}(\tau)$. □

One can also recover a generalized version of Theorem V.15 without assuming that the planner's discount function is exponential, which is the following:

Theorem V.16 (Feng, Ke, and McMillan; 2022, [111]). *Suppose for some* $\mathbf{N} = \{1, \ldots, N\}$ *and compact metric space* X*,* $N_t = N$ *and* $X_t = X$ *for every* $t \in \mathbf{T}$*. Assume that for some* $\theta \in (0,1)$*,* θ*-regular discount functions* $\delta_{i,t}$*'s with* $t \in \mathbf{T}$ *and* $i \in \mathbf{N}$*, and linearly independent instantaneous utility functions* $\{u_i\}_{i\in\mathbf{N}}$*, individuals' discounted utility functions take the following form:*

$$U_{i,t}(\mathbf{p}) = \sum_{\tau=t}^{T} \delta_{i,t}(\tau - t) u_i(p_\tau, \tau)$$

for any $t \in \mathbf{T}$ *and* $i \in \mathbf{N}$*. Moreover, for discount functions* δ_t*'s and* $\alpha_1, \ldots, \alpha_N \in [0,1]$ *such that* $\inf_{\tau\geq t} \beta_t(\tau) > \sup_{i\in\mathbf{N}, \tau\geq t} \beta_{i,t}(\tau)$ *for any* $t \in \mathbf{T}$ *and* $\sum_{i\in\mathbf{N}} \alpha_i = 1$*, we have*

$$U_t(\mathbf{p}) = \sum_{\tau=t}^{T} \delta_t(\tau - t) u(p_\tau, \tau)$$

for any $t \in \mathbf{T}$ *in which* $u = \sum_{i\in\mathbf{N}} \alpha_i u_i$*. Then, the planner's preference* $\{\succsim_t\}_{t\in\mathbf{T}}$ *is intergenerationally Pareto if and only if for any* $t \in \mathbf{T}$*,* $s \geq t$*, and* $i \in \mathbf{N}$*, the utilitarian weights satisfy the following recursive formula:*

$$\omega_t(i,s) = \begin{cases} \alpha_i, & \textit{if } s = t, \\ \sum_{\tau=t}^{s-1} \big[\frac{\delta_t(s-t)}{\delta_t(s-t-1)} \cdot \delta_{i,\tau}(s-1-\tau) - \delta_{i,\tau}(s-\tau)\big]\omega_t(i,\tau), & \textit{if } s > t. \end{cases} \tag{5.13}$$

Proof. As in the proof of Theorem V.15, we only prove the necessity part. Suppose the planner's preference $\{\succsim_t\}_{t\in\mathbf{T}}$ is intergenerationally Pareto. Fix any $t \in \mathbf{T}$. By Theorem V.5, there exists a unique nonnegative sequence $\{\omega_t(i,s)\}_{s\geq t, i\in\mathbf{N}} \in \ell^1$ such that

$$U_t = \sum_{s=t}^{\infty}\sum_{i=1}^{N} \omega_t(i,s)U_{i,s}. \tag{5.14}$$

We only need to verify (5.13).

First, consider any $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq t$. Equation (5.14) implies that

$$\sum_{i=1}^{N} \alpha_i u_i(p_t,t) = u(p_t,t) = \sum_{i=1}^{N} \omega_t(i,t)u_i(p_t,t)$$

for any $p_t \in \Delta(X_t)$. Since $\{u_i\}_{i\in\mathbf{N}}$ is linearly independent, we have $\omega_t(i,t) = \alpha_i$. Next, consider another $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq t+1$. Equation (5.14) implies that

$$\sum_{i=1}^{N} \delta_t(1)\alpha_i u_i(p_{t+1},t{+}1) = \delta_t(1)u(p_{t+1},t{+}1) = \sum_{i=1}^{N}[\omega_t(i,t)\delta_{i,t}(1){+}\omega_t(i,t{+}1)]u_i(p_{t+1},t{+}1)$$

for any $p_{t+1} \in \Delta(X_{t+1})$. Again, since $\{u_i\}_{i\in\mathbf{N}}$ is linearly independent, we have

$$\omega_t(i,t+1) = \delta_t(1)\alpha_i - \omega_t(i,t)\delta_{i,t}(1) = \big[\frac{\delta_t(1)}{\delta_t(0)}\delta_{i,t}(0) - \delta_{i,t}(1)\big]\omega_t(i,t).$$

Since $\delta_t(1) = \beta_t(1) > \beta_{i,t}(1) = \delta_{i,t}(1)$, $\omega_t(i,t+1)$ is nonnegative. Then, for any $s > t+1$, by considering any $\mathbf{p} \in \mathbf{P}$ such that $p_\tau = x_\tau^l$ whenever $\tau \neq s$, we can follow similar calculations to obtain that

$$\omega_t(i,s) = \sum_{\tau=t}^{s-1} \big[\frac{\delta_t(s-t)}{\delta_t(s-t-1)} \cdot \delta_{i,\tau}(s-1-\tau) - \delta_{i,\tau}(s-\tau)\big]\omega_t(i,\tau).$$

Again, $\omega_t(i,s)$ is nonnegative because $\inf_{\tau\geq t}\beta_t(\tau) > \sup_{i\in\mathbf{N},\tau\geq t}\beta_{i,t}(\tau)$. □

From Theorem V.15, it is clear that our observations about comparative statics and convergence from the example continue to hold when $T = +\infty$. A surprising finding from the example is that when the planner's discount factor δ converges to individual 2's discount factor, individual 2's utilitarian weight $\omega(2, s)$ turns out to converge to 0 for every $s \in \{2, \dots, T\}$. To understand this, suppose that in Theorem V.15 all individual discount functions are distinct and exponential. It follows that $\omega_t(i, t) = \alpha_i$ and $\omega_t(i, s) = \alpha_i \delta^{s-t-1}(\delta - \delta_i)$ for any $i \in \mathbf{N}$, $t \in \mathbf{T}$, and $s > t$. Therefore, it can be seen that future-generation individual i's utilitarian weights will be lower if δ is closer to δ_i and will converge to zero as δ goes to δ_i. More intuitively, suppose there is only one individual in each generation. If the planner wants her discount function to be essentially identical to the individual's, she can almost achieve it by taking only the current-generation individual's discounted utility function into account in her utilitarian aggregation.

In addition, fixing an arbitrary $t \in \mathbf{T}$, the planner assigns unique utilitarian weights $\{\alpha_i\}_{i \in \mathbf{N}}$ to generation-t individuals. These weights only depend on the social risk attitude characterized by the planner's and individuals' instantaneous utility functions. They do not depend on discounting.

These insights are summarized in the following corollary.

Corollary V.17 (Feng, Ke, and McMillan; 2022, [111])**.** *Suppose the assumptions of Theorem V.15 hold and, in addition, for some $\delta_1, \dots, \delta_N \in (0, 1)$ such that $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_{N-1} < \delta_N$, $\delta_{i,t}(\tau) = \delta_i^\tau$ for any $i \in \mathbf{N}$, $t \in \mathbf{T}$, and $\tau \in \mathbf{T}$. Then the following statements are true:*

1. *For any $\delta \in (\delta_N, 1)$, $\omega_t(i, t) = \alpha_i$ for any $i \in \mathbf{N}$ and $t \in \mathbf{T}$.*
2. *$\lim_{\delta \to \delta_N +} \omega_t(N, s) = 0$ for any $t \in \mathbf{T}$ and $s > t$.*

More importantly, Theorem V.15 allows us to study the asymptotic properties of the utilitarian weights. In particular, as Proposition V.18 shows, future-generation individuals' utilitarian weights diminish *exponentially* at a rate equal to the social discount rate $1 - \delta$.

Proposition V.18 (Feng, Ke, and McMillan; 2022, [111])**.** *Suppose $T = +\infty$ and the assumptions of Theorem V.15 hold. Then*

$$\lim_{s \to \infty} \frac{\omega_t(i, s+1)}{\omega_t(i, s)} = \delta$$

for each $i \in \mathbf{N}$ and $t \in \mathbf{T}$.

Proof. The utilitarian weight $\omega_t(i,s)$ can be written as a polynomial in the social discount factor δ and the relative discount factors:

$$\begin{aligned}\omega_t(i,s) &= \sum_{\tau=t}^{s-1}[\delta \cdot \delta_{i,\tau}(s-1-\tau) - \delta_{i,\tau}(s-\tau)]\omega_t(i,\tau) \\ &= \sum_{\tau=t}^{s-1}[\delta - \beta_{i,\tau}(s-\tau)]\delta_{i,\tau}(s-1-\tau)\omega_t(i,\tau) \\ &= \sum_{\tau=t}^{s-1}[\delta - \beta_{i,\tau}(s-\tau)]\prod_{k=0}^{s-\tau-1}\beta_{i,\tau}(k)\omega_t(i,\tau),\end{aligned}$$

in which $\beta_{i,\tau}(0) = 1$.

We claim that $\omega_t(i,s)$ is a homogeneous polynomial of degree $s-t$ and the coefficient of δ^{s-t} is always α_i. To verify this claim inductively, we first have $\omega_t(i,t) = \alpha_i$, which is a homogeneous polynomial of degree 0, and we also have $\omega_t(i,t+1) = \alpha_i\delta - \alpha_i\beta_{i,t}(1)$, which is a homogeneous polynomial of degree 1. Next, suppose the claim is true for $t,\ldots,s-1$. Note that $\omega_t(i,s) = \sum_{\tau=t}^{s-1}[\delta - \beta_{i,\tau}(s-\tau)]\omega_t(i,\tau)\prod_{k=0}^{s-\tau-1}\beta_{i,\tau}(k)$. Each term in the summation is the product of three components, among which $[\delta - \beta_{i,\tau}(s-\tau)]$ is of degree 1, $\omega_t(i,\tau)$ is of degree $\tau - t$, and $\prod_{k=0}^{s-\tau-1}\beta_{i,\tau}(k)$ is of degree $s-\tau-1$. Therefore, each term in the summation is of degree $s-t$. Since δ^{s-t} only shows up in the last term (when $\tau = s-1$) in the summation, $[\delta - \beta_{i,s-1}(1)]\omega_t(i,s-1)$, the coefficient of δ^{s-t} in $\omega_t(i,s)$ is the same as the coefficient of δ^{s-t-1} in $\omega_t(i,s-1)$, α_i.

Since the social discount factor $\delta < 1$ is greater than any individual relative discount factor $\beta_{i,t}(\tau)$, by L'Hospital's rule, we have $\lim_{s\to\infty}\dfrac{\omega_t(i,s)}{\alpha_i\delta^{s-t}} = 1$ for all $i \in \mathbf{N}$ and $t \in \mathbf{T}$. Hence, $\lim_{s\to\infty}\dfrac{\omega_t(i,s+1)}{\omega_t(i,s)} = \delta$. □

An easy way to see this is to note that when discount functions are all exponential, $\omega_t(i,t+1)/\omega_t(i,t) = \delta - \delta_i$ and $\omega_t(i,s+1)/\omega_t(i,s) = \delta$ for any $i \in \mathbf{N}$, $t \in \mathbf{T}$, and $s > t$. Hence, intuitively, a lower social discount factor is associated with higher intergenerational inequality measured by utilitarian weights.

CHAPTER VI

Discussion and Future Work

In this thesis, we study dynamic processes that arise in two domains of the academic literature—dynamical systems and social choice theory.

As for the novel contributions to the dynamical systems literature, we present our results in Chapter III. Firstly, we demonstrate the robustness of the auxiliary function method for long-time averages paired with SOS-SDP technology in order to compute sharp upper and lower bounds for time averaged quantities in the dynamical variables for both non-autonomous and certain forms of trigonometric dependence in nonlinear ODEs. Our procedure of augmenting a dynamical system with additional polynomial degrees of freedom appears robust. Additionally, we observe that sharp—within computer precision—bounds are often recovered for polynomial auxiliary functions of reasonably restricted degree. This appears to be the case not only in this work, but also various others [122, 45], so the SDP algorithm is able to concentrate the relative coefficients on potentially severely truncated polynomials for which sharp bounds are not guaranteed.

As for future directions, our technology is broadly applicable to a variety of disciplines in the applied sciences. There are a variety of applied science and engineering applications where moderately low dimensional ODE systems serve as central models for both conceptual and design purposes. These include energy harvesting [123, 124] where the challenge is to optimally extract power from vibrations of a continuously stimulated mechanical body where mathematical models often consist of periodically driven nonlinear oscillators [125]. Another area is the periodic operation of chemical and biochemical reactors [126] where the task is to optimize the time-average production of certain byproducts. Mass action and related kinetic models often consist of ODEs with polynomial vector fields. Circadian [127, 128] or seasonally forced [129, 130] models in biology, ecology and epidemiology are often described by such periodically driven ODEs with polynomial vector fields as well. Finally, we recognize the frontier for application of this auxiliary function approach and related numerical methods to systems described by partial differential equations. Recent work in this direction includes both fundamental theoretical results [49, 48] and revealing applications [47]. Still, many mathematical and computational questions remain for future research.

We also demonstrate that the auxiliary function method can serve as a viable tool in validating regions of stability as predicted by perturbative, asymptotic methods or establishing the true regions of stability in the face of failed perturbative, asymptotic methods. In §3.5, we investigate the higher order effects caused by coupling parameters on the stability region of a parametrically driven, coupled oscillator system across a broad range of modulation frequencies. We show that the stability region of a parametrically driven oscillator system as predicted by simplified, second order asymptotic methods ignoring the coupling terms between the oscillators can differ quite substantially at modulation frequencies away from parametric resonance frequencies. The simplified, asymptotic solution is un-conservative compared to a full, second order asymptotic solution when coupling terms are not ignored, which can in turn differ and still be non-trivially un-conservative in comparison to the true stability region at modulation frequencies away from the parametric resonance frequencies. The differences are caused by both neglecting the coupling terms and higher order effects.

This is a primary drawback of critical importance for asymptotic methods as the the validity of the asymptotic results depends crucially on one's choice of the approximations. However, there is currently no way to know which order will be sufficient to capture the true instability region, which depends on the strength of the coupling terms, g and r, as well as the phase of the parametric oscillator term ϕ. The solution is expected to be even more sensitive to these system parameters if the resonance frequency of the individual oscillations are different or if nonlinear terms are present. Hence, our results suggest that the auxiliary function method for long-time averages is an efficient and robust means of computing the true long-time averages and true regions of stability across all possible initial conditions without the need of ad hoc approximations. Moreover, this auxiliary function method has the advantage of being able to compute regions of stability both at and away from parametric resonance.

The differences between the true stability region and the approximate stability region are immaterial if one is operating withing a region of parameter space where both boundaries agree. However, if one is operating within a region of parameter space for which they disagree, this may be quite problematic for both experimental and real-world implementation—especially without the knowledge of operating within one of these regions of disagreement. This point is exacerbated by the fact that, at least in studying Equation (3.20), we discovered two, very narrow protruding tongues in parameter space for which the system was potentially unstable. Moreover, these tongues occur at very naturally occurring, non-pathological values of the driving frequency γ.

In the context of applications and future research directions, our results have profound implications on the reliance of asymptotic methods across a variety of applied sciences. Our results also illuminate future research directions with regards to the applicability of the auxiliary function method for long-time averages. For machine learning and environmental modeling applications, one is often interested in establishing a dynamical system via data-driven methods [131, 132, 133, 134], and many loss functions for the training of these machine learning models can be expressed in terms of long-time averages. This poses a potential connection and future direction of research between the training of machine learning models and this auxiliary function method. Moreover, if data sampling is performed on a relatively sparse, spatial or temporal grid, there may be narrow and protruding regions of parameter space, such as in Figure 3.10, for which the dynamics are unstable, and utilizing asymptotic methods may fail to accurately predict the underlying stability. Hence, we recommend this auxiliary function method as a viable method for checking system stability. For neuroscience and biological applications, studying synchrony behavior or global phase locking of neuronal firing or circadian rhythms via asymptotic methods may fail to capture sensitivity to a system's parameters. Moreover, with measures of synchrony being realized as time averages of the underlying oscillator's correlations [127, 128], this auxiliary function method may prove to be an indispensable tool in both capturing dynamic sensitivity to a system's parameters as well as concretely computing synchrony measures. For ecological applications, many ecologists have been trying to determine mechanisms that can stabilize ecosystems and support the biodiversity observed empirically as well as investigate the reactivity and subsequent stability of an ecological system [135, 136]. With previous work done on the applicability of this auxiliary function method to control problems [46] and with our results establishing this method's applicability to determining regions of stability, it may be that this method helps to illuminate these ecosystem stabilizing mechanisms as well as serve as an effective computation tool for determining parameter dependent stability.

As for the novel contributions to the social choice literature, we present our results in Chapter V. Firstly, we show that in a multi-generation setting with infinite time horizon—and hence countably infinitely many individuals (from all generations)—if we assume that the individuals' and the planner's preferences have continuous discounted expected utility representations, Harsanyi's (1955) utilitarianism theorem continues to hold under an additional boundedness assumption, and we provide two proofs of varying abstraction for our main result. We introduce new assumptions on the current generation individual's utility functions that ensure that the utilitarian weights are unique. This allows us to focus on utilitarian aggregation, and study how utilitarian weights change in the comparative static analysis and the limit of utilitarian weights for distant future generations. Among other findings, we find that less patient social discounting is associated with a more unequal, across generations assignment of utilitarian weights.

Our results are quite general as they follow the generality of our dynamic setting. The number of individuals in each generation may change, the set of consumption goods in each period may change, individuals' preferences are not necessarily related in any way, and individuals' and the planner's discount functions are not necessarily exponential. We have only considered intergenerational Pareto, which requires that the planner take all current- and future-generation individuals into account. It is immediate to extend our theorem to cases in which the Pareto condition only involves some, but not all, current- and future-generation individuals; for instance, we apply intergenerational Pareto to an auxiliary setting that only has the individuals involved in the Pareto condition. However, for our proof to go through, in each period t the planner must not completely ignore the current generation (generation-t individuals).

As for future directions, our results rely quite heavily on the previous work established by [6] and [84]. However, our results crucially depend on the expected, discounted utility form of both current- and future-generation individuals as well as the social planner. It is an open question as to whether a social planner's utility function can take on the form of utilitarian aggregation without these assumptions on the utility function forms. Additionally, as our setting considers discrete time horizons, it may be of future interest to explore utilitarian aggregation in a continuous time setting. Finally, recent work in operational engineering has studied the effect of multiple, temporal domains on dynamic choice [137]. Therefore, future work could consider the effects of these temporal domains on both the form of the individual utility functions as well as the form of the social planner's function, and moreover the natural analog of utilitarian aggregation in this setting.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] A.N. McMillan and C.R. Doering. Optimal time averages in non-autonomous nonlinear dynamical systems. *Journal of Pure and Applied Functional Analysis*, 208:382–386, 2018.

[2] Andrew N. McMillan and Yin Lu Young. Stability of a parametrically driven, coupled oscillator system: An auxiliary function method approach. *Journal of Applied Physics*, 131(15):154901, 2022.

[3] John von Neumann and Oskar Morgenstern. Theory of games and economic behavior. *Science and Society*, 9(4):366–369, 1944.

[4] Jeremy Bentham. *An Introduction to the Principles of Morals and Legislation*. Dover Publications, 1780.

[5] K. Arrow. *Social Choice and Justice*. Bellknap, 1984.

[6] J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321, 1955.

[7] Gyorgy Csaba and Wolfgang Porod. Coupled oscillators for computing: A review and perspective. *Applied Physics Reviews*, 7(1):011302, 2020.

[8] C. Wilson and J. Callaway. Coupled oscillator model of the dopaminergic neuron of the substantia nigra. *Journal of Neurophysiology*, 83:3084–3100, 2021.

[9] Timothy J. Lewis and Michael A. Schwemmer. *Weak Coupling Theory*, pages 3151–3159. Springer New York, New York, NY, 2015.

[10] Hirokazu Fujisaka and Tomoji Yamada. Stability Theory of Synchronized Motion in Coupled-Oscillator Systems: . *Progress of Theoretical Physics*, 69(1):32–47, 01 1983.

[11] J. F. Heagy, T. L. Carroll, and L. M. Pecora. Synchronous chaos in coupled oscillator systems. *Phys. Rev. E*, 50:1874–1885, Sep 1994.

[12] Steven H. Strogatz. *Nonlinear Dynamics and Chaos : with Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, 2015.

[13] Jaime G Carbonell, Ryszard S Michalski, and Tom M Mitchell. An overview of machine learning. *Machine learning*, pages 3–23, 1983.

[14] Yong Fang, Kin Keung Lai, and Shou-Yang Wang. Portfolio rebalancing model with transaction costs based on fuzzy decision theory. *European Journal of Operational Research*, 175(2):879–893, 2006.

[15] Gordon F Pitz and Natalie J Sachs. Judgment and decision: Theory and application. *Annual review of psychology*, 35(1):139–164, 1984.

[16] Joseph P La Salle. *The stability of dynamical systems*. SIAM, 1976.

[17] Yuri A Mitropolsky and Nguyen Van Dao. *Applied asymptotic methods in nonlinear oscillations*, volume 55. Springer Science & Business Media, 2013.

[18] Marcello Calvanese Strinati, Leon Bello, Avi Pe'er, and Emanuele G. Dalla Torre. Theory of coupled parametric oscillators beyond coupled ising spins. *Physical Review A 100*, 2019.

[19] Peter Kloeden and Martin Rasmussen. *Nonautonomous Dynamical Systems*, volume 176. 08 2011.

[20] Sudhir K. Goel and Dean A. Neumann. Completely unstable dynamical systems. *Transactions of the American Mathematical Society*, 291(2):639–668, 1985.

[21] Kamil Nar, Yuan Xue, and Andrew M. Dai. Learning unstable dynamical systems with time-weighted logarithmic loss. *CoRR*, abs/2007.05189, 2020.

[22] Roger B. Myerson. *Game theory analysis of conflict*. Harvard University Press, 1991.

[23] Ian Tobasco, David Goluskin, and Charles R. Doering. Optimal bounds and extremal trajectories for time averages in nonlinear dynamical systems. *Physics Letters A*, 382(6):382–386, 2018.

[24] P. Walters. *An Introduction to Ergodic Theory*. Springer, 1982.

[25] O. Jenkinson. Ergodic optimization. *Discrete & Continuous Dynamical Systems*, 15:197–224, 2006.

[26] J. Bochi. rgodic optimization of birkhoff averages and lyapunov exponents. *Proceedings of the International Congress of Mathematicians*, 3:1843–1864, 2018.

[27] OLIVER JENKINSON. Ergodic optimization in dynamical systems. *Ergodic Theory and Dynamical Systems*, 39(10):2593–2618, Jan 2018.

[28] B.R. Hunt and E. Ott. Optimal periodic orbits of chaotic systems occur at low period. *Physical Review E*, 54:328–337, 1996.

[29] C. Gregobi E. Ott and J.A. Yorke. Controlling chaos. *Physical Review Letters*, 64:1196–1199, 1990.

[30] J.-P. Conze and Y. Guivarc'h. Croissance des sommes ergodiques. circa 1993.

[31] K.G. Murty and S.N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

[32] P. A. Parrilo and R. R. Thomas. Sum of squares: Theory and applications. 2019.

[33] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[34] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Series B*, 96:293–320, 2003.

[35] Victoria Powers and Thorsten Wörmann. An algorithm for sums of squares of real polynomials. *Journal of Pure and Applied Algebra*, 127(1):99–104, 1998.

[36] N. Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23:731–734, 1987.

[37] K. Derinkuyu and M.Ç. Pinar. On the s-procedure and some variants. *Math Meth Oper Res*, 64:55–77, 2006.

[38] D. Hilbert. Uber die darstellung definiter formen als summe von formenquadraten. *Mathematische Annalen*, 32:342–350, 1888.

[39] J.B. Lasserre. A sum of squares approximation of nonnegative polynomials. *IFAC Proceedings Volumes*, 38:441–444, 2005.

[40] M. Lakshmi, G. Fantuzzi, J. Fernandez-Caballero, Y. Hwang, and S.I. Chernyshenko. Finding extremal periodic orbits with polynomial optimization, with application to a nine-mode model of shear flow. *arXiv:1906.04001*, 2020.

[41] M. Marshall. *Positive Polynomials and Sums of Squares*. Mathematical surveys and monographs. American Mathematical Society, 2008.

[42] J. Lofberg. Yalmip : A toolbox for modeling and optimization in matlab. *IEEE International Conference on Robotics and Automation*, 2004.

[43] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 2006.

[44] Matthew L. Olson, David Goluskin, William W. Schultz, and Charles R. Doering. Heat transport bounds for a truncated model of rayleigh–bénard convection via polynomial optimization. *Physica D: Nonlinear Phenomena*, 415:132748, 2021.

[45] G. Fantuzzi, D. Goluskin, D. Huang, and S. I. Chernyshenko. Bounds for deterministic and stochastic dynamical systems using sum-of-squares optimization. *SIAM Journal on Applied Dynamical Systems*, 15(4):1962–1988, 2016.

[46] Mayur V. Lakshmi, Giovanni Fantuzzi, Sergei I. Chernyshenko, and Davide Lasagna. Finding unstable periodic orbits: A hybrid approach with polynomial optimization. *Physica D: Nonlinear Phenomena*, 427:133009, 2021.

[47] D. Goluskin and G Fantuzzi. Bounds on mean energy in the kuramoto–sivashinsky equation computed using semidefinite programming. *Nonlinearity*, 32:1705–1730, 2019.

[48] Ricardo M.S. Rosa and Roger M. Temam. Optimal minimax bounds for time and ensemble averages of dissipative infinite-dimensional systems with applications to the incompressible navier-stokes equations. 2020.

[49] P.J. Goulart and S. Chernyshenko. Global stability analysis of fluid flows using sum-of-squares. *Physica D: Nonlinear Phenomena*, 241:692–704, 2021.

[50] Miguel F Anjos and Jean B Lasserre. *Handbook on semidefinite, conic and polynomial optimization*, volume 166. Springer Science & Business Media, 2011.

[51] Ury Passy and DJ Wilde. Generalized polynomial optimization. *SIAM Journal on Applied Mathematics*, 15(5):1344–1356, 1967.

[52] Jean B Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization*, 17(3):822–843, 2006.

[53] Azaria Paz and Shlomo Moran. Non deterministic polynomial optimization problems and their approximations. *Theoretical Computer Science*, 15(3):251–277, 1981.

[54] J. Parker. Can sum-of-squares programming tell us anything useful about hamiltonian chaos? *GFD Report (Woods Hole Oceanographic Institution)*, 2019.

[55] Michael J. Kovacic, Ivana; Brennan. *The Duffing Equation: Nonlinear Oscillators and their Behaviour*, volume 10.1002/9780470977859. John Wiley & Sons, Ltd, mar 2011.

[56] Tamás Kalmár-Nagy and Balakumar Balachandran. *Forced Harmonic Vibration of a Duffing Oscillator with Linear Viscous Damping*, chapter 5, pages 139–174. John Wiley & Sons, Ltd, 2011.

[57] Asok Kumar Mallik. *Forced Harmonic Vibration of a Duffing Oscillator with Different Damping Mechanisms*, chapter 6, pages 175–217. John Wiley & Sons, Ltd, 2011.

[58] Stefano Lenci and Giuseppe Rega. *Forced Harmonic Vibration in a Duffing Oscillator with Negative Linear Stiffness and Linear Viscous Damping*, chapter 7, pages 219–276. John Wiley & Sons, Ltd, 2011.

[59] Ivana Kovacic and Michael J. Brennan. *Forced Harmonic Vibration of an Asymmetric Duffing Oscillator*, chapter 8, pages 277–322. John Wiley & Sons, Ltd, 2011.

[60] Bena, I. and C. Van den Broeck. Coupled parametric oscillators. *Europhys. Lett.*, 48(5):498–504, 1999.

[61] Bruce Denardo, John Earwood, and Vera Sazonova. Parametric instability of two coupled nonlinear oscillators. *American Journal of Physics*, 67(3):187–195, 1999.

[62] William B. Case, Paul J. H. Tjossem, Katherine G. Abrams, and James F. St. Germaine-Fuller. Coupled oscillators driven with difference-frequency parametric position feedback. *Journal of Applied Physics*, 122(12):124905, 2017.

[63] R.J. Kinsey, D.L. Mingori, and R.H. Rand. Nonlinear controller to reduce resonance effects during despin of a dual-spin spacecraft through precession phase lock. In *[1992] Proceedings of the 31st IEEE Conference on Decision and Control*, pages 3025–3030 vol.4, 1992.

[64] H A Buckmaster and J C Dering. The application of phase-lock microwave frequency stabilizers to electron paramagnetic resonance spectrometers. *Journal of Scientific Instruments*, 43(8):554–557, aug 1966.

[65] Li-Qun Chen and Wen-An Jiang. Internal resonance energy harvesting. *Journal of Applied Mechanics*, 82(3), 2015.

[66] Vinod R Challa, MG Prasad, Yong Shi, and Frank T Fisher. A vibration energy harvesting device with bidirectional resonance frequency tunability. *Smart Materials and Structures*, 17(1):015035, 2008.

[67] Robert A York, Peter Liao, and Jonathan J Lynch. Oscillator array dynamics with broadband n-port coupling networks. *IEEE Transactions on Microwave Theory and Techniques*, 42(11):2040–2045, 1994.

[68] Kristine E Callan, Lucas Illing, Zheng Gao, Daniel J Gauthier, and Eckehard Schöll. Broadband chaos generated by an optoelectronic oscillator. *Physical review letters*, 104(11):113901, 2010.

[69] Young S Lee, Gaëtan Kerschen, D Michael McFarland, W Joel Hill, Chetan Nichkawde, Thomas W Strganac, Lawrence A Bergman, and Alexander F Vakakis. Suppressing aeroelastic instability using broadband passive targeted energy transfers, part 2: experiments. *AIAA journal*, 45(10):2391–2400, 2007.

[70] M. Beneke and V. I. Zakharov. Improving large-order perturbative expansions in quantum chromodynamics. *Phys. Rev. Lett.*, 69:2472–2474, Oct 1992.

[71] SA Fulling. The local geometric asymptotics of continuum eigenfunction expansions i. overview. *SIAM Journal on Mathematical Analysis*, 13(6):891–912, 1982.

[72] George A. Articolo. Chapter 7 - the wave equation in two spatial dimensions. In George A. Articolo, editor, *Partial Differential Equations & Boundary Value Problems with Maple (Second Edition)*, pages 409–476. Academic Press, Boston, second edition edition, 2009.

[73] Horst H. E. Leipholz. Use of galerkins method for vibration problems. *The Shock and Vibration Digest*, 8:3–18, 1976.

[74] Xiao-Jun Yang and H.M. Srivastava. An asymptotic perturbation solution for a linear oscillator of free damped vibrations in fractal medium described by local fractional derivatives. *Communications in Nonlinear Science and Numerical Simulation*, 29(1):499–504, 2015.

[75] William E Caswell. Accurate energy levels for the anharmonic oscillator and a summable series for the double-well potential in perturbation theory. *Annals of Physics*, 123(1):153–184, 1979.

[76] John Boyd. The devil's invention: Asymptotic, superasymptotic and hyperasymptotic series. *Acta Applicandae Mathematicae*, 56, 09 2000.

[77] Jayanta K. Ghosh. Higher order asymptotics. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 4:i–111, 1994.

[78] Peter Menck, Jobst Heitzig, Robert Marwan, and Juergen Kurths. How basin stability complements the linear-stability paradigm. *Nature Physics*, 9:89–92, 01 2013.

[79] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual.* 54 edition, 2015.

[80] J. Lofberg. Yalmip : A toolbox for modeling and optimization in matlab. *IEEE International Conference on Robotics and Automation*, 2004.

[81] Peter Miller. *Applied Asymptotic Analysis*. Graduate Studies in Mathematics. American Mathematical Society, 2006.

[82] Paul A. Samuelson. A note on measurement of utility. *The Review of Economic Studies*, 4(2):155–161, 1937.

[83] John C. Harsanyi. Rule utilitarianism and decision theory. *Erkenntnis (1975-)*, 11(1):25–53, 1977.

[84] L. Zhou. Harsanyi's Utilitarianism Theorems: General Societies. *Journal of Economic Theory*, 72(1):198–207, 1997.

[85] M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–546, 1953.

[86] A Tversky and D Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[87] Peter C. Fishburn. *Utility Theory and Decision Theory*, pages 303–312. Palgrave Macmillan UK, 1990.

[88] Tjalling C. Koopmans. Stationary ordinal utility and impatience. *Econometrica*, 28(2):287–309, 1960.

[89] Richard Thaler. Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8(3):201–207, 1981.

[90] E. S. Phelps and R. A. Pollak. On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2):185–199, 1968.

[91] George Loewenstein and Drazen Prelec. Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, 107(2):573–597, 1992.

[92] J. Pan, C.S. Webb, and H. Zank. An extension of quasi-hyperbolic discounting to continuous time. *Games and economic behavior.*, 89:43–55, January 2015.

[93] Pavlo Blavatskyy. A monotone model of intertemporal choice. *Economic Theory*, 62, 10 2016.

[94] G. E. Moore. *Principia Ethica*. Cambridge University Press, 1922.

[95] Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346, 1950.

[96] Amartya Sen. Social choice and justice: A review article. *Journal of Economic Literature*, 23(4):1764–1776, 1985.

[97] Peter A. Diamond. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy*, 75(5):765–766, 1967.

[98] Amartya Sen. *Social Choice Theory*. North Holland, Amsterdam, 1986.

[99] K. C. Border. More on harsanyi's utilitarian cardinal welfare theorem. *Social Choice and Welfare*, 1(4):279–281, 1985.

[100] John A. Weymark. Harsanyi's social aggregation theorem and the weak pareto principle. *Social Choice and Welfare*, 10(3):209–221, 1993.

[101] Z. Domoto. Ordered sum and tensor product of utility structures. *Theory Dec.*, pages 375–399, 1979.

[102] Bernard De Meyer and Philippe Mongin. A note on affine aggregation. *Economics Letters*, 47:177–183, 1995.

[103] P. Hammond. Harsanyi's utilitarian theorem : a simpler proof and some ethical connotations. *EUI ECO*, 1991.

[104] David McCarthy, Kalle Mikkola, and Teruji Thomas. Aggregation for Potentially Infinite Populations without Continuity or Completeness. arXiv preprint arXiv:1911.00872, 2019.

[105] Nabil I Al-Najjar and Luciano Pomatto. Aggregate Risk and the Pareto Principle. *Journal of Economic Theory*, 189:105084, 2020.

[106] P. Mongin. The Paradox of the Bayesian Experts and State-Dependent Utility Theory. *Journal of Mathematical Economics*, 29(3):331–361, 1998.

[107] Thierry Coulhon and Philippe Mongin. Social Choice Theory in The Case of von Neumann-Morgenstern Utilities. *Social Choice and Welfare*, 6(3):175–187, 1989.

[108] P. Fishburn. On Harsanyi's Utilitarian Cardinal Welfare Theorem. *Theory and Decision*, 17(1):21–28, 1984.

[109] J. Weymark. Harsanyi's Social Aggregation Theorem with Alternative Pareto Principles. In Eichhorn W., editor, *Models and Measurement of Welfare and Inequality*, pages 869–887. Springer, Berlin, Heidelberg, 1994.

[110] T. Börgers and Y. Choo. Revealed Relative Utilitarianism. Working Paper, University of Michigan, 2017.

[111] Tangren Feng, Shaowei Ke, and Andrew McMillan. Utilitarianism and social discounting with countably many generations. *Journal of Mathematical Economics*, 98:102576, 2022.

[112] T. Feng and S. Ke. Social Discounting and Intergenerational Pareto. *Econometrica*, 86(5):1537–1567, 2018.

[113] S. Zuber. Can Social Preferences Be Both Stationary and Paretian? *Annals of Economics and Statistics*, (101/102):347–360, 2011.

[114] M. Jackson and L. Yariv. Collective Dynamic Choice: The Necessity of Time Inconsistency. *American Economic Journal: Microeconomics*, 7(4):150–178, 2015.

[115] C. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin, Heidelberg, 3 edition, 2006.

[116] K. Yosida and E. Hewitt. Finitely Additive Measures. *Transactions of the American Mathematical Society*, 72(1):46–66, 1952.

[117] B. Craven and J. Koliha. Generalizations of Farkas' Theorem. *SIAM Journal on Mathematical Analysis*, 8(6):983–997, 1977.

[118] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2010.

[119] Y. Abramovich and C. Aliprantis. *An Invitation to Operator Theory*. American Mathematical Society, 2002.

[120] J. Diestel. *Sequences and Series in Banach Spaces*. Springer-Verlag, 1984.

[121] R. Buck. Bounded Continuous Functions on a Locally Compact Space. *Michigan Mathematical Journal*, 5(2):95–104, 1958.

[122] Polynomial sum of squares in fluid dynamics: a review with a look ahead. *Philosophical Transactions of the Royal Society A*, 372, 2014.

[123] Mann B. P. and Sims N.D. Energy harvesting from the nonlinear oscillations of magnetic levitation. *Journal of Sound and Vibration*, 319:515–530, 2009.

[124] Wei C. and Xingjian J. A comprehensive review on vibration energy harvesting: modelling and realization. *Renewable and Sustainable Energy Reviews*, 74:1–18, 2017.

[125] D.J. Inman A. Erturk. Broadband piezoelectric power generation on high-energy orbits of the bistable duffing oscillator with electromechanical coupling. *Journal of Sound and Vibration*, 330:2339–2353, 2011.

[126] Silveston P.L. and Hudgins R.R. Periodic operation of reactors. 2013.

[127] Gonze D, Bernard S, Waltermann C, Kramer A, and Herzel H. Spontaneous synchronization of coupled circadian oscillators. *J. Biophysics*, 2005.

[128] N. Komin, Murza A. C., Hernandez-Garcia E., and Toral R. Synchronization and entrainment of coupled circadian oscillators. *Interface Focus*, 1(1):167–176, 2011.

[129] Greenman J.V. and Norman R.A. Environmental forcing, invasion and control of ecological and epidemiological systems. *Journal of Theoretical Biology*, 247:492–506, 2007.

[130] Taylor R.A., Sherratt J.A., and White A. Seasonal forcing and multi-year cycles in interacting populations: Lessons from a predator-prey model. *Journal of Mathematical Biology*, 67:1741–1764, 2012.

[131] Jianming Liu and Steven L Grant. Proportionate adaptive filtering for block-sparse system identification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(4):623–630, 2015.

[132] Salar Fattahi and Somayeh Sojoudi. Data-driven sparse system identification. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 462–469. IEEE, 2018.

[133] Nicholas Kalouptsidis, Gerasimos Mileounis, Behtash Babadi, and Vahid Tarokh. Adaptive algorithms for sparse system identification. *Signal Processing*, 91(8):1910–1919, 2011.

[134] J. N. Kutz E. Kaiser and S. L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, (2219), 2018.

[135] Si Tang and Stefano Allesina. Reactivity and stability of large ecosystems. *Frontiers in Ecology and Evolution*, 2, 2014.

[136] Mark Emmerson and Jon Yearsley. Weak interactions, omnivory and emergent food-web properties. *Proceedings. Biological sciences / The Royal Society*, 271:397–405, 03 2004.

[137] Hyeong Soo Chang, P.J. Fard, S.I. Marcus, and M. Shayman. Multitime scale markov decision processes. *IEEE Transactions on Automatic Control*, 48(6):976–987, 2003.