Finite-Dimensional Approximation in Dual Domain with Applications in Opinion Dynamics and Dynamic Programming

Sharifi Kolarijani, M.A.

DOI
10.4233/uuid:d096378a-c676-492c-a409-a5fd3b6e0474

Publication date
2022

Document Version
Final published version

Citation (APA)
Sharifi Kolarijani, M. A. (2022). *Finite-Dimensional Approximation in Dual Domain: with Applications in Opinion Dynamics and Dynamic Programming*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:d096378a-c676-492c-a409-a5fd3b6e0474

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

FINITE-DIMENSIONAL APPROXIMATION IN DUAL DOMAIN

WITH APPLICATIONS IN OPINION DYNAMICS AND DYNAMIC PROGRAMMING

FINITE-DIMENSIONAL APPROXIMATION IN DUAL DOMAIN

WITH APPLICATIONS IN OPINION DYNAMICS AND DYNAMIC PROGRAMMING

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on Monday 20 June 2022 at 15:00 o'clock.

by

Mohamad Amin SHARIFI KOLARIJANI

Master of Science in Biomedical Engineering, Amirkabir University of Technology, Iran,
born in Ghaemshahr, Iran.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	Chairperson
Prof.dr.ir. T. Keviczky,	Delft University of Technology, promotor
Dr. P. Mohajerin Esfahani,	Delft University of Technology, promotor
Independent members:	
Prof.dr.ir. M.C. Veraar,	Delft University of Technology
Prof.dr.ir. K.I. Aardal,	Delft University of Technology
Prof.dr. W.M. McEneaney,	University of California San Diego, USA
Prof.dr.ir. M. Cao,	University of Groningen

ISBN 978-94-6366-526-1

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Blessed is the gambler who has lost everything,
except the desire to gamble once more.

Rumi

To Avin and Anisa.

CONTENTS

SUMMARY

This thesis is comprised of two main parts. In the first part of the thesis, we study the nonlinear Fokker-Planck (FP) equation that arises as a mean-field (macroscopic) approximation of the bounded confidence opinion dynamics, where opinions are influenced by environmental noises and opinions of radicals (stubborn individuals). The distribution of radical opinions serves as an infinite-dimensional exogenous input to the FP equation, visibly influencing the steady opinion profile. We first establish the mathematical properties of the FP equation. In particular, we (i) show the well-posedness of the dynamic equation, (ii) provide existence result accompanied by a quantitative global estimate for the corresponding stationary solution, and, (iii) establish an explicit lower bound on the noise level that guarantees exponential convergence of the dynamics to stationary state. Combining the results in (ii) and (iii) readily yields the input-output stability of the system for sufficiently large noises. Next, using Fourier analysis, the structure of opinion clusters under the uniform initial distribution is examined. Specifically, two numerical schemes for (i) identification of order-disorder transition and (ii) characterization of initial clustering behavior are provided. The results of the analysis are validated through several numerical simulations of the continuum-agent model (partial differential equation) and the corresponding discrete-agent model (interacting stochastic differential equations) for a particular distribution of radicals.

In the second part of the thesis, we focus on the value iteration algorithm for solving optimal control problems. We propose two novel numerical schemes for approximate implementation of the dynamic programming (DP) operation concerned with finite-horizon, optimal control of deterministic, discrete-time systems with input-affine dynamics. The proposed algorithms involve discretization of the state and input spaces and are based on an alternative path that solves the dual problem corresponding to the DP operation. We provide error bounds for the proposed algorithms, along with detailed analyses of their computational complexity. In particular, for a specific class of problems with separable data in the state and input variables, the proposed approach can reduce the typical time complexity of the DP operation from $\mathcal{O}(XU)$ to $\mathcal{O}(X+U)$, where X and U denote the size of the discrete state and input spaces, respectively. We next discuss the extensions of the proposed conjugate value iteration algorithm for problems with separable data. The extensions are three-fold: We consider (i) infinite-horizon, discounted cost problems with (ii) stochastic dynamics, while (iii) computing the conjugate of input cost numerically. In particular, we analyze the convergence, complexity, and error of the proposed algorithm under these extensions. The theoretical results are validated through multiple numerical examples.

SAMENVATTING

Dit proefschrift bestaat uit twee hoofddelen. In het eerste deel van het proefschrift bestuderen wij de niet-lineaire Fokker-Planck (FP) vergelijking die ontstaat als macroscopische benadering van de begrensde vertrouwen opinie dynamiek, waarbij opinies worden beïnvloed door omgevingsruis en opinies van radicalen (koppige individuen). De verdeling van radicale opinies dient als een oneindigdimensionale exogene invoer voor de FP vergelijking, en heeft een zichtbare invloed op het stabiele opinieprofiel. Wij stellen eerst de wiskundige eigenschappen van de FP vergelijking. In het bijzonder, (i) tonen wij de goed-gestelde van de dynamische vergelijking; (ii) bieden wij het bestaansresultaat vergezeld van een kwantitatieve globale schatting voor de overeenkomstige stationaire oplossing; en, (iii) stellen wij een expliciete ondergrens voor het ruisniveau vast dat exponentiële convergentie van de dynamiek naar stationaire toestand garandeert. Het combineren van de resultaten in (ii) en (iii) levert de invoer-uitvoer stabiliteit van het systeem op voor voldoende grote ruis. Vervolgens wordt met behulp van Fourier-analyse de structuur van opinieclusters onder de uniforme initiële verdeling onderzocht. Meer bepaald worden twee numerieke schema's verstrekt voor (i) identificatie van de ordewanorde overgang en (ii) karakterisering van het initiële clustergedrag. De resultaten van de analyse worden gevalideerd door middel van verschillende numerieke simulaties van het continuüm-agent model (partiële differentiaalvergelijking) en het overeenkomstige discrete-agent model (interacterende stochastische differentiaalvergelijkingen) voor een bepaalde verdeling van radicalen.

In het tweede deel van het proefschrift richten wij ons op het waarde iteratie algoritme voor het oplossen van optimale controle problemen. Wij stellen twee nieuwe numerieke schema's voor een benaderende implementatie van de dynamische programmering (DP) operatie met betrekking tot de eindige-horizon, optimale controle van deterministische en discrete-tijd systemen met invoer-affiene dynamiek. De voorgestelde algoritmen omvatten discretisatie van de toestands- en invoerruimte en zijn gebaseerd op een alternatief pad dat het duale probleem oplost dat overeenkomt met de DP operatie. Wij geven foutgrenzen voor de voorgestelde algoritmen, samen met een gedetailleerde analyse van hun computationele complexiteit. In het bijzonder, voor een specifieke klasse van problemen met scheidbare gegevens in de toestands- en invoervariabele, kan de voorgestelde aanpak de typische tijdscomplexiteit van de DP operatie verminderen van $\mathcal{O}(XU)$ tot $\mathcal{O}(X+U)$, waarbij X en U respectievelijk de grootte van de discrete toestand- en invoerruimte aanduiden. Wij bespreken vervolgens de uitbreidingen van het voorgestelde algoritme voor problemen met scheidbare gegevens. De uitbreidingen zijn drieledig: Wij beschouwen (i) oneindige-horizon, verdisconteerde-kosten problemen met (ii) stochastische dynamiek, terwijl (iii) de geconjugeerde invoer kosten numeriek worden berekend. In het bijzonder analyseren wij de convergentie, complexiteit en fout van het voorgestelde algoritme onder deze uitbreidingen. De theoretische resultaten worden gevalideerd aan de hand van meerdere numerieke voorbeelden.

ACKNOWLEDGEMENTS

First, I have to thank my supervisors Peyman Mohajerin Esfahani and Tamás Keviczky for their trust and support. Special thanks to Peyman for sharing his invaluable experiences without any reservation over the past four years. Working with him, I ended up adopting a fundamental point of view towards scientific research. I am also very grateful to Tamás for providing me with a kind of support that I believe to be rare. He truly leads by example which has helped me grow both professionally and personally.

I would like to thank my colleagues Anton Proskurnikov and Gyula Max; parts of this dissertation are the result of collaborations with them. My sincere gratitude to the defense committee members, Mark Veraar, Karen Aardel, William MacEneaney, and Ming Cao, for accepting to be part of the committee and also for their feedback that improved the quality of this dissertation. I am also thankful to Gijs for helping me with the Dutch translation of the summary of the dissertation.

Many thanks to the amazing staff at DCSC, particularly, Heleen, Marieke, Francy, and Erica. I would like to thank Charalampos and Francisca for the great job they did during their master's projects. A big thanks to Giannis and Daniel for the remarkable experience during our TAship for the course Control Theory. I have to also thank the members of Peyman's group, Zhingwei, Pedro Z., Max, Pedro F., Rayyan, Shabnam, and Reza. A warm thanks to Tomas, Vittorio, and Amin for all the pleasant chats over these years.

Finally, I would like to use this opportunity to thank my family and friends. To my parents, and to Iman, Ehsan, Shafagh, and Maryam: There are no words to express my gratitude and appreciation to all of you. To Yaser and Iman: Thank you for being a part of my life; you guys are and always will be like family to me. To Manyu: Thank you for the long talks (therapy sessions!); your friendship is one of the most valuable pay-offs of my time at DCSC. To Aylin: Thank you for the never-ending love and support; I cannot believe how kind and understanding you have been. Last but not least, to Arman: Thank you for everything; I could not have done this without your help.

Amin Kolarijani
Delft, March 2022

1

INTRODUCTION

1

Finite-dimensional approximation of infinite-dimensional objects is an indispensable part of modern practice in science and technology. Indeed, any physical implementation of a numerical algorithm requires such an approximation due to finite machine precision, i.e., finite number of bits (dimensions) available for representing a possibly irrational (infinite-dimensional) real number. To make the matters worse, it is quite common that the object of interest is itself a function living in an infinite-dimensional space (requiring infinite, possibly uncountable, number of real numbers for full representation). Take, for example, the solution to an ordinary differential equation (ODE)

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x),$$

where $f : \mathbb{R} \to \mathbb{R}$ is a given Lipschitz-continuous function. Any numerical algorithm for solving this ODE for, say, $t \in [0, T]$ and initial condition $x(0) = x_0$, works with a finite discretization of the independent variable t, i.e., a finite-dimensional approximation of the infinite-dimensional object $x : [0, T] \to \mathbb{R}$. Naturally, similar issues arise in numerical algorithms for solving partial differential equations (PDEs). As a second example, consider the partial optimization problem

$$f^\star(x) = \min_y f(x, y),$$

where $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a given function. Once again, the optimal value $f^\star$ is an infinite-dimensional object. Therefore, unless the minimization problem has an analytic solution, one has no choice but to settle for a finite-dimensional approximation of this problem by, e.g., solving the problem for a finite number of the independent variable x.

Arguably, the most essential aspect of any function approximation technique is the proper choice of the parameterization scheme leading to a compact representation of the true function. For ODEs and PDEs, a particularly well-established class of approximation schemes is the so-called spectral method, where the true function is approximated as a linear combination of a finite set of global basis functions. Here, the underlying assumption is that the true function belongs to a certain function space with a known (countable) basis. For example, for problems with periodic geometry, the Fourier series is the proper choice. The numerical algorithm then involves finding the corresponding coefficients for a truncated (finite) Fourier expansion of the solution that satisfies a (weak) reformulation of the original differential equation. The fundamental idea here is to exploit the geometry of the problem, and use a more efficient representation of the solution in the frequency ("dual") domain.

When it comes to optimization problems, convex geometry is undoubtedly the most important type of geometry out there. Utilizing again the spectral method, for infinite-dimensional minimization problems with convex geometry, one can approximate the solution as a linear combination of a finite number of basis functions. However, the proper function space, in this case, is a max-plus space. Precisely, the compact representation of the true function is constructed in the slope ("dual") domain, as a max-plus linear combination of the basis functions. Furthermore, this dual representation potentially allows us to exploit the operational duality of infimal convolution and addition with respect to the conjugate transform: For two functions $f_1, f_2 : \mathbb{R}^n \to [-\infty, \infty]$, we have

$$(f_1 \square f_2)^* = f_1^* + f_2^*,$$

where

$$f_1 \square f_2(x) := \inf\{f_1(x_1) + f_2(x_2) : x_1 + x_2 = x\},$$

is the infimal convolution of f_1 and f_2, and

$$f_1^*(y) := \max_x y^\top x - f_1(x), \quad y \in \mathbb{R}^n,$$

is the convex conjugate (also known as Legendre-Fenchel transform) of f_1. This is analogous to the well-known operational duality of convolution and multiplication with respect to the Fourier transform. Actually, the Legendre-Fenchel transform plays a similar role as Fourier transform when the underlying algebra is the max-plus algebra, as opposed to the conventional plus-times algebra.

In this thesis, we aim to use this concept of finite-dimensional approximation in the dual domain in the context of two problems:

- for analysis and numerical simulation of a highly nonlinear PDE arising as the macroscopic model of opinion dynamics, and,
- for developing fast numerical algorithms for solving infinite-dimensional minimization problems arising in optimal control of discrete-time systems.

In what follows, we provide a summary of these two main parts along with an overview of the main results presented in the corresponding chapters.

Part One (Chapters 2 and 3)

In the first part of the thesis, we advance the theory of macroscopic modeling of bounded confidence opinion dynamics. Bounded confidence models stipulate that a social actor is insensitive to opinions beyond its bounded confidence set (usually, this set is an open or closed ball, centered at the actor's own opinion), which makes the graph of interactions among the agents distance-dependent. These models exhibit convergence of the opinions to some steady values, which can reach consensus or split into several disjoint clusters. Opinions in real social groups, however, usually do not terminate at steady values yet oscillate, which is usually explained by two factors. The first reason explaining opinion fluctuation is exogenous influence, which can be interpreted as some "truth" available to some individuals or a position shared by a group of close-minded opinion leaders ("radicals"). The second culprit of this fluctuation is uncertainty in the opinion dynamics, usually modeled as a random drift of each opinion. Whereas these models are still waiting for clear sociopsychological interpretation, they are broadly adopted in statistical physics to study phase transitions in systems of interacting particles.

Despite some progress in the analysis of noisy bounded confidence models, in particular, the interplay of confidence ranges and noise levels, all consequences of noise and exogenous influence in nonlinearly coupled networks are far from being understood. Even for the classical models, disclosing the relationship between the initial and the terminal opinion profiles remains a challenging problem. This motivates the examination of the corresponding mean-field models with an infinite number of actors. The arising macroscopic approximations of microscopic models describe the evolution of the distribution (a probability measure or a density) of opinion over some domain.

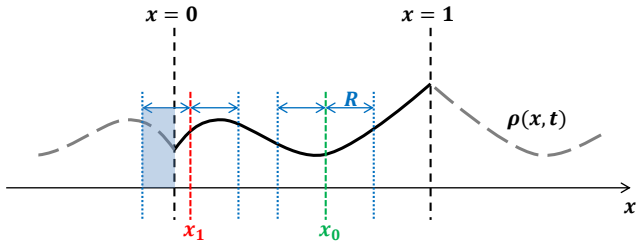


Figure 1.1: The even 2-periodic extension of the system. The opinion value is assumed to belong to the set $X = [0,1]$ without loss of generality. This basic opinion domain is first extended evenly to $\tilde{X} = [-1,1]$ and then periodically to $\mathbb{R}$. As can be seen, this particular extension leads to an *almost* reflective boundary condition. The opinion value $x_0 \in [R, 1-R]$ effectively experiences a reflective boundary condition, while for the opinion value $x_1 \in [0, R]$ there is also a boundary effect due to the even extension. In particular, the influence of more extreme neighbors of x_1 is reinforced by introducing artificial ones (the shaded area in blue). The same boundary effect exists for opinion values in $[1-R, 1]$.

The continuous-time model for opinion dynamics considered in this thesis is the following (even) 2-periodic nonlinear Fokker-Planck (FP) equation (the subscripts x, xx, and t denote the corresponding partial derivatives)

$$\begin{cases} \rho_t = (\rho\, G_\rho)_x + \frac{\sigma^2}{2}\rho_{xx} & \text{in} \quad \tilde{X} \times (0,T) \\ \rho(\cdot+2,t) = \rho(\cdot,t) & \text{on} \quad \partial\tilde{X} \times (0,T) \\ \rho(x,\cdot) = \rho_0(x) & \text{on} \quad \tilde{X} \times \{t=0\}, \end{cases} \tag{1.1}$$

where

$$G_\rho(x,t) := w(x) \star \big(\rho(x,t) + M\rho_r(x)\big). \tag{1.2}$$

Above, $\rho(x,t)$ denotes the even extension of the density of the opinions from $X = [0,1]$ to $\tilde{X} = [-1,1]$, while $\rho_0(x)$ is the corresponding extension of the initial opinion profile; see Figure 1.1 for a visualization of the even 2-periodic extension of the system. The function

$$w(x) = \begin{cases} x, & |x| \le R, \\ 0, & \text{o.w.}, \end{cases}$$

is the interaction kernel corresponding to the confidence bound of radius $R < 1$. In particular, note that the opinions are influenced by environmental noises (modeled by the diffusion term with $\sigma > 0$ being the noise level) and opinions of a group of radicals (modeled by adding the even extension of the radical opinion density ρ_r and its relative mass $M \le 1$ in the drift term). The distribution of radical opinions indeed serves as an infinite-dimensional exogenous input to the FP equation and is shown to visibly influence the steady state opinion profile. This macroscopic model is validated by comparing the numerical solution of PDE (1.1) with the numerical solution of the corresponding microscopic model described by a coupled system of stochastic differential equations.

Chapter 2: We first establish the mathematical properties of the FP equation (1.1). In particular, we show the well-posedness of PDE (1.1), i.e., the existence and uniqueness

of a classical solution $\rho \in C^1(0,\infty;C^2(\tilde{X}))$ for sufficiently smooth initial density ρ_0 and radical density ρ_r. Moreover, we establish an explicit lower bound on the noise level that guarantees exponential convergence of the dynamics to stationary state. To be precise, we show that $\rho(\cdot,t)$ converges to a stationary state $\rho^s \in C^2(\tilde{X})$ exponentially in L^2 as $t \to \infty$ if $\sigma > \sigma_s$, where $\sigma_s > 0$ uniquely solves

$$\sigma_s^2 = \frac{4R(3+M)}{\pi} + \frac{4R^2}{\pi\sqrt{3}} \exp\left(\frac{8R(1+M)}{\sigma_s^2}\right).$$

We then focus on the stationary state of the system, i.e., the solution to

$$\frac{\sigma^2}{2}\rho_{xx} + (\rho\, G_\rho)_x = 0.$$

In particular, the existence of a classical stationary solution $\rho^s \in C^2(\tilde{X})$ is shown for a sufficiently smooth radical density ρ_r. Moreover, a global estimate is provided that bounds the deviation of the stationary state from the uniform distribution. Precisely, we show that for any $\eta > 0$, if $\sigma^2 > \sigma_b^2 + \eta c_b$, then $\|\rho^s - 1\|_{L^2} \le \frac{1}{\eta}\|\rho_r\|_{L^2}$, where

$$\sigma_b^2 := \frac{4R}{\pi}\left(M + \frac{R}{\sqrt{3}} + 2\right) \text{ and } c_b := \frac{4R^2 M}{\pi\sqrt{3}}.$$

As we will see, combining the preceding result with the exponential stability of the dynamics yields the input-output stability of the system for sufficiently large noises.

Chapter 3: Next, we exploit the periodicity and evenness of the model (in space $\tilde{X}$) and use Fourier analysis to examine the structure of the opinion clusters under the uniform initial distribution $\rho_0 = 1$. This is where we work with a finite-dimensional approximation of PDE (1.1) in the dual (Fourier) domain. To be precise, we consider the finite Fourier expansions of ρ and ρ_r and obtain a system of quadratic ODEs

$$\dot{p}_n = c_n + b_n^T p + p^T Q_n p, \quad n = 1,\dots,N_f, \tag{1.3}$$

which describe the time evolution of the Fourier coefficients $p_n(t)$ of

$$\rho(x,t) = 1 + \sum_{n=1}^{N_f} p_n(t)\cdot\cos(\pi n x).$$

Then, we carry out a linear stability analysis on this finite-dimensional system of ODEs for identification of the so-called order-disorder transition in the system. (In the study of systems of noisy interacting particles, "order" refers to clustering behaviors and "disorder" refers to uniform behaviors). This is done by approximating the critical noise level at which this transition occurs. To be precise, we linearize the system at $t = 0$ to obtain the linear ODEs

$$\dot{p} = c + Bp, \tag{1.4}$$

with $p = (p_1,\dots,p_{N_f})^\top$. We then numerically compute the critical noise level above which this linear system is stable and converges to a stationary state close to uniform

distribution. We also provide another approximation scheme for characterizing the initial clustering behavior of the system including the number and the timing of possible clusters. This is done by further simplifying the linearized model. Precisely, we ignore the interactions between different frequencies in (1.4), and consider the equations

$$\dot{p}_n = c_n + \gamma_n p_n, \quad n = 1, \dots, N_f,$$

for the initial evolution of each Fourier coefficient p_n. As we will see, these simple numerical schemes, lead to a reasonably accurate prediction of the behavior of the system without the need to solve the equations describing the dynamics of the system.

PART TWO (CHAPTERS 4 AND 5)

The second part of the thesis revolves around the value iteration (VI) algorithm for solving optimal control problems of discrete-time systems with continuous state and input spaces. The VI algorithm simply involves the consecutive applications of the dynamic programming (DP) operator

$$\mathcal{T} J(x) = \min_u \{C(x,u) + \gamma \mathbb{E}[J(x^+)]\}, \tag{1.5}$$

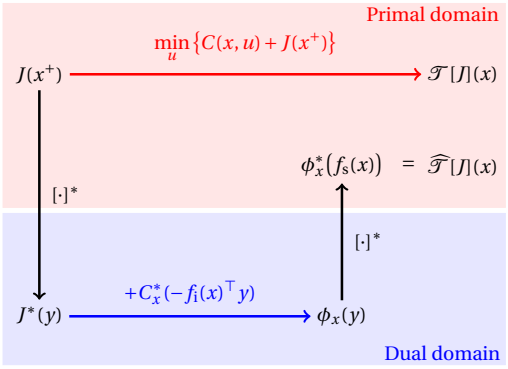
where $C(x,u)$ is the cost of taking the control action $u \in \mathbb{R}^m$ at the state $x \in \mathbb{R}^n$, and $\gamma \in (0,1)$ is the discount factor (in discounted cost problems). Arguably, the most important drawback of VI algorithm is its high computational cost for large-scale finite state spaces. For problems with a continuous state space, the DP operation becomes an infinite-dimensional optimization problem, rendering the exact implementation of VI impossible in most cases. A common approach is to incorporate function approximation techniques and compute the output of the DP operator for a finite sample (i.e., a discretization) of the underlying continuous state space. This approximation again suffers from a high computational cost for fine discretizations of the state space, particularly, in high-dimensional problems.

For some DP problems, however, it is possible to reduce this complexity by using duality, i.e., approaching the minimization problem (1.5) in the conjugate domain. E.g., for the deterministic linear dynamics $x^+ = Ax + Bu$ with the separable cost $C(x,u) = C_s(x) + C_i(u)$, we have

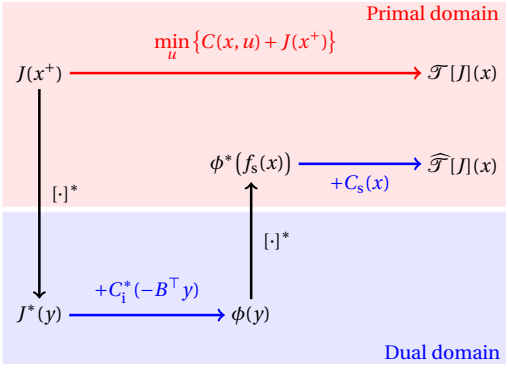
$$\mathcal{T} J(x) \geq C_s(x) + \left[C_i^*(-B^\top \cdot) + (\gamma J)^*\right]^*(Ax),$$

where the operator $[\cdot]^*$ is the (convex) conjugate transform. In particular, notice how the minimization in the primal domain in the DP operation can be transformed to a simple addition in the dual (conjugate) domain, at the expense of three conjugate transforms. Fundamentally, we will be exploiting the operational duality of infimal convolution and addition with respect to the conjugate transform.

In this part, we use duality and propose multiple conjugate VI (ConjVI) algorithms that involve a sample-based approximation using a finite subset $\mathbb{X}^g$ (the superscript g denotes *grid-like* finite sets) of the underlying continuous state space. These algorithms are based on an alternative path that solves the dual problem corresponding to the DP operation, by utilizing the linear-time Legendre transform (LLT) algorithm for discrete conjugation. In particular, the proposed approaches involve incorporating a



(a) First setting with dynamics $x^+ = f_s(x) + f_i(x) \cdot u$ and cost $C(x,u)$.



(b) Second setting with dynamics $x^+ = f_s(x) + B \cdot u$ and cost $C(x,u) = C_s(x) + C_i(u)$.

Figure 1.2: Sketch of the proposed ConjVI algorithms for deterministic dynamics – the standard DP operation in the primal domain (upper red paths) and the conjugate DP (CDP) operation through the dual domain (bottom blue paths).

finite-dimensional approximation of the value function in the dual domain, which leads to a convex, max-plus linear approximation (namely, the maximum of affine functions). Figure 1.2 shows the sketch of the proposed algorithms for the deterministic dynamics.
Chapter 4: We begin with the presentation and analysis of the basic algorithms for finite-horizon optimal control of deterministic systems. In particular, we introduce the discrete conjugate DP (d-CDP) operator for problems with input-affine dynamics

$$x^+ = f_s(x) + f_i(x) \cdot u.$$

See Figure 1.2a for the sketch of this operator. Precisely, the d-CDP operator $\widehat{\mathcal{T}}^d$ reads as (the superscript d denotes finite (discrete) sets and functions)

$$\begin{cases} J^{d*d}(y) = \max_{x \in \mathbb{X}^g} \{\langle y, x \rangle - J^d(x)\}, & y \in \mathbb{Y}^g, \\ \varphi_x^d(y) = C_x^*(-f_i(x)^\top y) + J^{d*d}(y), & y \in \mathbb{Y}^g, \\ \widehat{\mathcal{T}}^d J^d(x) = \varphi_x^{d*}\big(f_s(x)\big), & x \in \mathbb{X}^g, \end{cases}$$

where $C_x^*(v) := \max_u \{\langle v, u\rangle - C(x,u)\}$ is the conjugate of the cost with respect to the input variable and assumed to be analytically available. Note that d-CDP operator takes the discrete function $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ as the input, and outputs another discrete function $\widehat{\mathcal{T}}^{\mathrm{d}} J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. We also note that the operation $[\cdot]^{\mathrm{d}*\mathrm{d}}$ is the discrete conjugate operation that can be efficiently handled via the LLT algorithm for gridded dual domains. Here, we are particularly using the linearity of the dynamics in the input to effectively incorporate the operational duality of addition and infimal convolution, and transform the minimization of the DP operation into a simple addition at the expense of two discrete conjugate transforms. This, in turn, leads to a computational cost of $\mathcal{O}(XY)$ for the d-CDP operation, where X and Y denote the size of the discrete primal state space $\mathbb{X}^{\mathrm{g}}$ and discrete dual state space $\mathbb{Y}^{\mathrm{g}}$, respectively. We then modify the proposed d-CDP operator and reduce its time complexity for a subclass of problems with "separable" data in the state and input variables; see Figure 1.2b for the sketch of the modified operator. This subclass is most importantly identified by a state-independent input dynamics $f_{\mathrm{i}}(\cdot) = B \in \mathbb{R}^{n\times n}$ and separable cost $C(x,u) = C_{\mathrm{s}}(x) + C_{\mathrm{i}}(u)$. The *modified* d-CDP operator $\widehat{\mathcal{T}}_{\mathrm{m}}^{\mathrm{d}}$ reads as

$$\begin{cases} J^{\mathrm{d}*\mathrm{d}}(y) = \max_{x\in\mathbb{X}^{\mathrm{g}}} \{\langle y, x\rangle - J^{\mathrm{d}}(x)\}, & y \in \mathbb{Y}^{\mathrm{g}}, \\ \varphi^{\mathrm{d}}(y) := C_{\mathrm{i}}^*(-B^\top y) + J^{\mathrm{d}*\mathrm{d}}(y), & y \in \mathbb{Y}^{\mathrm{g}}. \\ \varphi^{\mathrm{d}*\mathrm{d}}(z) = \max_{y\in\mathbb{Y}^{\mathrm{g}}} \{\langle z, y\rangle - \varphi(y)\}, & z \in \mathbb{Z}^{\mathrm{g}}, \\ \widehat{\mathcal{T}}_{\mathrm{m}}^{\mathrm{d}} J^{\mathrm{d}}(x) = C_{\mathrm{s}}(x) + \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big), & x \in \mathbb{X}^{\mathrm{g}}, \end{cases}$$

where $\overline{[\cdot]}$ denotes the multi-linear interpolation operator. (Here, again, we are assuming the conjugate of input cost, $C_{\mathrm{i}}^*(v) := \max_u \{\langle v, u\rangle - C_{\mathrm{i}}(u)\}$, is analytically available). In particular, for this subclass, the time complexity of the DP operation reduces to $\widetilde{\mathcal{O}}(X + Y + Z)$, where Z is the size of the grid $\mathbb{Z}^{\mathrm{g}}$. (The notation $\widetilde{\mathcal{O}}$ hides the logarithmic terms). This, in turn, points to the possibility of a huge reduction in the computational cost for grids $\mathbb{Y}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$ of proper size. One of the most important aspects of our development is the error analysis of the proposed d-CDP operator and its modification. In particular, we use the results of our error analysis to provide concrete guidelines for the construction of the grids $\mathbb{Y}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$.

Chapter 5: We next discuss three extensions of the proposed (modified) d-CDP operator for the subclass of problems with separable data. First, we consider stochastic dynamics

$$x^+ = f_{\mathrm{s}}(x) + Bu + w,$$

where $w \in \mathbb{R}^n$ is an additive disturbance with a finite support $\mathbb{W}^{\mathrm{d}}$ of size W and a given probability mass function $p : \mathbb{W}^{\mathrm{d}} \to [0,1]$. Second, we propose a numerical scheme for computing the conjugate of the input cost C_{i}, using the discretization $C_{\mathrm{i}}^{\mathrm{d}} : \mathbb{U}^{\mathrm{g}} \to \mathbb{R}$ of this function over a grid-like discretization $\mathbb{U}^{\mathrm{g}}$ of the input space. (Recall that in Chapter 4, we assume C_{i}^* is analytically available). In particular, we consider the implications of these extensions on the complexity and the error of the d-CDP operation. Finally, we consider solving the infinite-horizon, discounted cost optimal control problem, which involves finding the fixed-point of the DP operator. This, in turn, requires us to provide a set of sufficient conditions for the convergence of the corresponding ConjVI algorithm, and to further extend our error analysis by considering the difference between the output of

this algorithm (after a finite number of iterations) with the true optimal value function. The *extended* (modified) d-CDP operator $\widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}}$ precisely reads as

$$\begin{cases} \varepsilon^{\mathrm{d}}(x) = \gamma \cdot \sum_{w \in \mathbb{W}^{\mathrm{d}}} p(w) \cdot \widetilde{J^{\mathrm{d}}}(x+w), & x \in \mathbb{X}^{\mathrm{g}}, \\ \varepsilon^{\mathrm{d}*\mathrm{d}}(y) = \max_{x \in \mathbb{X}^{\mathrm{g}}} \left\{ \langle x, y \rangle - \varepsilon^{\mathrm{d}}(x) \right\}, & y \in \mathbb{Y}^{\mathrm{g}}, \\ C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}(v) = \max_{u \in \mathbb{U}^{\mathrm{g}}} \left\{ \langle u, v \rangle - C_{\mathrm{i}}^{\mathrm{d}}(u) \right\}, & v \in \mathbb{V}^{\mathrm{g}}, \\ \varphi^{\mathrm{d}}(y) = \overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}(-B^{\top} y) + \varepsilon^{\mathrm{d}*\mathrm{d}}(y), & y \in \mathbb{Y}^{\mathrm{g}}, \\ \varphi^{\mathrm{d}*\mathrm{d}}(z) = \max_{y \in \mathbb{Y}^{\mathrm{g}}} \left\{ \langle y, z \rangle - \varphi^{\mathrm{d}}(y) \right\}, & z \in \mathbb{Z}^{\mathrm{g}}, \\ \widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}}(x) = C_{\mathrm{s}}(x) + \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big), & x \in \mathbb{X}^{\mathrm{g}}, \end{cases}$$

where $\widetilde{[\cdot]}$ is a generic extension of a discrete function. We show that, under some conditions on the sizes of the grids $\mathbb{Y}^{\mathrm{g}}$, $\mathbb{V}^{\mathrm{g}}$, and $\mathbb{Z}^{\mathrm{g}}$, the ConjVI algorithm that utilizes the extended d-CDP operator, has a one-time compilation complexity of $\mathscr{O}(X+U)$ and a per-iteration complexity of $\widetilde{\mathscr{O}}(XWE)$, where E denotes the cost of each evaluation of the extension operator $\widetilde{[\cdot]}$. Moreover, we again use the results of the error analysis to provide concrete guidelines for the construction of the grids $\mathbb{Y}^{\mathrm{g}}$, $\mathbb{V}^{\mathrm{g}}$, and $\mathbb{Z}^{\mathrm{g}}$.

Chapter 6 concludes the thesis by providing some remarks on the limitations of the proposed models/approaches. In this final chapter, we also discuss some interesting future research directions.

PART ONE

A MACROSCOPIC MODEL FOR OPINION DYNAMICS

2

Model, Well-posedness, and Stability

Parts of this chapter have been published in IEEE Transactions on Automatic Control **66**, 3 (2021) [1].

2

In this chapter, we introduce and study the nonlinear Fokker-Planck (FP) equation that arises as a mean-field (macroscopic) approximation of the bounded confidence opinion dynamics, where opinions are influenced by environmental noises and radical (stubborn) individuals. In particular, we focus on mathematical properties of the FP equation such as well-posedness and stability. The chapter is organized as follows. A review of the related literature along with the motivation is provided in Section 2.1. We then present our general notational conventions and some preliminaries on function spaces in Section 2.2. The macroscopic opinion dynamics model in question is then introduced in Section 2.3. We next present our main theoretical results regarding the well-posedness and stability of the model in Section 2.4. The final section of this chapter concerns the technical proofs of these results.

2.1. MOTIVATION AND LITERATURE REVIEW

Recent decades have witnessed enormous progress in the study of complex systems and their system-theoretic properties [2, 3]. The main effort has been invested in the study of "self-organization" and "spontaneous order" phenomena [4] that have inspired the development of synchronization and consensus theory [5, 6]. Paradoxically, these regular behaviors arising from local interactions between subsystems (agents, nodes) of a complex system are studied much better than various "irregular" dynamic effects such as persistent disagreement and clustering, exhibited by many real-world systems. Although some culprits of this asynchrony and dissent (e.g. symmetries and other special structures in the coupling mechanisms, exogenous forces acting on some nodes, heterogeneous dynamics of nodes, etc.) have been revealed in the literature [7–11], only a few mathematical models have been proposed that are sufficiently "rich" to capture the diversity of clustering behaviors in real-world networks and, at the same time, admit rigorous analysis. Long before the recent "boom" in complex systems, the lack of such models was realized in mathematical sociology. The problem of disclosing mechanisms preventing consensus and maintaining enduring disagreement between individuals [12] is nowadays referred to as the community cleavage problem or Abelson's diversity puzzle [13, 14]. The interdisciplinary area of sociodynamical modeling [14–21] has attracted enormous attention of the research community and is primarily concerned with mechanisms of opinion formation under social influence.

Only a few models, proposed in the literature to describe opinion formation processes, have been secured by experimental evidence. Such models, however, play an important role and contribute, in various aspects, in comprehending complex systems' behaviors such as birth, death, and evolution of clusters in systems of interacting particles, and in developing algorithms for control of these behaviors. This explains the explosion of interest in models of opinion formation in systems and control literature. From the control-theoretic prospect, most of these models are simply networks of interacting agents, obeying the first-order integrator model. However, the term "opinion" is now widespread and used to denote the scalar or multi-dimensional state of an agent, even if this state does not have a clear sociological interpretation[1] (belonging, e.g., to an

[1] From the sociological viewpoint, opinions are cognitive orientations of individuals towards some objects or topics [14].

abstract manifold [22]). The opinion is thus some value of interest, held by an agent and updated, based on displayed opinions of the other agents.

Linear models of opinion dynamics, extending the classical French-DeGroot system in various directions (allowing, e.g., stubborn agents, asynchronous interactions, and repulsion of opinions [14, 18, 23, 24]) have been thoroughly studied. These models are sufficient to explain consensus and disagreement in social groups, as well as the formation of special opinion profiles (e.g., bimodal distributions, standing for opinion polarization), however, general mechanisms leading to emergence and destruction of unequal clusters are still far from being well understood. To explain them, more complicated nonlinear models have been proposed, mimicking some important features of social influence. One feature observed in social and biological systems is the *homophily* [25], or the tendency of individuals to bond with similar ones. Homophily is related to *biased assimilation* [26] effects: individuals readily accept opinions consistent with their views and tend to dismiss and discount opinions contradicting their own views. Mathematically, coupling between close opinions is stronger than that of distant opinions, which is modeled by introducing opinion-dependent influence weights. Although the possibility of such nonlinearities in opinion dynamics models was mentioned in the pioneering work [12], substantial progress has been primarily achieved in the analysis of bounded confidence models proposed several decades later as extensions of the deterministic [27] and randomized gossip-based [28] consensus algorithms for multi-agent networks. Bounded confidence models stipulate that a social actor is insensitive to opinions beyond its bounded confidence set (usually, this set is an open or closed ball, centered at the actor's own opinion), which makes the graph of interactions among the agents distance-dependent. A detailed survey of bounded confidence models and relevant mathematical results can be found in [19]. Bounded confidence models exhibit convergence of the opinions to some steady values, which can reach consensus or split into several disjoint clusters. If the state-dependent interaction graph of the system is symmetric, this follows from general properties of iterative averaging procedures, and can alternatively be proved by exploring a special Lyapunov function ("kinetic energy") [19, 29, 30]. In the general case of asymmetric interaction graphs, such a convergence has been proved only in special situations [30, 31], but seems to be a generic behavior [31–33].

Opinions in real social groups, however, usually do not terminate at steady values yet oscillate, which is usually explained by two factors. The first reason explaining opinion fluctuation is exogenous influence, which can be interpreted as some "truth" available to some individuals [34] or a position shared by a group of close-minded opinion leaders or stubborn individuals ("radicals") [35–37]. Important results on the stability of the Hegselmann-Krause (HK) model with radicals and more general "inertial" bounded confidence models were obtained in [31]. Typically, the exogenous signal is supposed to change slowly compared to the opinion evolution and is thus replaced by a constant; the main concern is the dependence between the constant input and the resulting opinion profile. Numerical results, reported in [35, 36] demonstrate high sensitivity of the opinion clusters to the radical's opinion and reveal some counter-intuitive effects, e.g., an increase in the number of radicals sometimes decreases the number of their followers. The second culprit of persistent opinion fluctuation is uncertainty in the opinion

dynamics, usually modeled as a random drift of each opinion. The presence of a random excitation can be interpreted as "free will" and unpredictability of a human's decision [38]; besides this, randomized opinion dynamics models are broadly adopted in statistical physics [39–42] to study phase transitions in systems of interacting particles.

Even for the classical models from [27, 28], disclosing the relationship between the initial and the terminal opinion profiles remains a challenging problem (including, e.g., the $2R$-conjecture [43, 44]). In presence of noise, the analysis becomes even more difficult; some progress in the study of the interplay between confidence range and noise level have been achieved in recent works [45, 46]. One of the important directions in the analysis of bounded confidence models is the examination of their asymptotic properties as the number of social actors becomes very large ($N \to \infty$) and their individual opinions are replaced by infinitesimal "elements". The arising macroscopic approximations of agent-based models describe the evolution of the distribution of opinion (usually supposed to have a density) and are referred to as density-based [47], continuum-agent [48, 49], Eulerian [50, 51], kinetic [52], hydrodynamical [29] or mean-field [44, 53] models of opinion formation. In the continuous-time situation, the density obeys a nonlinear FP equation. To study the clustering behavior of the macroscopic bounded confidence models, efficient numerical methods have been proposed that are based on Fourier analysis [41, 44, 54].

From a practical viewpoint, it is convenient to consider opinions staying in a predefined interval, e.g., $[0,1]$. The HK and Deffuant-Weisbuch (DW) models, as well as their continuous-time counterparts [19], imply that starting within the interval, opinions never escape from it. This property, however, is destroyed by arbitrarily small noises. To keep the opinions bounded, some boundary conditions are usually introduced. The absorbing boundary condition assumes that the opinions are saturated at the extreme values 0 and 1 [41, 46]; an important result from [46] demonstrates that arbitrarily small noises in this situation destroy clusters and lead to approximate consensus (the maximal deviation of opinions is proportional to the noise level). More interesting are opinion dynamics with the periodic boundary condition, wrapping the interval $[0,1]$ into a circle. The opinion density on the circle corresponds to a 1-*periodic* solution of the FP equation on the real line [44, 54, 55]. A disadvantage of the simple periodic boundary condition is the merging of two extreme opinion values 0 and 1. To distinguish between these extreme opinions, we incorporate an "almost" reflective (precisely, an *even* 2-*periodic*) boundary condition. Dealing with the macroscopic FP equation, the opinion density is then conveniently represented by an even 2-periodic solution on the real line. We are primarily concerned with the mathematical properties of such solutions.

In the first part of the thesis (Chapters 2 and 3), we advance the theory of macroscopic modeling of bounded confidence dynamics. We consider a bounded confidence model with environmental noise which also includes radical opinions, which are not concentrated at a single point (as in [34, 35, 50]) but rather distributed. The FP equation acquires an (infinite-dimensional) exogenous input, describing the density and total mass of the radical opinions. This setup allows us to consider the interplay between the noise and the distributed radicals concerning the behavior of the system.

2.2. NOTATIONS AND PRELIMINARIES

2.2.1. GENERAL NOTATIONS

The convolution of two functions f and g is denoted by $f \star g = \int f(x)\ g(y-x)\ \mathrm{d}y$. We note that in our case one of the functions has a compact support, so the integral always exists. For a function $f(t,x)$ we use f_x (respectively, f_t) to denote the derivatives with respect to x (respectively, t), so that f_{xx} is the second partial derivative with respect to x. We also use the notation $\partial_x^i f$ for the i-th order derivative with respect to x.

Let $X = [0,1]$ and $\tilde{X} = [-1,1]$. We use $\mathscr{P}(X)$ to denote the the space of probability densities on X. That is, $\rho \in \mathscr{P}(X)$ if $\int_X \rho(x)\ \mathrm{d}x = 1$ and $\rho(x) \geq 0$ for all $x \in X$. We also use $\mathscr{P}_e(\tilde{X})$ to denote the space of probability densities on X, extended evenly to $\tilde{X}$. That is, $\mathscr{P}_e(\tilde{X})$ is the space of all functions $\rho : \tilde{X} \to [0,\infty)$ such that $\int_X \rho(x)\ \mathrm{d}x = 1$ and $\rho(x) = \rho(-x) \geq 0$ for all $x \in \tilde{X}$.

2.2.2. REVIEW OF FUNCTION SPACES

The definitions provided here are mostly borrowed from [56]. Let $\{f_k\}_{k=1}^{\infty}$ be a sequence in a Banach space B with norm $\|\cdot\|_B$. The *strong* convergence $f_k \to f$ implies $\|f_k - f\|_B \to 0$, while the *weak* convergence $f_k \rightharpoonup f$ implies $g(f_k) \to g(f)$ for all bounded linear functionals $g : B \to \mathbb{R}$.

Let $f : \tilde{X} \to \mathbb{R}$ be a measurable function on $\tilde{X} = (-1,1)$. The L^p-norm of f is

$$\|f\|_{L^p(\tilde{X})} = \begin{cases} \left(\int_{\tilde{X}} |f(x)|^p\right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \operatorname{ess\,sup}_{\tilde{X}} |f(x)|, & p = \infty. \end{cases}$$

Then, $L^p(\tilde{X})$ denotes the Banach space of all measurable functions $f : \tilde{X} \to \mathbb{R}$ for which $\|f\|_{L^p(\tilde{X})} < \infty$. Let $f, g \in L^1_{loc}(\tilde{X})$ be locally summable functions (i.e., f, g have a finite integral over every compact subset of $\tilde{X}$). We say that g is the k-th weak (partial) derivative of f, if

$$\int_{\tilde{X}} f\ \partial_x^k \phi\ \mathrm{d}x = (-1)^k \int_{\tilde{X}} g\ \phi\ \mathrm{d}x,$$

for all test functions $\phi \in C_c^{\infty}(\tilde{X})$ (infinitely differentiable functions $\phi : \tilde{X} \to \mathbb{R}$ with compact support in $\tilde{X}$). $H^k(\tilde{X})$ for $k \in \mathbb{N}$ is used to denote the Sobolev space $W^{k,2}(\tilde{X})$ consisting of functions $f \in L^2(\tilde{X})$ whose weak derivatives up to order k exist and belong to $L^2(\tilde{X})$. Note that $H^k(\tilde{X})$ is a Hilbert space. We use the subscript *per* to denote the closed subspace of *periodic* functions in the corresponding function space, e.g.,

$$L^p_{per}(\tilde{X}) = \{f \in L^p(\tilde{X}) : f(-1) = f(1)\},$$
$$H^k_{per}(\tilde{X}) = \{f \in H^k(\tilde{X}) : f(-1) = f(1)\}.$$

Similarly, we use the subscript *ep* to denote the closed subspace of *even periodic* functions in the corresponding function space, e.g.,

$$L^p_{ep}(\tilde{X}) = \{f \in L^p_{per}(\tilde{X}) : f(-x) = f(x),\ \forall x \in \tilde{X}\},$$
$$H^k_{ep}(\tilde{X}) = \{f \in H^k_{per}(\tilde{X}) : f(-x) = f(x),\ \forall x \in \tilde{X}\}.$$

We denote the dual space of $H^1_{per}(\tilde{X})$ by $H^{-1}_{per}(\tilde{X})$, that is, the space of bounded linear functionals on $H^1_{per}(\tilde{X})$. Moreover, we use $\langle\cdot,\cdot\rangle$ to denote the corresponding paring of $H^1_{per}(\tilde{X})$ and $H^{-1}_{per}(\tilde{X})$. That is, for $f \in H^1_{per}(\tilde{X})$ and $g \in H^{-1}_{per}(\tilde{X})$, we use $\langle g, f\rangle$ to denote the real number $g(f)$. Since periodic boundary condition allows for integration by parts without extra terms, $H^{-1}_{per}(\tilde{X})$ has most of the properties of the space $H^{-1}(\tilde{X})$, the dual space of $H^{-1}_0(\tilde{X})$; see [56, Sec. 5.9.1] for a detailed description of the space $H^{-1}(\tilde{X})$. In particular, one can extend the result in [56, Sec. 5.9, Thm. 3] to derive [55, Thm. 3.8]. For the reader's convenience, the corresponding theorem is presented below.

Theorem. *[55, Thm. 3.8] Let the function $f : \tilde{X} \times [0,T] \to \mathbb{R}$ be such that*

$$f \in L^2(0,T; H^1_{per}(\tilde{X})) \text{ and } f_t \in L^2(0,T; H^{-1}_{per}(\tilde{X})).$$

Then, $f \in C(0,T; L^2_{per}(\tilde{X}))$ after possibly being redefined on a set of measure zero. Moreover, the mapping $t \mapsto \|f(t)\|^2_{L^2(\tilde{X})}$ is absolutely continuous, with

$$\frac{\mathrm{d}}{\mathrm{d}t}\|f(t)\|^2_{L^2(\tilde{X})} = 2\langle f_t, f\rangle,$$

for almost every $t \in [0,T]$.

2.3. MACROSCOPIC MODEL OF OPINION FORMATION

The conventional bounded confidence model describes opinion formation process in a network of $N > 1$ agents. All agents have the same *confidence range* $R > 0$. Agent i's opinion at time $t \geq 0$, denoted by $x_i(t) \in \mathbb{R}$, is (directly) influenced only by the opinions of agents j, such that $|x_j(t) - x_j(t)| \leq R$. One of the simplest continuous-time bounded confidence models is [29]

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1}^{N} w\big(x_j(t) - x_i(t)\big), \quad w(\xi) = \begin{cases} \xi, & |\xi| \leq R \\ 0, & |\xi| > R. \end{cases} \tag{2.1}$$

It can be shown [19] that the opinions obeying the model (2.1) always converge: $x_i(t) \to x_i^s$ as $t \to \infty$, with $w(x_i^s - x_j^s) = 0$ for all i, j. This corresponds to either consensus ($x_i^s = x_j^s$ for all i, j) of the terminal opinions or their splitting into clusters, comprising one or several coincident opinions. In the latter situation, the distance between every two clusters is greater than R.

Dynamics of real opinions (and other physical processes, portrayed by "opinion dynamics" models) often do not exhibit convergence to steady values, and the fluctuation of opinions persists. In order to capture this effect, random uncertainties can be introduced into the model mimicking "free will" and the unpredictability of a human's decision [38]. The simplest of these uncertainties is an additive random noise. The model (2.1) is then replaced by the system of nonlinear stochastic differential equations (SDEs)

$$\mathrm{d}x_i(t) = \frac{1}{N}\sum_{j=1}^{N} w\big(x_j(t) - x_i(t)\big)\mathrm{d}t + \sigma \mathrm{d}W_i(t), \tag{2.2}$$

where W_i are independent standard Wiener processes and $\sigma > 0$ is the noise level.

Since the dynamics of the stochastic system (2.2) becomes quite complicated as the number of agents grows, the standard approach to examine it is the mean-field (or macroscopic) approximation, considering the opinion profile $(x_i(t))_{i=1}^N$ as a *random sampling* drawn from some (time-varying) probability distribution of the opinion. Precisely, it can be shown [57–59] that empirical distributions $N^{-1}\sum_{i=1}^N \delta_{x_i(t)}$ converge (in the weak sense) as $N \to \infty$ to a distribution, whose density $\rho(t,x)$ obeys the FP equation

$$\rho_t = \big[\rho\,(w \star \rho)\big]_x + \tfrac{\sigma^2}{2}\rho_{xx}, \quad t \geq 0,\ x \in \mathbb{R}. \tag{2.3}$$

An extension of the bounded confidence dynamics allows the presence of $N_r \geq 1$ "radicals" (stubborn agents, zealots) that do not assimilate others' opinions, however, influence them directly or indirectly. Typically, the radicals' opinions are supposed to be constant (or changing very slowly compared to the opinion formation of "normal" agents). Indexing the normal individuals 1 through N and the radicals $(N+1)$ through $(N+N_r)$, the opinion dynamics becomes

$$\begin{aligned} \mathrm{d}x_i(t) &= \frac{1}{N}\sum_{j=1}^{N+N_r} w\big(x_j(t) - x_i(t)\big)\mathrm{d}t + \sigma \mathrm{d}W_i(t), \quad i = 1,\dots,N \\ \dot{x}_i(t) &= 0, \quad i = N+1,\dots,N+N_r. \end{aligned} \tag{2.4}$$

Often it is supposed that the radicals share a common opinion $x_i \equiv T$ for $i = N+1,\dots,N+N_r$, which may also be considered as some "truth" perceived by some individuals [34] or, more generally, an exogenous signal [35]. The ratio $M = N_r/N$ can be treated as the relative "weight" or "strength" of this external opinion. More generally, one can assume that the radicals' opinions are spread over $\mathbb{R}$. Supposing that $N, N_r \to \infty$, the relative mass of the radicals M remains constant, and their empirical distribution $N_r^{-1}\sum_{i=1}^{N_r}\delta_{x_{N+i}}$ converges (in the weak sense) to a distribution with sufficiently smooth density ρ_r, the density of the normal opinions obeys the modified FP equation

$$\rho_t = \big[\rho\,(w \star (\rho + M\rho_r))\big]_x + \tfrac{\sigma^2}{2}\rho_{xx}, \quad t \geq 0,\ x \in \mathbb{R}. \tag{2.5}$$

Note that the classical bounded confidence dynamics (2.1), being a special case of continuous-time consensus protocol, has an important property: the minimal opinion $\min_i x_i(t)$ and the maximal opinion $\max_i x_i(t)$ are, respectively, non-decreasing and non-increasing. In particular, if the initial opinions are confined to some predefined interval, e.g., $x_i(0) \in [0,1]$, then one has $x_i(t) \in [0,1]$ for all $t \geq 0$. The additive noise leads to random drift of the opinion profile, thus destroying the latter important property. Since in practice bounded ranges of opinions are usually considered, the dynamics (2.2) and (2.4) are usually complemented by boundary conditions [41], preventing the opinions from escaping from the predefined range.

A typical boundary condition is the *periodic* condition, where the opinion domain $[0,1]$ is wrapped on a circle of circumference 1 (formally, replacing a real opinion value $x \in \mathbb{R}$ by its fractional part $\{x\} = x - \lfloor x \rfloor = x \mod 1$). A disadvantage of the periodic boundary condition is that there is no distinction between the extreme opinions 0 and 1. We address this issue by considering another type of boundary condition, which we

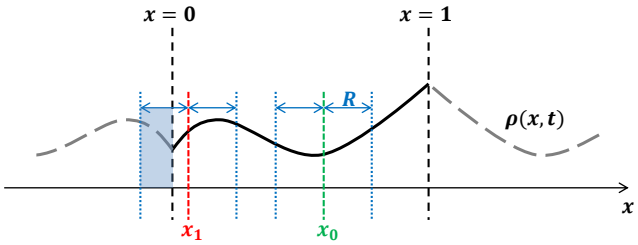


Figure 2.1: The even 2-periodic extension of the system. The opinion value $x_0 \in [R, 1-R]$ effectively experiences a reflective boundary condition, while for the opinion value $x_1 \in [0, R]$ there is also a boundary effect due to the even extension. In particular, the influence of more extreme neighbors of x_1 is reinforced by introducing artificial ones (the shaded area in blue). The same boundary effect exists for opinion values in $[1-R, 1]$.

call *even 2-periodic*. Precisely, a real opinion $x \in \mathbb{R}$ is replaced by $f(x)$, where f is an even 2-periodic function, such that $f(x) = x$ on $[0,1]$ (and hence $f(x) = -x$ for $x \in [-1,0]$, $f(x) = 2-x$ for $x \in [1,2]$ and so on). In other words, we first evenly extend the opinion domain $[0,1]$ into the interval $[-1,1]$ and then wrap it on a circle of circumference 2 so that the extreme opinions 0 and 1 correspond to the antipodes of this circle. We note that with this even 2-periodic extension, the "effective" boundary condition experienced by the agents is an "almost" reflective one, that is, when an agent leaves the opinion domain from one end, it is reflected back into the domain from the same end. This is different from the behavior under simple periodic boundary condition where the agents leaving the domain from one end, enter the domain from the other end. However, the introduced boundary condition is almost reflective since the even extension causes some boundary effects: the influence of more extreme neighbors of opinion values in the R-neighborhood of extreme opinions 0 and 1 is reinforced. This is due to the even extension which introduces more extreme "artificial" neighbors; see Fig. 2.1.

As discussed in [44, 54, 55], the FP equation (2.3) under the periodic conditions retains its validity, however, $\rho(t,x)$ is not a probability density on $\mathbb{R}$ but a 1-periodic function $\rho(t,x+1) = \rho(t,x) \geq 0$, such that $\int_0^1 \rho(t,x)\mathrm{d}x = 1$ (that is, $\rho(t,\cdot)$ serves as a density on the interval $[0,1]$). Similarly, for the even 2-periodic boundary condition, the equation (2.3) retains its validity when we replace the probability density $\rho(t,x)$ with an even 2-periodic function, that is, $\rho(t,-x) = \rho(t,x)$ and $\rho(t,x+2) = \rho(t,x)$. On the interval $[0,1]$, the function $\rho(t,\cdot)$ again serves as a probability density: $\int_0^1 \rho(t,x)\mathrm{d}x = 1$. We also assume that the initial density $\rho_0(x) = \rho(0,x)$ and the density of radical opinions $\rho_r(x)$, defined on $[0,1]$, are extended (in the unique possible way) to even 2-periodic functions.

Without loss of generality, we take $X = [0,1]$ and $\tilde{X} = [-1,1]$ to be the bounded opinion domain and its even extension, respectively. To summarize the discussion above, the macroscopic model for opinion dynamics is fully described by the following partial differential equation (PDE)

$$\begin{cases} \rho_t = (\rho\, G_\rho)_x + \frac{\sigma^2}{2}\rho_{xx} & \text{in} \quad \tilde{X} \times (0,T) \\ \rho(\cdot+2,t) = \rho(\cdot,t) & \text{on} \quad \partial\tilde{X} \times (0,T) \\ \rho(x,\cdot) = \rho_0(x) & \text{on} \quad \tilde{X} \times \{t=0\}, \end{cases} \tag{2.6}$$

where

$$G_\rho(x,t) := w(x) \star \big(\rho(x,t) + M\rho_r(x)\big). \tag{2.7}$$

Note that in (2.6), we are considering the dynamics over a finite time horizon T for the sake of analysis, however, T can be chosen arbitrarily large. We again emphasize that the initial density ρ_0 and the radical density ρ_r are the unique even 2-periodic extensions of the corresponding densities from X to $\tilde{X}$. In essence, we are considering the same dynamics as in [55] with the extra requirement for ρ_0 (and the newly introduced density ρ_r) to be even. Finally, we note that [60] also provides a detailed treatment of these dynamics (without radicals) for a class of interaction potentials on a torus in higher dimensions.

2.4. MAIN THEORETICAL RESULTS

To recapitulate, we are interested in even 2-periodic solutions of PDE (2.6), where ρ_0 and ρ_r are even 2-periodic. A natural question arises whether the model is well-posed in the sense that every (sufficiently smooth) initial condition ρ_0 and input ρ_r correspond to a unique solution. The affirmative answer is given in the following theorem.

Theorem 2.4.1 (Well-posedness of dynamics)**.** *Let the initial density of normal opinions and the radical opinions density satisfy $\rho_0 \in H^3_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ and $\rho_r \in H^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$, respectively. Then, PDE (2.6) has a unique, even, strictly positive, classical solution $\rho \in C^1(0,\infty; C^2_{ep}(\tilde{X}))$ such that $\rho(t) \in \mathscr{P}_e(\tilde{X})$ for all $t > 0$.*

This result implies that $\rho(t) := \rho(\cdot, t)$ is a (strictly positive) probability density on $X = [0,1]$ for all $t > 0$, as required. For the autonomous systems (without radicals), [55, 60] provide a sufficient condition for exponential convergence of the dynamics towards uniform distribution $\rho = 1$ as an equilibrium of the system. Unlike those studies, the uniform distribution is not an equilibrium of the model that we consider. However, it is possible to extend this stability result to our model including the exogenous input, i.e., the radicals. To this end, we first consider the stationary equation corresponding to PDE (2.6) given by

$$\frac{\sigma^2}{2}\rho_{xx} + (\rho\ G_\rho)_x = 0. \tag{2.8}$$

We are particularly interested in even stationary solutions $\rho^s \in \mathscr{P}_e(\tilde{X})$ of (2.8). Our next result characterizes the stationary state of the system.

Theorem 2.4.2 (Stationary behavior)**.** *Let $\rho_r \in H^1_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ be the radical density.*

- **Existence.** *The stationary equation (2.8) has an even, strictly positive, classical solution $\rho^s \in C^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$.*
- **Estimate.** *For any $\eta > 0$, if $\sigma^2 > \sigma_b^2 + \eta c_b$, then $\|\rho^s - 1\|_{L^2} \le \frac{1}{\eta}\|\rho_r\|_{L^2}$, where*

$$\sigma_b^2 := \frac{4R}{\pi}\left(M + \frac{R}{\sqrt{3}} + 2\right) \quad \text{and} \quad c_b := \frac{4R^2M}{\pi\sqrt{3}}. \tag{2.9}$$

Notice how the global estimate in the preceding theorem bounds the difference between the stationary solution and the uniform distribution. This result shows that, even

in presence of radical opinions, the stationary solution can be made arbitrarily close to the uniform distribution by increasing the noise level beyond a minimum level σ_b. We note that the minimum noise level σ_b is directly related to the confidence range R and the relative mass M of radicals. Also, as the "energy" $M\|\rho_r\|_{L^2}$ of the radicals increases, in order to counteract their effect and keep the stationary profile in a (close to) uniform state, one must increase the noise level further beyond σ_b.

With this result in hand, we can now consider the asymptotic stability of stationary state. The next result provides a sufficient condition for exponential convergence of the dynamics to stationary state for arbitrary (and sufficiently smooth) initial density ρ_0 and radical density ρ_r.

Theorem 2.4.3 (Stability)**.** *Let $\rho_0 \in H^3_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ be the initial density of normal opinions and $\rho_r \in H^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ be the radical opinions density. Also, let $\rho \in C^1(0,\infty; C^2_{ep}(\tilde{X}))$ with $\rho(t) \in \mathscr{P}_e(\tilde{X})$ be the solution to the dynamic equation (2.6). Then, $\rho(t)$ converges to a stationary state $\rho^s \in C^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ exponentially in L^2 as $t \to \infty$ if $\sigma > \sigma_s$, where $\sigma_s > 0$ uniquely solves*

$$\sigma_s^2 = \frac{4R(3+M)}{\pi} + \frac{4R^2}{\pi\sqrt{3}} \exp\left(\frac{8R(1+M)}{\sigma_s^2}\right). \tag{2.10}$$

An immediate result of Theorems 2.4.2 and 2.4.3 is that for sufficiently large noises, the dynamics will converge to a stationary state that can be made arbitrarily close to uniform distribution by increasing the noise level.

Corollary 2.4.4 (Input-output stability)**.** *For any $\eta > 0$, if $\sigma^2 > \max\{\sigma_b^2 + \eta c_b, \sigma_s^2\}$, where σ_b and c_b are defined in (2.9) and $\sigma_s > 0$ uniquely solves (2.10), then it holds that*

$$\|\rho(t) - 1\|_{L^2} \leq \beta e^{-\lambda t} + \frac{1}{\eta}\|\rho_r\|_{L^2}, \tag{2.11}$$

where the constant $\beta > 0$ depends on ρ_0 and ρ_r and the convergence rate $\lambda > 0$ depends on σ, R, and M.

Remark 2.4.5 (Connection to existing works)**.** *The stability result of Corollary 2.4.4 corresponds to the result reported in [55, Thm. 2.3] on the global stability of uniform distribution $\rho = 1$ for sufficiently large noises in the autonomous system* without *radicals. In particular, by setting $M = 0$ in the estimate given in Theorem 2.4.2, one has $c_b = 0$, hence $\rho^s = 1$ is the unique stationary state of the system for $\sigma^2 > \sigma_b^2 = \frac{4R}{\pi}\left(2 + R/\sqrt{3}\right)$. We note that σ_b is the same minimum noise level given in [55, Thm. 2.3], taking into account a multiplicative factor of two due to the even extension in our model. However, direct application of Theorem 2.4.3 for stability of $\rho^s = 1$ leads to a sufficient minimum noise level $\sigma_s > \sigma_b$. This is because this result is based on conservative estimates for ρ^s. Indeed, if one incorporates the fact that $\rho^s = 1$ and modifies some of the arguments provided in the proof of Theorem 2.4.3 in Section 2.5.3, then one can show that, in the absence of radical agents, the uniform distribution $\rho^s = 1$ is also globally exponentially stable for $\sigma > \sigma_b$, reproducing the result of [55, Thm. 2.3].*

Finally, we note that, based on the results provided in [60], the input-output stability result of Corollary 2.4.4 can be generalized to multi-dimensional first-order stochastic interacting particle systems for a particular class of interaction potentials.

In the remainder of this chapter, we provide the technical proofs of the theoretical results listed above.

2.5. TECHNICAL PROOFS

2.5.1. WELL-POSEDNESS OF DYNAMICS

This section is devoted to the proof of Theorem 2.4.1 concerning the well-posedness of the dynamics (2.6). Throughout this section, all the norms are with respect to $\tilde{X} = [-1,1]$ (as opposed to $X = [0,1]$), unless indicated otherwise. We use $C, C_0, C_1, \dots$ to represent a generic constant (depending on the model parameters) whose actual values may change from line to line. In the case these constants depend on a particular object of interest, say θ, this dependence is explicitly indicated by $C[\theta]$.

Let us first note that because of periodicity, the mass is preserved in (2.6), that is,

$$\int_{\tilde{X}} \rho(x,t)\,\mathrm{d}x = \int_{\tilde{X}} \rho_0(x)\,\mathrm{d}x = 2,$$

for all $t \geq 0$. In particular, we have

$$\|\rho(t)\|_{L^1} \geq \int_{\tilde{X}} \rho(x,t)\,\mathrm{d}x = 2 > 0.$$

We will be using this property in the sequel.

We start by presenting some useful estimates for the object G_ρ defined in (2.7) that make it possible to extend the results provided by [55] to our model.

Lemma 2.5.1 (Estimates for G_ρ)**.** *Let G_ρ be the function defined in (2.7) with $\rho_r \in \mathscr{P}_e(\tilde{X})$. If $\rho(t) \in L^1_{per}(\tilde{X})$, then*

$$\|G_\rho\|_{L^\infty} \leq R\big(\|\rho(t)\|_{L^1} + 2M\big). \tag{2.12}$$

If, moreover, $\|\rho(t)\|_{L^1} > 0$, then

$$\|G_\rho\|_{L^\infty} \leq C\,\|\rho(t)\|_{L^1} \leq C\,\|\rho(t)\|_{L^2}. \tag{2.13}$$

Proof. Notice

$$\begin{aligned}
|G_\rho(x,t)| &= \left| \int (x-y)\,\mathbf{1}_{|x-y|\leq R}\,\big(\rho(y,t) + M\rho_r(y)\big)\,\mathrm{d}y \right| \\
&\leq \int |x-y|\,\mathbf{1}_{|x-y|\leq R}\,\big|\rho(y,t) + M\rho_r(y)\big|\,\mathrm{d}y \\
&\leq R \int_{\tilde{X}} \big|\rho(y,t) + M\rho_r(y)\big|\,\mathrm{d}y \\
&\leq R\left(\int_{\tilde{X}} |\rho(y,t)|\,\mathrm{d}y + 2M \right),
\end{aligned}$$

from which we can conclude the inequality (2.12). The first inequality in (2.13) then immediately follows from (2.12) and the assumption $\|\rho(t)\|_{L^1} > 0$. For the second inequality in (2.13) notice that since $\tilde{X}$ is of finite measure $\mu(\tilde{X}) = \mu([-1,1]) = 2$, for any measurable function v we have

$$\|v\|_{L^p(\tilde{X})} \leq \mu(\tilde{X})^{\frac{1}{p}-\frac{1}{q}}\,\|v\|_{L^q(\tilde{X})}, \tag{2.14}$$

where $1 \leq p \leq q \leq \infty$. □

Using estimate (2.13) in Lemma 2.5.1, one can follow similar arguments as in [55, Lem. 2.1] to show $\|\rho(t)\|_{L^1} = 2$ and $\rho(t) \geq 0$ for all $t \geq 0$; see also [55, Cor. 2.2]. Specifically, assuming PDE (2.6) has a solution $\rho \in C^1(0,T;C^2_{per}(\tilde{X}))$, one can derive a priori estimate which in turn implies that the solution is non-negative so that $\rho(t)$ is a probability distribution on $X = [0,1]$ for all $t \geq 0$.

Lemma 2.5.2 (Estimates for $\partial_x^k G_\rho$)**.** *Consider G_ρ in (2.7) with $\rho_r \in \mathscr{P}_e(\tilde{X})$.*

- *For $1 \leq p \leq \infty$, if $\rho(t), \rho_r \in L^p_{per}(\tilde{X})$ with $\|\rho(t)\|_{L^1} > 0$, then*

$$\|(G_\rho)_x\|_{L^p} \leq C_1 \|\rho(t)\|_{L^p} + C_2 \|\rho_r\|_{L^p} \leq C[\|\rho_r\|_{L^p}] \|\rho(t)\|_{L^p}. \tag{2.15}$$

- *For $k \geq 2$, if $\rho(t), \rho_r \in H^{k-1}_{per}(\tilde{X})$ with $\|\rho(t)\|_{L^1} > 0$, then*

$$\|\partial_x^k G_\rho\|_{L^2} \leq C[\|\rho_r\|_{H^{k-1}}] \|\rho(t)\|_{H^{k-1}}. \tag{2.16}$$

Proof. We have

$$\begin{aligned}
\big(G_\rho(x,t)\big)_x =& \partial_x \left(\int (x-y)\, \mathbf{1}_{|x-y|\leq R} \left(\rho(y,t) + M\rho_r(y)\right) \mathrm{d}y \right) \\
=& \partial_x \left(\int_{x-R}^{x+R} (x-y) \left(\rho(y,t) + M\rho_r(y)\right) \mathrm{d}y \right) \\
=& -R\big(\rho(x+R,t) + \rho(x-R,t) + M\rho_r(x+R) + M\rho_r(x-R)\big) \\
&+ \int_{x-R}^{x+R} \big(\rho(y,t) + M\rho_r(y,t)\big)\, \mathrm{d}y,
\end{aligned} \tag{2.17}$$

which leads to the first inequality in (2.15). Using the fact that $\|\rho(t)\|_{L^2} \geq C\|\rho(t)\|_{L^1} > 0$ (see (2.14)), we have the second inequality in (2.15). Computing the higher-order derivatives with respect to x, we obtain for $k \geq 2$

$$\begin{aligned}
\partial_x^k G_\rho =& -R\Big(\partial_x^{k-1}\rho(x+R,t) + \partial_x^{k-1}\rho(x-R,t) + M\partial_x^{k-1}\rho_r(x+R,t) + M\partial_x^{k-1}\rho_r(x-R,t)\Big) \\
&+ \partial_x^{k-2}\rho(x+R,t) - \partial_x^{k-2}\rho(x-R,t) + M\partial_x^{k-2}\rho_r(x+R,t) - M\partial_x^{k-2}\rho_r(x-R,t).
\end{aligned}$$

Hence,

$$\begin{aligned}
\|\partial_x^k G_\rho\|_{L^2} &\leq C\Big(\|\partial_x^{k-1}\rho(t)\|_{L^2} + \|\partial_x^{k-2}\rho(t)\|_{L^2} + \|\partial_x^{k-1}\rho_r\|_{L^2} + \|\partial_x^{k-2}\rho_r\|_{L^2}\Big) \\
&\leq C\left(\|\rho(t)\|_{H^{k-1}} + \|\rho_r\|_{H^{k-1}}\right) \\
&\leq C[\|\rho_r\|_{H^{k-1}}]\, \|\rho(t)\|_{H^{k-1}},
\end{aligned}$$

where for the last inequality we used the fact that

$$\|\rho(t)\|_{H^{k-1}} \geq \|\rho(t)\|_{L^2} \geq C\|\rho(t)\|_{L^1} > 0.$$

□

Lemma 2.5.3 (More estimates for G_ρ)**.** *Let $v \in H^k_{per}(\tilde{X})$, $\rho_r \in H^{k-1}_{per}(\tilde{X}) \cap \mathscr{P}_e(X)$, and $\rho(t) \in H^{k-1}_{per}(\tilde{X})$ with $\|\rho(t)\|_{L^1} > 0$. Then for $k \geq 2$*

$$\|vG_\rho\|_{H^k} \leq C[\|\rho_r\|_{H^{k-1}}]\, \|v\|_{H^k}\, \|\rho(t)\|_{H^{k-1}}. \tag{2.18}$$

Proof. Notice

$$\|vG_\rho\|_{H^k} \leq C\left(\|vG_\rho\|_{L^2} + \|\partial_x^k(vG_\rho)\|_{L^2}\right). \tag{2.19}$$

For the first term on the right-hand side of (2.19), we have

$$\|vG_\rho\|_{L^2} \leq \|v\|_{L^2}\, \|G_\rho\|_{L^\infty} \leq C\, \|v\|_{L^2}\, \|\rho(t)\|_{L^2} \leq C\, \|v\|_{H^k}\, \|\rho(t)\|_{H^{k-1}},$$

where for the second inequality we used (2.13). Also, using Leibniz rule, for the second term on the right-hand side of (2.19), we can write

$$\begin{aligned}
\|\partial_x^k(vG_\rho)\|^2_{L^2} &= \left\| \sum_{i=0}^{k} C_i\, \partial_x^{k-i} v\, \partial_x^i G_\rho \right\|^2_{L^2} \\
&\leq C_0\, \|\partial_x^k v\|^2_{L^2}\, \|G_\rho\|^2_{L^\infty} + \sum_{i=1}^{k} C_i\, \|\partial_x^{k-i} v\|^2_{L^\infty}\, \|\partial_x^i G_\rho\|^2_{L^2} \\
&\leq C_0\, \|v\|^2_{H^k}\, \|\rho\|^2_{L^2} + \sum_{i=1}^{k} C_i\, \|\partial_x^{k-i} v\|^2_{H^1}\, \|\partial_x^i G_\rho\|^2_{L^2},
\end{aligned}$$

where for the last inequality we used Morrey's inequality which implies

$$\|\partial_x^{k-i} v\|_{L^\infty} \leq C\, \|\partial_x^{k-i} v\|_{H^1}.$$

Now, from (2.15) we have for $i = 1$

$$\|\partial_x^i G_\rho\|^2_{L^2} \leq C[\|\rho_r\|_{L^2}]\, \|\rho(t)\|^2_{L^2},$$

and from (2.16) we have for $i \geq 2$

$$\|\partial_x^i G_\rho\|^2_{L^2} \leq C[\|\rho_r\|_{H^{i-1}}]\, \|\rho(t)\|^2_{H^{i-1}}.$$

Combining these estimates while keeping only the highest Sobolev norms, we have

$$\begin{aligned}
\|vG_\rho\|_{H^k} &\leq C_1\, \|v\|_{H^k}\, \|\rho(t)\|_{H^{k-1}} + C_2\, \|v\|_{H^k}\, \|\rho(t)\|_{L^2} + C_3[\|\rho_r\|_{H^{k-1}}]\, \|v\|_{H^1}\, \|\rho(t)\|_{H^{k-1}} \\
&\leq C[\|\rho_r\|_{H^{k-1}}]\, \|v\|_{H^k}\, \|\rho(t)\|_{H^{k-1}}.
\end{aligned}$$

□

Remark 2.5.4 (Connection to existing works)**.** *The result of Lemma 2.5.3 is an extension of [55, Prop. 4.1].*

With these estimates in hand, we can follow the same arguments as in [55] to show the well-posedness of the dynamics described by PDE (2.6).

Sketch of proof of Theorem 2.4.1. Consider the following sequence of PDEs

$$\begin{cases} \partial_t \rho_n = \partial_x(\rho_n\, G_{\rho_{n-1}}) + \frac{\sigma^2}{2}\partial_{xx}\rho_n & \text{in} \quad \tilde{X}\times(0,T) \\ \rho_n(\cdot+2,t) = \rho_n(\cdot,t) & \text{on} \quad \partial\tilde{X}\times(0,T) \\ \rho_n(x,\cdot) = \rho_0(x) & \text{on} \quad \tilde{X}\times\{t=0\}, \end{cases} \tag{2.20}$$

with smooth initial and radical distributions $\rho_0, \rho_r \in C^\infty_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ for now. By standard results on linear parabolic PDEs [56, Ch. 7], there exists a sequence $\{\rho_n : n \geq 0\}$ in $C^\infty(0,T; C^\infty_{per}(\tilde{X}))$ that satisfies (2.20). Furthermore, using estimate (2.13), one can follow the same procedure provided in [55, Prop. 3.1] to show that $\|\rho_n(t)\|_{L^1} = \|\rho_n(0)\|_{L^1} = 2$, and hence, $\rho_n(t) \geq 0$ for all $n \geq 1$ and $t \geq 0$; see also [55, Cor. 3.2].

Remark 2.5.5 (Evenness of ρ_n)**.** *One can use the evenness of ρ_0 and ρ_r to show that the unique solutions ρ_n to PDEs (2.20) are also even in x for all $t \geq 0$. However, since this property will not be used for the existence, uniqueness, and regularity results provided below, we will postpone this argument to later when we deal with the evenness of the unique solution to PDE (2.6).*

Existence with smooth data. Using Lemmas 2.5.1 and 2.5.2 and following a similar idea as in [55, Lemm. 3.5 and 3.7], we can obtain the following convergence results

$$\rho_n \rightharpoonup \bar{\rho} \quad \text{in } L^1(0,T; L^1_{per}(\tilde{X})), \tag{2.21a}$$

$$\rho_{n_k} \rightharpoonup \bar{\rho} \quad \text{in } L^2(0,T; H^1_{per}(\tilde{X})), \tag{2.21b}$$

$$\partial_t \rho_{n_k} \rightharpoonup \bar{\rho}_t \quad \text{in } L^2(0,T; H^{-1}_{per}(\tilde{X})), \tag{2.21c}$$

for a limiting object $\bar{\rho}$, where n_k denotes a subsequence. Moreover, we have the following estimate for $\{\rho_n : n \geq 1\}$ and $\bar{\rho}$

$$\|\rho\|_{L^\infty(0,T;L^2)} + \|\rho\|_{L^2(0,T;H^1)} + \|\rho_t\|_{L^2(0,T;H^{-1})} \leq C[T]\, \|\rho_0\|_{L^2}. \tag{2.22}$$

We claim that $\bar{\rho}$ is the unique weak solution to (2.6). That is, $\bar{\rho}$ solves the weak formulation of (2.6) defined as

$$\int_0^T \langle \eta, \rho_t \rangle \, \mathrm{d}t + \int_0^T \int_{\tilde{X}} \left(\frac{\sigma^2}{2}\rho_x + \rho\, G_\rho \right) \eta_x \, \mathrm{d}x\mathrm{d}t = 0, \tag{2.23}$$

for any $\eta \in L^2(0,T; H^1_{per}(\tilde{X}))$. To show this, we multiply (2.20) by η with $n = n_k$ and integrate to obtain

$$\int_0^T \langle \eta, \partial_t \rho_{n_k} \rangle \, \mathrm{d}t + \frac{\sigma^2}{2} \int_0^T \int_{\tilde{X}} \partial_x \rho_{n_k}\, \eta_x \, \mathrm{d}x\mathrm{d}t + \int_0^T \int_{\tilde{X}} \rho_{n_k}\, G_{\rho_{n_k-1}}\, \eta_x \, \mathrm{d}x\mathrm{d}t = 0. \tag{2.24}$$

For the first two terms in (2.24), using convergence results (2.21c) and (2.21b), we have

$$\int_0^T \langle \eta, \partial_t \rho_{n_k} \rangle \, \mathrm{d}t \to \int_0^T \langle \eta, \bar{\rho}_t \rangle \, \mathrm{d}t,$$

and

$$\int_0^T \int_{\tilde{X}} \partial_x \rho_{n_k}\, \eta_x \, \mathrm{d}x\mathrm{d}t \to \int_0^T \int_{\tilde{X}} \bar{\rho}_x \eta_x \, \mathrm{d}x\mathrm{d}t,$$

as $k \to \infty$. Also, the last term in (2.24) can be written as

$$\int_0^T \int_{\tilde{X}} (\rho_{n_k} - \bar{\rho})\, G_{\rho_{n_k-1}}\, \eta_x \,\mathrm{d}x\mathrm{d}t \tag{2.25a}$$

$$+ \int_0^T \int_{\tilde{X}} \bar{\rho}\, (w \star (\rho_{n_k-1} - \bar{\rho}))\, \eta_x \,\mathrm{d}x\mathrm{d}t \tag{2.25b}$$

$$+ \int_0^T \int_{\tilde{X}} \bar{\rho}\, G_{\bar{\rho}}\, \eta_x \,\mathrm{d}x\mathrm{d}t,$$

where the limits of (2.25a) and (2.25b) are zero as $k \to \infty$. Indeed, in (2.25a), $G_{\rho_{n_k-1}}$ is bounded by the inequality (2.12) in Lemma 2.5.1, hence, $\eta_x G_{\rho_{n_k-1}} \in L^2(0,T;L^2_{per}(\tilde{X}))$, while (2.21b) implies $\rho_{n_k} \rightharpoonup \bar{\rho}$ in $L^2(0,T;L^2_{per}(\tilde{X}))$. Moreover, for (2.25a), we have

$$\begin{aligned}
&\int_0^T \int_{\tilde{X}} \bar{\rho}\, (w \star (\rho_{n_k-1} - \bar{\rho}))\, \eta_x \,\mathrm{d}x\mathrm{d}t \\
&\qquad \le \|\bar{\rho}\|_{L^\infty(0,T;L^2)}\, \|\eta_x\|_{L^2(0,T;L^2)}\, \|w \star (\rho_{n_k-1} - \bar{\rho})\|_{L^2(0,T;L^2)} \\
&\qquad \le C[T]\, \|\rho_0\|_{L^2}\, \|\eta\|_{L^2(0,T;H^1)} \left(\int_0^T \|\rho_{n_k-1} - \bar{\rho}\|^2_{L^1} \,\mathrm{d}t \right)^{\frac{1}{2}},
\end{aligned}$$

where for the second inequality we used (2.22) and the fact that

$$|w \star (\rho_{n_k-1} - \bar{\rho})| \le C\, \|\rho_{n_k-1} - \bar{\rho}\|_{L^1},$$

by Lemma 2.7 (set $M = 0$ in (2.12)). Now, notice

$$\|\rho_{n_k-1} - \bar{\rho}\|_{L^1(\tilde{X})} \le \|\rho_{n_k-1}\|_{L^1(\tilde{X})} + \|\bar{\rho}\|_{L^1(\tilde{X})} \le 4.$$

Hence,

$$\int_0^T \|\rho_{n_k-1} - \bar{\rho}\|^2_{L^1} \,\mathrm{d}t \le 4 \int_0^T \|\rho_{n_k-1} - \bar{\rho}\|_{L^1} \,\mathrm{d}t = 4\, \|\rho_{n_k-1} - \bar{\rho}\|_{L^1(0,T;L^1)} \to 0,$$

as $k \to \infty$ by the strong convergence (2.21a). Putting all these results together, we see that $\bar{\rho}$ indeed satisfies the weak formulation (2.23).

To complete the existence result, we have to show $\bar{\rho}(x,0) = \rho_0(x)$. This condition makes sense since $\bar{\rho} \in C(0,T;L^2_{per}(\tilde{X}))$ by [55, Thm. 3.8] and the convergence results (2.21b) and (2.21c). Pick some $\eta \in C^1(0,T;H^1_{per}(\tilde{X}))$ with $\eta(T) = 0$ and rewrite the weak formulation (2.23) as

$$-\int_0^T \langle \bar{\rho}, \eta_t \rangle \,\mathrm{d}t + \int_0^T \int_{\tilde{X}} \left(\frac{\sigma^2}{2} \bar{\rho}_x + \bar{\rho}\, G_{\bar{\rho}} \right) \eta_x \,\mathrm{d}x\mathrm{d}t = \int_{\tilde{X}} \bar{\rho}(x,0)\, \eta(x,0) \,\mathrm{d}x. \tag{2.26}$$

Similarly, since $\rho_{n_k}(x,0) = \rho_0(x)$, we have

$$-\int_0^T \langle \rho_{n_k}, \eta_t \rangle \,\mathrm{d}t + \int_0^T \int_{\tilde{X}} \left(\frac{\sigma^2}{2} \partial_x \rho_{n_k} + \rho_{n_k} G_{\rho_{n_k}} \right) \eta_x \,\mathrm{d}x\mathrm{d}t = \int_{\tilde{X}} \rho_0(x)\, \eta(x,0) \,\mathrm{d}x. \tag{2.27}$$

Let $k \to \infty$ in (2.27), so for arbitrary $\eta(x,0)$ we obtain from (2.27) and (2.26) that

$$\int_{\tilde{X}} \bar{\rho}(x,0)\, \eta(x,0)\, \mathrm{d}x = \int_{\tilde{X}} \rho_0(x)\, \eta(x,0)\, \mathrm{d}x,$$

which implies $\bar{\rho}(x,0) = \rho_0(x)$.

Relaxed regularity on data. In order to relax regularity assumption on data to $\rho_0, \rho_r \in L^2_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$, we can use the mollified version of the distributions, i.e., $\rho_0^\epsilon = \phi_\epsilon \star \rho_0$ and $\rho_r^\epsilon = \phi_\epsilon \star \rho_r$ with the standard positive mollifier ϕ_ϵ, follow the same procedure and take the limit $\epsilon \to 0$ at the end. See also [55, Thm. 3.12] for the details of this process.

Uniquness. Let $\xi = \bar{\rho}_1 - \bar{\rho}_2$ where $\bar{\rho}_1$ and $\bar{\rho}_2$ are two weak solutions to (2.6) with $\rho_0, \rho_r \in L^2_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$. Then, for every $\eta \in L^2(0,T; H^1_{per}(\tilde{X}))$ we have

$$\int_0^T \langle \eta, \xi_t \rangle \, \mathrm{d}t + \frac{\sigma^2}{2} \int_0^T \int_{\tilde{X}} \xi_x\, \eta_x \, \mathrm{d}x\mathrm{d}t + \int_0^T \int_{\tilde{X}} (\bar{\rho}_1\, G_{\bar{\rho}_1} - \bar{\rho}_2\, G_{\bar{\rho}_2})\, \eta_x \, \mathrm{d}x\mathrm{d}t = 0.$$

We can rewrite the last integrand as

$$\begin{aligned} \bar{\rho}_1\, G_{\bar{\rho}_1} - \bar{\rho}_2\, G_{\bar{\rho}_2} &= \bar{\rho}_1 (w \star (\bar{\rho}_1 + M\rho_r)) - \bar{\rho}_2 (w \star (\bar{\rho}_2 + M\rho_r)) \\ &= (\bar{\rho}_1 - \bar{\rho}_2)(w \star (\bar{\rho}_1 + M\rho_r)) + \bar{\rho}_2 (w \star (\bar{\rho}_1 - \bar{\rho}_2)) \\ &= \xi\, G_{\bar{\rho}_1} + \bar{\rho}_2\, (w \star \xi), \end{aligned}$$

to obtain

$$\begin{aligned} \int_0^T \langle \eta, \xi_t \rangle \, \mathrm{d}t + \frac{\sigma^2}{2} \int_0^T \int_{\tilde{X}} \xi_x\, \eta_x \, \mathrm{d}x\mathrm{d}t = \\ - \int_0^T \int_{\tilde{X}} \xi\, G_{\bar{\rho}_1}\, \eta_x \, \mathrm{d}x\mathrm{d}t - \int_0^T \int_{\tilde{X}} \bar{\rho}_2\, (w \star \xi)\, \eta_x \, \mathrm{d}x\mathrm{d}t. \end{aligned} \tag{2.28}$$

Now, for the first integral on the right-hand side of (2.28), we have

$$\begin{aligned} \left| \int_0^T \int_{\tilde{X}} \xi\, G_{\bar{\rho}_1}\, \eta_x \, \mathrm{d}x\mathrm{d}t \right| &\le 2R(1+M)\, \|\xi\|_{L^2(0,T;L^2)}\, \|\eta_x\|_{L^2(0,T;L^2)} \\ &\le \frac{\sigma^2}{4}\, \|\eta_x\|^2_{L^2(0,T;L^2)} + C_1\, \|\xi\|^2_{L^2(0,T;L^2)}, \end{aligned} \tag{2.29}$$

where for the first inequality we used (2.12) in Lemma 2.5.1 and Cauchy-Schwarz inequality, and for the second inequality we used Young's inequality. Similarly, for the second integral on the right-hand side of (2.28), we have

$$\begin{aligned} \left| \int_0^T \int_{\tilde{X}} \bar{\rho}_2\, (w \star \xi)\, \eta_x \, \mathrm{d}x\mathrm{d}t \right| &\le \|\bar{\rho}_2\|_{L^\infty(0,T;L^2)}\, \|\eta_x\|_{L^2(0,T;L^2)}\, \|w \star \xi\|_{L^2(0,T;L^2)} \\ &\le C_2[T]\, \|\rho_0\|_{L^2}\, \|\eta_x\|_{L^2(0,T;L^2)}\, \|\xi\|_{L^2(0,T;L^2)} \\ &\le \frac{\sigma^2}{4}\, \|\eta_x\|^2_{L^2(0,T;L^2)} + C_2[T]\, \|\rho_0\|^2_{L^2}\, \|\xi\|^2_{L^2(0,T;L^2)}, \end{aligned} \tag{2.30}$$

where for the second inequality we used (2.22) and Lemma 2.5.1 (see (2.13) and (2.14)). Using (2.29) and (2.30) for (2.28) and setting $\eta = \xi$, we obtain

$$\int_0^T \langle \xi, \xi_t \rangle \, \mathrm{d}t \le \left(C_1 + C_2[T]\, \|\rho_0\|^2_{L^2}\right)\, \|\xi\|^2_{L^2(0,T;L^2)}.$$

By [55, Thm. 3.8], we know

$$\langle \xi, \xi_t \rangle = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t)\|_{L^2}^2.$$

Thus, for all T, we have

$$\frac{1}{2}\int_0^T \frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t)\|_{L^2}^2 \,\mathrm{d}t \le \left(C_1 + C_2[T]\,\|\rho_0\|_{L^2}^2\right)\int_0^T \|\xi(t)\|_{L^2}^2 \mathrm{d}t.$$

This implies, for a.e. $t \in [0,T]$

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t)\|_{L^2}^2 \le C[T,\rho_0]\,\|\xi(t)\|_{L^2}^2.$$

Hence, by Grönwall's inequality,

$$\|\xi(t)\|_{L^2}^2 \le C[T,\rho_0]\,\|\xi(0)\|_{L^2}^2.$$

This implies $\|\xi(t)\|_{L^2} = \|\bar{\rho}_1(t) - \bar{\rho}_2(t)\|_{L^2} = 0$ since $\xi(0) = \rho_0 - \rho_0 = 0$. Then, from continuity of $\bar{\rho}_1$ and $\bar{\rho}_2$ in time (by [55, Thm. 3.8]), we obtain uniqueness. That is, $\bar{\rho}_1 = \bar{\rho}_2$ for all $t \in [0,T]$.

Regularity. Here, we first mollify the problem data ρ_0 and ρ_r with the standard positive mollifier ϕ_ϵ so that the solutions $\{\rho_n : n \ge 0\}$ to (2.20) are all smooth. This allows us to take derivatives of (2.20) to any order. We then take the limit $\epsilon \to \infty$ at the end. For simplicity, we omit the arguments for this last step and drop the subscript ϵ.

Employing Lemma 2.5.3, we can extend the improved regularity results in space in [55, Thm. 4.2]. That is, for $\rho_0 \in H^k_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ and $\rho_r \in H^{k-1}_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$, we have

$$\bar{\rho} \in L^2(0,T;H^{k+1}_{per}(\tilde{X})) \cap L^\infty(0,T;H^k_{per}(\tilde{X})). \tag{2.31}$$

Moreover, since ρ_r is constant in time, we can also employ the results on improved regularity in time provided by [55, Thm. 4.3] for our model. This means, for $\rho_0 \in H^{2k}_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ and $\rho_r \in L^2_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$, we have for $i \le k$

$$\partial_t^i \bar{\rho} \in L^2(0,T;H^{2k-2i+1}_{per}(\tilde{X})) \cap L^\infty(0,T;H^{2k-2i}_{per}(\tilde{X})), \tag{2.32}$$

and

$$\partial_t^{k+1}\bar{\rho} \in L^2(0,T;H^{-1}_{per}(\tilde{X})). \tag{2.33}$$

With these regularity results in space and time, we can derive the required regularity on the solution as stated in Theorem 2.4.1. Let $\rho_0 \in H^3_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ and also let $\rho_r \in H^2_{per}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ and $\bar{\rho}$ be the unique weak solution to PDE (2.6). Then, by (2.31), we have $\bar{\rho} \in L^\infty(0,T;H^3_{per}(\tilde{X}))$. Hence, by Sobolev embedding theorem [61, Sec. 4.12], we have $\bar{\rho}(t) \in C^2_{per}(\tilde{X})$ (after possibly being redefined on a set of measure zero). This gives the required regularity in space. Also, (2.32) and (2.33) imply that $\bar{\rho}_t \in L^2(0,T;H^1_{per}(\tilde{X}))$ and $\bar{\rho}_{tt} \in L^2(0,T;H^{-1}_{per}(\tilde{X}))$. Hence, by [55, Thm. 3.8], we have $\bar{\rho}_t \in C(0,T;L^2_{per}(\tilde{X}))$ (after possibly being redefined on a set of measure zero). This gives the required regularity in time. Putting these results together, we have $\bar{\rho} \in C^1(0,T;C^2_{per}(\tilde{X}))$.

Evenness. The evenness imposed on ρ_0 and ρ_r implies that if $\rho(x,t)$ is a solution of (2.6), then $\rho(-x,t)$ is also a solution. Indeed, from (2.6) we obtain

$$\begin{aligned}\partial_t\rho(-x,t)-\frac{\sigma^2}{2}\partial_x^2\rho(-x,t) &= \partial_x\left(\rho(-x,t)\int w(-x-y)\,(\rho(y,t)+M\rho_r(y))\,\mathrm{d}y\right)\\ &= \partial_x\left(\rho(-x,t)\int w(-x+y)\,(\rho(-y,t)+M\rho_r(-y))\,(-\mathrm{d}y)\right)\\ &= \partial_x\left(\rho(-x,t)\int -w(-x+y)\,(\rho(-y,t)+M\rho_r(y))\,\mathrm{d}y\right)\\ &= \partial_x\left(\rho(-x,t)\int w(x-y)\,(\rho(-y,t)+M\rho_r(y))\,\mathrm{d}y\right),\end{aligned}$$

where for that last equality we used the fact that w is an odd function. Then, assuming $\rho_0\in H^3_{ep}(\tilde{X})\cap\mathscr{P}_e(\tilde{X})$ and $\rho_r\in H^2_{ep}(\tilde{X})\cap\mathscr{P}_e(\tilde{X})$ (notice that $H^k_{ep}(\tilde{X})\subset H^k_{per}(\tilde{X})$), the *uniqueness* of the solution $\bar{\rho}\in C^1(0,T;C^2_{per}(\tilde{X}))$ to PDE (2.6) implies that the solution is even, that is, $\bar{\rho}\in C^1(0,T;C^2_{ep}(\tilde{X}))$.

Positivity. Using the same approach as in [60], we consider the following version of (2.6) in the unknown function ρ, with $\bar{\rho}$ being the non-negative weak solution

$$\rho_t=(\rho\; G_{\bar{\rho}})_x+\frac{\sigma^2}{2}\rho_{xx}.$$

This is a linear parabolic PDE with smooth and bounded coefficients (by Lemmas 2.5.1 and 2.5.2) for which $\bar{\rho}$ is a classic non-negative solution. Thus, by the parabolic Harnack inequality [56, Sec. 7.1.4, Thm. 10], we have

$$\sup_{x\in\tilde{X}}\bar{\rho}(x,t_1)\le c\inf_{x\in\tilde{X}}\bar{\rho}(x,t_2),$$

for $0<t_1<t_2<\infty$ and some positive constant c. Non-negativity of $\bar{\rho}(x,t)$ implies that $\inf_{x\in\tilde{X}}\bar{\rho}(x,t)$ and hence $\bar{\rho}(x,t)$ is strictly positive for all $t>0$. □

2.5.2. STATIONARY SOLUTION

EXISTENCE OF STATIONARY SOLUTION

This section mainly concerns the proof of *existence* result in Theorem 2.4.2 for stationary equation (2.8). All the norms in this section are with respect to $X=[0,1]$ (as opposed to $\tilde{X}=[-1,1]$), unless indicated otherwise. We note that norms on the even 2-periodic spaces computed with respect to to X and $\tilde{X}$ differ by a multiplicative constant, e.g., $\|u\|_{L^p(\tilde{X})}=2^{\frac{1}{p}}\|u\|_{L^p(X)}$. We again use $C,C_0,C_1,\dots$ to represent a generic constant (depending on the model parameters) whose actual values may change from line to line. In the case these constants depend on a particular object of interest, say θ, this dependence is explicitly indicated by $C[\theta]$.

Let us begin with providing a fixed point characterization of the solution to stationary equation (2.8). We note that, corresponding to the solution to dynamic equation (2.6), we are particularly interested in *even* solutions $\rho^s\in\mathscr{P}_e(\tilde{X})$ of stationary equation (2.8).

Lemma 2.5.6 (Fixed point characterization). *$\rho^s \in C^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$ is a solution of stationary equation (2.8) if and only if ρ^s is a fixed point of the operator $\mathscr{T} : \mathscr{P}_e(\tilde{X}) \to \mathscr{P}_e(\tilde{X})$ defined by*

$$\mathscr{T}\rho := \frac{1}{K} \exp\left(-\frac{2}{\sigma^2}\int_0^x G_\rho(z)\,\mathrm{d}z\right), \tag{2.34}$$

where the constant K is determined by the normalizing condition

$$K = \int_0^1 \exp\left(-\frac{2}{\sigma^2}\int_0^x G_\rho(z)\,\mathrm{d}z\right)\mathrm{d}x.$$

Proof. The "if" part is clear since any fixed point $\rho^s \in C^2_{ep}(\tilde{X})$ of $\mathscr{T}$ satisfies the stationary equation (2.8). For the "only if" part, note that integrating (2.8) once, we have

$$\frac{\sigma^2}{2}\rho_x + \rho\, G_\rho = C. \tag{2.35}$$

Now notice that we can set $C = 0$ since we are interested in *even* solutions to (2.35). Indeed, from (2.35) we have

$$\frac{\sigma^2}{2}\rho_x(-x) + \rho(-x)[w(-x) \star (\rho(-x) + M\rho_r(-x))] = C.$$

Hence, for an even solution, we obtain

$$-\frac{\sigma^2}{2}\rho_x(x) - \rho(x)[w(x) \star (\rho(x) + M\rho_r(x))] = C,$$

where we used the fact that w is an odd function. This implies $C = 0$. Rearranging and integrating (2.35) once again, we have

$$\rho(x) = \frac{1}{K} \exp\left(-\frac{2}{\sigma^2}\int_0^x G_\rho(z)\,\mathrm{d}z\right), \tag{2.36}$$

where the normalizing condition gives the constant K as

$$K = \int_0^1 \exp\left(-\frac{2}{\sigma^2}\int_0^x G_\rho(z)\,\mathrm{d}z\right)\mathrm{d}x.$$

This completes the proof. □

This characterization allows us to use tools from operator theory. To be precise, we will use Schauder fixed point theorem to derive the existence result for the stationary solution. Before that, we present some preliminary results for the operator $\mathscr{T}$.

Lemma 2.5.7 (Estimates for $\mathscr{T}$). *Let $\mathscr{T}$ be the operator on $\mathscr{P}_e(\tilde{X})$ defined by (2.34).*

- *If $\rho, \rho_r \in \mathscr{P}_e(\tilde{X})$, then*

$$\|\mathscr{T}\rho\|_{L^\infty} \le \exp\left(\frac{8R(1+M)}{\sigma^2}\right), \tag{2.37}$$

and

$$\|\partial_x \mathscr{T}\rho\|_{L^\infty} \le \frac{4R(1+M)}{\sigma^2} \exp\left(\frac{8R(1+M)}{\sigma^2}\right). \tag{2.38}$$

- *If* $\rho, \rho_r \in L^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$, *then*

$$\|\mathscr{T}\rho\|_{H^2} \le C[\|\rho_r\|_{L^2}]\,\|\rho\|_{L^2}. \tag{2.39}$$

- *If* $\rho, \rho_r \in H^{k-2}_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X})$, *then for* $k \ge 3$

$$\|\mathscr{T}\rho\|_{H^k} \le \sum_{i=1}^{k-1} C_i[\|\rho_r\|_{H^{k-2}}]\,\|\rho\|^i_{H^{k-2}}. \tag{2.40}$$

Proof. From the definition (2.34) and the inequality (2.12) in Lemma 2.5.1 we obtain

$$|\mathscr{T}\rho| = \frac{\exp\left\{-\frac{2}{\sigma^2}\int_0^x G_\rho(z)\,\mathrm{d}z\right\}}{\int_0^1 \exp\left\{-\frac{2}{\sigma^2}\int_0^x G_\rho(z)\,\mathrm{d}z\right\}\mathrm{d}x} \le \frac{\exp\left\{\frac{4R(1+M)}{\sigma^2}\right\}}{\exp\left\{-\frac{4R(1+M)}{\sigma^2}\right\}} = \exp\left\{\frac{8R(1+M)}{\sigma^2}\right\},$$

which gives estimate (2.37).

Now, observe

$$\|\partial_x\mathscr{T}\rho\|_{L^\infty} = \left\|-\frac{2}{\sigma^2}\,G_\rho\,\mathscr{T}\rho\right\|_{L^\infty} \le \frac{2}{\sigma^2}\,\|G_\rho\|_{L^\infty}\,\|\mathscr{T}\rho\|_{L^\infty}.$$

Using (2.12) in Lemma 2.5.1 and (2.37), we obtain the inequality (2.38).

For the inequality (2.39), first, notice

$$\|\mathscr{T}\rho\|_{H^2} \le C\left(\|\mathscr{T}\rho\|_{L^2} + \|\partial_x^2\mathscr{T}\rho\|_{L^2}\right) \le C_1 + C_2\,\|\partial_x^2\mathscr{T}\rho\|_{L^2}, \tag{2.41}$$

where for the second inequality we used the fact that $\|\mathscr{T}\rho\|_{L^2} \le C\,\|\mathscr{T}\rho\|_{L^\infty}$ is bounded by (2.37). Also, we have

$$\begin{aligned}
\|\partial_x^2\mathscr{T}\rho\|_{L^2} &= \left\|-\frac{2}{\sigma^2}\left(\mathscr{T}\rho\,\partial_x G_\rho + G_\rho\,\partial_x\mathscr{T}\rho\right)\right\|_{L^2} \\
&\le C\left(\|\mathscr{T}\rho\|_{L^\infty}\,\|\partial_x G_\rho\|_{L^2} + \|G_\rho\|_{L^\infty}\,\|\partial_x\mathscr{T}\rho\|_{L^2}\right) \\
&\le C[\|\rho_r\|_{L^2}]\,\|\rho\|_{L^2} + C_2 \\
&\le C[\|\rho_r\|_{L^2}]\,\|\rho\|_{L^2},
\end{aligned}$$

where for the second inequality we used (2.15) in Lemma 2.5.2 and the last inequality follows from the fact that $\|\rho\|_{L^2} \ge \|\rho\|_{L^1} > 0$ (see (2.14)). Inserting this result in (2.41), we obtain the inequality (2.39).

Similarly, for $k \ge 3$, we have (see (2.41))

$$\|\mathscr{T}\rho\|_{H^k} \le C_1 + C_2\,\|\partial_x^k\mathscr{T}\rho\|_{L^2}. \tag{2.42}$$

Now, notice

$$\begin{aligned}
\|\partial_x^k\mathscr{T}\rho\|_{L^2} = \|\partial_x^{k-1}\partial_x\mathscr{T}\rho\|_{L^2} &= \left\|\partial_x^{k-1}\left(-\frac{2}{\sigma^2}\,G_\rho\,\mathscr{T}\rho\right)\right\|_{L^2} = \frac{2}{\sigma^2}\,\|\partial_x^{k-1}\left(\mathscr{T}\rho\,G_\rho\right)\|_{L^2} \\
&\le C\,\|\mathscr{T}\rho\,G_\rho\|_{H^{k-1}} \le C[\|\rho_r\|_{H^{k-2}}]\,\|\rho\|_{H^{k-2}}\,\|\mathscr{T}\rho\|_{H^{k-1}},
\end{aligned}$$

where for the last inequality we used Lemma 2.5.3. Combining this result with (2.42), we derive a recursive inequality. Performing the recursive computations while keeping the highest Sobolev norms, we obtain

$$\|\mathscr{T}\rho\|_{H^k} \leq C_0 + \sum_{i=1}^{k-1} C_i [\|\rho_r\|_{H^{k-2}}] \|\rho\|^i_{H^{k-2}}.$$

Then, since $\|\rho\|^i_{H^{k-2}} \geq \|\rho\|_{L^2} \geq C\|\rho\|_{L^1} > 0$, we can remove the constant C_0 and consider its effect in constants C_i. This gives the desired inequality (2.40). □

Proposition 2.5.8 (Lipschitz continuity of $\mathscr{T}$)**.** *Let $\mathscr{T}$ be the operator on $\mathscr{P}_e(\tilde{X})$ defined by (2.34) with $\rho_r \in \mathscr{P}_e(\tilde{X})$. Then, $\mathscr{T}$ is Lipschitz continuous in L^p for $1 \leq p < \infty$ with Lipschitz constant*

$$L_{\mathscr{T}} = \frac{1}{2} \exp\left\{\left(\frac{8R(1+M)}{\sigma^2}\right)\left(1-\frac{1}{p}\right)\right\}\left(\exp\left\{\frac{16R}{\sigma^2}\right\}-1\right). \tag{2.43}$$

Proof. We use a similar argument to the one provided by [53]. Let $\rho_1, \rho_2 \in \mathscr{P}_e(\tilde{X})$. Using estimate (2.37) in Lemma 2.5.7, we have for $1 \leq p < \infty$

$$\begin{aligned}\|\mathscr{T}\rho_2 - \mathscr{T}\rho_1\|_{L^p} &= \left\|\mathscr{T}\rho_1\left(\frac{\mathscr{T}\rho_2}{\mathscr{T}\rho_1}-1\right)\right\|_{L^p} \leq \|\mathscr{T}\rho_1\|_{L^p}\left\|\frac{\mathscr{T}\rho_2}{\mathscr{T}\rho_1}-1\right\|_{L^\infty} \\ &\leq \|\mathscr{T}\rho_1\|^{1-\frac{1}{p}}_{L^\infty}\left\|\frac{K_1}{K_2}\exp\left\{-\frac{2}{\sigma^2}\int_0^x w\star(\rho_2-\rho_1)\,\mathrm{d}z\right\}-1\right\|_{L^\infty},\end{aligned} \tag{2.44}$$

where for the last inequality we used $\|\mathscr{T}\rho\|_{L^1(X)} = 1$. Now, define

$$\Gamma(\rho_1-\rho_2) := \frac{2}{\sigma^2}\int_0^x w\star(\rho_1-\rho_2)\,\mathrm{d}z,$$

and observe

$$\begin{aligned}\left|\Gamma(\rho_2-\rho_1)\right| &= \frac{2}{\sigma^2}\left|\int_0^x\int (z-y)\,\mathbf{1}_{|z-y|\leq R}\,(\rho_2(y)-\rho_1(y))\,\mathrm{d}y\mathrm{d}z\right| \\ &\leq \frac{2}{\sigma^2}\int_0^x\int |(z-y)|\,\mathbf{1}_{|z-y|\leq R}\,|\rho_2(y)-\rho_1(y)|\,\mathrm{d}y\mathrm{d}z \\ &\leq \frac{2R}{\sigma^2}\int_0^x\int_{\tilde{X}} |\rho_2(y)-\rho_1(y)|\,\mathrm{d}y\mathrm{d}z \leq \frac{4R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}.\end{aligned} \tag{2.45}$$

Similarly, we can write the normalizing constant K_1 as

$$K_1 = \int_0^1 \exp\left(-\frac{2}{\sigma^2}\int_0^x G_{\rho_1}\,\mathrm{d}z\right)\mathrm{d}x = \int_0^1 \exp\left(-\frac{2}{\sigma^2}\int_0^x G_{\rho_2}\,\mathrm{d}z\right)\exp\left\{-\Gamma(\rho_1-\rho_2)\right\}\,\mathrm{d}x.$$

From (2.45), it follows

$$K_1 \leq \int_0^1 \exp\left(-\frac{2}{\sigma^2}\int_0^x G_{\rho_2}\,\mathrm{d}z\right)\exp\left(\frac{4R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}\right)\,\mathrm{d}x = K_2\exp\left(\frac{4R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}\right),$$

and

$$K_1 \geq \int_0^1 \exp\left(-\frac{2}{\sigma^2}\int_0^x G_{\rho_2}\,\mathrm{d}z\right)\exp\left(-\frac{4R}{\sigma^2}\,\|\rho_2-\rho_1\|_{L^1}\right)\mathrm{d}x = K_2\exp\left(-\frac{4R}{\sigma^2}\,\|\rho_2-\rho_1\|_{L^1}\right).$$

Hence,

$$\exp\left(-\frac{4R}{\sigma^2}\,\|\rho_2-\rho_1\|_{L^1}\right) \leq \frac{K_1}{K_2} \leq \exp\left(\frac{4R}{\sigma^2}\,\|\rho_2-\rho_1\|_{L^1}\right). \tag{2.46}$$

Using (2.45) and (2.46), we can rewrite (2.44) as

$$\|\mathcal{T}\rho_2-\mathcal{T}\rho_1\|_{L^p} \leq \|\mathcal{T}\rho_1\|_{L^\infty}^{1-\frac{1}{p}}\,\max\left\{\exp\left(\frac{8R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}\right)-1,\ 1-\exp\left(-\frac{8R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}\right)\right\}.$$

Hence,

$$\|\mathcal{T}\rho_2-\mathcal{T}\rho_1\|_{L^p} \leq \|\mathcal{T}\rho_1\|_{L^\infty}^{1-\frac{1}{p}}\left(\exp\left(\frac{8R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}\right)-1\right). \tag{2.47}$$

Now, notice that

$$\|\rho_2-\rho_1\|_{L^1} \leq \|\rho_2\|_{L^1}+\|\rho_1\|_{L^1} = 2,$$

(recall that norms are defined over X) and for $a>0$

$$\mathrm{e}^{ax}-1 \leq \frac{1}{2}(\mathrm{e}^{2a}-1)x, \quad \forall x\in[0,2].$$

Thus, we have

$$\exp\left(\frac{8R}{\sigma^2}\|\rho_2-\rho_1\|_{L^1}\right)-1 \leq \frac{1}{2}\left(\mathrm{e}^{\frac{16R}{\sigma^2}}-1\right)\|\rho_2-\rho_1\|_{L^1}. \tag{2.48}$$

Combining (2.47) and (2.48), we obtain

$$\|\mathcal{T}\rho_2-\mathcal{T}\rho_1\|_{L^p} \leq \frac{1}{2}\,\|\mathcal{T}\rho_1\|_{L^\infty}^{1-\frac{1}{p}}\left(\mathrm{e}^{\frac{16R}{\sigma^2}}-1\right)\|\rho_2-\rho_1\|_{L^1}.$$

Finally, using (2.37) in Lemma 2.5.7 and the inequality (2.14) which relates norms over domains of finite measure, we have

$$\|\mathcal{T}\rho_2-\mathcal{T}\rho_1\|_{L^p} \leq L_{\mathcal{T}}\,\|\rho_2-\rho_1\|_{L^p},$$

where the constant $L_{\mathcal{T}}$ is given by (2.43). □

With these preliminary results in hand, we next move on to the proof of existence result in Theorem 2.4.2.

Proof of Theorem 2.4.2 (Existence). Following a similar argument as in [60, Thm. 2.3] and using Lemma 2.5.6, we can present the existence result for the stationary solution as the fixed point of the operator $\mathcal{T}$. First note that using estimate (2.37) in Lemma 2.5.7, we have $\|\mathcal{T}\rho\|_{L^2} \leq C\,\|\mathcal{T}\rho\|_{L^\infty} \leq c$ for some positive constant c. Thus, for the purpose of finding the fixed points of $\mathcal{T}$, we can restrict $\mathcal{T}$ to act on the closed and convex set $E :=$

$\left\{\rho \in L^2_{ep}(\tilde{X}) \cap \mathscr{P}_e(\tilde{X}) : \|\rho\|_{L^2} \le c\right\}$. Now, notice that using inequalities (2.37) and (2.38) in Lemma 2.5.7, we have for any $\rho \in E$

$$\|\mathscr{T}\rho\|^2_{H^1} \le \|\mathscr{T}\rho\|^2_{L^2} + \|\partial_x \mathscr{T}\rho\|^2_{L^2} \le C_1\|\mathscr{T}\rho\|^2_{L^\infty} + C_2\|\partial_x \mathscr{T}\rho\|^2_{L^\infty} \le c', \tag{2.49}$$

for some constant $c' > 0$. That is, $\mathscr{T}(E) \subset E$ is uniformly bounded in $H^1_{ep}(\tilde{X})$. Thus, by the Rellich-Kondrachov compactness theorem [56, Sec. 5.7, Thm. 1], $\mathscr{T}(E)$ is precompact in $L^2_{ep}(\tilde{X})$. Since $E \subset L^2_{ep}(\tilde{X})$ is closed, this implies $\mathscr{T}(E)$ is also precompact in E. Also, $\mathscr{T}$ is Lipschitz continuous by Proposition 2.5.8. Hence, by Schauder fixed point theorem [56, Sec. 9.2.2, Thm. 3], it has a fixed point $\rho^s \in E$ which belongs to $H^1_{ep}(\tilde{X})$ by (2.49).

Regularity. Estimate (2.40) in Lemma (2.5.7) implies that if $\rho_r \in H^{k-2}_{ep}(\tilde{X})$, then the fixed point $\rho^s = \mathscr{T}\rho^s \in H^k_{ep}(\tilde{X})$. In particular, if $\rho_r \in H^1_{ep}(\tilde{X})$, then $\rho^s \in H^3_{ep}(\tilde{X})$. Hence, by Sobolev embedding theorem [61, Sec. 4.12], $\rho \in C^2_{ep}(\tilde{X})$ (after possibly being redefined on a set of measure zero).

Positivity. The positivity of the fixed point follows from the representation (2.34). □

Remark 2.5.9 (Uniqueness)**.** *By Proposition 2.5.8, $\mathscr{T}$ is Lipschitz continuous in L^p with Lipschitz constant $L_{\mathscr{T}}$ given by (2.43), and thus, is a contraction for $L_{\mathscr{T}} < 1$. Hence, by Banach fixed-point theorem [56, Sec. 9.2.1, Thm. 1], $\mathscr{T}$ has a unique fixed point for $L_{\mathscr{T}} < 1$. Setting $p = 1$ in (2.43) gives the sufficient condition $\sigma^2 > \frac{16R}{\ln 3}$ for uniqueness of stationary solution. This result corresponds to the sufficient condition provided in [53, Thm. 2].*

We finish this section with a remark on the shape of the stationary opinion clusters for a highly concentrated radical opinion distribution, by providing an approximate solution to the stationary equation (2.8). To this end, we assume radicals are highly concentrated around a particular opinion value $x = A$. To be precise, we assume that the average opinion of radicals is $A = \int_X x\ \rho_r(x)\ \mathrm{d}x$ and the variance of radicals $\sigma_r^2 = \int_X (x-A)^2\ \rho_r(x)\ \mathrm{d}x$ is much smaller than the confidence range R. It helps to think of the limit being a point mass of radicals located at opinion value $x = A$. We further assume that the noise level σ is also much smaller than R so that the inter-cluster influences (from other possible clusters) can be ignored. Using these assumptions, we can expect this particular cluster of normal agents to be concentrated around A. This implies that to evaluate the integral in (2.36), we only need to consider values of y near A. Under these assumptions, for $R < A < L - R$, we can write

$$\begin{aligned}\int_0^x w \star (\rho + M\rho_r)\ \mathrm{d}z &= \int_0^x \int (z-y)\ \mathbf{1}_{|y-z|\le R}\ (\rho(y) + M\rho_r(y))\ \mathrm{d}y\mathrm{d}z \\ &\approx \int_0^x \int_{A-R}^{A+R} (z-A)\ \mathbf{1}_{|z-A|\le R}\ (\rho(y) + M\rho_r(y))\ \mathrm{d}y\mathrm{d}z.\end{aligned}$$

We can now handle the two integrations separately and obtain

$$\begin{aligned}\int_0^x w \star (\rho + M\rho_r)\ \mathrm{d}z &= \int_0^x (z-A)\mathbf{1}_{|z-A|\le R}\ \mathrm{d}z \int_{A-R}^{A+R} (\rho(y) + M\rho_r(y))\ \mathrm{d}y \\ &= \frac{1}{2}\left((x-A)^2 - R^2\right)\mathbf{1}_{|x-A|\le R} \int_{A-R}^{A+R} (\rho(y) + M\rho_r(y))\ \mathrm{d}y \\ &\approx \frac{M+1}{2}\ \left((x-A)^2 - R^2\right)\ \mathbf{1}_{|x-A|\le R}.\end{aligned}$$

Inserting this result in (2.36), we have

$$\rho^s(x) = \frac{1}{K}\exp\left\{-\frac{M+1}{\sigma^2}\left((x-A)^2 - R^2\right)\mathbf{1}_{|x-A|\leq R}\right\},$$

which can also be expressed as (by modifying the normalizing constant K)

$$\rho^s(x) = \frac{1}{K}\exp\left\{-\frac{M+1}{\sigma^2}\min\left\{(x-A)^2,\, R^2\right\}\right\}. \tag{2.50}$$

This result is an extension of the approximate solution provided by [44, Sec. 5.2]. In particular, one can reproduce the same result by setting $M = 0$ and $A = 0$. Equation (2.50) shows that for highly concentrated radicals the possible accumulation of normals around the average radical opinion A in the stationary state is semi-Gaussian with variance $\frac{\sigma^2}{2(M+1)}$. Note that, as argued in [44], other clusters centered at opinion values other than $x = A$ may also exist. As long as these clusters are well-separated so that inter-cluster influences can be ignored, one can use the same approximation to derive a semi-Gaussian profile for the shape of these clusters (set $M = 0$ and $A = x_0$ in (2.50) where x_0 denotes the center of the corresponding cluster). This analysis shows that M affects the shape of the possible cluster formed at the average radical opinion A in the stationary state. We will examine the provided approximate solution in our numerical simulations in Chapter 3.

GLOBAL ESTIMATE FOR STATIONARY SOLUTION

This section is devoted to the proof of the *estimate* given in Theorem 2.4.2. In this section, all the norms are with respect to the domain $\tilde{X} = [-1, 1]$, unless indicated otherwise.

Proof of Theorem 2.4.2 (Estimate). Let $\psi = \rho^s - 1$ so that $\int_X \psi(x)\,\mathrm{d}x = 0$. From the stationary equation (2.8) we obtain

$$-\frac{\sigma^2}{2}\psi_{xx} = \left[(\psi+1)\,G_{\psi+1}\right]_x = \left[(\psi+1)\,(w\star 1 + G_\psi)\right]_x = \left[(\psi+1)\,G_\psi\right]_x = \left[\psi\,G_\psi\right]_x + [G_\psi]_x,$$

where we used the fact that $w\star 1 = 0$. Next, we multiply this last equation by ψ and integrate by part over $\tilde{X}$ to derive

$$\frac{\sigma^2}{2}\|\psi_x\|_{L^2}^2 = -\int_{\tilde{X}}\psi_x\,\psi\,G_\psi\,\mathrm{d}x - \int_{\tilde{X}}\psi_x\,G_\psi\,\mathrm{d}x.$$

The extra terms are zero due to periodicity. Thus,

$$\begin{aligned}\frac{\sigma^2}{2}\|\psi_x\|_{L^2}^2 &\leq \left|\int_{\tilde{X}}\psi_x\,\psi\,G_\psi\,\mathrm{d}x\right| + \left|\int_{\tilde{X}}\psi_x\,G_\psi\,\mathrm{d}x\right| \\ &\leq \|G_\psi\|_{L^\infty}\,\|\psi_x\|_{L^2}\,\|\psi\|_{L^2} + \|\psi_x\|_{L^2}\,\|G_\psi\|_{L^2}.\end{aligned} \tag{2.51}$$

Now, using the inequality (2.12) in Lemma 2.5.1, we obtain

$$\begin{aligned}\|G_\psi\|_{L^\infty} &\leq 2R\left(\|\psi\|_{L^1(X)} + M\right) = 2R\left(\|\rho - 1\|_{L^1(X)} + M\right) \\ &\leq 2R\left(\|\rho\|_{L^1(X)} + 1 + M\right) \leq 2R(M+2).\end{aligned} \tag{2.52}$$

Also, we have

$$
\begin{aligned}
|G_\psi(x)|^2 &= \left(\int w(x-y)\,(\psi(y)+M\rho_r(y))\,\mathrm{d}y \right)^2 \\
&= \left(\int_{x-R}^{x+R} (x-y)\,(\psi(y)+M\rho_r(y))\,\mathrm{d}y \right)^2 \\
&\le \int_{x-R}^{x+R} (x-y)^2\,\mathrm{d}y \int_{x-R}^{x+R} (\psi(y)+M\rho_r(y))^2\,\mathrm{d}y \\
&\le \frac{2}{3}R^3 \int_{x-R}^{x+R} (\psi(y)+M\rho_r(y))^2\,\mathrm{d}y. \qquad (2.53)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\|G_\psi\|_{L^2}^2 &\le \frac{2}{3}R^3 \int_{\tilde{X}} \int_{x-R}^{x+R} (\psi(y)+M\rho_r(y))^2\,\mathrm{d}y\mathrm{d}x \\
&= \frac{2}{3}R^3 \int_{\tilde{X}} \int_{-R}^{R} (\psi(x+y)+M\rho_r(x+y))^2\,\mathrm{d}y\mathrm{d}x \\
&= \frac{2}{3}R^3 \int_{-R}^{R} \int_{\tilde{X}} (\psi(x+y)+M\rho_r(x+y))^2\,\mathrm{d}x\mathrm{d}y \\
&= \frac{4}{3}R^4 \|\psi+M\rho_r\|_{L^2}^2. \qquad (2.54)
\end{aligned}
$$

Using estimates (2.52) and (2.54), we can obtain form (2.51) (recall that uniform distribution is not an equilibrium of the system and hence $\|\psi_x\|_{L^2} \neq 0$)

$$
\begin{aligned}
\frac{\sigma^2}{2}\|\psi_x\|_{L^2} &\le 2R(M+2)\|\psi\|_{L^2} + \frac{2R^2}{\sqrt{3}}\,\|\psi+M\rho_r\|_{L^2} \\
&\le 2R(M+2)\|\psi\|_{L^2} + \frac{2R^2}{\sqrt{3}}\left(\|\psi\|_{L^2} + M\|\rho_r\|_{L^2}\right) \\
&= 2R\left(M+\frac{R}{\sqrt{3}}+2\right)\|\psi\|_{L^2} + \frac{2R^2M}{\sqrt{3}}\|\rho_r\|_{L^2}. \qquad (2.55)
\end{aligned}
$$

Now, since $\int_X \psi(x)\,\mathrm{d}x = 0$, we can employ the Poincaré inequality [56, Sec. 5.8.1, Thm. 1] to obtain $\|\psi\|_{L^2} \le C\,\|\psi_x\|_{L^2}$. The optimal value for the Poincaré constant for $\tilde{X} = [-1,1]$ is $C = \frac{1}{\pi}$. Combining this result with the inequality (2.55), we have

$$
\left(\sigma^2 - \frac{4R}{\pi}\left(M+\frac{R}{\sqrt{3}}+2\right)\right)\|\psi\|_{L^2} \le \frac{4R^2M}{\pi\sqrt{3}}\|\rho_r\|_{L^2}. \qquad (2.56)
$$

Defining σ_b and c_b as in (2.9) gives the inequality $\|\psi\|_{L^2} \le \frac{1}{\eta}\|\rho_r\|_{L^2}$, where $\eta = \frac{\sigma^2-\sigma_b^2}{c_b}$. □

2.5.3. STABILITY OF STATIONARY STATE

This section is devoted to the proof of Theorem 2.4.3 concerning the stability of stationary state. All the norms in this subsection are with respect to the domain $\tilde{X} = [-1,1]$ (as opposed to $X = [0,1]$), unless indicated otherwise.

Proof of Theorem 2.4.3. We follow similar arguments as the ones in [55], except we consider a general stationary state ρ^s (instead of the uniform distribution considered in [55]). Let $\psi = \rho - \rho^s$ so that $\int_X \psi(x)\,\mathrm{d}x = 0$. From the dynamic equation (2.6), we obtain

$$\begin{aligned}
\psi_t &= \left[(\psi+\rho^s)\,G_{\psi+\rho^s}\right]_x + \frac{\sigma^2}{2}\,[\psi+\rho^s]_{xx} \\
&= \left[(\psi+\rho^s)\,(w\star\psi + G_{\rho^s})\right]_x + \frac{\sigma^2}{2}\,[\psi+\rho^s]_{xx} \\
&= \left[\psi\,(w\star\psi+G_{\rho^s})\right]_x + \left[\rho^s\,(w\star\psi)\right]_x + \left[\rho^s\,G_{\rho^s}\right]_x + \frac{\sigma^2}{2}\,\psi_{xx} + \frac{\sigma^2}{2}\,\rho^s{}_{xx} \\
&= \left[\psi\,(w\star\psi+G_{\rho^s})\right]_x + \left[\rho^s\,(w\star\psi)\right]_x + \frac{\sigma^2}{2}\,\psi_{xx},
\end{aligned} \tag{2.57}$$

where for the last equality we used the fact that ρ^s is a solution to the stationary equation (2.8), that is,

$$\left[\rho^s\,G_{\rho^s}\right]_x + \frac{\sigma^2}{2}\,\rho^s_{xx} = 0.$$

Multiplying (2.57) by ψ and integrating by part over $\tilde{X}$ we obtain (the extra terms are zero due to periodicity)

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\psi\|_{L^2}^2 + \frac{\sigma^2}{2}\|\psi_x\|_{L^2}^2 \\
&\qquad \le \left|\int_{\tilde{X}} \psi_x\,\psi\,(w\star\psi+G_{\rho^s})\,\mathrm{d}x\right| + \left|\int_{\tilde{X}} \psi_x\,\rho^s\,(w\star\psi)\,\mathrm{d}x\right| \\
&\qquad \le \left(\|w\star\psi\|_{L^\infty} + \|G_{\rho^s}\|_{L^\infty}\right)\,\|\psi_x\|_{L^2}\,\|\psi\|_{L^2} + \|\rho^s\|_{L^\infty}\,\|\psi_x\|_{L^2}\,\|w\star\psi\|_{L^2},
\end{aligned} \tag{2.58}$$

Now, from the inequality (2.12) in Lemma 2.5.1, we have

$$\|w\star\psi\|_{L^\infty} \le 2R\,\|\psi\|_{L^1(X)} = 2R\,\|\rho-\rho^s\|_{L^1(X)} \le 2R\,\left(\|\rho\|_{L^1(X)} + \|\rho^s\|_{L^1(X)}\right) = 4R,$$

and

$$\|G_{\rho^s}\|_{L^\infty} \le 2R\left(\|\rho^s\|_{L^1(X)} + M\right) = 2R(1+M).$$

Also, following a similar procedure as in (2.53) and (2.54) with $M=0$, we obtain

$$\|w\star\psi\|_{L^2} \le \frac{2}{\sqrt{3}}R^2\|\psi\|_{L^2}.$$

Finally, from (2.37) in Lemma 2.5.7, we have $\|\rho^s\|_{L^\infty} \le \exp\left(8R(1+M)/\sigma^2\right)$. Using these estimates and the Young's inequality, we can rewrite (2.58) as

$$\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\psi\|_{L^2}^2 + \frac{\sigma^2}{2}\|\psi_x\|_{L^2}^2 &\le \left(2R(3+M) + \frac{2R^2}{\sqrt{3}}\,\exp\left(\frac{8R(1+M)}{\sigma^2}\right)\right)\,\|\psi_x\|_{L^2}\,\|\psi\|_{L^2} \\
&\le \frac{1}{\sigma^2}\left(2R(3+M) + \frac{2R^2}{\sqrt{3}}\,\exp\left(\frac{8R(1+M)}{\sigma^2}\right)\right)^2\,\|\psi\|_{L^2}^2 + \frac{\sigma^2}{4}\,\|\psi_x\|_{L^2}^2.
\end{aligned}$$

Hence,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\psi\|_{L^2}^2 \leq \frac{1}{\sigma^2}\left(2R(3+M)+\frac{2R^2}{\sqrt{3}}\exp\left(\frac{8R(1+M)}{\sigma^2}\right)\right)^2\|\psi\|_{L^2}^2-\frac{\sigma^2}{4}\|\psi_x\|_{L^2}^2.$$

Once again, since $\int_X \psi(x)\,\mathrm{d}x = 0$, we can employ the Poincaré inequality [56, Sec. 5.8.1, Thm. 1] $\|\psi\|_{L^2} \leq C\|\psi_x\|_{L^2}$ with the optimal Poincaré constant $C = \frac{1}{\pi}$ to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\psi\|_{L^2}^2 \leq \left\{\frac{2}{\sigma^2}\left(2R(3+M)+\frac{2R^2}{\sqrt{3}}\exp\left(\frac{8R(1+M)}{\sigma^2}\right)\right)^2-\frac{\pi^2\sigma^2}{2}\right\}\|\psi\|_{L^2}^2.$$

Then, by Grönwall's inequality, we have

$$\|\psi(t)\|_{L^2}^2 \leq \|\psi(0)\|_{L^2}^2\exp\left[\left\{\frac{2}{\sigma^2}\left(2R(3+M)+\frac{2R^2}{\sqrt{3}}\exp\left(\frac{8R(1+M)}{\sigma^2}\right)\right)^2-\frac{\pi^2\sigma^2}{2}\right\}t\right].$$

Now, notice that $\|\psi(0)\|_{L^2} \leq \|\rho_0\|_{L^2} + \|\rho^s\|_{L^2}$ is finite. Thus, if the constant factor in the exponential is negative, then $\|\psi(t)\|_{L^2}^2 \to 0$ exponentially fast as $t \to \infty$. Negativity of the this constant factor corresponds to the condition $\sigma > \sigma_s$, where σ_s solves (2.10). □

3

CHARACTERIZATION OF SOLUTION IN FOURIER DOMAIN

Parts of this chapter have been published in IEEE Transactions on Automatic Control **66**, 3 (2021) [1] and in proceedings of 58th IEEE Conference on Decision and Control (2019) [62].

In this chapter, we continue the study of the macroscopic model developed in Chapter 2. In particular, developing ideas from [41, 44, 54], we use Fourier analysis to characterize the clustering behavior of the model under the uniform initial distribution. This chapter is organized as follows. In Section 3.1, two numerical schemes are presented to analyze the so-called order-disorder transition in the system and also the initial clustering behavior of the system. These general schemes are then employed in Section 3.2 for a particular distribution of the radical opinions and verified via numerical simulations of both the agent-based and the macroscopic models.

3.1. CHARACTERIZATION OF SOLUTION: FOURIER ANALYSIS

In this section, we exploit the periodic nature of the system and use Fourier analysis to study the behavior of the solution to

$$\begin{cases} \rho_t = (\rho\, G)_x + \frac{\sigma^2}{2}\rho_{xx} & \text{in} \quad \tilde{X}\times(0,T) \\ \rho(\cdot+2,t) = \rho(\cdot,t) & \text{on} \quad \partial\tilde{X}\times(0,T) \\ \rho(x,\cdot) = \rho_0(x) & \text{on} \quad \tilde{X}\times\{t=0\}, \end{cases} \tag{3.1}$$

where

$$G(x,t) := w(x) \star \big(\rho(x,t) + M\rho_r(x)\big), \tag{3.2}$$

with *uniform initial condition*, i.e., $\rho_0 = 1$. To this end, we derive a system of ordinary differential equations (ODEs) describing the evolution of Fourier coefficients of the normal opinion density ρ. Then, these ODEs are used for the identification of the order-disorder transition. Precisely, a numerical scheme is presented for approximating the critical noise level at which this transition occurs. Moreover, we use these ODEs to provide another approximation scheme for characterizing the initial clustering behavior of the system including the number and the timing of possible clusters. These numerical schemes are in essence similar to the linear stability analysis previously employed by [38, 41, 44, 54, 63] for analysis of noisy bounded confidence models without radicals.

3.1.1. FOURIER ODES FOR MACROSCOPIC MODEL

Notice that the set $\{\cos(\pi n x)\}_{n=0}^{\infty}$ is an orthogonal basis for the space $L^2_{ep}(\tilde{X})$ containing even 2-periodic functions on $\tilde{X} = [-1,1]$. Then, the even 2-periodic extension in the model allows us to consider the Fourier expansions of ρ and ρ_r in the form of

$$\rho(x,t) = \sum_{n=0}^{\infty} p_n(t)\,\cos(\pi n x) \quad \text{and} \quad \rho_r(x) = \sum_{n=0}^{\infty} q_n\,\cos(\pi n x). \tag{3.3}$$

By inserting the expansions (3.3) into (3.1) and setting the inner product of the residual with elements of the basis to zero (i.e., taking inverse Fourier transform), we can obtain a system of quadratic ODEs describing the evolution of Fourier coefficients. Considering the first N_f frequency components, these ODEs are

$$\dot{p}_n = c_n + b_n^T p + p^T Q_n p, \quad n = 1,\dots,N_f, \tag{3.4}$$

where $p = (p_1, p_2, \dots, p_{N_f})^T \in \mathbb{R}^{N_f}$. Note that for $n = 0$, i.e., the constant term in the Fourier expansion, we obtain $\dot{p}_0 = 0$. This is due to the periodic nature of the system that preserves the zeroth moment. The coefficients in (3.4) are given by $(n,k,l = 1,\dots,N_f)$

$$
\begin{aligned}
c_n &= 2MR\, f_n\, q_n, \\
(b_n)_k &= \begin{cases} 2R\, f_n + \frac{MR}{2}\, f_{2n}\, q_{2n} - \frac{\pi^2\sigma^2 n^2}{2}, & k = n \\ nMR\left\{\frac{q_{n+k}\, f_{n+k}}{n+k} + \frac{q_{|n-k|}\, f_{n-k}}{n-k}\right\}, & k \neq n, \end{cases} \\
(Q_n)_{k,l} &= \begin{cases} nR\frac{f_k}{k}, & l = n-k > 1 \\ nR\left\{\frac{f_k}{k} + \frac{f_{n-k}}{n-k}\right\}, & l = k-n > 1 \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}
\tag{3.5}
$$

where

$$
f_n := -\cos(\pi n R) + \operatorname{sinc}(\pi n R), \tag{3.6}
$$

with $\operatorname{sinc} x = \frac{\sin x}{x}$. Recall that $q_n,\ n \in \mathbb{N}$, are the Fourier coefficients of ρ_r.

Interestingly, one notices that the interaction between different frequency components in the quadratic terms is limited to those that are in a sense complements of each other. That is, each frequency n of ρ is affected by the frequency pairs (n_1, n_2) such that either $n_1 + n_2 = n$ or $|n_1 - n_2| = n$. This, in turn, leads to a particular structure for the matrix Q_n in the quadratic terms. As expected, a similar behavior is seen in the linear terms: the effect of each frequency k of ρ on a given frequency n of ρ is modulated by the frequency components $n+k$ and $|n-k|$ of ρ_r.

3.1.2. ORDER-DISORDER TRANSITION

A common behavior in noisy interactive particle systems is the order-disorder transition. Here, "order" refers to a clustered behavior, while "disorder" refers to a uniform (or close to uniform) opinion profile. For large values of σ, the effect of the diffusion process can overcome the attracting forces among agents preventing the system from forming any cluster. This behavior has been analyzed and observed in several noisy bounded confidence models for opinion dynamics. Pineda et. al. used linear stability analysis in [38, 63] to compute the critical noise level above which the clustering behavior disappears for a modified version of Defuant model [28]. This technique was also used in [44, 54] to compute the critical noise level for a noisy HK system similar to our model, except without radicals.

Here, we provide a method for approximating the critical noise level σ_c at which the transition occurs. To this end, we linearize the systems at $t = 0$ to obtain a system of linear ODEs expressed as

$$
\dot{p} = c + Bp. \tag{3.7}
$$

The vector $c \in \mathbb{R}^{N_f}$ and the matrix $B \in \mathbb{R}^{N_f \times N_f}$ are defined accordingly using the objects c_n and b_n in (3.5). We emphasize that the linearization (3.7) is for a uniform initial condition, i.e., $p_n(0) = 0$ for $n = 1, \ldots, N_f$.

Looking at coefficients c_n and b_n in (3.5), we notice that the noise level σ only appears in the diagonal entries of B such that by increasing σ, these diagonal entries decrease and eventually become negative. That is, for a large enough σ, the matrix B is

Hurwitz (all its eigenvalues have negative real parts) and the linearized system (3.7) is stable. This will be our first criterion for determining the critical noise level σ_c: the noise level above which B is Hurwitz. In order to consider the effect of the constant linear growth rates c in (3.7), we further require the stationary values $\bar{p}_n, n = 1,\dots,N_f$, of the linearized system (3.7) (i.e., the solution to the equation $c + B\bar{p} = 0$) to be relatively small. In other words, taking the equilibrium of the linearized system, $1 + \sum_{n=0}^{N_f} \bar{p}_n \cos(\pi n x)$, as an approximation of the stationary state ρ^s, we require ρ^s to be close to uniform distribution $\rho = 1$, representing disorder. Similar to the theoretical estimate of Theorem 2.4.2, we quantify this criterion by using Parseval's identity and setting

$$\|\rho^s - 1\|_{L^2}^2 \approx \|\bar{p}\|_2^2 < \gamma, \tag{3.8}$$

where the constant $\gamma > 0$ determines the level of similarity between ρ^s and uniform distribution. To sum up, for a given $\gamma > 0$, we solve numerically for the minimum level of noise for which B is Hurwitz and the inequality (3.8) holds.

3.1.3. INITIAL CLUSTERING BEHAVIOR

For noises smaller than the critical noise level σ_c, we expect to see a clustering behavior. In order to characterize the initial clustering behavior, we make use of the *exponential growth rate* $\gamma_n := (b_n)_n$ and *linear growth rate* c_n given in (3.5). The proposed numerical method is as follows. We ignore the interactions between different frequencies in (3.4), that is, for each frequency $n = 1,\dots,N_f$, we consider the equation $\dot{p}_n = c_n + \gamma_n p_n$ with $p_n(0) = 0$ (corresponding to uniform initial distribution) for initial evolution of the Fourier coefficient p_n. Then, for a given set of model parameters (σ, R, M) and radical opinions density ρ_r, we numerically compute the *dominant wave-number* $n^* := \operatorname{argmax}_n \gamma_n$ with $\gamma_{n^*} > 0$, that is, the unstable mode with the largest exponential growth rate. We speculate that the corresponding trigonometric term $p_{n^*}\cos(\pi n^* x)$ is the dominant component of the initial clustering behavior. The sign of p_{n^*} depends on the linear growth rate c_{n^*}: we have $p_{n^*} > 0$ if $c_{n^*} > 0$, and $p_{n^*} < 0$ otherwise.

Considering the even 2-periodic extension of the model, the dominant waveform must be interpreted on the interval $\tilde{X} = [-1, 1]$. Then, the *number of initial clusters* n_{clu} in the interval $X = [0, 1]$ resulting from the waveform $1 + p_{n^*}\cos(\pi n^* x)$ is given by

$$n_{\text{clu}} := \begin{cases} \lfloor \frac{n^*}{2} \rfloor + 1, & c_{n^*} > 0 \\ \lceil \frac{n^*}{2} \rceil, & c_{n^*} < 0, \end{cases} \tag{3.9}$$

where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are the floor and ceiling functions, respectively. The timing of this initial clustering behavior is also expected to be inversely related to γ_{n^*}. Indeed, by solving for the time for which the solution to the equation $\dot{p}_n = c_n + \gamma_n p_n$ is equal to ± 1, we can approximate the *time to initial clustering* by

$$t_{\text{clu}} := \frac{1}{\gamma_{n^*}} \ln\left(1 + \frac{\gamma_{n^*}}{|c_{n^*}|}\right). \tag{3.10}$$

A similar approximation has been used in [54] to derive the time to the initial clustering using fluctuation theory.

3.2. NUMERICAL STUDY

In this section, we provide a numerical study of the model at hand for a particular distribution of radical agents/opinions through simulations of the corresponding discrete- and continuum-agent models. Furthermore, we validate the result of our Fourier analysis for identification of order-disorder transition (Section 3.1.2) and characterization of initial clustering behavior (Section 3.1.3).

The particular radical distribution considered in this section is a triangular distribution with average A and width $2S$, i.e.,

$$\rho_r(x) = \begin{cases} \frac{1}{S^2}(S - |x - A|), & |x - A| \le S \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

Although this choice may seem specific, it is rich enough for our purposes. In particular, with this choice, the zeroth, first and second moments of the radical opinions density are simply captured by the parameters M, A, and S, respectively. Moreover, we assume that the radicals are concentrated around their average opinion, that is, we consider small values of S (with respect to the confidence range R).

In the sequel, we make use of the *order parameter*

$$Q_d(t) = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{1}_{|x_i(t) - x_j(t)| \le R},$$

introduced by [44] and its continuum counterpart

$$Q_c(t) = \int_{X^2} \rho(x,t)\, \rho(y,t)\, \mathbf{1}_{|x-y| \le R}\, \mathrm{d}x\mathrm{d}y,$$

to quantify orderedness in the clustering behavior of the model. In words, the order parameter Q is the (normalized) number/mass of agents that are in the R-neighborhood of each other and hence interacting. In particular, in the continuum case, we have $Q_c = 2R$ for a uniform distribution of opinions (complete disorder) while $Q_c = 1$ for a single-cluster distribution with all agents residing in an interval of width R or less (complete order). In the case of a clustered behavior, roughly speaking, the inverse of the order parameter is equal to the number of clusters. We also use the evolution of order parameter to characterize the timing of the clustering behavior.

In all the simulation results reported in this section the width of radicals distribution and the confidence range are fixed at $S = 0.1$ and $R = 0.1$, respectively.

3.2.1. SIMULATION OF MODELS

Discrete-agent model: For the discrete-agent model, the SDEs (2.4) are solved numerically using the Euler-Maruyama method for $N = 500$ normal agents, with time step $\Delta t = 0.01$. To be precise, we solve the following SDEs

$$\begin{cases} \mathrm{d}x_i = -\frac{1}{N}\left(\sum_{j \in \mathcal{N}_i}(x_i - x_j^{\text{ext}}) + \sum_{j \in \mathcal{N}_i}(x_i - x_{r_j}^{\text{ext}})\right)\mathrm{d}t + \sigma\, \mathrm{d}W_t^i, \\ x_i(0) = x_{i_0}. \end{cases} \tag{3.12}$$

where x_i^{ext}, $i = 1, \dots N$, are the opinions of normal agents and $x_{r_i}^{\text{ext}}$, $i = 1, \dots N_r$, are the opinions of radical agents with $N_r = MN$. The superscript "ext" corresponds to the

Algorithm 1 Euler-Maruyama method for even 2-periodic extension of SDE (3.12)

Step 0. $\mathbf{x}_r = (x_{r_1}, x_{r_2}, \cdots, x_{r_{N_r}})^T \sim \rho_r(x)$;
$\mathbf{x}_r^{\text{ext}} = [\mathbf{x}_r;\ -\mathbf{x}_r;\ 2-\mathbf{x}_r]$;
for $t = 0$ **to** $t = \frac{T}{\Delta t} - 1$: **do**
Step 1. $\mathbf{x}^{\text{ext}}(t) = [\mathbf{x}(t);\ -\mathbf{x}(t);\ 2-\mathbf{x}(t)]$, where $\mathbf{x}(t) = (x_1(t), x_2(t), \cdots, x_N(t))^T$;
Step 2. $\dot{x}_i(t) = -\frac{1}{N}\left(\sum_{j\in\mathcal{N}_i}(x_i - x_j^{\text{ext}}) + \sum_{j\in\mathcal{N}_i}(x_i - x_{r_j}^{\text{ext}})\right)$;
Step 3. $dW_t^i = z_i\sqrt{\Delta t}$, where $z_i \sim \mathrm{N}(0,1)$;
Step 4. $x_i(t+1) = x_i(t) + \dot{x}_i(t)\cdot\Delta t + \sigma\ dW_t^i$;
Step 5. $x_i(t+1) = x_i(t+1) \bmod (2L)$;
if $x_i(t+1) > L$, then $x_i(t+1) = 2 - x_i(t+1)$.
end for

even 2-periodic extension as we explain shortly. Algorithm 1 summarizes the numerical scheme for solving (3.12). As described above, we assume that the radicals have a triangular distribution centered at A with width $2S$. That is, we produce a random sample of radicals with size N_r from the triangular distribution (3.11) (Step 0). In particular, for complete correspondence between the discrete- and continuum- agent models, we also consider the effect of even 2-periodic extension in our simulations. To this end, we use even 2-periodic extensions of $\mathbf{x}$ and $\mathbf{x}_r$ for calculating the sum on the right-hand side of (3.12) (vectors denoted by $\mathbf{x}^{\text{ext}}$ and $\mathbf{x}_r^{\text{ext}}$ in Steps 0, 1 and 2). Also, because of periodicity, in each iteration, the opinion values outside the support $X = [0,1]$ are *reflected* back to X (Step 5).

Continumm-agent model: To solve the continuum-agent model described by PDF (3.1) numerically, we use the Fourier ODEs (3.4) to compute the coefficients of Fourier expansion of normal opinion density ρ using the first N_f terms of the expansion. However, regarding the radical opinion density, one notices that the considered triangular distribution does not satisfy the conditions of Theorem 2.4.1 for well-posedness of the dynamics, that is, $\rho_r \notin H^2_{ep}(\tilde{X})$. This will not be an issue since we will be working with the projection of the proposed ρ_r in the Hilbert space $L^2_{ep}(\tilde{X})$. That is, we use the Fourier coefficients of ρ_r in (3.4) which for the triangular distribution (3.11) are given by

$$q_n = 2\cos(n\pi A)\,\mathrm{sinc}^2(n\pi S/2). \tag{3.13}$$

To be precise, we need the Fourier coefficients q_n of ρ_r for $1 \le n \le 2N_f$, that is, twice the length of Fourier expansion of ρ; see the linear terms of (3.4). For the initial condition, we again consider uniform distribution $\rho_0 = 1$, which corresponds to $p_0 = 1$ and $p_n(0) = 0$ for the Fourier coefficients.

Alternatively, we can employ a semi-explicit pseudo-spectral method, similar to the one provided by [44], for numerically solving (3.1). To be precise, using the first N_f terms of Fourier expansions of ρ and ρ_r, we can write

$$\rho(x,t) + M\rho_r(x) = \sum_{k=-N_f}^{N_f}\left(\hat{\rho}_k(t) + M\hat{\rho}_{r_k}\right)e^{i\pi kx}.$$

Algorithm 2 Pseudo-spectral method for PDE (3.1)

Step 0. for $x \in [-1,0]$ set $\rho_r(x) = \rho_r(-x)$;
$\hat{\rho}_{r_k} = \text{FFT}\left[\rho_r(x)\right]$;
for $t = 0$ **to** $t = \frac{T}{\Delta t} - 1$: **do**
Step 1. for $x \in [-1,0]$ set $\rho(x,t) = \rho(-x,t)$;
Step 2. $\hat{\rho}_k(t) = \text{FFT}\left[\rho(x,t)\right]$;
Step 3. $\hat{G}_k(t) = -\frac{2iR}{\pi k} f_k\,(\hat{\rho}_k(t) + M\hat{\rho}_{r_k})$, $\hat{G}_0(t) = 0$;
$G(x,t) = \text{iFFT}\left[\hat{G}_k(t)\right]$;
Step 4. $h(x,t) = \rho(x,t)\; G(x,t)$;
$\hat{h}_k(t) = \text{FFT}\,[h(x,t)]$;
Step 5. $\hat{\rho}_k(t+1) = \left(i\pi k\hat{h}_k(t) - \frac{\pi^2\sigma^2k^2}{2}\hat{\rho}_k(t+1)\right)\cdot\Delta t + \hat{\rho}_k(t)$;
$\hat{\rho}_0(t+1) = \hat{\rho}_0(t)$;
$\rho(x,t) = \text{iFFT}\left[\hat{\rho}_k(t)\right]$;
end for

Inserting this into (3.2), we obtain (we are dropping the subscript ρ for convenience)

$$G(x,t) \quad = \quad \sum_{-N_f \le k \le N_f, k\neq 0} -\frac{2iR}{\pi k} f_k\left(\hat{\rho}_k(t) + M\hat{\rho}_{r_k}\right) e^{i\pi kx},$$

where f_k is given by (3.6). Hence,

$$\hat{G}_k(t) = \begin{cases} -\frac{2iR}{\pi k} f_k\left(\hat{\rho}_k(t) + M\hat{\rho}_{r_k}\right), & k \neq 0 \\ 0, & k = 0. \end{cases}$$

With Fourier coefficients of G in terms of Fourier coefficients of ρ in hand, we can apply the pseudo-spectral method for solving (3.1) as described in Algorithm 2. As shown, the multiplication $h = \rho\; G$ on the right-hand side of the first equation in (3.1) is performed in the time domain (Step 4), while the differentiations with respect to x are performed in the frequency domain (Step 5). Note that the symmetric nature of the solution is preserved in the algorithm (Step 1). Also, preservation of mass is satisfied by setting $\hat{\rho}_0(t+1) = \hat{\rho}_0(t)$ (Step 5). Finally, we note that the algorithm is semi-explicit (see the first equation in Step 5).

The main difference between the two methods is that the pseudo-spectral method solves the PDE for a set of discrete points in the opinion space ($x \in X$) while solving the Fourier ODEs gives an approximation of the solution in terms of a finite basis for the corresponding Hilbert space. These two methods (if both converge) result in the same solution. Fig. 3.1 compares the result of numerical simulations of the model using these two methods for a particular combination of system data. Note that, in these simulations, the number of points for the spatial discretization in the pseudo-spectral method is twice the the number N_f of frequencies in the Fourier ODEs so that the methods are compatible, i.e., both include the same set of frequency components. The left panel of Fig. 3.1 shows a similar result using these two methods for $N_f = 32$ frequencies. However, as the number of frequencies considered in the simulations is decreased, we see that the pseudo-spectral method starts to diverge while the Fourier ODEs are still stable.

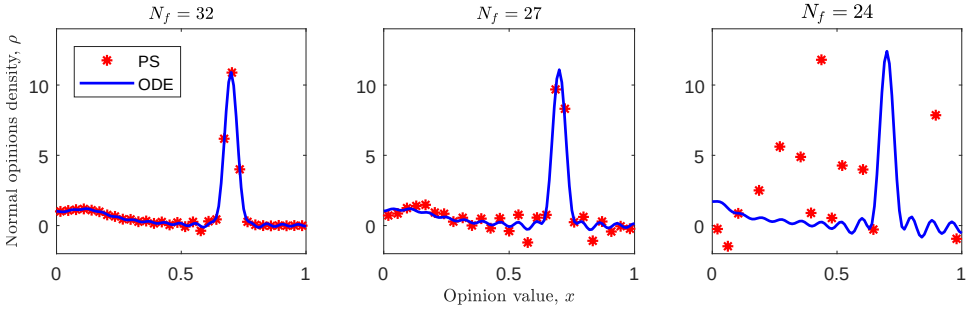


Figure 3.1: Comparison of the pseudo-spectral method (PS) with $\Delta t = 0.01$ and the Fourier ODEs (ODE) for numerical simulation of the continuum-agent model (3.1). The results are for $t = 400$ with system data $(\sigma, M, A) = (0.03, 0.1, 0.7)$. In the right panel, some of the points in the solution of the pseudo-spectral method are outside the limits of the vertical axis.

In the remainder of this section, we use the Fourier ODEs (3.4) with $N_f = 128$ for numerical simulation of the continuum-agent model.

3.2.2. ORDER-DISORDER TRANSITION

In this section, we numerically study the order-disorder transition in the model. In particular, we consider the effect of the relative mass M of radicals on the critical noise level σ_c at which this transition occurs. Furthermore, we use our simulation results to examine the approximation scheme presented in Section 3.1.2. In this regard, we note that the interplay between the confidence range R and the critical noise level σ_c have been studied in [44]. There, the authors showed that as R increases, the critical noise level σ_c also increases in such a way that for small values of R, we observe a first-order transition.

ILLUSTRATIVE EXAMPLE

Our model exhibits the same order-disorder transition previously reported for similar noisy HK systems [41, 44, 54]. Fig. 3.2 shows this effect for a particular combination of system data in the discrete- and continuum-agent models. Notice that for σ larger than a critical level the clustering behavior almost disappears (see the lower panel corresponding to $\sigma = 0.05$ in Fig. 3.2a). To be more precise, a higher level of noise decreases the lifetime of clustering behaviors with a larger number of clusters. This effect can be particularly seen in the evolution of the order parameter in Fig. 3.2b. In this regard, notice that for noises smaller than the critical noise level (here $\sigma < 0.05$) the horizontal parts in the order parameter in Fig. 3.2b correspond to a clustered behavior, where the number of clusters is equal to the inverse of the order parameter. To illustrate, observe that for $\sigma = 0.03$ and $\sigma = 0.04$, the system reaches a single-cluster profile around the average radical opinion $A = 0.7$. Notice, however, for $\sigma = 0.03$ the system first goes through a 2-cluster profile corresponding to the horizontal part in the blue solid line at height 0.5 in Fig. 3.2b. On the other hand, for $\sigma = 0.02$, we observe a 2-cluster profile at $t = 10^4$ in Fig. 3.2a. Notice, however, how the system goes through 4-cluster and 3-cluster profiles as depicted in Fig. 3.2b (the horizontal parts in the order parameter). Finally, for $\sigma = 0.01$,

we observe a very fast emergence of a 4-cluster profile (Fig. 3.2b) that has survived until $t = 10^4$ as shown in Fig. 3.2a. Here, we also notice that the exact position of clusters in the discrete- and continuum-agent models differ. This particular difference between mean-field and agent-based models has been also mentioned in [38, 63]. Indeed, our numerical simulations show that even the number of clusters resulting from mean-field and agent-based models may differ; this also has been reported and explained previously in [44]. Finally, we note that for $M = 0.1$, the approximation scheme explained in Section 3.1.2 results in $\sigma_c = 0.043$ for $\gamma = 1$ and $\sigma_c = 0.051$ for $\gamma = 0.1$ (see (3.8) for influence of γ).

EFFECT OF M ON σ_c

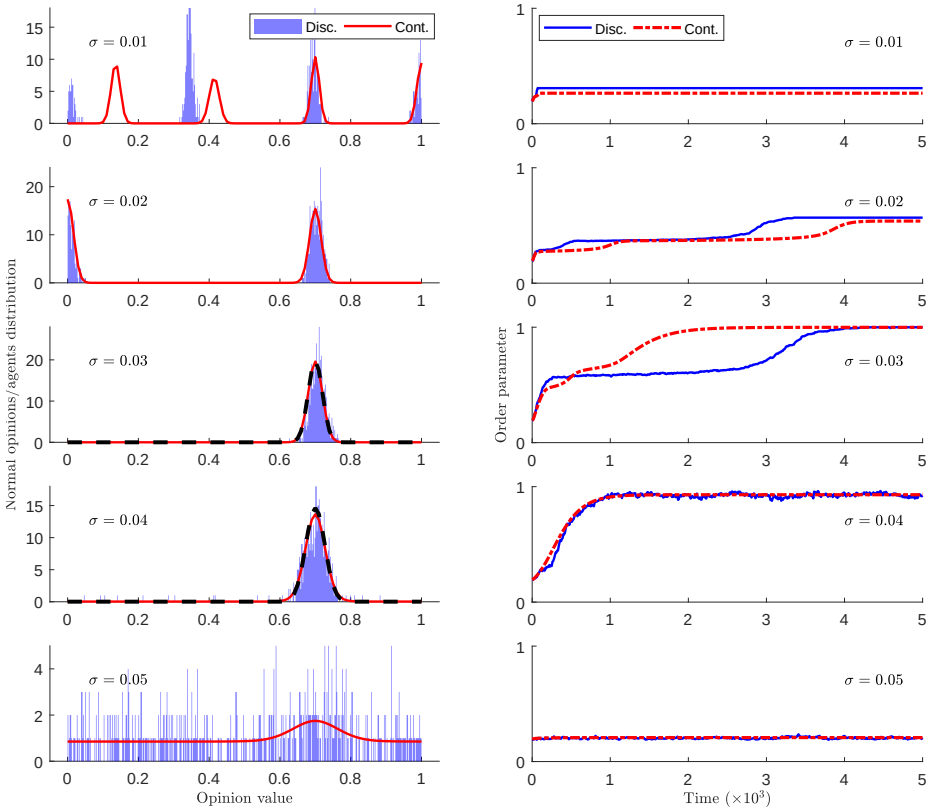
Fig. 3.3 shows the order parameter derived numerically by simulating the continuum- and discrete-agent models. Notice how for each M, as noise increases, the system experiences a transition from order (with $Q \approx 1$ in the yellow stripe) to disorder (with $Q \approx 0.2$ in the dark blue area in the upper part of the plots). Also, we note that the blue stripe in the lower part of plots in Fig. 3.3 represents clustering behaviors with a larger number of clusters (similar to the behavior seen for $\sigma = 0.01$ in Fig. 3.2).

This result shows that as the relative mass of radicals M increases, the corresponding critical noise level σ_c, above which the system is in a disordered state, also increases. The dependence of σ_c on M is in the form of a concave function. Furthermore, for small values of M, the transition seems to be discrete, signaling a first-order transition. However, for large values of M the transition becomes blurry. This phenomenon was also reported in [44] for the dependence of the critical noise level on the confidence range R. In this regard, note that as M increases, the required noise level for disordered behavior also increases, which leads to wider clusters. This, in turn, makes it difficult to differentiate order from disorder; see, e.g., the panels corresponding to $\sigma = 0.04, 0.05$ in Fig. 3.2.

Also shown in Fig. 3.3 (red lines) is the result of the scheme provided in Section 3.1.2 for approximating the critical noise level. As can be seen, the scheme indeed provides a good approximation of the critical noise level. In particular, the dashed red line (for $\gamma = 1$) almost perfectly separates the two phases of order and disorder.

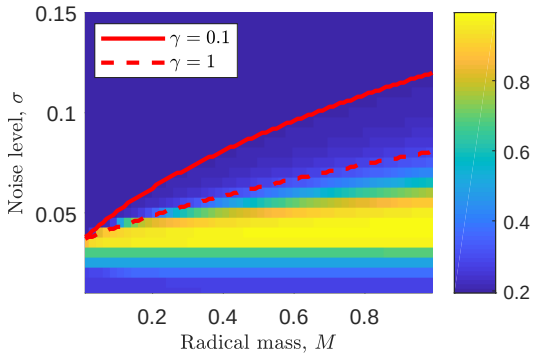
3.2.3. INITIAL CLUSTERING BEHAVIOR

For noises smaller than the critical noise level, agents start to form clusters; see Fig. 3.2. In particular, we observe a cluster of normal agents around the average A of the radical opinion due to the force field generated by the radicals. Generally, three types of clusters may form: (1) the cluster at the average radical opinion A, (2) the cluster(s) at the extreme opinions $x = 0$ and/or $x = 1$, and (3) the cluster(s) around opinion values other than $x = 0, 1, A$. The third type of cluster is expected to perform a random walk with their center of mass moving like a Brownian motion (assuming clusters do not interact). The effective diffusivity of these Brownian motions is inversely related to the size of the cluster, i.e., the number of agents in the cluster. This will result in a process of consecutive merging between these clusters until the complete disappearance of them. Detailed descriptions of this process are provided in [44, 54]. Notice however that this description does not apply to cluster(s) formed at $x = A$ and $x = 0, 1$. These clusters are affected by forces other than the normal attractions among the agents within the cluster. The cluster formed at

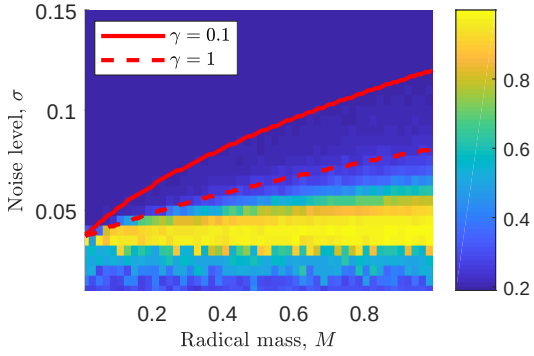


(a) Distribution of opinions/agents at $t = 10^4$ (b) Evolution of order parameter

Figure 3.2: Numerical simulation of the discrete-agent model (Disc.) and continuum-agent model (Cont.) for different values of noise σ with system data $(M, A) = (0.1, 0.7)$. As noise increases the number of clusters decreases so that for a large enough noise the clustering behavior disappears (see Section 3.2.2). The black dashed lines in left panels for $\sigma = 0.03, 0.04$ are the approximate stationary solutions (2.50). This result shows that the approximate solution is indeed a good approximation as it almost perfectly matches the numerical solution of the continuum-agent model.



(a) Continuum-agent model



(b) Discrete-agent model

Figure 3.3: The order parameter at $t = 10^3$ from numerical simulation of the continuum- and discrete-agent models starting form uniform initial distribution. For the discrete-agent model, the average of order parameter over the time window $[900, 1000]$ is reported. The plot covers the region $\sigma \times M \in [0.01, 0.15] \times [0.01, 1]$ with step sizes $\Delta\sigma = 0.005$ and $\Delta M = 0.02$. The red lines show the result of the numerical scheme described in Section 3.1.2 for approximating the critical noise level for different values of γ with respect to the second criterion (3.8). See Section 3.2.2 for details.

$x = A$ is under influence of radicals, and the possible clusters at the extreme opinions $x = 0, 1$ are reinforced due to the even 2-periodic extension considered in our model. The behavior of these clusters (survival or dissolution) depends on their size, the exogenous force acting on them, and the effect of other clusters in their neighborhood.

In this section, we use the analysis scheme provided in Section 3.1.3 to investigate the effect of the zeroth and first moment of radicals (M and A, respectively) on the initial clustering behavior of the model for noises smaller than the critical level. In particular, we investigate the effect of M and A on the number, position, and timing of initial clusters for different values of σ. We again emphasize that we are considering a concentrated triangular distribution for radical agents and a uniform initial distribution for normal agents. Let us begin with illustrating how the objects introduced in Section 3.1.3, namely, exponential and linear growth rates and the dominant wavenumber, can be used to characterize the initial clustering behavior.

ILLUSTRATIVE EXAMPLE

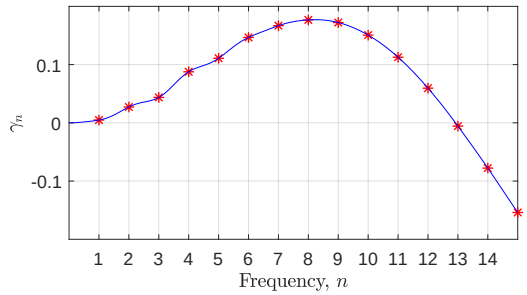
Consider the system data $(\sigma, M, A) = (0.01, 0.1, 0.7)$. Fig. 3.4 depicts the values of the exponential growth rate γ_n and the linear growth rate c_n for different frequencies. In Fig. 3.4a, we observe that the unstable mode with the maximum exponential growth rate is $n^* = 8$ with $\gamma_{n^*} = 0.177$. Fig. 3.4b shows that the linear coefficient corresponding to this frequency is $c_{n^*} = 0.007 > 0$. Then, (3.9) implies that the initial clustering behavior is expected to have $n_{\text{clu}} = 5$ clusters. Also, using (3.10), we obtain $t_{\text{clu}} = 18.16$ for the time to initial clustering.

Fig. 3.5 shows the time evolution of the distribution of normal opinions/agents for the system data corresponding to Fig. 3.4. For the continuum-agent model, we can see a 5-cluster profile corresponding to the speculated waveform as depicted in Fig. 3.5a. A similar clustering behavior is observed in the Monte Carlo simulation of the discrete-agent model in Fig. 3.5b. Here, we observe three clear clusters: the cluster at average radical opinion $A = 0.7$ and the two clusters at extreme opinions $x = 0, 1$. However, we observe an almost uniform distribution of normal agents in the opinion range $[0.1, 0.5]$. This is because the exact position of the corresponding clusters formed in the discrete-agent model varies within this range. Individual realizations of the discrete model show one, two, or three clusters in this range with two clusters being the most frequent behavior as expected. This effect has been also reported by [38] in Monte Carlo simulations of a noisy Defuant model. Furthermore, we notice that the timing object $t^* = 18.16$ also gives a good approximation for the onset of the corresponding clustering behavior for both continuum- and discrete-agent systems.

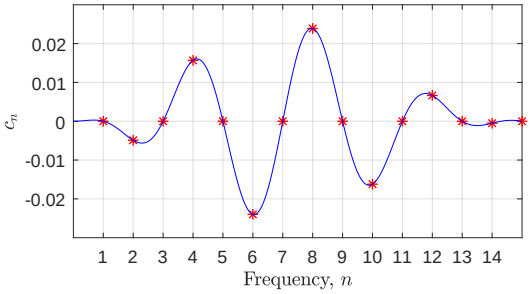
EFFECT OF M AND A ON INITIAL CLUSTERING

Performing a similar analysis to the one provided in the example above, we can compute the dominant wave-number (n^*), the number of initial clusters (n_{clu}) and time to initial clustering (t_{clu}) for a general combination of system data. Fig. 3.6 shows the result of this analysis for different values of M and A at three different noise levels σ. Here, we only considered the values $A < 1 - R = 0.9$ since for $1 - R < A < 1$ the boundary effect due to even 2-periodic extension comes into play.

Comparing the left, middle, and right panels of Fig. 3.6 corresponding to different levels of noise, we observe that as the level of noise increases, the number of clusters in

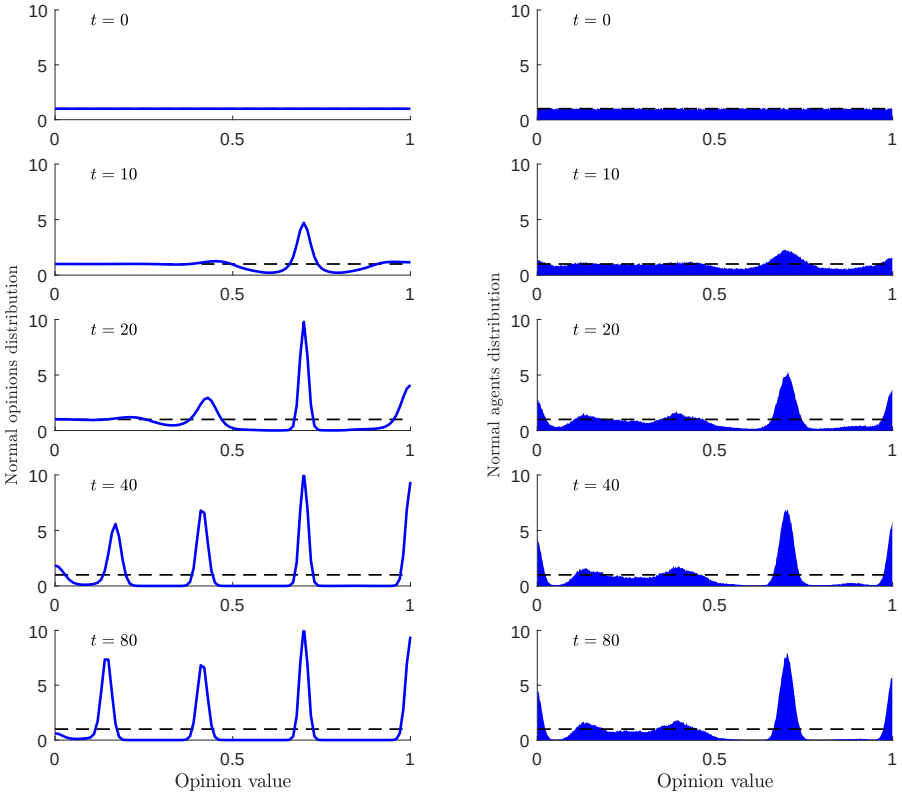


(a) Exponential growth rate



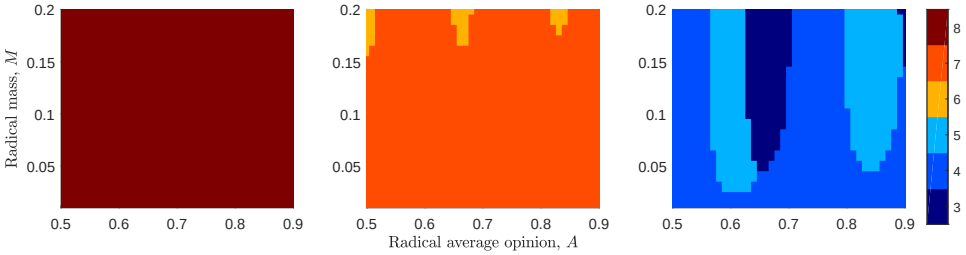
(b) Linear growth rate

Figure 3.4: Exponential and linear growth rates for system data $(\sigma, M, A) = (0.01, 0.1, 0.7)$ for different frequencies. On the left panel we see the maximum exponential growth corresponds to $n^* = 8$ with $\gamma_{n^*} = 0.177$. On the right panel we see $c_{n^*} = 0.007 > 0$. This implies that the waveform $p_8 \cos(8\pi x)$ with $p_8 > 0$ is the dominant component of the initial clustering behavior.

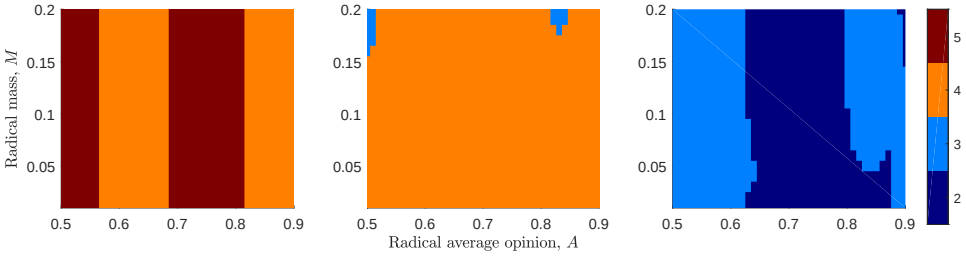


(a) Continuum-agent model (b) Discrete-agent model

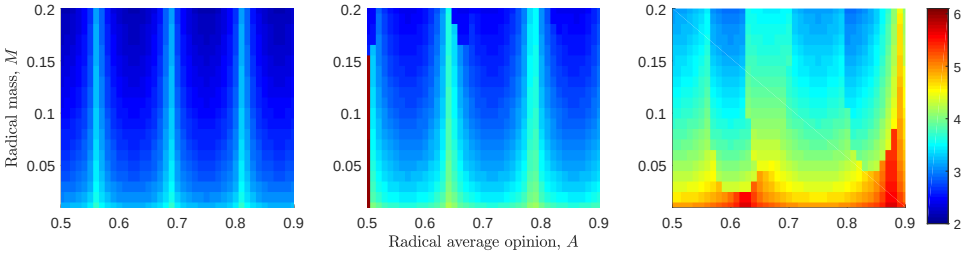
Figure 3.5: Evolution of distribution of normal opinions/agents during the initial clustering behavior for system data $(\sigma, M, A) = (0.01, 0.1, 0.7)$ corresponding to Fig. 3.4. The distributions shown for the discrete-agent model are the average profiles of 300 realizations. The onset of a 5-cluster behavior is observed from approximately $t = 20$ corresponding to the waveform $1 + \cos(8\pi x)$ speculated for the initial clustering behavior with $t_{\text{clu}} = 18.16$.



(a) Dominant wave-number: n^*



(b) Number of initial clusters: n_{clu}



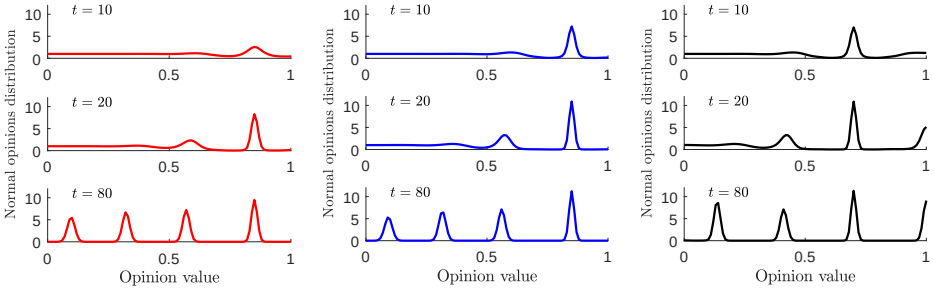
(c) Time to initial clustering: $\ln(t_{\text{clu}})$

Figure 3.6: Characterization of the initial clustering behavior based on the dominant wave-number in the Fourier expansion of the continuum-agent model for different values of M and A, with noise levels $\sigma = 0.01$ (left), $\sigma = 0.02$ (middle), and $\sigma = 0.03$ (right).

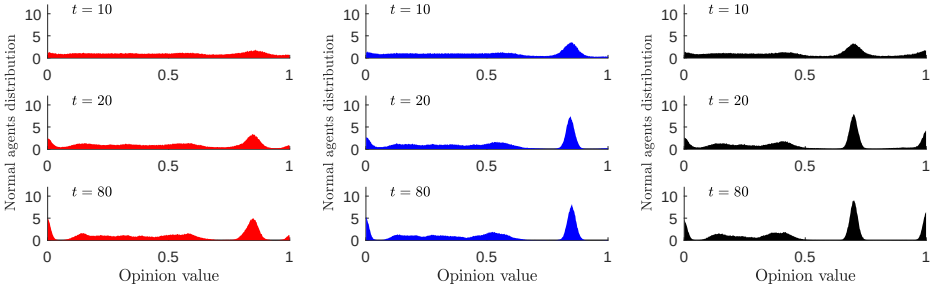
the possible clustering behavior of the system decreases (see Fig. 3.6b), while the timing experiences a general increase (see Fig. 3.6c). This effect has been already shown in Fig. 3.2. In particular, concerning the timing, we notice that as the level of noise decreases, the *initial* clustered profile emerges faster; see Fig. 3.2b.

For low levels of noise, e.g., $\sigma = 0.01$ (see the left panels in Fig. 3.6), the dominant wave-number does not depend on the M or A. In this case, the most important effect of the first moment A of radical opinions density is on the position of clusters. That is, the clustered profile emerges in a way that we observe a particular cluster formed at the average radical opinion A. The parameter A also affects the timing of the clustering behavior in a periodic fashion. On the other hand, the zeroth moment of radical opinions density M only affects the timing of the clustering behavior: as M increases, t_{clu} decreases. Fig. 3.7 shows the simulation results for $\sigma = 0.01$ and compares the evolution of opinions for different values of M and A. For the continuum model in the top panels of Fig. 3.7 we observe that indeed a 4-cluster profile has emerged in all systems. Comparing Figs. 3.7a and 3.7b shows that M only affects the timing of clustering behavior. This effect is better seen in Fig. 3.7g where we observe a faster convergence of order parameter for $\mathscr{S}_2$ with larger M. On the other hand, comparing Figs. 3.7b and 3.7c corresponding to $A = 0.85$ and $A = 0.7$, respectively, we observe a change in the positioning of the clusters. Monte Carlo simulations of the discrete-agent model reveal that the same general description also holds for this system. This is particularly seen in the time evolution of the order parameter in the discrete-agent model as depicted in Fig. 3.7h. However, we once again note that there are differences between the behavior of the continuum- and discrete-agent models. In particular, the evolution of order parameter in Fig. 3.7g shows that the continuum-agent model has seemingly converged to steady-state with four clusters, while this is not the case for the discrete-agent model as can be seen in Fig. 3.7h. Indeed, in the discrete-agent model, as described at the beginning of this section, all the possible clusters formed around opinion values other than $x = 0, 1, A$ will necessarily disappear in the steady state profile, where the time required for their disappearance depends on the noise level and particularly the size of these clusters. Hence, unlike the discrete-agent model, for the continuum-agent model (in the limit $N \to \infty$), the system may require *infinite time* for this merging of the clusters to occur. This, in turn, can lead to different behaviors in the discrete- and continuum-agent models over exponentially large times [44]; see also the evolution of order parameter in Fig. 3.2b.

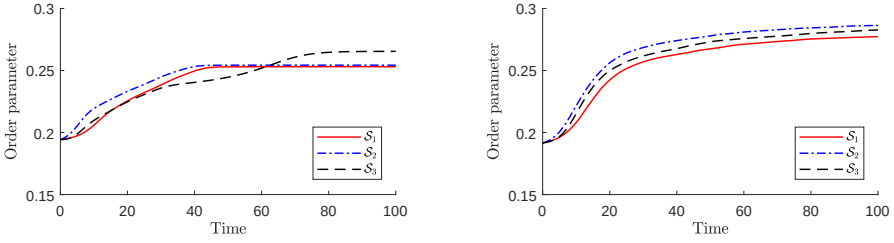
As shown in Fig. 3.6, for higher levels of noise, e.g., $\sigma = 0.03$, we observe nonlinear effects: M and A start to affect the dominant wave-number (see the middle and right panels of Fig. 3.6a). Nevertheless, these effects are limited as the number of clusters is still 3 or 4 for $\sigma = 0.02$, and 2 or 3 for $\sigma = 0.03$. Besides, we still observe a general increase in the timing of the clustering behavior as M decreases. Fig. 3.8 shows the evolution of normal opinions/agents distribution and the corresponding order parameter for three different combinations of M and A at the noise level $\sigma = 0.03$. Once again, in the continuum-agent model, we observe a 2-cluster profile for all combinations as shown in the top panels of Fig. 3.8. For the discrete-agent model, we observe a 3-cluster behavior in which the cluster formed between the two clusters at $x = 0$ and $x = A$ has already disappeared for $\mathscr{S}_3$ in Fig. 3.8f at $t = 400$. Indeed, our simulations for $\sigma = 0.03$ reveal a single cluster around the average radical opinion after a large enough time; see Fig. 3.2.



(a) Continuum - $\mathscr{S}_1$ (b) Continuum - $\mathscr{S}_2$ (c) Continuum - $\mathscr{S}_3$

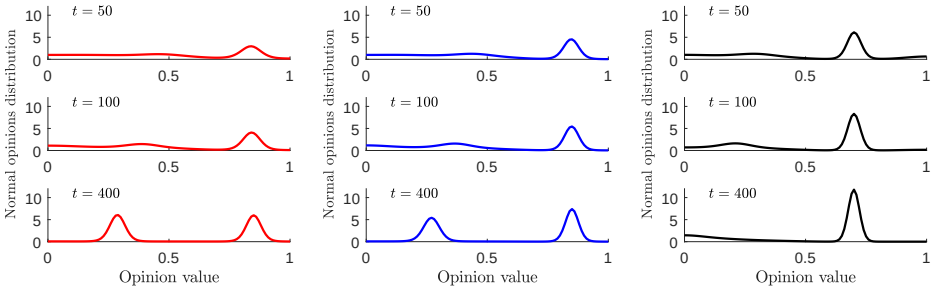


(d) Discrete - $\mathscr{S}_1$ (e) Discrete - $\mathscr{S}_2$ (f) Discrete - $\mathscr{S}_3$

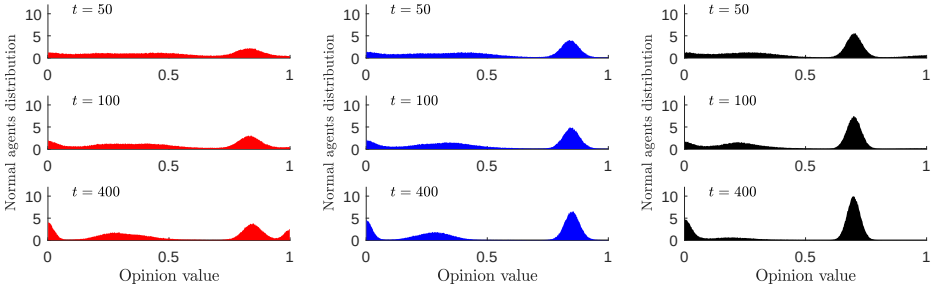


(g) Continuum-agent model (h) Discrete-agent model

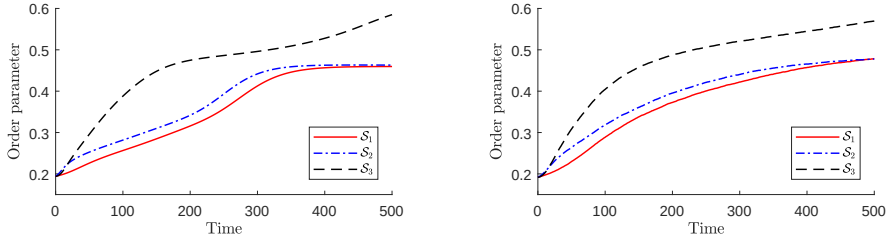
Figure 3.7: Numerical simulation of the model with $\sigma = 0.01$ for different values of (M, A), namely, $\mathscr{S}_1 : (0.05, 0.85)$, $\mathscr{S}_2 : (0.15, 0.85)$, and $\mathscr{S}_3 : (0.15, 0.7)$. The upper panels (a, b, and c) show the opinion distribution for continuum-agent model. The middle panels (d, e, and f) show the the result of Monte Carlo simulation (average of 300 realizations) of discrete-agent model. The lower panels (g and h) show the evolution of order parameter for these systems.



(a) Continuum - $\mathscr{S}_1$ (b) Continuum - $\mathscr{S}_2$ (c) Continuum - $\mathscr{S}_3$



(d) Discrete - $\mathscr{S}_1$ (e) Discrete - $\mathscr{S}_2$ (f) Discrete - $\mathscr{S}_3$



(g) Continuum-agent model (h) Discrete-agent model

Figure 3.8: Numerical simulation of the model with $\sigma = 0.03$ for different values of (M, A), namely, $\mathscr{S}_1 : (0.05, 0.85)$, $\mathscr{S}_2 : (0.15, 0.85)$, and $\mathscr{S}_3 : (0.15, 0.7)$. The upper panels (a, b, and c) show the opinion distribution for continuum-agent model. The middle panels (d, e, and f) show the the result of Monte Carlo simulation (average of 300 realizations) of discrete-agent model. The lower panels (g and h) show the evolution of order parameter for these systems.

PART TWO

DYNAMIC PROGRAMMING IN CONJUGATE DOMAIN

4

FINITE-HORIZON PROBLEM WITH DETERMINISTIC DYNAMICS

With this chapter, we start the second part of this thesis, focusing on the value iteration (VI) algorithm for solving optimal control problems. In particular, in this chapter, we propose two novel numerical schemes for approximate implementation of the dynamic programming (DP) operation concerned with finite-horizon, optimal control of deterministic, discrete-time systems with input-affine dynamics. The chapter is organized as follows. We begin with the literature review in Section 4.1. After presenting some preliminaries in Section 4.2, we provide the problem statement and its standard solution via VI algorithm (in the primal domain) in Section 4.3. Sections 4.4 and 4.5 contain our main results on the proposed alternative approach for solving the DP problem in the conjugate domain. In particular, we provide error bounds for the proposed algorithms, along with a detailed analysis of their computational complexity. In Section 4.6, we validate our theoretical results and compare the performance of the proposed algorithms with the benchmark VI algorithm through a synthetic numerical example. The chapter concludes with Section 4.7 including all the technical proofs. To facilitate the application of the proposed algorithms, we provide a MATLAB package [64]. We note that the numerical example of Section 4.6 is also included in the package and reproducible.

4.1. MOTIVATION AND LITERATURE REVIEW

Value iteration (VI) is one of the most basic and widespread algorithms employed for tackling problems in reinforcement learning (RL) and optimal control [65, 66], formulated as Markov decision processes (MDPs). The VI algorithm simply involves the consecutive applications of the DP operator

$$J_t(x_t) = \min_{u_t} \{C(x_t, u_t) + J_{t+1}(x_{t+1})\}, \tag{4.1}$$

backward in time t, for the costs-to-go J_t, where $C(x_t, u_t)$ is the cost of taking the control action u_t at the state x_t. Arguably, the most important drawback of VI is in its high computational cost in solving problems with a large scale *finite* state space. Indeed, in [67], the authors show that for a finite-horizon MDP, the problem of determining whether a control action u_0 is an optimal action at a given initial state x_0 using value iteration is EXPTIME-complete. For problems with a *continuous* state space, which is commonly the case in engineering applications, solving the DP operation requires solving an infinite number of optimization problems. This usually renders the exact implementation of the DP operation impossible, except for a few cases with an available closed-form solution, e.g., linear quadratic regulator [68, Sec. 4.1]. To address this issue, various schemes have been introduced, commonly known as *approximate* dynamic programming; see, e.g., [65, 69]. A common scheme is to use a sample-based approach accompanied by some form of function approximation. This usually amounts to deploying a brute force search over the discretizations/abstractions of the state and input spaces, leading to a time complexity of at least $\mathcal{O}(XU)$, where X and U are the cardinalities of the discrete state and input spaces, respectively.

For some DP problems, it is possible to reduce this complexity by using duality, i.e., approaching the minimization problem in (4.1) in the conjugate domain. For instance, for the dynamics $x_{t+1} = Ax_t + Bu_t$ and cost $C(x_t, u_t) = C_s(x_t) + C_i(u_t)$, we have

$$J_t(x_t) \geq C_s(x_t) + \left[C_i^*(-B^\top \cdot) + J_{t+1}^*\right]^*(Ax_t), \tag{4.2}$$

where the operator $[\cdot]^*$ denotes the Legendre-Fenchel transform, also known as (convex) conjugate transform. Under some technical assumptions (including, among others, convexity of the functions C_i and J_{t+1}), we have equality in (4.2); see [70, Prop. 5.3.1]. Notice how the minimization operator of (4.1) in the primal domain transforms into a simple addition in (4.2) in the conjugate ("dual") domain. This observation signals the possibility of a significant reduction in the time complexity of solving the DP operation, at least for particular classes of problems.

Approaching the DP problem via conjugate duality goes back to Bellman [71]. Further applications of this idea for reducing the computational complexity were later explored in [72] and [73]. Fundamentally, these approaches exploit the operational duality of infimal convolution and addition with respect to the conjugate transform [74]: For two functions $f_1, f_2 : \mathbb{R}^n \to [-\infty, +\infty]$, we have $(f_1 \square f_2)^* = f_1^* + f_2^*$, where

$$f_1 \square f_2(w) := \inf_{w_1, w_2} \{f_1(w_1) + f_2(w_2) : w_1 + w_2 = w\}, \tag{4.3}$$

is the infimal convolution of f_1 and f_2. This is analogous to the well-known operational duality of convolution and multiplication with respect to the Fourier transform. Actually, the Legendre-Fenchel transform plays a similar role as the Fourier transform when the underlying algebra is the max-plus algebra, as opposed to the conventional plus-times algebra. Much like the extensive application of the latter operational duality upon the introduction of the fast Fourier transform, "fast" numerical algorithms for conjugate transform can facilitate efficient applications of the former one. Interestingly, the first fast algorithm for computing (discrete) conjugate functions, known as fast Legendre transform, was inspired by fast Fourier transform, and enjoys the same *log-linear* complexity in the number of data points; see [75, 76] and the references therein. Later, this complexity was reduced by introducing a *linear-time* algorithm known as linear-time Legendre transform (LLT) [77]. We refer the interested reader to [78] for an extensive review of these algorithms (and other similar algorithms) and their applications. In this regard, we also note that recently, in [79], the authors introduced a quantum algorithm for computing the (discrete) conjugate of convex functions, which achieves a *poly-logarithmic* time complexity in the number of data points.

One of the first and most widespread applications of these fast algorithms has been in solving the Hamilton-Jacobi equation [75, 80, 81]. Another interesting area of application is image processing, where the Legendre-Fenchel transform is commonly known as "distance transform" [82, 83]. Recently, in [84], the authors used these algorithms to tackle the optimal transport problem with strictly convex costs, with applications in image processing and in numerical methods for solving partial differential equations. However, surprisingly, the application of these fast algorithms in solving discrete-time optimal control problems seems to remain largely unexplored. An exception is [85], where the authors use LLT to propose the "fast value iteration" algorithm for computing the fixed-point of the DP operator arising from a specific class of infinite-horizon, discrete-time problems. Indeed, the setup in [85] corresponds to a subclass of problems that we consider that allows for a "perfect" transformation of the minimization in the DP operation in the primal domain to an addition in the dual domain; this connection will be discussed in detail in Section 4.5.4. Let us also note that the algorithms developed in [82, 83] for distance transform can also potentially tackle the (discretized) optimal

control problems similar to the ones considered in this chapter. In particular, these algorithms require the stage cost to be reformulated as a convex distance function of the current and next states. While the this property might arise naturally, it can generally be restrictive as it is in our case.

Another line of work, closely related to ours, involves algorithms that utilize max-plus algebra in solving, continuous-time, continuous-space, deterministic optimal control problems; see, e.g., [86–88]. These works exploit the compatibility of the DP operation with max-plus operations and approximate the value function as a max-plus linear combination. In particular, recently in [89, 90], the authors used this idea to propose an approximate value iteration algorithm for deterministic MDPs with continuous state space. In this regard, we note that the proposed algorithms in this chapter also implicitly involve representing cost functions as max-plus linear combinations, yielding piecewise affine approximations. The key difference of the proposed algorithms is however to choose a grid-like (factorized) set of slopes in the dual space in order to reduce the computational cost; we will discuss this point in more detail in Section 4.4.2.

In this part of the thesis (Chapters 4 and 5), we use duality and propose multiple alternative VI algorithms that involve a sample-based approximation using a finite subset of the underlying continuous state space. These algorithms are based on a path that solves the dual problem corresponding to the DP operation, by utilizing the LLT algorithm for discrete conjugation. In particular, the proposed approaches involve incorporating a finite-dimensional approximation of the value function in the dual domain. Figure 4.1 shows the sketch of the proposed algorithms in this chapter.

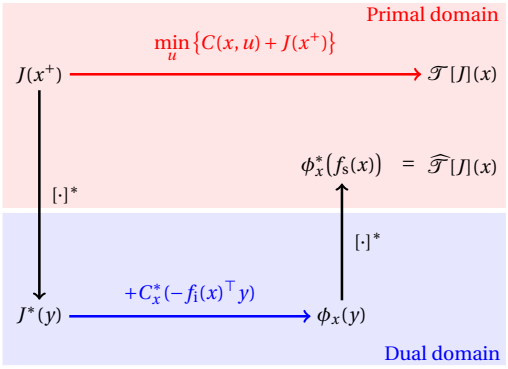
4.2. NOTATIONS AND PRELIMINARIES

4.2.1. GENERAL NOTATIONS

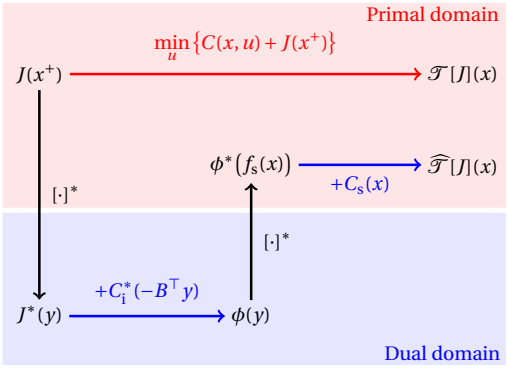
We use $\mathbb{R}$ to denote the real line and $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$, $\underline{\overline{\mathbb{R}}} = \mathbb{R} \cup \{\pm\infty\}$ to denote its extensions. The standard inner product in $\mathbb{R}^n$ and the corresponding induced 2-norm are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_2$, respectively, and the infinity-norm is denoted by $\|\cdot\|_\infty$. We also use $\|\cdot\|_2$ to denote the operator norm (with respect to the 2-norm) of a matrix; i.e., for $A \in \mathbb{R}^{m\times n}$, we denote $\|A\|_2 = \sup\{\|Ax\|_2 : \|x\|_2 = 1\}$. We use the common convention in optimization whereby the optimal value of an infeasible minimization (respectively, maximization) problem is set to $+\infty$ (respectively, $-\infty$).

Continuous (infinite, uncountable) sets are denoted as $\mathbb{X}, \mathbb{Y}, \ldots$. We use the superscript d as in $\mathbb{X}^{\mathrm{d}}$ to denote the *finite* discretization of a continuous set $\mathbb{X}$. Moreover, we use the superscript g to differentiate *grid-like* (factorized) discretizations. Precisely, a grid $\mathbb{X}^{\mathrm{g}} \subset \mathbb{R}^n$ is the Cartesian product $\mathbb{X}^{\mathrm{g}} = \Pi_{i=1}^n \mathbb{X}_i^{\mathrm{g}} = \mathbb{X}_1^{\mathrm{g}} \times \ldots \times \mathbb{X}_n^{\mathrm{g}}$, where $\mathbb{X}_i^{\mathrm{g}}$ is a finite set of real numbers $x_i^1 < x_i^2 < \ldots < x_i^{X_i}$. Assuming $X_i \geq 3$ for all $i = 1, \ldots, n$, we define $\mathbb{X}_{\mathrm{sub}}^{\mathrm{g}} := \Pi_{i=1}^n \mathbb{X}_{\mathrm{sub}\,i}^{\mathrm{g}}$, where $\mathbb{X}_{\mathrm{sub}\,i}^{\mathrm{g}} = \mathbb{X}_i^{\mathrm{g}} \setminus \{x_i^1, x_i^{X_i}\}$; that is, $\mathbb{X}_{\mathrm{sub}}^{\mathrm{g}}$ is the *sub-grid* derived by omitting the smallest and largest elements of $\mathbb{X}^{\mathrm{g}}$ in each dimension. The cardinality of a finite set $\mathbb{X}^{\mathrm{d}}$ (or $\mathbb{X}^{\mathrm{g}}$) is denoted by X. Let $\mathbb{X}, \mathbb{Y}$ be two arbitrary sets in $\mathbb{R}^n$. The convex hull of $\mathbb{X}$ is denoted by $\mathrm{co}(\mathbb{X})$. The diameter of $\mathbb{X}$ is defined as $\Delta_{\mathbb{X}} := \sup_{x,y\in\mathbb{X}} \|x - y\|_2$. We use $\mathrm{d}(\mathbb{X}, \mathbb{Y}) := \inf_{x\in\mathbb{X}, y\in\mathbb{Y}} \|x - y\|_2$ to denote the distance between $\mathbb{X}$ and $\mathbb{Y}$. The one-sided Hausdorff distance *from* $\mathbb{X}$ *to* $\mathbb{Y}$ is defined as $\mathrm{d_H}(\mathbb{X}, \mathbb{Y}) := \sup_{x\in\mathbb{X}} \inf_{y\in\mathbb{Y}} \|x - y\|_2$.

For an extended real-valued function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, the effective domain of h is defined



(a) Setting 4.4.1: dynamics $x^+ = f_s(x) + f_i(x) \cdot u$ and cost $C(x,u)$.



(b) Setting 4.5.1: dynamics $x^+ = f_s(x) + B \cdot u$ and cost $C(x,u) = C_s(x) + C_i(u)$.

Figure 4.1: Sketch of the proposed conjugate VI (ConjVI) algorithms for deterministic dynamics – the standard DP operation in the primal domain (upper red paths) and the conjugate DP (CDP) operation through the dual domain (bottom blue paths).

by $\mathrm{dom}(h) := \{x \in \mathbb{R}^n : h(x) < +\infty\}$, and the range of h is defined as

$$\mathrm{rng}(h) = \max_{x \in \mathrm{dom}(h)} h(x) - \min_{x \in \mathrm{dom}(h)} h(x).$$

The Lipschtiz constant of h over a set $\mathbb{X} \subset \mathrm{dom}(h)$ is denoted by

$$\mathrm{L}(h;\mathbb{X}) := \sup_{x,y \in \mathbb{X}} \frac{|h(x) - h(y)|}{\|x - y\|_2}.$$

We also denote $\mathrm{L}(h) := \mathrm{L}\big(h;\mathrm{dom}(h)\big)$ and $\mathbb{L}(h) := \Pi_{i=1}^n \big[\mathrm{L}_i^-(h), \mathrm{L}_i^+(h)\big]$, where $\mathrm{L}_i^+(h)$ (respectively, $\mathrm{L}_i^-(h)$) is the maximum (respectively, minimum) slope of the function h along

the i-th dimension, i.e.,

$$\mathrm{L}_i^+(h) := \sup\left\{\frac{h(x)-h(y)}{x_i-y_i} : x, y \in \mathrm{dom}(h),\ x_i > y_i,\ x_j = y_j\ (j \neq i)\right\},$$
$$\mathrm{L}_i^-(h) := \inf\left\{\frac{h(x)-h(y)}{x_i-y_i} : x, y \in \mathrm{dom}(h),\ x_i > y_i,\ x_j = y_j\ (j \neq i)\right\}.$$

The subdifferential of h at a point $x \in \mathbb{R}^n$ is defined as

$$\partial h(x) := \left\{y \in \mathbb{R}^n : h(\tilde{x}) \geq h(x) + \left\langle y, \tilde{x} - x\right\rangle, \forall \tilde{x} \in \mathrm{dom}(h)\right\}.$$

Note that $\partial h(x) \subseteq \mathbb{L}(h)$ for all $x \in \mathbb{X}$; in particular, $\mathbb{L}(h) = \cup_{x\in\mathbb{X}} \partial h(x)$ if h is convex.

We report the complexities using the standard big O notations $\mathcal{O}$ and $\widetilde{\mathcal{O}}$, where the latter hides the logarithmic factors. We are mainly concerned with the dependence of the computational complexities on *the size of the finite sets* involved (discretization of the primal and dual domains). In particular, we ignore the possible dependence of the computational complexities on the dimension of the variables, unless they appear in the power of the size of those discrete sets; e.g., the complexity of a single evaluation of an analytically available function is taken to be of $\mathcal{O}(1)$, regardless of the dimension of its input and output arguments. For the reader's convenience, we also provide the list of the most important objects used in this chapter in Table 4.1.

Table 4.1: List of the most important notational conventions.

Notation	Description	Definition
h^{d}	Discretization of the function h	–
$\widetilde{h^{\mathrm{d}}}$	Extension of the discrete function h^{d}	–
$\overline{h^{\mathrm{d}}}$	LERP extension of the discrete function h^{d} (with grid-like domain)	–
h^*	Conjugate of h	(4.4)
$h^{\mathrm{d}*}$	Discrete conjugate of h (conjugate of h^{d})	(4.5)
h^{**}	Biconjugate of h	(4.6)
$h^{\mathrm{d}*\mathrm{d}*}$	Discrete biconjugate of h	(4.7)
$\mathscr{T}$	Dynamic Programming (DP) operator	(4.17) & (4.27)
$\mathscr{T}^{\mathrm{d}}$	Discrete DP (d-DP) operator	(4.18)
$\widehat{\mathscr{T}}$	Conjugate DP (CDP) operator	(4.21)
$\widehat{\mathscr{T}}^{\mathrm{d}}$	Discrete CDP (d-CDP) operator	(4.22) & (4.28)
$\widehat{\mathscr{T}}^{\mathrm{d}}_{\mathrm{m}}$	Modified d-CDP operator [for Setting 4.5.1]	(4.29)

4.2.2. EXTENSION OF DISCRETE FUNCTIONS

Consider an extended real-valued function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, and its discretization $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \overline{\mathbb{R}}$, where $\mathbb{X}^{\mathrm{d}}$ is a finite subset of $\mathbb{R}^n$. We use the superscript d, as in h^{d}, to denote the discretization of h. We particularly use this notation in combination with a second operation to emphasize that the second operation is applied on the discretized version of

the operand. In particular, we use $\widetilde{h^{\mathrm{d}}} : \mathbb{R}^n \to \overline{\mathbb{R}}$ to denote the extension of the discrete function $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \overline{\mathbb{R}}$. The extension can be considered as a generic parametric approximation $\widetilde{h^{\mathrm{d}}}[\theta] : \mathbb{R}^n \to \overline{\mathbb{R}}$, where the parameters θ are computed using regression, i.e., by fitting $\widetilde{h^{\mathrm{d}}}[\theta]$ to the data points $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \overline{\mathbb{R}}$.

Remark 4.2.1 (Complexity of extension operation)**.** *We use E to denote the complexity of a generic extension operator. That is, for each $x \in \mathbb{R}^n$, the time complexity of the single evaluation $\widetilde{h^{\mathrm{d}}}(x)$ is assumed to be of $\mathcal{O}(E)$, with E (possibly) being a function of X.*

For example, for the linear approximation $\widetilde{h^{\mathrm{d}}}(x) = \sum_{i=1}^{B} \theta_i \cdot b_i(x)$, we have $E = B$ (the size of the basis), while for the kernel-based approximation $\widetilde{h^{\mathrm{d}}}(x) = \sum_{\tilde{x} \in \mathbb{X}^{\mathrm{d}}} \theta_{\tilde{x}} \cdot r(x, \tilde{x})$, we generally have $E \leq X$. A kernel-based approximator of interest in the following sections is the *multilinear interpolation & extrapolation* (LERP) of a discrete function with a *grid-like* domain. Hence, we denote this operation with the different notation $\overline{h^{\mathrm{d}}} : \mathbb{R}^n \to \overline{\mathbb{R}}$ for the discrete function $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \overline{\mathbb{R}}$. Notice that the LERP extension preserves the value of the function at the discrete points, i.e, $\overline{h^{\mathrm{d}}}(x) = h^{\mathrm{d}}(x)$ for all $x \in \mathbb{X}^{\mathrm{g}}$. To facilitate our complexity analysis in subsequent sections, we discuss the computational complexity of LERP in the following remark.

Remark 4.2.2 (Complexity of LERP)**.** *Given a discrete function $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ with a grid-like domain $\mathbb{X}^{\mathrm{g}} \subset \mathbb{R}^n$, the time complexity of a single evaluation of the LERP extension $\overline{h^{\mathrm{d}}}$ at a point $x \in \mathbb{R}^n$ is of $\mathcal{O}(2^n + \log X) = \widetilde{\mathcal{O}}(1)$ if $\mathbb{X}^{\mathrm{g}}$ is non-uniform, and of $\mathcal{O}(2^n) = \mathcal{O}(1)$ if $\mathbb{X}^{\mathrm{g}}$ is uniform. To see this, note that, in the case $\mathbb{X}^{\mathrm{g}}$ is non-uniform, LERP requires $\mathcal{O}(\log X)$ operations to find the position of x with respect to the grid points, using binary search. If $\mathbb{X}^{\mathrm{g}}$ is a uniform grid, this can be done in $\mathcal{O}(n)$ time. Upon finding the position of x, LERP then involves a series of one-dimensional linear interpolations or extrapolations along each dimension, which takes $\mathcal{O}(2^n)$ operations.*

For a convex function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, we have $\partial h(x) \neq \emptyset$ for all x in the relative interior of $\mathrm{dom}(h)$ [70, Prop. 5.4.1]. This characterization of convexity can be extended to discrete functions. A discrete function $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$ is called *convex-extensible* if $\partial h^{\mathrm{d}}(x) \neq \emptyset$ for all $x \in \mathrm{dom}(h) = \mathbb{X}^{\mathrm{d}}$. Equivalently, h^{d} is convex-extensible, if it can be extended to a convex function $\widetilde{h^{\mathrm{d}}} : \mathbb{R}^n \to \overline{\mathbb{R}}$ such that $\widetilde{h^{\mathrm{d}}}(x) = h^{\mathrm{d}}(x)$ for all $x \in \mathbb{X}^{\mathrm{d}}$; we refer the reader to, e.g., [91] for different extensions of the notion of convexity to discrete functions.

4.2.3. LEGENDRE-FENCHEL TRANSFORM

Consider an extended-real-valued function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, with a nonempty effective domain $\mathrm{dom}(h) = \mathbb{X}$. The Legendre-Fenchel transform (convex conjugate transform) of h is the function

$$h^* : \mathbb{R}^n \to \overline{\mathbb{R}} : y \mapsto \sup_{x \in \mathbb{X}} \{\langle y, x \rangle - h(x)\}. \tag{4.4}$$

Note that the conjugate function h^* is convex by construction. We particularly consider *discrete* conjugation, which involves computing the conjugate function using the discretized version $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \overline{\mathbb{R}}$ of the function h, where $\mathbb{X}^{\mathrm{d}} \cap \mathbb{X} \neq \emptyset$. We use the notation

$[\cdot]^{\mathrm{d}*}$, as opposed the standard notation $[\cdot]^*$, for discrete conjugation; that is,

$$h^{\mathrm{d}*} = [h^{\mathrm{d}}]^* : \mathbb{R}^n \to \mathbb{R} : y \mapsto \max_{x \in \mathbb{X}^{\mathrm{d}}} \left\{ \langle y, x \rangle - h^{\mathrm{d}}(x) \right\}. \tag{4.5}$$

The biconjugate of h is the function

$$h^{**} = [h^*]^* : \mathbb{R}^n \to \overline{\mathbb{R}} : x \mapsto \sup_{y \in \mathbb{R}^n} \left\{ \langle x, y \rangle - h^*(y) \right\} = \sup_{y \in \mathbb{R}^n} \inf_{z \in \mathbb{X}} \left\{ \langle x - z, y \rangle + h(z) \right\}. \tag{4.6}$$

Using the notion of discrete conjugation $[\cdot]^{\mathrm{d}*}$, we also define the discrete biconjugate

$$h^{\mathrm{d}*\mathrm{d}*} = [h^{\mathrm{d}*}]^{\mathrm{d}*} : \mathbb{R}^n \to \mathbb{R} : x \mapsto \max_{y \in \mathbb{Y}^{\mathrm{d}}} \left\{ \langle x, y \rangle - h^{\mathrm{d}*\mathrm{d}}(y) \right\} = \max_{y \in \mathbb{Y}^{\mathrm{d}}} \min_{z \in \mathbb{X}^{\mathrm{d}}} \left\{ \langle x - z, y \rangle + h^{\mathrm{d}}(z) \right\}, \tag{4.7}$$

where $\mathbb{X}^{\mathrm{d}}$ and $\mathbb{Y}^{\mathrm{d}}$ are finite subsets of $\mathbb{R}^n$ such that $\mathbb{X}^{\mathrm{d}} \cap \mathbb{X} \neq \emptyset$.

The Linear-time Legendre Transform (LLT) is an efficient algorithm for computing the discrete conjugate over a finite *grid-like* dual domain. Precisely, to compute the conjugate of the function $h : \mathbb{X} \to \mathbb{R}$, LLT takes its discretization $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$ as an input, and outputs $h^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$, for the grid-like dual domain $\mathbb{Y}^{\mathrm{g}}$. That is, LLT is equivalent to the operation $[\cdot]^{\mathrm{d}*\mathrm{d}}$. We refer the interested reader to [77] for a detailed description of the LLT algorithm. We will use the following result for analyzing the computational complexity of the proposed algorithms.

Remark 4.2.3 (Complexity of LLT)**.** *Consider a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and its discretization over the grid $\mathbb{X}^{\mathrm{g}} \subset \mathbb{R}^n$ such that $\mathbb{X}^{\mathrm{g}} \cap \mathrm{dom}(h) \neq \emptyset$. LLT computes the discrete conjugate function $h^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ using the data points $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \overline{\mathbb{R}}$, with a time complexity of $\mathcal{O}\left(\Pi_{i=1}^n (X_i + Y_i)\right)$, where X_i (respectively, Y_i) is the cardinality of the i-th dimension of the grid $\mathbb{X}^{\mathrm{g}}$ (respectively, $\mathbb{Y}^{\mathrm{g}}$). If the grids $\mathbb{X}^{\mathrm{g}}$ and $\mathbb{Y}^{\mathrm{g}}$ have approximately the same cardinality in each dimension, then the time complexity of LLT is of $\mathcal{O}(X + Y)$ [77, Cor. 5].*

Hereafter, to simplify the exposition, we consider the following assumption.

Assumption 4.2.4 (Grid sizes in LLT)**.** *The primal and dual grids used for LLT operation have approximately the same cardinality in each dimension.*

4.2.4. PRELIMINARY RESULTS ON CONJUGATE TRANSFORM

We now provide two preliminary lemmas on the error of discrete conjugate transform and its approximate version. To this end, we recall some of the notations introduced so far. For a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ with nonempty effective domain $\mathbb{X} = \mathrm{dom}(h)$, let $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$ be the discretization of h where $\mathbb{X}^{\mathrm{d}} \subset \mathbb{X}$, $h^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ be the conjugate (4.4) of h, and $h^{\mathrm{d}*} : \mathbb{R}^n \to \mathbb{R}$ be the discrete conjugate (4.5) of h using the primal discrete domain $\mathbb{X}^{\mathrm{d}}$.

Lemma 4.2.5 (Conjugate vs. discrete conjugate)**.** *Let h be proper, closed, and convex. For each $y \in \mathbb{R}^n$, it holds that*

$$0 \leq h^*(y) - h^{\mathrm{d}*}(y) \leq \min_{x \in \partial h^*(y)} \left\{ \left[\|y\|_2 + \mathrm{L}\left(h; \{x\} \cup \mathbb{X}^{\mathrm{d}}\right) \right] \cdot \mathrm{d}(x, \mathbb{X}^{\mathrm{d}}) \right\} =: \widetilde{e}_1(y, h, \mathbb{X}^{\mathrm{d}}). \tag{4.8}$$

If, moreover, $\mathbb{X}$ is compact and h is Lipschitz continuous, then for each $y \in \mathbb{R}^n$,

$$0 \leq h^*(y) - h^{\mathrm{d}*}(y) \leq \left[\|y\|_2 + \mathrm{L}(h) \right] \cdot \mathrm{d}_{\mathrm{H}}(\mathbb{X}, \mathbb{X}^{\mathrm{d}}) =: \widetilde{e}_2(y, h, \mathbb{X}^{\mathrm{d}}). \tag{4.9}$$

The preceding lemma indicates that discrete conjugate transform leads to an under-approximation of the conjugate function, with the error depending on the discrete representation $\mathbb{X}^{\mathrm{d}}$ of the primal domain $\mathbb{X}$. In particular, the inequality (4.8) implies that for $y \in \mathbb{R}^n$, if $\mathbb{X}^{\mathrm{d}}$ contains $x \in \partial h^*(y)$, which is equivalent to $y \in \partial h(x)$ by the assumptions, then $h^{\mathrm{d}*}(y) = h^*(y)$.

We next present another preliminary however vital result on approximate conjugation. Let $h^{*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ be the discretization of h^* over the grid-like dual domain $\mathbb{Y}^{\mathrm{g}} \subset \mathrm{dom}(h^*) \subseteq \mathbb{R}^n$. Also, let $\overline{h^{*\mathrm{d}}} : \mathbb{R}^n \to \mathbb{R}$ be the extension of $h^{*\mathrm{d}}$ using LERP. The approximate conjugation is then simply the approximation of $h^*(y)$ via $\overline{h^{*\mathrm{d}}}(y)$ for $y \in \mathbb{R}^n$. This approximation introduces a one-sided error:

Lemma 4.2.6 (Approximate conjugation using LERP)**.** *Let $\mathbb{X} = \mathrm{dom}(h)$ be compact. Then,*

$$0 \le \overline{h^{*\mathrm{d}}}(y) - h^*(y) \le \Delta_{\mathbb{X}} \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}), \quad \forall\, y \in \mathrm{co}(\mathbb{Y}^{\mathrm{g}}). \tag{4.10}$$

If, moreover, the dual grid $\mathbb{Y}^{\mathrm{g}}$ is such that $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(h)$, then

$$0 \le \overline{h^{*\mathrm{d}}}(y) - h^*(y) \le \Delta_{\mathbb{X}} \cdot \mathrm{d}_{\mathrm{H}}\big(\mathrm{co}(\mathbb{Y}^{\mathrm{g}}), \mathbb{Y}^{\mathrm{g}}\big), \quad \forall\, y \in \mathbb{R}^n. \tag{4.11}$$

As expected, the error due to the discretization $\mathbb{Y}^{\mathrm{g}}$ of the dual domain $\mathbb{Y}$ depends on the resolution of the discrete dual domain. We also note that the condition $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(h)$ in the second part of the preceding lemma, essentially requires the dual grid $\mathbb{Y}^{\mathrm{g}}$ to "more than cover the range of slopes" of the function h.

The algorithms developed in this chapter use LLT to compute discrete conjugate functions. However, as we will see, we sometimes require the value of the conjugate function at points other than the dual grid points used in LLT. To solve this issue, we use the same approximation described above, but now for discrete conjugation. In this regard, we note that the result of Lemme 4.2.6 also holds for discrete conjugation. To be precise, consider the discrete function $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$. Let $h^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ be the discretization of $h^{\mathrm{d}*}$ over the grid-like dual domain $\mathbb{Y}^{\mathrm{g}} \subset \mathbb{R}^n$, and $\overline{h^{\mathrm{d}*\mathrm{d}}} : \mathbb{R}^n \to \mathbb{R}$ be the extension of $h^{\mathrm{d}*\mathrm{d}}$ using LERP.

Corollary 4.2.7 (Approximate discrete conjugation using LERP)**.** *We have*

$$0 \le \overline{h^{\mathrm{d}*\mathrm{d}}}(y) - h^{\mathrm{d}*}(y) \le \Delta_{\mathbb{X}^{\mathrm{d}}} \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}), \quad \forall\, y \in \mathrm{co}(\mathbb{Y}^{\mathrm{g}}). \tag{4.12}$$

If, moreover, the grid $\mathbb{Y}^{\mathrm{g}}$ is such that $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(h^{\mathrm{d}})$, then

$$0 \le \overline{h^{\mathrm{d}*\mathrm{d}}}(y) - h^{\mathrm{d}*}(y) \le \Delta_{\mathbb{X}^{\mathrm{d}}} \cdot \mathrm{d}_{\mathrm{H}}\big(\mathrm{co}(\mathbb{Y}^{\mathrm{g}}), \mathbb{Y}^{\mathrm{g}}\big) \quad \forall\, y \in \mathbb{R}^n. \tag{4.13}$$

4.3. PROBLEM STATEMENT AND STANDARD SOLUTION

In this chapter, we consider the optimal control of deterministic, discrete-time systems

$$x_{t+1} = f(x_t, u_t), \quad t = 0, \ldots, T-1, \tag{4.14}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ describes the dynamics, and $T \in \mathbb{N}$ is the finite horizon. We also consider state and input constraints of the form

$$\begin{cases} x_t \in \mathbb{X} \subset \mathbb{R}^n & \text{for} \quad t \in \{0, \dots, T\}, \\ u_t \in \mathbb{U} \subset \mathbb{R}^m & \text{for} \quad t \in \{0, \dots, T-1\}. \end{cases} \tag{4.15}$$

Let $C : \mathbb{X} \times \mathbb{U} \to \overline{\mathbb{R}}$ and $C_T : \mathbb{X} \to \mathbb{R}$ be the stage and terminal costs, respectively. Note that we let the stage cost C take $+\infty$ for $(x, u) \in \mathbb{X} \times \mathbb{U}$ so that it can embed the *state-dependent input constraints*. For an initial state $x_0 \in \mathbb{X}$, the cost incurred by the state trajectory $\mathbf{x} = (x_0, \dots, x_T)$ in response to the input sequence $\mathbf{u} = (u_0, \dots, u_{T-1})$ is

$$J(x_0, \mathbf{u}) = \textstyle\sum_{t=0}^{T-1} C(x_t, u_t) + C_T(x_T).$$

4

The problem of interest is then to find an optimal control sequence $\mathbf{u}_\star(x_0)$, that is, a solution to the minimization problem

$$J_\star(x_0) = \min_{\mathbf{u}} \big\{ J(x_0, \mathbf{u}) : (4.14) \;\&\; (4.15) \big\}. \tag{4.16}$$

In this chapter, we assume that the problem data satisfy the following conditions.

Assumption 4.3.1 (Problem data)**.** *Assume:*

(i) **Dynamics.** *The mapping $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is locally Lipschitz continuous.*

(ii) **Constraints.** *The sets $\mathbb{X}$ and $\mathbb{U}$ are compact. Moreover, the set of admissible inputs $\mathbb{U}(x) \coloneqq \{u \in \mathbb{U} : C(x, u) < +\infty,\ f(x, u) \in \mathbb{X}\}$ is nonempty for all $x \in \mathbb{X}$.*

(iii) **Cost functions.** *$C : \mathbb{X} \times \mathbb{U} \to \overline{\mathbb{R}}$ has a compact effective domain. Moreover, C and C_T are Lipschitz continuous.*

The properties laid out in Assumption 4.3.1 imply that the set $\mathbb{U}(x)$ of admissible inputs is nonempty and compact, and the objective in (4.16) is continuous (compactness of $\mathbb{U}(x)$ follows from compactness of $\operatorname{dom}(C)$ and $\mathbb{X}$, and continuity of f). Hence, the optimal value in (4.16) is achieved. To solve this problem using VI, we have to solve the DP equation

$$J_t(x_t) = \min_u \big\{ C(x_t, u_t) + J_{t+1}(x_{t+1}) : (4.14) \;\&\; (4.15) \big\}, \quad x_t \in \mathbb{X},$$

backward in time $t = T-1, \dots, 0$, initialized by $J_T = C_T$. The iteration finally outputs $J_0 = J_\star$ [68, Prop. 1.3.1]. To simplify the exposition, let us embed the state and input constraints in the cost functions (C and J_t) by extending them to infinity outside their effective domain. Let us also drop the time subscript t and focus on a single step of the recursion by defining the DP operator

$$\mathscr{T} J(x) \coloneqq \min_u \big\{ C(x, u) + J\big(f(x, u)\big) \big\}, \quad x \in \mathbb{X}, \tag{4.17}$$

so that $J_t = \mathscr{T} J_{t+1} = \mathscr{T}^{(T-t)} J_T$ for $t = T-1, \dots, 0$.

Notice that the DP operation (4.17) requires solving an infinite number of optimization problems for all $x \in \mathbb{X}$. Except for a few cases with an available closed-form solution, the exact implementation of DP operation is impossible. A standard approximation scheme is then to incorporate function approximation techniques and solve (4.17) for a finite sample (i.e., a discretization) of the underlying continuous state space. Precisely, we consider solving the optimization in (4.17) for a finite number of $x \in \mathbb{X}^{\mathrm{g}}$, where $\mathbb{X}^{\mathrm{g}} \subset \mathbb{X}$ is a grid-like discretization of the state space. The *T-step VI problem* then involves finding the discrete costs-to-go $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ for $t = 0, 1, \ldots, T-1$. Notice that the DP operator $\mathscr{T}$ now takes the discrete function $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ as an input. However, in order to compute the output $[\mathscr{T} J]^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$, we require evaluating J at points $f(x,u)$ for $(x,u) \in \mathbb{X}^{\mathrm{g}} \times \mathbb{U}$, which do not necessarily belong to the discrete state space $\mathbb{X}^{\mathrm{g}}$. Hence, along with the discretization of the state space, we also need to consider some form of function approximation for the cost-to-go function, that is, an extension $\widetilde{J^{\mathrm{d}}} : \mathbb{X} \to \mathbb{R}$ of the function $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. Next to be addressed is the issue of solving the minimization

$$\min_{u \in \mathbb{U}} \left\{ C(x,u) + \widetilde{J^{\mathrm{d}}}\big(f(x,u)\big) \right\},$$

for each $x \in \mathbb{X}^{\mathrm{g}}$, where the next step cost-to-go is approximated by the extension $\widetilde{J^{\mathrm{d}}}$. This minimization problem is often a difficult, non-convex problem. Again, a common approximation involves enumeration over a proper discretization $\mathbb{U}^{\mathrm{d}} \subset \mathbb{U}$ of the inputs space.[1] Incorporating these approximations, we can introduce the *discrete* DP (d-DP) operator as follows

$$\mathscr{T}^{\mathrm{d}} J^{\mathrm{d}}(x) := \min_{u \in \mathbb{U}^{\mathrm{d}}} \left\{ C(x,u) + \widetilde{J^{\mathrm{d}}}\big(f(x,u)\big) \right\}, \quad x \in \mathbb{X}^{\mathrm{g}}. \tag{4.18}$$

Under some regularity assumptions, the error corresponding to these approximations depends on the discretization of the state and input spaces and the extension operation:

Proposition 4.3.2 (Error of d-DP)**.** *Consider the DP operator $\mathscr{T}$ (4.17) and the d-DP operator $\mathscr{T}^{\mathrm{d}}$ (4.18). Assume that the functions J and $\widetilde{J^{\mathrm{d}}}$ are Lipschtiz continuous, and $\widetilde{J^{\mathrm{d}}}(x) = J(x)$ for all $x \in \mathbb{X}^{\mathrm{g}}$. Then,*

$$-e_1 \le \mathscr{T}^{\mathrm{d}} J^{\mathrm{d}}(x) - \mathscr{T} J(x) \le e_1 + e_2(x), \quad \forall x \in \mathbb{X}^{\mathrm{g}},$$

where

$$e_1 = \big[\mathrm{L}(J) + \mathrm{L}(\widetilde{J^{\mathrm{d}}})\big] \cdot \mathrm{d_H}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}),$$
$$e_2(x) = \big[\mathrm{L}(J) + \mathrm{L}(C)\big] \cdot \mathrm{d_H}\big(\mathbb{U}(x), \mathbb{U}^{\mathrm{d}}(x)\big).$$

The VI algorithm that utilizes the d-DP operator (4.18) will be our benchmark for evaluating the performance of the proposed algorithms. To this end, we discuss the time complexity of the d-DP operation in the following remark.

[1] We assume that the joint discretization of the state-input space is "proper" in the sense that the feasibility condition of Assumption 4.3.1-(ii) holds for the discrete state-input space, i.e., $\mathbb{U}^{\mathrm{d}}(x) := \mathbb{U}(x) \cap \mathbb{U}^{\mathrm{d}}$ is nonempty for all $x \in \mathbb{X}^{\mathrm{g}}$.

Remark 4.3.3 (Complexity of d-DP)**.** *Let the time complexity of a single evaluation of the extension operator* $\widehat{[\cdot]}$ *in* (4.18) *be of* $\mathcal{O}(E)$*. Then, the time complexity of the d-DP operation* (4.18) *is of* $\mathcal{O}\left(XUE\right)$*. Moreover, for solving the* T*-step VI problem, the time complexity increases linearly with the horizon* T*.*

Let us clarify that the scheme described above essentially involves approximating a continuous-state/action MDP with a finite-state/action MDP, and then applying the VI algorithm. In this regard, we note that $\mathcal{O}(XU)$ is the best existing time-complexity in the literature for finite MDPs; see, e.g., [89, 92]. Indeed, regardless of the problem data, the d-DP algorithm involves solving a minimization problem for each $x \in \mathbb{X}^{\mathrm{g}}$, via enumeration over $u \in \mathbb{U}^{\mathrm{d}}$. However, as we will see in the subsequent sections, for certain classes of problems, it is possible to exploit the structure of the underlying continuous setup to avoid the minimization over the input and achieve a lower time complexity.

4

4.4. FROM MINIMIZATION TO ADDITION

We now introduce a general class of problems that allows us to employ conjugate duality for the DP problem and hence propose an alternative path for implementing the corresponding operator. In particular, we show that the linearity of dynamics in the input is the key property in developing the alternative solution, whereby the minimization in the primal domain is transformed to an addition in the dual domain at the expense of three conjugate transforms. The problem class of interest is as follows:

Setting 4.4.1. *The dynamics are input-affine, that is,* $f(x,u) = f_{\mathrm{s}}(x) + f_{\mathrm{i}}(x) \cdot u$*, where* $f_{\mathrm{s}} : \mathbb{R}^n \to \mathbb{R}^n$ *is the "state" dynamics, and* $f_{\mathrm{i}} : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ *is the "input" dynamics.*

4.4.1. THE d-CDP OPERATOR

Alternatively, we can approach the optimization problem in the DP operation (4.17) in the dual domain. To this end, let us fix $x \in \mathbb{X}$, and consider the following reformulation of the problem (4.17)

$$\mathscr{T} J(x) = \min_{u,z} \left\{ C(x,u) + J(z) : z = f(x,u) \right\}.$$

Notice how for input-affine dynamics of Setting 4.4.1, this formulation resembles the infimal convolution (4.3) (by taking $w_1 = z$ and $w_2 = u$, the equality constraint becomes $w_1 - f_{\mathrm{i}}(x) \cdot w_2 = f_{\mathrm{s}}(x)$). In this regard, consider the corresponding dual problem

$$\widehat{\mathscr{T}} J(x) := \max_{y} \min_{u,z} \left\{ C(x,u) + J(z) + \left\langle y, f(x,u) - z \right\rangle \right\}, \tag{4.19}$$

where $y \in \mathbb{R}^n$ is the dual variable. Indeed, for input-affine dynamics, we can derive an equivalent formulation for the dual problem (4.19), which forms the basis for the proposed algorithms.

Lemma 4.4.2 (CDP operator)**.** *Let*

$$C_x^*(v) := \max_{u} \left\{ \langle v, u \rangle - C(x,u) \right\}, \quad v \in \mathbb{R}^m, \tag{4.20}$$

denote the partial conjugate of the stage cost with respect to the input variable u. Then, for the input-affine dynamics of Setting 4.4.1, the operator $\widehat{\mathscr{T}}$ (4.19) equivalently reads as

$$\phi_x(y) \coloneqq C_x^*(-f_{\mathrm{i}}(x)^\top y) + J^*(y), \qquad y \in \mathbb{R}^n, \tag{4.21a}$$

$$\widehat{\mathscr{T}} J(x) = \phi_x^*\big(f_{\mathrm{s}}(x)\big), \qquad x \in \mathbb{X}. \tag{4.21b}$$

As we mentioned, the construction above suggests an alternative path for computing the output of the DP operator through the conjugate domain. We call this alternative approach *conjugate* DP (CDP). Figure 4.1a characterizes this alternative path schematically. Numerical implementation of CDP operation requires the computation of conjugate functions. In particular, as shown in Figure 4.1a, CDP operation involves three conjugate transforms. In this chapter, we assume that the partial conjugate C_x^* of the stage cost in (4.20) is analytically available.

Assumption 4.4.3 (Conjugate of stage cost)**.** *The conjugate function C_x^* (4.20) is analytically available. That is, the complexity of evaluating $C_x^*(v)$ for each $v \in \mathbb{R}^m$ is of $\mathscr{O}(1)$.*

The two remaining conjugate operations of the CDP path in Figure 4.1a are handled numerically. In particular, we again take a sample-based approach and compute $\widehat{\mathscr{T}} J$ for a finite number of states $x \in \mathbb{X}^{\mathrm{g}}$. To be precise, for a grid-like discretization $\mathbb{Y}^{\mathrm{g}}$ of the dual domain, we employ LLT to compute $J^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ using the data points $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. Proper construction of $\mathbb{Y}^{\mathrm{g}}$ will be discussed in Section 4.4.3. Now, let

$$\varphi_x^{\mathrm{d}}(y) \coloneqq C_x^*(-f_{\mathrm{i}}(x)^\top y) + J^{\mathrm{d}*\mathrm{d}}(y), \quad y \in \mathbb{Y}^{\mathrm{g}},$$

be a discrete *approximation* of ϕ_x in (4.21a). The approximation stems from the fact that we used the discrete conjugate $J^{\mathrm{d}*}$ instead of the conjugate J^*. Using this object, we can also handle the last conjugate transform in Figure 4.1a numerically, and approximate $\phi_x^*\big(f_{\mathrm{s}}(x)\big)$ in (4.21b) by

$$\varphi_x^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) = \max_{y \in \mathbb{Y}^{\mathrm{g}}} \Big\{ \big\langle f_{\mathrm{s}}(x), y \big\rangle - \varphi_x^{\mathrm{d}}(y) \Big\},$$

via enumeration over $y \in \mathbb{Y}^{\mathrm{g}}$. Based on the construction described above, we can introduce the *discrete* CDP (d-CDP) operator as follows

$$J^{\mathrm{d}*\mathrm{d}}(y) = \max_{x \in \mathbb{X}^{\mathrm{g}}} \Big\{ \big\langle y, x \big\rangle - J^{\mathrm{d}}(x) \Big\}, \qquad y \in \mathbb{Y}^{\mathrm{g}}, \tag{4.22a}$$

$$\varphi_x^{\mathrm{d}}(y) = C_x^*(-f_{\mathrm{i}}(x)^\top y) + J^{\mathrm{d}*\mathrm{d}}(y), \qquad y \in \mathbb{Y}^{\mathrm{g}}, \tag{4.22b}$$

$$\widehat{\mathscr{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) \coloneqq \varphi_x^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big), \qquad x \in \mathbb{X}^{\mathrm{g}}. \tag{4.22c}$$

Algorithm 3 provides the pseudo-code for the numerical implementation of the T-step *conjugate* VI (ConjVI) algorithm that utilizes the d-CDP operation (4.22). Next, we analyze the complexity and error of the d-CDP operation.

4.4.2. ANALYSIS OF d-CDP OPERATOR

We begin with the computational complexity of the d-CDP operator.

Algorithm 3 ConjVI algorithm via d-CDP operator (4.22) for Setting 4.4.1.

Input: dynamics $f_{\mathrm{s}} : \mathbb{R}^n \to \mathbb{R}^n$, $f_{\mathrm{i}} : \mathbb{R}^n \to \mathbb{R}^{n\times m}$; discrete state space $\mathbb{X}^{\mathrm{g}} \subset \mathbb{X}$; conjugate of stage cost $C_x^* : \mathbb{R}^m \to \mathbb{R}$ for $x \in \mathbb{X}^{\mathrm{g}}$; discrete terminal cost $C_T^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$.

Output: discrete costs-to-go $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$, $t = 0, 1, \dots, T$.

initialization:

$J_T^{\mathrm{d}}(x) \leftarrow C_T^{\mathrm{d}}(x)$ for $x \in \mathbb{X}^{\mathrm{g}}$;

backward iteration:

for $t = T, \dots, 1$ **do**

  construct the grid $\mathbb{Y}^{\mathrm{g}}$;

  d-CDP operation:

  use LLT to compute $J_t^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ from $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$;

  for each $x \in \mathbb{X}^{\mathrm{g}}$ **do**

    $\varphi_x^{\mathrm{d}}(y) \leftarrow C_x^*(-f_{\mathrm{i}}(x)^\top y) + J_t^{\mathrm{d}*\mathrm{d}}(y)$ for $y \in \mathbb{Y}^{\mathrm{g}}$;

    $J_{t-1}^{\mathrm{d}}(x) \leftarrow \widehat{\mathscr{T}}^{\mathrm{d}} J_t^{\mathrm{d}}(x) = \max_{y\in\mathbb{Y}^{\mathrm{g}}} \left\{ \langle f_{\mathrm{s}}(x), y\rangle - \varphi_x^{\mathrm{d}}(y) \right\}$;

  end for

end for

Theorem 4.4.4 (Complexity of d-CDP)**.** *Let Assumptions 4.2.4 and 4.4.3 hold. Then, the implementation of the d-CDP operator (4.22) in Algorithm 3 requires $\mathscr{O}(XY)$ operations.*

Recall that the time complexity of the d-DP operator (4.18) is of $\mathscr{O}(XUE)$; see Remark 4.3.3. Comparing this complexity to the one reported in Theorem 4.4.4, points to a basic characteristic of the proposed approach: CDP avoids the minimization over the control input in DP and casts it as a simple addition in the dual domain at the expense of three conjugate transforms. Consequently, the time complexity is transferred from the primal input domain $\mathbb{U}^{\mathrm{d}}$ into the dual state domain $\mathbb{Y}^{\mathrm{g}}$. This observation implies that if $Y < UE$, then d-CDP is expected to computationally outperform d-DP. We also note that the complexity of the ConjVI Algorithm 3 for solving the T-step VI problem increases linearly with the horizon T (assuming that the dual grid $\mathbb{Y}^{\mathrm{g}}$ can be constructed with at most $\mathscr{O}(X)$ operations; see Remark 4.4.7).

We now consider the error introduced by the d-CDP operator (4.22) with respect to the DP operator (4.17). Let us begin with presenting an alternative representation of the d-CDP operator that sheds some light on the main sources of error.

Proposition 4.4.5 (d-CDP reformulation)**.** *Assume that the stage cost $C : \mathbb{X} \times \mathbb{U} \to \overline{\mathbb{R}}$ is convex in the input variable. The d-CDP operator (4.22) equivalently reads as*

$$\widehat{\mathscr{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) = \min_u \left\{ C(x,u) + J^{\mathrm{d}*\mathrm{d}*}\big(f(x,u)\big) \right\}, \quad x \in \mathbb{X}^{\mathrm{g}}, \tag{4.23}$$

where $J^{\mathrm{d}*\mathrm{d}*}$ is the discrete biconjugate of J, using the primal grid $\mathbb{X}^{\mathrm{g}}$ and the dual grid $\mathbb{Y}^{\mathrm{g}}$.

First, note that

$$J^{\mathrm{d}*\mathrm{d}*}(x) = \max_{y\in\mathbb{Y}^{\mathrm{g}}} \left\{ \langle x, y\rangle - J^{\mathrm{d}*\mathrm{d}}(y) \right\}, \tag{4.24}$$

is a max-plus linear combination using the basis functions $\{\langle \cdot, y\rangle : y \in \mathbb{Y}^{\mathrm{g}}\}$ and coefficients $\{J^{\mathrm{d}*\mathrm{d}}(y) : y \in \mathbb{Y}^{\mathrm{g}}\}$. That is, ConjVI algorithm, similarly to the approximate VI algorithms in [89, 90], employs a max-plus approximation of J. The key difference in the proposed algorithm is however that by choosing a grid-like dual domain $\mathbb{Y}^{\mathrm{g}}$, we can incorporate the linear-time complexity of the LLT in our advantage in computing the coefficients $\{J^{\mathrm{d}*\mathrm{d}}(y) : y \in \mathbb{Y}^{\mathrm{g}}\}$. Moreover, as we discuss below, instead of using a fixed basis, we incorporate a dynamic basis by updating the grid $\mathbb{Y}^{\mathrm{g}}$ at each iteration in order to reduce the error of the algorithm.

Comparing the representations (4.17) and (4.23), we also note that the d-CDP operator $\widehat{\mathscr{T}}^{\mathrm{d}}$ differs from the DP operator $\mathscr{T}$ in that it uses $J^{\mathrm{d}*\mathrm{d}*}$ as an approximation of J. This observation points to two main sources of error in the proposed approach, namely, dualization and discretization. Indeed, $\widehat{\mathscr{T}}^{\mathrm{d}}$ is a discretized version of the dual problem (4.19). Regarding the dualization error, we note that the d-CDP operator is "blind" to non-convexity; that is, it essentially replaces the cost-to-go J by its convex envelope (the greatest convex function that supports J from below). The discretization error, on the other hand, depends on the choice of the finite primal and dual domains $\mathbb{X}^{\mathrm{g}}$ and $\mathbb{Y}^{\mathrm{g}}$. In particular, by a proper choice of $\mathbb{Y}^{\mathrm{g}}$, it is indeed possible to eliminate the corresponding error due to discretization of the dual domain. To illustrate, let J^{d} be a one-dimensional, discrete, convex-extensible function with domain $\mathbb{X}^{\mathrm{g}} = \{x^i\}_{i=1}^N \subset \mathbb{R}$, where $x^i < x^{i+1}$. Also, choose $\mathbb{Y}^{\mathrm{g}} = \{y^i\}_{i=1}^{N-1} \subset \mathbb{R}$ with $y^i = \frac{J^{\mathrm{d}}(x^{i+1}) - J^{\mathrm{d}}(x^i)}{x^{i+1} - x^i}$ as the discrete dual domain. Then, for all $x \in \mathrm{co}(\mathbb{X}^{\mathrm{g}}) = [x^1, x^N]$, we have $J^{\mathrm{d}*\mathrm{d}*}(x) = \overline{J^{\mathrm{d}}}(x)$, where $\overline{[\cdot]}$ is the LERP extension. Hence, the only source of error under such construction is the discretization of the primal state space (i.e., approximation of the true J via $\overline{J^{\mathrm{d}}}$). However, a similar construction of $\mathbb{Y}^{\mathrm{g}}$ in dimensions $n \geq 2$ can lead to dual grids of size $Y = \mathscr{O}(X^n)$, which makes the proposed algorithm computationally inefficient; see Theorem 4.4.4. The following result provides us with specific bounds on the discretization error that point to a more practical way for construction of $\mathbb{Y}^{\mathrm{g}}$.

Theorem 4.4.6 (Error of d-CDP)**.** *Consider the DP operator $\mathscr{T}$ (4.17) and the d-CDP operator $\widehat{\mathscr{T}}^{\mathrm{d}}$ (4.22). Assume that $C : \mathbb{X} \times \mathbb{U} \to \overline{\mathbb{R}}$ is convex in the input variable. Also assume that $J : \mathbb{X} \to \mathbb{R}$ is a Lipschitz continuous, convex function. Then,*

$$-e_{\mathrm{x}} \leq \mathscr{T} J(x) - \widehat{\mathscr{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) \leq e_{\mathrm{y}}(x), \quad \forall x \in \mathbb{X}^{\mathrm{g}}, \tag{4.25}$$

where

$$e_{\mathrm{y}}(x) = \left[\left\| f_{\mathrm{s}}(x) \right\|_2 + \left\| f_{\mathrm{i}}(x) \right\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}} \right] \cdot \mathrm{d}\left(\partial \mathscr{T} J(x), \mathbb{Y}^{\mathrm{g}}\right), \tag{4.26a}$$

$$e_{\mathrm{x}} = [\Delta_{\mathbb{Y}^{\mathrm{g}}} + \mathrm{L}(J)] \cdot \mathrm{d}_{\mathrm{H}}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}). \tag{4.26b}$$

4.4.3. CONSTRUCTION OF $\mathbb{Y}^{\mathrm{g}}$

We now use the result of our error analysis in Theorem 4.4.6 to provide a computationally efficient numerical scheme for construction of the grid $\mathbb{Y}^{\mathrm{g}}$ in Algorithm 3. Notice how the two terms e_{y} and e_{x} in(4.26) capture the errors due to the discretization of the dual state space ($\mathbb{Y}$) and the primal state space ($\mathbb{X}$), respectively. In particular, the first error term suggests that we choose $\mathbb{Y}^{\mathrm{g}}$ such that $\partial \mathscr{T} J(x) \cap \mathbb{Y}^{\mathrm{g}} \neq \emptyset$ for all $x \in \mathbb{X}^{\mathrm{g}}$. Even if we had access

to $\mathscr{T}J$, satisfying such a condition can again lead to dual grids of size $Y = \mathscr{O}(X^n)$. A more realistic objective is then to choose $\mathbb{Y}^{\mathrm{g}}$ such that $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}) \cap \partial\mathscr{T}J(x) \neq \emptyset$ for all $x \in \mathbb{X}^{\mathrm{g}}$. With such a construction, the distance $\mathrm{d}\big(\partial\mathscr{T}J(x), \mathbb{Y}^{\mathrm{g}}\big)$ and hence e_{y} decrease by using finer grids for the dual domain. The latter condition is satisfied if $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}) \supseteq \mathbb{L}(\mathscr{T}J)$. Hence, we need to approximate "the range of slopes" of the function $\mathscr{T}J$. Notice, however, that we do not have access to $\mathscr{T}J$ since it is the *output* of the d-CDP operation in Algorithm 3. What we have at our disposal as *inputs* are the stage cost C and the next step (discrete) cost-to-go J^{d}. A coarse way to approximate the range of slopes of $\mathscr{T}J$ is then to use the extrema of the functions C and J^{d}, and the diameter of $\mathbb{X}^{\mathrm{g}}$ in each dimension. The following remark explains such an approximation for the construction of $\mathbb{Y}^{\mathrm{g}}$.

Remark 4.4.7 (Construction of $\mathbb{Y}^{\mathrm{g}}$)**.** *Let*

$$\mathrm{rng}(C) = \max_{(x,u)\in\mathrm{dom}(C)} C(x,u) \; - \min_{(x,u)\in\mathrm{dom}(C)} C(x,u).$$

Compute

$$\mathrm{rng}(J^{\mathrm{d}}) = \max_{x\in\mathbb{X}^{\mathrm{g}}} J^{\mathrm{d}}(x) \; - \min_{x\in\mathbb{X}^{\mathrm{g}}} J^{\mathrm{d}}(x),$$

and then choose $\mathbb{Y}^{\mathrm{g}} = \Pi_{i=1}^{n} \mathbb{Y}_i^{\mathrm{g}} \subset \mathbb{R}^n$ *such that for each dimension* $i = 1, \ldots, n$*, we have*

$$\pm\alpha \cdot \frac{\mathrm{rng}(C) + \mathrm{rng}(J^{\mathrm{d}})}{\Delta_{\mathbb{X}_i^{\mathrm{g}}}} \in \mathrm{co}(\mathbb{Y}_i^{\mathrm{g}}).$$

Here, $\alpha > 0$ *is a scaling factor mainly depending on the dimension* n *of the state space. Construction of* $\mathbb{Y}^{\mathrm{g}}$ *as described above requires* $\mathscr{O}(X)$ *operations* per iteration *(for computing* $\mathrm{rng}(J^{\mathrm{d}})$ *via enumeration).*

4.5. FROM QUADRATIC TO LINEAR COMPLEXITY

In this section, we focus on a specific subclass of the optimal control problems considered in this study. In particular, we exploit the problem structure in this subclass to reduce the computational cost of the d-CDP operation. In this regard, a closer look to Algorithm 3 reveals a computational bottleneck in its numerical implementation: the computation of the objects $\varphi_x^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$, $x \in \mathbb{X}^{\mathrm{g}}$, and their conjugates which requires working in the product space $\mathbb{X}^{\mathrm{g}} \times \mathbb{Y}^{\mathrm{g}}$. This step is indeed the dominating factor in the time complexity of $\mathscr{O}(XY)$ of the d-CDP operation; see the proof of Theorem 4.4.4. Hence, if the structure of the problem allows for the complete decomposition of these objects, then a significant reduction in the time complexity is achievable. This is indeed possible for problems with separable data:

Setting 4.5.1. ***(i)*** *The dynamics are input-affine with state-independent input dynamics, i.e.,* $f(x,u) = f_{\mathrm{s}}(x) + B \cdot u$*, where* $f_{\mathrm{s}} : \mathbb{R}^n \to \mathbb{R}^n$ *and* $B \in \mathbb{R}^{n\times m}$*.* ***(ii)*** *The stage cost is separable in state and input, i.e.,* $C(x,u) = C_{\mathrm{s}}(x) + C_{\mathrm{i}}(u)$*, where* $C_{\mathrm{s}} : \mathbb{X} \to \mathbb{R}$ *and* $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ *are the state and input costs, respectively.*

Note that the separability of the stage cost C implies that the constraints are also separable, i.e, there are no state-dependent input constraints.

4.5.1. MODIFIED d-CDP OPERATOR

For the separable cost of Setting 4.5.1, the state cost (C_s) can be taken out of the minimization in the DP operator (4.17) as follows

$$\mathscr{T} J(x) = C_\mathrm{s}(x) + \min_u \{C_\mathrm{i}(u) + J\big(f(x,u)\big)\}, \quad x \in \mathbb{X}. \tag{4.27}$$

Following a similar dualization and discretization procedure described in Section 4.4.1, we can derive the corresponding d-CDP operator

$$J^{\mathrm{d}*\mathrm{d}}(y) = \max_{x \in \mathbb{X}^\mathrm{g}} \Big\{ \langle y, x \rangle - J^\mathrm{d}(x) \Big\}, \qquad y \in \mathbb{Y}^\mathrm{g}, \tag{4.28a}$$

$$\varphi^\mathrm{d}(y) := C_\mathrm{i}^*(-B^\top y) + J^{\mathrm{d}*\mathrm{d}}(y), \qquad y \in \mathbb{Y}^\mathrm{g}, \tag{4.28b}$$

$$\widehat{\mathscr{T}}^\mathrm{d} J^\mathrm{d}(x) = C_\mathrm{s}(x) + \varphi^{\mathrm{d}*}\big(f_\mathrm{s}(x)\big), \qquad x \in \mathbb{X}^\mathrm{g}. \tag{4.28c}$$

Here, again, we assume that the conjugate of the input cost is analytically available (similar to Assumption 4.4.3, now in the context posed by Setting 4.5.1).

Assumption 4.5.2 (Conjugate of input cost)**.** *The conjugate function*

$$C_\mathrm{i}^*(v) = \max_u \{\langle v, u \rangle - C_\mathrm{i}(u)\}, \quad v \in \mathbb{R}^m,$$

is analytically available, i.e., the complexity of evaluating $C_\mathrm{i}^*(v)$ for each $v \in \mathbb{R}^m$ is of $\mathscr{O}(1)$.

Notice how the function φ^d in (4.28b) is now independent of the state variable x. This means that the computation of φ^d requires $\mathscr{O}(X+Y)$ operations, as opposed to $\mathscr{O}(XY)$ for the computation of φ_x^d in Algorithm 3. What remains to be addressed is the computation of the conjugate function $\varphi^{\mathrm{d}*}\big(f_\mathrm{s}(x)\big) = \max_{y \in \mathbb{Y}^\mathrm{g}} \{\langle f_\mathrm{s}(x), y \rangle - \varphi(y)\}$ for $x \in \mathbb{X}^\mathrm{g}$ in (4.28c). The straightforward maximization via enumeration over $y \in \mathbb{Y}^\mathrm{g}$ for each $x \in \mathbb{X}^\mathrm{g}$ (as in Algorithm 3) again leads to a time complexity of $\mathscr{O}(XY)$. The key idea here is to use *approximate discrete conjugation*:

- Use LLT to compute $\varphi^{\mathrm{d}*\mathrm{d}} : \mathbb{Z}^\mathrm{g} \to \mathbb{R}$ from the data points $\varphi^\mathrm{d} : \mathbb{Y}^\mathrm{g} \to \mathbb{R}$ for a grid $\mathbb{Z}^\mathrm{g}$;
- For each $x \in \mathbb{X}^\mathrm{g}$, use LERP to compute $\overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_\mathrm{s}(x)\big)$ using $\varphi^{\mathrm{d}*\mathrm{d}} : \mathbb{Z}^\mathrm{g} \to \mathbb{R}$.

Proper construction of the grid $\mathbb{Z}^\mathrm{g}$ will be discussed in Section 4.5.3. With such an approximation, the d-CDP operator (4.28) *modifies* to

$$J^{\mathrm{d}*\mathrm{d}}(y) = \max_{x \in \mathbb{X}^\mathrm{g}} \Big\{ \langle y, x \rangle - J^\mathrm{d}(x) \Big\}, \qquad y \in \mathbb{Y}^\mathrm{g}, \tag{4.29a}$$

$$\varphi^\mathrm{d}(y) = C_\mathrm{i}^*(-B^\top y) + J^{\mathrm{d}*\mathrm{d}}(y), \qquad y \in \mathbb{Y}^\mathrm{g}, \tag{4.29b}$$

$$\varphi^{\mathrm{d}*\mathrm{d}}(z) = \max_{y \in \mathbb{Y}^\mathrm{g}} \Big\{ \langle z, y \rangle - \varphi^\mathrm{d}(y) \Big\}, \qquad z \in \mathbb{Z}^\mathrm{g}, \tag{4.29c}$$

$$\widehat{\mathscr{T}}_\mathrm{m}^\mathrm{d} J^\mathrm{d}(x) := C_\mathrm{s}(x) + \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_\mathrm{s}(x)\big), \qquad x \in \mathbb{X}^\mathrm{g}. \tag{4.29d}$$

Algorithm 4 provides the pseudo-code for the ConjVI algorithm that utilizes the modified d-CDP operator.

Algorithm 4 ConjVI algorithm via modified d-CDP operator (4.29) for Setting 4.5.1.

Input: dynamics $f_{\mathrm{s}} : \mathbb{R}^n \to \mathbb{R}^n$, $B \in \mathbb{R}^{n\times m}$; discrete state space $\mathbb{X}^{\mathrm{g}} \subset \mathbb{X}$; discrete state cost $C_{\mathrm{s}}^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$; conjugate of input cost $C_{\mathrm{i}}^* : \mathbb{R}^m \to \mathbb{R}$; discrete terminal cost $C_T^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$.

Output: discrete costs-to-go $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$, $t = 0, 1, \ldots, T$.

initialization:

construct the grid $\mathbb{Z}^{\mathrm{g}}$;

$J_T^{\mathrm{d}}(x) \leftarrow C_T^{\mathrm{d}}(x)$ for $x \in \mathbb{X}^{\mathrm{g}}$;

backward iteration:

for $t = T, \ldots, 1$ **do**

  construct the grid $\mathbb{Y}^{\mathrm{g}}$;

  modified d-CDP operation:

  use LLT to compute $J_t^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ from $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$;

  $\varphi^{\mathrm{d}}(y) \leftarrow C_{\mathrm{i}}^*(-B^\top y) + J_t^{\mathrm{d}*\mathrm{d}}(y)$ for $y \in \mathbb{Y}^{\mathrm{g}}$;

  use LLT to compute $\varphi^{\mathrm{d}*\mathrm{d}} : \mathbb{Z}^{\mathrm{g}} \to \mathbb{R}$ from $\varphi^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$;

  for each $x \in \mathbb{X}^{\mathrm{d}}$ **do**

    use LERP to compute $\overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big)$ from $\varphi^{\mathrm{d}*\mathrm{d}} : \mathbb{Z}^{\mathrm{g}} \to \mathbb{R}$;

    $J_{t-1}^{\mathrm{d}}(x) \leftarrow \widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}} J_t^{\mathrm{d}}(x) = C_{\mathrm{s}}^{\mathrm{d}}(x) + \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big)$;

  end for

end for

4.5.2. ANALYSIS OF MODIFIED d-CDP OPERATOR

We again begin with the time complexity of the modified d-CDP operator.

Theorem 4.5.3 (Complexity of modified d-CDP)**.** *Let Assumptions 4.2.4 and 4.5.2 hold. Then, the computation of the modified d-CDP operator (4.29) in Algorithm 4 has a time complexity of* $\widetilde{\mathscr{O}}(X + Y + Z)$.

Once again, we note that in the application of the ConjVI Algorithm 4 for solving the T-step VI problem, the time complexity increases linearly with the horizon T (assuming that the grids $\mathbb{Y}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$ can be constructed with at most $\mathscr{O}(X)$ operations; see Remarks 4.4.7 and 4.5.5).

Comparing the time complexity of the modified d-CDP operator $\widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}}$ (4.29) with that of d-DP operator $\mathscr{T}^{\mathrm{d}}$ (4.18) and d-CDP operator $\widehat{\mathscr{T}}^{\mathrm{d}}$ (4.22) (i.e., $\mathscr{O}(XUE)$ and $\mathscr{O}(XY)$, respectively), we observe a reduction from quadratic complexity to (log-)linear complexity. To illustrate, let us assume that all of the involved grids ($\mathbb{X}^{\mathrm{g}}$, $\mathbb{Y}^{\mathrm{g}}$, and $\mathbb{Z}^{\mathrm{g}}$) are of the same size, i.e., $Y, Z = X$ (this is also consistent with Assumption 4.2.4). Then, the complexity of $\widehat{\mathscr{T}}^{\mathrm{d}}$ is of $\mathscr{O}(X^2)$, while the complexity of $\widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}}$ is of $\widetilde{\mathscr{O}}(X)$.

We next consider the error of the proposed algorithm by providing a bound on the difference between the modified d-CDP operator (4.29) and the DP operator (4.27).

Theorem 4.5.4 (Error of modified d-CDP)**.** *Consider the DP operator $\mathscr{T}$ (4.27) and the modified d-CDP operator $\widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}}$ (4.29). Assume that the input cost $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ is convex, and the function $J : \mathbb{X} \to \mathbb{R}$ is a Lipschitz continuous, convex function. Also, assume that the grid $\mathbb{Z}^{\mathrm{g}}$ is such that* $\mathrm{co}(\mathbb{Z}^{\mathrm{g}}) \supseteq f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$. *Then,*

$$-\big(e_{\mathrm{x}} + e_{\mathrm{z}}\big) \le \mathscr{T} J(x) - \widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}} J^{\mathrm{d}}(x) \le e_{\mathrm{y}}^m(x), \quad \forall x \in \mathbb{X}^{\mathrm{g}}, \tag{4.30}$$

where

$$e_{\mathrm{y}}^{m}(x) = \left[\left\| f_{\mathrm{s}}(x) \right\|_2 + \|B\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}} \right] \cdot \mathrm{d}\left(\partial\left(\mathcal{T} J - C_{\mathrm{s}}\right)(x), \mathbb{Y}^{\mathrm{g}}\right), \tag{4.31a}$$

$$e_{\mathrm{x}} = [\Delta_{\mathbb{Y}^{\mathrm{g}}} + \mathrm{L}(J)] \cdot \mathrm{d}_{\mathrm{H}}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}), \tag{4.31b}$$

$$e_{\mathrm{z}} = \Delta_{\mathbb{Y}^{\mathrm{g}}} \cdot \mathrm{d}_{\mathrm{H}}\left(f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\right). \tag{4.31c}$$

4.5.3. CONSTRUCTION OF $\mathbb{Y}^{\mathrm{g}}$ AND $\mathbb{Z}^{\mathrm{g}}$

We now provide specific guidelines for proper construction of the grids $\mathbb{Y}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$ using the result of our error analysis in Theorem 4.5.4. Once again, the three terms capture the errors due to discretization in y, x, and z, respectively. Concerning the grid $\mathbb{Y}^{\mathrm{g}}$, because of the error term e_{y}^{m} (4.31a), similar guidelines to the ones provided in Section 4.4.3 apply here. In particular, notice that e_{y}^{m} now depends on $\mathrm{d}\left(\partial\left(\mathcal{T} J - C_{\mathrm{s}}\right)(x), \mathbb{Y}^{\mathrm{g}}\right)$, and hence in the construction of $\mathbb{Y}^{\mathrm{g}}$, we need to consider the range of slopes of $\mathcal{T} J - C_{\mathrm{s}}$. This essentially means using $\mathrm{rng}(C_{\mathrm{i}})$ instead of $\mathrm{rng}(C)$ in Remark 4.4.7.

Next to be addressed is the construction of the grid $\mathbb{Z}^{\mathrm{g}}$. Here, we are dealing with the issue of constructing the dual grid for approximate discrete conjugation. Then, by Corollary 4.2.7, we can

- either construct a *fixed* grid $\mathbb{Z}^{\mathrm{g}}$ such that $\mathrm{co}(\mathbb{Z}^{\mathrm{g}}) \supseteq f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$,
- or construct $\mathbb{Z}^{\mathrm{g}}$ *dynamically* such that $\mathrm{co}(\mathbb{Z}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(\varphi^{\mathrm{d}})$ at each iteration.

The former has a *one-time* computational cost of $\mathcal{O}(X)$, while the latter requires $\mathcal{O}(Y)$ operations *per iteration*. For this reason, as also assumed in Theorem 4.5.4, we use the first method to construction $\mathbb{Z}^{\mathrm{g}}$. The following remark summarizes this discussion.

Remark 4.5.5 (Construction of $\mathbb{Z}^{\mathrm{g}}$)**.** *Construct the grid $\mathbb{Z}^{\mathrm{g}}$ such that* $\mathrm{co}(\mathbb{Z}^{\mathrm{g}}) \supseteq f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$. *This can be done by finding the vertices of the smallest hyper-rectangle that contains the set $f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$. Such a construction has a* one-time *computational cost of $\mathcal{O}(X)$.*

4.5.4. PERFECT TRANSFORMATION

Let us first note that the developed algorithms involve two conjugate transforms at the beginning and end of each step (see, e.g., lines 5 and 7 in Algorithm 4). Hence, the possibility of a "perfect" transformation of the minimization in the primal domain to a simple addition in the conjugate domain is interesting since it allows for performing the value iteration completely in the conjugate domain for the *conjugate* of the costs-to-go. In other words, we can stay in the conjugate domain over multiple steps in time, and avoid the conjugate operation at the beginning of the intermediate steps. This, in turn, leads to a lower computational cost in multistep implementations. However, for such a perfect transformation to be possible, we need to impose further restrictions on the problem data. To be precise, on top of the properties laid out in Setting 4.5.1, we need

- the dynamics to be *linear*, i.e., $f(x,u) = Ax + Bu$, with *invertible* $A \in \mathbb{R}^{n \times n}$,
- and the cost to be *state-independent*, i.e., $C_{\mathrm{s}}(x) = 0$, and hence $C(x,u) = C_{\mathrm{i}}(u)$.

For systems satisfying these conditions, the DP operator reads as

$$\mathscr{T}J(x) = \min_u \{C_{\mathrm{i}}(u) + J(Ax+Bu)\}, \quad x \in \mathbb{X},$$

and its conjugate can be shown to be given by

$$[\mathscr{T}J]^*(y) = C_{\mathrm{i}}^*(-B^\top A^{-\top} y) + J^*(A^{-\top} y), \quad y \in \mathbb{R}^n.$$

Notice how the minimization in the DP operator in the primal domain is perfectly transformed to an addition in the dual domain. This property indeed allows us to stay in the dual domain over multiple steps in time, while only computing the conjugate of the costs in the intermediate steps. The possibility of such a perfect transformation, accompanied by the application of LLT for better time complexity, was first noticed in [85]. Indeed, there, the authors introduced the "fast value iteration" algorithm for a more restricted class of DP problems (besides the properties discussed above, they required, among other conditions, the state matrix A to be non-negative and monotone). In this regard, we also note that, as in [85], the possibility of staying in the conjugate domain over multiple steps is particularly interesting for infinite-horizon problems.

4

4.5.5. TOTAL COMPLEXITY OF SOLVING THE OPTIMAL CONTROL PROBLEM

We finish this section with some remarks on using the output of the backward value iteration for finding a suboptimal control sequence $\mathbf{u}_\star(x_0)$ for a given instance of the optimal control problem with initial state x_0. Having the discrete costs-to-go $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$, $t = 0, 1, \dots, T-1$, at our disposal (the output of the VI or ConjVI algorithms), at each time step, we can use *the greedy action* with respect to the next step's cost-to-go, i.e.,

$$u_t^\star \in \operatorname*{argmin}_{u_t \in \mathbb{U}^{\mathrm{d}}} \left\{ C(x_t, u_t) + \widetilde{J_{t+1}^{\mathrm{d}}}\big(f(x_t, u_t)\big) \right\}, \quad t = 0, 1, \dots, T-1, \tag{4.32}$$

for a proper discrete input space $\mathbb{U}^{\mathrm{d}}$. Assuming these minimization problems are handled via enumeration, they lead to an additional computational burden of $\mathscr{O}(UE)$ per iteration, where E represents the complexity of the extension operation in (4.32). Then, the *total* time complexity of solving a T-step optimal control problem (i.e., the time requirement of backward value iteration for finding J_t^{d}, $t = 0, 1, \dots, T-1$, plus the time requirement of forward iteration for finding $u_t^\star$, $t = 0, 1, \dots, T-1$) of the three algorithms can be summarized as follows.[2]

[2] We note that the standard VI algorithm that utilizes the d-DP operation (4.18) also provides us with control laws $\mu_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{U}^{\mathrm{g}}$, $t = 0, 1, \dots, T-1$. However, the ConjVI algorithms only provide us with the costs J_t^{d}, $t = 0, 1, \dots, T-1$. Hence, when the standard VI algorithm is applied, we can alternatively use the control laws, accompanied by a proper extension operator, to produce a suboptimal control sequence, i.e.,

$$u_t^\star(x_t) = \widetilde{\mu_t^{\mathrm{d}}}(x_t), \quad t = 0, 1, \dots, T-1.$$

This method has a time complexity of $\mathscr{O}(E)$, where E represents the complexity of the extension operation used above. This complexity can be particularly lower than that of generating greedy actions with respect to the computed costs in (4.32). However, generating control actions using the control laws has a higher memory complexity for systems with multiple inputs, and is also usually more sensitive to modeling errors due to its completely open-loop nature. Moreover, we note that the *total* time complexity of solving an instance of the optimal control problem, i.e., backward iteration for computing the costs J_t^{d} and control laws μ_t^{d}, and forward iteration for computing the control sequence $\mathbf{u}_\star(x_0)$, is in both methods of $\mathscr{O}(TXUE)$. That is, computationally, the backward value iteration is the dominating factor.

Remark 4.5.6 (Comparison of total complexities)**.** *The total time complexity of solving a T-step optimal control problem for a given initial state, where the control input is generated using the greedy policy* (4.32)*, is of*

- $\mathcal{O}(TXUE)$ *for the VI algorithm that utilizes the d-DP operation* (4.18)*,*
- $\mathcal{O}\big(T(XY+UE)\big)$ *for the ConjVI Algorithm 3,*
- $\widetilde{\mathcal{O}}\big(T(X+Y+Z+UE)\big)$ *for the ConjVI Algorithm 4,*

where E represents the complexity of the extension operation in (4.18) *and* (4.32)*.*

Once again, we see a reduction from quadratic to linear complexity in the modified d-CDP Algorithm 4 compared to both the d-DP algorithm and the d-CDP Algorithm 3.

4

4.6. NUMERICAL EXPERIMENTS

In this section, we examine the performance of the ConjVI Algorithms 3 and 4 (referred to as ConjVI 3 and ConjVI-m 4, respectively, in this section) in comparison with the standard VI algorithm in the primal domain that utilizes the d-DP operation (4.18) (referred to as VI in this section). In particular, we use a synthetic numerical example to verify our theoretical results on the complexity and error of the proposed algorithms. All the simulations presented in this chapter were implemented via MATLAB version R2017b, on a PC with an Intel Xeon 3.60 GHz processor and 16 GB RAM. We also note that the presented numerical example is also included in the d-CDP MATLAB package [64].

We consider a linear system with two states and two inputs described by

$$x_{t+1} = \begin{bmatrix} -0.5 & 2 \\ 1 & 3 \end{bmatrix} x_t + \begin{bmatrix} 1 & 0.5 \\ 1 & 1 \end{bmatrix} u_t,$$

over the finite horizon $T = 10$, with the state and input constraints $x_t \in \mathbb{X} = [-1,1]^2 \subset \mathbb{R}^2$ and $u_t \in \mathbb{U} = [-2,2]^2 \subset \mathbb{R}^2$, respectively. Moreover, we consider *quadratic state cost* $C_{\mathrm{s}}(x) = C_T(x) = \|x\|_2^2$ and *exponential input cost* $C_{\mathrm{i}}(u) = e^{|u_1|} + e^{|u_2|} - 2$. Note that the conjugate of the input cost is indeed analytically available and given by

$$C_{\mathrm{i}}^*(v) = 1 + \langle \hat{u}, v\rangle - e^{|\hat{u}_1|} - e^{|\hat{u}_2|}, \quad v \in \mathbb{R}^2,$$

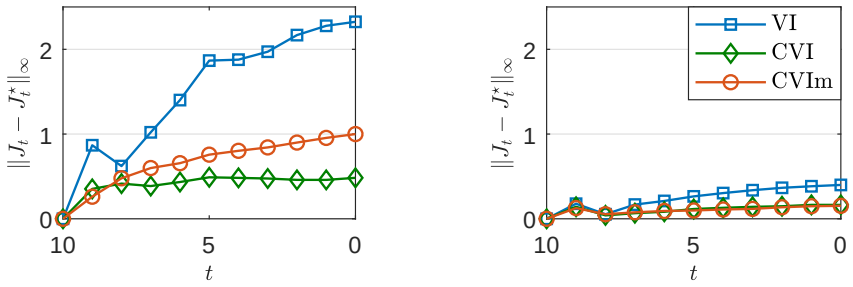
where

$$\hat{u}_i = \begin{cases} \max\big\{-2,\ \min\big\{2,\ \mathrm{sgn}(v_i)\ln|v_i|\big\}\big\}, & v_i \neq 0, \\ 0, & v_i = 0, \end{cases} \qquad i = 1,2.$$

Moreover, corresponding to the notation of Section 4.4, the stage cost and its conjugate are given by

$$C_x(u) = C(x,u) = \|x\|_2^2 + e^{|u_1|} + e^{|u_2|} - 2, \quad (x,u) \in \mathbb{X} \times \mathbb{U},$$
$$C_x^*(v) = C_{\mathrm{i}}^*(v) - \|x\|_2^2, \quad (x,v) \in \mathbb{X} \times \mathbb{R}^2.$$

We use *uniform* grid-like discretizations $\mathbb{X}^{\mathrm{g}}$ and $\mathbb{U}^{\mathrm{g}}$ for the state and input spaces, such that $\mathrm{co}(\mathbb{X}^{\mathrm{g}}) = \mathbb{X}$ and $\mathrm{co}(\mathbb{U}^{\mathrm{g}}) = \mathbb{U}$. The grids $\mathbb{Y}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$ involved in ConjVI algorithms



(a) $N = 11^2$ (b) $N = 41^2$

Figure 4.2: Error of the computed discrete costs $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ using VI, ConjVI 3 (CVI), and ConjVI-m 4 (CVIm) for grid sizes $X, U, Y, Z = N$. Notice that the time axis is backward.

4

are also constructed *uniformly*, according to the guidelines provided in Remarks 4.4.7 and 4.5.5 (with $\alpha = 1$). We are interested in the performance (error and time complexity) of ConjVI algorithms in comparison with VI, as the size of these discrete sets increases. Since all the discrete sets are uniform grids, and we use LERP for all the extension operations (particularly, for the extension of the discrete cost functions in the d-DP operation (4.18) and for generating the greedy control actions in (4.32)), the complexity of a single evaluation of all extensions is of $\mathcal{O}(E) = \mathcal{O}(1)$; see Remark 4.2.2.

We begin with examining the error in VI and ConjVI algorithms with respect to the "reference" costs-to-go $J_t^\star : \mathbb{X} \to \mathbb{R}$. Since the problem does not have a closed-form solution, $J_t^\star$ is computed numerically via a high-resolution application of VI with $X, U = 81^2$. Figure 4.2 depicts the maximum absolute error in the discrete cost functions J_t^{d} computed using these algorithms over the horizon. As expected and in line with our error analysis (Theorems 4.4.6 and 4.5.4 and Proposition 4.3.2), using a finer discretization scheme with larger $X, U, Y, Z = N$, leads to a smaller error. Moreover, over the time steps in the backward iteration, a general increase is seen in the error which is due to the accumulation of error.

For further illustration, Figure 4.3 shows the corresponding costs-to-go at $t = 9$ and $t = 0$, with $N = 21^2$. Notice that, since the stage and terminal costs are convex and the dynamics are linear, the costs-to-go are also convex. As can be seen in Figure 4.3, while ConjVI 3 preserves the convexity, VI and ConjVI-m 4 output non-convex costs-to-go (due to the application of LERP in these algorithms). In particular, notice how J_0^{CVI} is convex-extensible while J_0^{VI} and J_0^{CVIm} are not.

We next compare the performance of the three algorithms in solving instances of the optimal control problem, using the cost functions derived from the backward value iteration. To this end, we apply the greedy control input (4.32) with respect to the computed discrete costs-to-go J_t^{d} using VI and ConjVI algorithms, and the same discrete input space $\mathbb{U}^{\mathrm{g}}$ as the one in VI. Let us first consider the complexity of VI and ConjVI algorithms. Figure 4.4a reports the *total* run-time of a random problem instance for different grid sizes (i.e., the time requirement of backward value iteration for finding J_t^{d}, $t = 0, 1, \ldots, T-1$, plus the time requirement of forward iteration for finding $u_t^\star$, $t =$

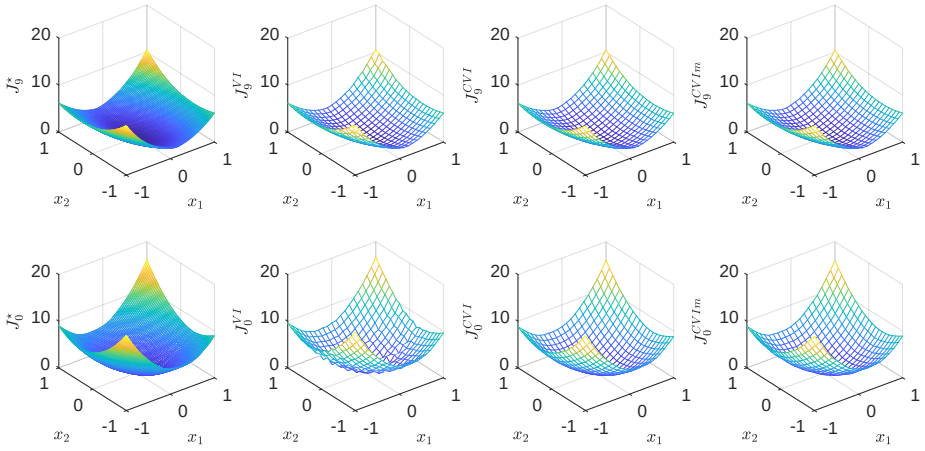
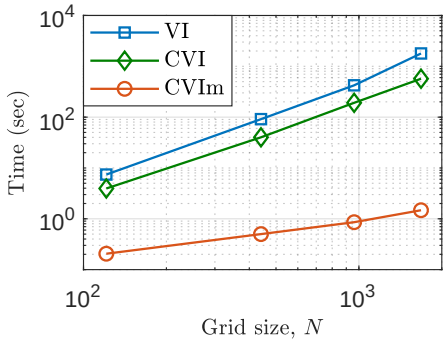


Figure 4.3: Computed discrete costs $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ using VI, ConjVI 3 (CVI), and ConjVI-m 4 (CVIm) for grid sizes $X, U, Y, Z = 21^2$ at $t = 9$ (top) and $t = 0$ (bottom).

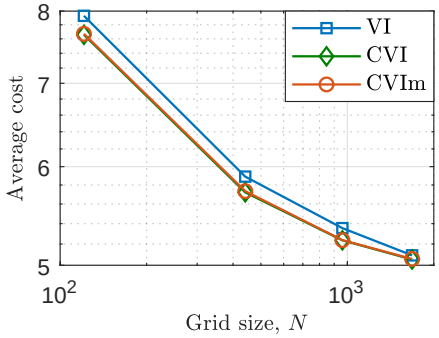
$0, 1, \ldots, T-1$). Regarding the reported running times, note that they correspond to the given complexities in Theorems 4.4.4 and 4.5.3 and Remark 4.5.6: For our numerical example, the running time is of $\mathcal{O}(TN^2)$ for VI and ConjVI 3, and of $\mathcal{O}(TN)$ for ConjVI-m 4. The difference can be readily seen in the slope of the corresponding lines in Figure 4.4a as N increases. In this regard, we also note that the backward value iteration is the dominant factor in the reported running times. (Effectively, the reported numbers can be taken to be the run-time of the backward value iteration). In Figure 4.4b, we also report the average cost of the controlled trajectories over 100 instances of the optimal control problem with random initial conditions, chosen uniformly from $\mathbb{X} = [-1, 1]^2$.

Looking at Figure 4.4, one notices that ConjVI-m 4, compared to VI, has a similar performance when it comes to the quality of greedy control actions, however, with a significant reduction in the running time. In particular, notice how the lower complexity of ConjVI-m 4 allows us to increase the size of the grids to $N = 41^2$, while keeping the running time at the same order as that of VI with $N = 11^2$.

Comparing the performance of ConjVI 3 with VI, on the other hand, one notices that they show effectively the same performance with respect to the considered measures. ConjVI 3, however, gives us an extra degree of freedom for the size Y of the dual grid. In particular, if the cost functions are "compactly representable" in the dual domain (i.e., via their slopes), we can reduce the time complexity of ConjVI 3 by using a more coarse grid $\mathbb{Y}^{\mathrm{g}}$, with a limited effect on the "quality" of computed cost functions. This effect is illustrated in Table 4.2: For solving the same optimal control problem with $X, U, Y = 41^2$, we can reduce the size of the dual grid by a factor of 4 to $Y = 21^2$, and hence reduce the running time of ConjVI 3, while achieving the same average cost in the controlled trajectories.



(a) Total running time for solving a random problem instance



(b) Average cost of 100 random instances of the optimal control problem.

Figure 4.4: Performance of VI, ConjVI 3 (CVI), ConjVI-m 4 (CVIm) for different grid sizes $X, U, Y, Z = N$.

4.7. TECHNICAL PROOFS

PROOF OF LEMMA 4.2.5

Let $y \in \mathbb{R}^n$, and observe that (recall that $h^{\mathrm{d}}(x) = h(x)$ for all $x \in \mathbb{X}^{\mathrm{d}} \subset \mathbb{X}$)

$$h^{\mathrm{d}*}(y) = \max_{x \in \mathbb{X}^{\mathrm{d}}} \{\langle y, x\rangle - h^{\mathrm{d}}(x)\} \leq \max_{x \in \mathbb{X}} \{\langle y, x\rangle - h(x)\} = h^*(y).$$

This settles the first inequality in (4.8) and (4.9). Also, observe that if $\partial h^*(y) = \emptyset$, then the upper bound in (4.8) becomes trivial, i.e., $h^*(y) = +\infty$, $h^{\mathrm{d}*}(y) < +\infty$, and $\widetilde{e}_1 = +\infty$. Now, assume that $\partial h^*(y) \neq \emptyset$, and let $x \in \partial h^*(y)$ so that $h(x) + h^*(y) = \langle y, x\rangle$ [70, Prop. 5.4.3]. Also, let $\tilde{x} \in \operatorname{argmin}_{z \in \mathbb{X}^{\mathrm{d}}} \|x - z\|_2$, and note that $h^{\mathrm{d}*}(y) \geq \langle y, \tilde{x}\rangle - h^{\mathrm{d}}(\tilde{x})$. Then,

$$\begin{aligned} h^*(y) - h^{\mathrm{d}*}(y) &\leq \langle y, x - \tilde{x}\rangle - h(x) + h^{\mathrm{d}}(\tilde{x}) \\ &\leq \left[\|y\|_2 + \mathrm{L}\left(h; \{x\} \cup \mathbb{X}^{\mathrm{d}}\right)\right] \cdot \|x - \tilde{x}\|_2 \\ &= \left[\|y\|_2 + \mathrm{L}\left(h; \{x\} \cup \mathbb{X}^{\mathrm{d}}\right)\right] \cdot \mathrm{d}(x, \mathbb{X}^{\mathrm{d}}). \end{aligned}$$

Table 4.2: Performance VI and ConjVI 3 for grid sizes $X, U = 41^2$ and Y. The first two rows correspond to the rightmost data points in Figure 4.4.

Algorithm	Total run-time (sec)	Avg. cost (100 runs)
VI	1790	5.09
ConjVI 3 with $Y = 41^2$	570	5.05
ConjVI 3 with $Y = 21^2$	187	5.05

Hence, by minimizing over $x \in \partial h^*(y)$, we derive the upper bound provided in (4.8). Finally, the additional constraint of compactness of $\mathbb{X} = \mathrm{dom}(h)$ implies that $\partial h^*(y) \cap \mathbb{X} \neq \emptyset$. Hence, we can choose $x \in \partial h^*(y) \cap \mathbb{X}$ and use Lipschitz-continuity of h to write

$$\begin{aligned} h^*(y) - h^{\mathrm{d}*}(y) &\leq \left[\|y\|_2 + \mathrm{L}\left(h; \{x\} \cup \mathbb{X}^{\mathrm{d}}\right) \right] \cdot \mathrm{d}(x, \mathbb{X}^{\mathrm{d}}) \\ &\leq \left[\|y\|_2 + \mathrm{L}(h) \right] \cdot \max_{z \in \mathbb{X}} \mathrm{d}(z, \mathbb{X}^{\mathrm{d}}) = \widetilde{e}_2(y, h, \mathbb{X}^{\mathrm{d}}). \end{aligned}$$

PROOF OF LEMMA 4.2.6

Let us first consider the case $y \in \mathrm{co}(\mathbb{Y}^{\mathrm{g}})$. The value of the multi-linear interpolation $\overline{h^{*\mathrm{d}}}(y)$ is a convex combination of $h^{*\mathrm{d}}(y^{(k)}) = h^*(y^{(k)})$ over the grid points $y^{(k)} \in \mathbb{Y}^{\mathrm{g}}$, $k \in 1, \ldots, 2^n$, located at the vertices of the hyper-rectangular cell that contains y such that

$$y = \textstyle\sum_k \alpha^{(k)}\, y^{(k)} \quad \text{and} \quad \overline{h^{*\mathrm{d}}}(y) = \textstyle\sum_k \alpha^{(k)}\, h^*(y^{(k)}),$$

where $\sum_k \alpha^{(k)} = 1$ and $\alpha^{(k)} \in [0,1]$. Then,

$$h^*(y) = h^*\left(\textstyle\sum_k \alpha^{(k)}\, y^{(k)}\right) \leq \textstyle\sum_k \alpha^{(k)}\, h^*(y^{(k)}) = \overline{h^{*\mathrm{d}}}(y), \tag{4.33}$$

where the inequality follows from the convexity of h^*. Also, notice that

$$\begin{aligned} \overline{h^{*\mathrm{d}}}(y) &= \textstyle\sum_k \alpha^{(k)}\, h^*(y^{(k)}) = \sum_k \alpha^{(k)} \max_{x \in \mathbb{X}} \left\{ \langle y^{(k)}, x \rangle - h(x) \right\} \\ &= \textstyle\sum_k \alpha^{(k)} \max_{x \in \mathbb{X}} \left\{ \langle y, x \rangle - h(x) + \langle y^{(k)} - y, x \rangle \right\} \\ &\leq \textstyle\sum_k \alpha^{(k)} \max_{x \in \mathbb{X}} \left\{ \langle y, x \rangle - h(x) + \|y^{(k)} - y\|_2 \cdot \|x\|_2 \right\} \\ &\leq \textstyle\sum_k \alpha^{(k)} \max_{x \in \mathbb{X}} \left\{ \langle y, x \rangle - h(x) + \Delta_{\mathbb{X}} \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}) \right\}. \end{aligned}$$

Then, using $\sum_k \alpha^k = 1$, we have

$$\overline{h^{*\mathrm{d}}}(y) \leq \max_{x \in \mathbb{X}} \left\{ \langle y, x \rangle - h(x) \right\} + \Delta_{\mathbb{X}} \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}) = h^*(y) + \Delta_{\mathbb{X}} \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}). \tag{4.34}$$

Combining the two inequalities (4.33) and (4.34) gives us the inequality (4.10).

We next consider the case $y \notin \mathrm{co}(\mathbb{Y}^{\mathrm{g}})$ under the extra assumption $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(h)$. Note that this assumption implies that:

- $\mathbb{L}(h)$ is bounded (h is Lipschitz continuous); and,
- $y_i^1 < y_i^2 \leq \mathrm{L}_i^-(h)$ and $\mathrm{L}_i^+(h) \leq y_i^{Y_i-1} < y_i^{Y_i}$ for all $i \in \{1,\dots,n\}$.

To simplify the exposition, we consider the two-dimensional case ($n = 2$), while noting that the provided arguments can be generalized to higher dimensions. So, let $\mathbb{Y}^{\mathrm{g}} = \mathbb{Y}_1^{\mathrm{g}} \times \mathbb{Y}_2^{\mathrm{g}}$, where $\mathbb{Y}_i^{\mathrm{g}}$ ($i = 1,2$) is the finite set of real numbers $y_i^1 < y_i^2 < \ldots < y_i^{Y_i}$ with $Y_i \geq 3$. Let us further simplify the argument by letting $y = (y_1, y_2) \notin \mathrm{co}(\mathbb{Y}^{\mathrm{g}})$ be such that $y_1 < y_1^1$ and $y_2^1 \leq y_2 \leq y_2^2$, so that computing $\overline{h^{*\mathrm{d}}}(y)$ involves extrapolation in the first dimension and interpolation in the second dimension; see Figure 4.5a for a visualization of this instantiation. Since the extension uses LERP, using the points depicted in Figure 4.5a, we can write

$$\overline{h^{*\mathrm{d}}}(y) = \alpha\, \overline{h^{*\mathrm{d}}}(y') + (1-\alpha)\, \overline{h^{*\mathrm{d}}}(y''), \tag{4.35}$$

where $\alpha = (y_1^2 - y_1)/(y_1^2 - y_1^1)$, and (recall that $h^{*\mathrm{d}}(y) = h^*(y)$ for $y \in \mathbb{Y}^{\mathrm{g}}$)

$$\begin{aligned} \overline{h^{*\mathrm{d}}}(y') &= \beta\, h^{*\mathrm{d}}(y^{1,1}) + (1-\beta)\, h^{*\mathrm{d}}(y^{1,2}) = \beta\, h^*(y^{1,1}) + (1-\beta)\, h^*(y^{1,2}), \\ \overline{h^{*\mathrm{d}}}(y'') &= \beta\, h^{*\mathrm{d}}(y^{1,2}) + (1-\beta)\, h^{*\mathrm{d}}(y^{2,2}) = \beta\, h^*(y^{1,2}) + (1-\beta)\, h^*(y^{2,2}), \end{aligned} \tag{4.36}$$

where $\beta = (y_2^2 - y_2)/(y_2^2 - y_2^1)$. In Figure 4.5a, we have also paired each of the points of interest in the dual domain with its corresponding maximizer in the primal domain. That is, for

$$\xi = y, y', y'', y^{1,1}, y^{1,2}, y^{1,2}, y^{2,2},$$

we have respectively identified

$$\eta = x, x', x'', x^{1,1}, x^{1,2}, x^{1,2}, x^{2,2} \in \mathbb{X},$$

where $\xi \in \partial h(\eta)$ so that

$$h^*(\xi) = \langle \eta, \xi \rangle - h(\eta). \tag{4.37}$$

We now list the *implications* of the assumption $y_1^1 < y_1^2 \leq \mathrm{L}_1^-(h)$; Figure 4.5b illustrates these implications in the one-dimensional case:

I.1. We have $h^*(y) = \alpha\, h^*(y') + (1-\alpha)\, h^*(y'')$.

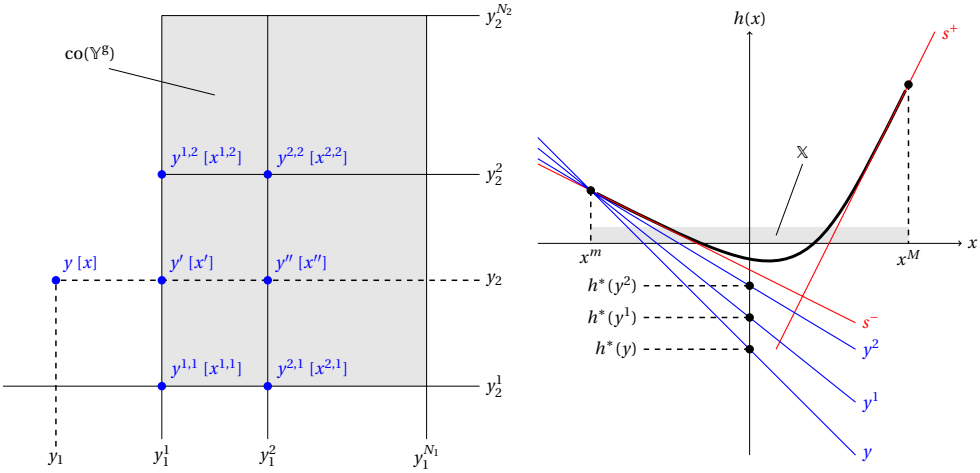
I.2. We can choose the maximizers in the primal domain such that

I.2.1. $x^{1,1} = x^{2,1}$, $x^{1,2} = x^{2,2}$, and $x = x' = x''$;

I.2.2. $x_1^{1,1} = x_1^{1,2} = x_1 = \min_{(z_1,z_2)\in\mathbb{X}} z_1$.

With these preparatory discussions, we can now consider the error of extrapolative discrete conjugation at the point y. In this regard, first note that $\{y', y''\} \subset \mathrm{co}(\mathbb{Y}^{\mathrm{g}})$, and hence we can use the result of the first part of the lemma to write

$$\overline{h^{*\mathrm{d}}}(y') = h^*(y') + e', \quad \overline{h^{*\mathrm{d}}}(y'') = h^*(y'') + e'', \tag{4.38}$$



(a) Position of the point y w.r.t. the grid $\mathbb{Y}^{\mathrm{g}}$

(b) Implications of the assumption

Figure 4.5: Illustration of the proof of Lemma 4.2.6. (a) The grid $\mathbb{Y}^{\mathrm{g}}$ and the position of the point y with respect to the grid. The blue dots show the points of interest and their corresponding maximizer in the primal domain. E.g., "y $[x]$" implies that $y \in \partial h(x)$, where $x \in \mathbb{X}$, so that $\langle x, y \rangle = h(x) + h^*(y)$. (b) Illustration of the implications of the assumption $y^1 < y^2 \leq s^- = \mathrm{L}^-(h)$ in the one-dimensional case. The colored (red and blue) variables denote the slope of the corresponding lines. Note that $\{y, y^1, y^2\} \subset \partial h(x^m)$, where $x^m = \min_{x \in \mathbb{X}} x$. Indeed, for all $y \leq s^-$, the conjugate $h^*(y) = \langle x^m, y \rangle - h(x^m)$ is a linear function with slope x^m. In particular, for $y < y^1$, we have $h^*(y) = \alpha h^*(y^1) + (1-\alpha) h^*(y^2)$, where $\alpha = (y^2 - y)/(y^2 - y^1)$.

where $\{e', e''\} \subset [0, \Delta_{\mathbb{X}} \cdot \mathrm{d_H}(\{y', y''\}, \mathbb{Y}^{\mathrm{g}})]$. We claim that these error terms are equal. Indeed, from (4.36) and (4.38), we have

$$e' - e'' = \beta \left[h^*(y^{1,1}) - h^*(y^{2,1}) \right] + (1-\beta) \left[h^*(y^{1,2}) - h^*(y^{2,2}) \right] + h^*(y'') - h^*(y').$$

Then, using the pairings in (4.37) and the implication I.2, we can write

$$\begin{aligned} e' - e'' &\overset{(I.2.1)}{=} \beta \langle x^{1,1}, y^{1,1} - y^{2,1} \rangle + (1-\beta) \langle x^{1,2}, y^{1,2} - y^{2,2} \rangle + \langle x, y'' - y' \rangle \\ &= \beta \langle x^{1,1}, (y_1^1 - y_1^2, 0) \rangle + (1-\beta) \langle x^{1,2}, (y_1^1 - y_1^2, 0) \rangle + \langle x, (y_1^2 - y_1^1, 0) \rangle \\ &= \left(\beta x_1^{1,1} + (1-\beta) x_1^{1,2} - x_1 \right) (y_1^1 - y_1^2) \overset{(I.2.2)}{=} 0. \end{aligned}$$

With this result at hand, we can employ (4.35) and the implication I.1 to write

$$\overline{h^{*\mathrm{d}}}(y) - h^*(y) = \alpha \left[\overline{h^{*\mathrm{d}}}(y') - h^*(y') \right] + (1-\alpha) \left[\overline{h^{*\mathrm{d}}}(y'') - h^*(y'') \right] = \alpha e' + (1-\alpha) e'' = e'.$$

That is,

$$0 \leq \overline{h^{*\mathrm{d}}}(y) - h^*(y) \leq \Delta_{\mathbb{X}} \cdot \mathrm{d_H}(\{y', y''\}, \mathbb{Y}^{\mathrm{g}}) \leq \Delta_{\mathbb{X}} \cdot \mathrm{d_H}\left(\mathrm{co}(\mathbb{Y}^{\mathrm{g}}), \mathbb{Y}^{\mathrm{g}} \right),$$

where for the last inequality we used the fact that $\{y', y''\} \subset \mathrm{co}(\mathbb{Y}^{\mathrm{g}})$.

PROOF OF COROLLARY 4.2.7

The first statement follows from Lemma 4.2.6 since the finite set $\mathbb{X}^{\mathrm{d}}$ is compact. For the second statement, the extra condition $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(h)$ has the same implications as the ones provided in the proof of Lemma 4.2.6. Hence, following the same arguments, we can show that provided bounds hold for all $y \in \mathbb{R}^n$ under the given condition.

PROOF OF PROPOSITION 4.3.2

Define $Q_x(u) := C(x,u) + J\big(f(x,u)\big)$ and $\widetilde{Q}_x(u) := C(x,u) + \widetilde{J^{\mathrm{d}}}\big(f(x,u)\big)$. Let us fix $x \in \mathbb{X}^{\mathrm{g}}$. In what follows, we consider the effect of (i) replacing J with $\widetilde{J^{\mathrm{d}}}$, and (ii) minimizing over $\mathbb{U}^{\mathrm{d}}$ instead of $\mathbb{U}(x)$, separately. To this end, we define the *intermediate* DP operator

$$\mathscr{T}^{\mathrm{i}} J(x) := \min_u \widetilde{Q}_x(u), \quad x \in \mathbb{X}^{\mathrm{g}}.$$

4

(i) Difference between $\mathscr{T}$ and $\mathscr{T}^{\mathrm{i}}$: Let $u^\star \in \operatorname{argmin}_u Q(x,u) \subseteq \mathbb{U}(x)$, so that $\mathscr{T} J(x) = Q(x,u^\star)$ and $\mathscr{T}^{\mathrm{i}} J(x) \leq \widetilde{Q}(x,u^\star)$. Also, let $z^\star \in \operatorname{argmin}_{z\in\mathbb{X}^{\mathrm{g}}} \big\|z - f(x,u^\star)\big\|_2$. Then,

$$\begin{aligned}\mathscr{T}^{\mathrm{i}} J(x) - \mathscr{T} J(x) &\leq \widetilde{Q}(x,u^\star) - Q(x,u^\star) \\ &= \widetilde{J^{\mathrm{d}}}\big(f(x,u^\star)\big) - \widetilde{J^{\mathrm{d}}}(z^\star) + J(z^\star) - J\big(f(x,u^\star)\big),\end{aligned}$$

where we used the assumption that $\widetilde{J^{\mathrm{d}}}(z^\star) = J(z^\star)$ for $z^\star \in \mathbb{X}^{\mathrm{g}}$. Hence,

$$\begin{aligned}\mathscr{T}^{\mathrm{i}} J(x) - \mathscr{T} J(x) &\leq \big[\mathrm{L}(J) + \mathrm{L}(\widetilde{J^{\mathrm{d}}})\big] \cdot \big\|z^\star - f(x,u^\star)\big\|_2 \\ &= \big[\mathrm{L}(J) + \mathrm{L}(\widetilde{J^{\mathrm{d}}})\big] \cdot \min_{z\in\mathbb{X}^{\mathrm{g}}} \big\|z - f(x,u^\star)\big\|_2 \\ &\leq \big[\mathrm{L}(J) + \mathrm{L}(\widetilde{J^{\mathrm{d}}})\big] \cdot \max_{z'\in\mathbb{X}} \min_{z\in\mathbb{X}^{\mathrm{g}}} \big\|z - z'\big\|_2 \\ &= \big[\mathrm{L}(J) + \mathrm{L}(\widetilde{J^{\mathrm{d}}})\big] \cdot \mathrm{d_H}(\mathbb{X},\mathbb{X}^{\mathrm{g}}) = e_1,\end{aligned}$$

where for the second inequality we used the fact that $f(x,u^\star) \in \mathbb{X}$. We can use the same line of arguments by defining $\tilde{u}^\star \in \operatorname{argmin}_u \widetilde{Q}(x,u)$, and $\tilde{z}^\star \in \operatorname{argmin}_{z\in\mathbb{X}^{\mathrm{g}}} \big\|z - f(x,\tilde{u}^\star)\big\|_2$ to show that $\mathscr{T}^{\mathrm{i}} J(x) - \mathscr{T} J(x) \leq e_1$. Combining these results, we have

$$-e_1 \leq \mathscr{T}^{\mathrm{i}} J(x) - \mathscr{T} J(x) \leq e_1. \tag{4.39}$$

(ii) Difference between $\mathscr{T}^{\mathrm{i}}$ and $\mathscr{T}^{\mathrm{d}}$: First note that, by construction, we have $\mathscr{T}^{\mathrm{i}} J(x) \leq \mathscr{T}^{\mathrm{d}} J^{\mathrm{d}}(x)$. Now, let $\tilde{u}^\star \in \operatorname{argmin}_u \widetilde{Q}(x,u) \subseteq \mathbb{U}(x)$, so that $\mathscr{T}^{\mathrm{i}} J(x) = \widetilde{Q}(x,\tilde{u}^\star)$. Also, let $\bar{u}^\star \in \operatorname{argmin}_{u\in\mathbb{U}^{\mathrm{d}}(x)} \big\|u - \tilde{u}^\star\big\|_2$, and note that $\mathscr{T}^{\mathrm{d}} J^{\mathrm{d}}(x) \leq \widetilde{Q}(x,\bar{u}^\star)$. Then, since $\widetilde{Q}$ is Lipschitz continuous, we have

$$\begin{aligned}0 \leq \mathscr{T}^{\mathrm{d}} J^{\mathrm{d}}(x) - \mathscr{T}^{\mathrm{i}} J(x) &\leq \widetilde{Q}(x,\bar{u}^\star) - \widetilde{Q}(x,\tilde{u}^\star) \leq \mathrm{L}(\widetilde{Q}_x) \cdot \big\|\bar{u}^\star - \tilde{u}^\star\big\|_2 \\ &\leq \big[\mathrm{L}(J) + \mathrm{L}(C)\big] \cdot \min_{u\in\mathbb{U}^{\mathrm{d}}(x)} \big\|u - \tilde{u}^\star\big\|_2 \\ &\leq \big[\mathrm{L}(J) + \mathrm{L}(C)\big] \cdot \max_{u'\in\mathbb{U}(x)} \min_{u\in\mathbb{U}^{\mathrm{d}}(x)} \big\|u - u'\big\|_2 \\ &= \big[\mathrm{L}(J) + \mathrm{L}(C)\big] \cdot \mathrm{d_H}\big(\mathbb{U}(x),\mathbb{U}^{\mathrm{d}}(x)\big) = e_2(x),\end{aligned}$$

Combining this last result with the inequality (4.39), we derive the reported bounds.

PROOF OF LEMMA 4.4.2

Using the definition of conjugate transform, we have

$$\begin{aligned}\widehat{\mathcal{T}}J(x) &= \max_{y\in\mathbb{R}^n}\min_{u,z\in\mathbb{R}^n}\{C(x,u)+J(z)+\langle y, f_{\mathrm{s}}(x)+f_{\mathrm{i}}(x)u-z\rangle\}\\ &= \max_{y}\Big\{\langle y, f_{\mathrm{s}}(x)\rangle - \max_{u}\big[\langle -f_{\mathrm{i}}(x)^\top y, u\rangle - C(x,u)\big] - \max_{z}\big[\langle y, z\rangle - J(z)\big]\Big\}\\ &= \max_{y}\big\{\langle y, f_{\mathrm{s}}(x)\rangle - C_x^*(-f_{\mathrm{i}}(x)^\top y) - J^*(y)\big\}\\ &= \max_{y}\big\{\langle y, f_{\mathrm{s}}(x)\rangle - \phi_x(y)\big\} = \phi_x^*\big(f_{\mathrm{s}}(x)\big).\end{aligned}$$

PROOF OF THEOREM 4.4.4

In what follows, we provide the time complexity of each line of the d-CDP operation in Algorithm 3. The LLT of line 4 requires $\mathcal{O}(X+Y)$ operations; see Remark 4.2.3. By Assumption 4.4.3, computing φ_x^{d} in line 6 has a complexity of $\mathcal{O}(Y)$. The minimization via enumeration in line 7 also has a complexity of $\mathcal{O}(Y)$. This, in turn, implies that the `for loop` over $x\in\mathbb{X}^{\mathrm{g}}$ requires $\mathcal{O}(XY)$ operations. Hence, the total complexity is of $\mathcal{O}(XY)$.

PROOF OF PROPOSITION 4.4.5

We can use the representation (4.22) and the definition (4.20) to obtain

$$\begin{aligned}\widehat{\mathcal{T}}^{\mathrm{d}}J^{\mathrm{d}}(x) &= \max_{y\in\mathbb{Y}^{\mathrm{g}}}\{\langle f_{\mathrm{s}}(x), y\rangle - \varphi_x^{\mathrm{d}}(y)\}\\ &= \max_{y\in\mathbb{Y}^{\mathrm{g}}}\Big\{\langle f_{\mathrm{s}}(x), y\rangle - C_x^*(-f_{\mathrm{i}}(x)^\top y) - J^{\mathrm{d}*\mathrm{d}}(y)\Big\}\\ &= \max_{y\in\mathbb{Y}^{\mathrm{g}}}\Big\{\langle f_{\mathrm{s}}(x), y\rangle - \max_{u\in\operatorname{dom}C(x,\cdot)}\big[\langle -f_{\mathrm{i}}(x)^\top y, u\rangle - C(x,u)\big] - J^{\mathrm{d}*\mathrm{d}}(y)\Big\}\\ &= \max_{y\in\mathbb{Y}^{\mathrm{g}}}\min_{u\in\operatorname{dom}C(x,\cdot)}\Big\{C(x,u)+\langle y, f(x,u)\rangle - J^{\mathrm{d}*\mathrm{d}}(y)\Big\},\end{aligned}$$

Since C is convex in u and the map f is affine in u, the objective function of this maximin problem is convex in u, with $\operatorname{dom}\big(C(x,\cdot)\big)$ being compact. Moreover, the objective is Ky Fan concave in y, which follows from the convexity of $J^{\mathrm{d}*}$. Then, by the Ky Fan's Minimax Theorem (see, e.g., [93, Thm. A]), we can swap the maximization and minimization operators to obtain

$$\begin{aligned}\widehat{\mathcal{T}}^{\mathrm{d}}J^{\mathrm{d}}(x) &= \min_{u\in\operatorname{dom}C(x,\cdot)}\max_{y\in\mathbb{Y}^{\mathrm{g}}}\Big\{C(x,u)+\langle y, f(x,u)\rangle - J^{\mathrm{d}*\mathrm{d}}(y)\Big\}\\ &= \min_{u}\Big\{C(x,u)+J^{\mathrm{d}*\mathrm{d}*}\big(f(x,u)\big)\Big\}.\end{aligned}$$

PROOF OF THEOREM 4.4.6

Fix $x\in\mathbb{X}^{\mathrm{g}}$ and observe that

$$\mathcal{T}J(x) - \widehat{\mathcal{T}}^{\mathrm{d}}J^{\mathrm{d}}(x) = \Big[\mathcal{T}J(x) - \widehat{\mathcal{T}}J(x)\Big] + \Big[\widehat{\mathcal{T}}J(x) - \widehat{\mathcal{T}}^{\mathrm{d}}J^{\mathrm{d}}(x)\Big]. \tag{4.40}$$

Let us first note that the convexity $C:\mathbb{X}\times\mathbb{U}\to\overline{\mathbb{R}}$ (in u) and $J:\mathbb{X}\to\mathbb{R}$ implies that the duality gap $\mathcal{T}J-\widehat{\mathcal{T}}J$ in (4.40) is zero. Indeed, following a similar argument as the one

provided in the proof of Proposition 4.4.5, and using Sion's Minimax Theorem (see, e.g., [94, Thm. 3]), we can show that

$$\widehat{\mathcal{T}} J(x) = \min_u \{C(x,u) + J^{**}(f(x,u))\}, \quad x \in \mathbb{X}.$$

Then, since J is a proper, closed, convex function, we have $J^{**} = J$, and hence $\widehat{\mathcal{T}} J = \mathcal{T} J$. We next consider the discretization error $\widehat{\mathcal{T}} J - \widehat{\mathcal{T}}^{\mathrm{d}} J^{\mathrm{d}}$ in (4.40). From (4.21b) and (4.22c), we have

$$\begin{aligned}\widehat{\mathcal{T}} J(x) - \widehat{\mathcal{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) &= \phi_x^*(f_{\mathrm{s}}(x)) - \varphi_x^{\mathrm{d}*}(f_{\mathrm{s}}(x)) \\ &= \left[\phi_x^*(f_{\mathrm{s}}(x)) - \phi_x^{\mathrm{d}*}(f_{\mathrm{s}}(x))\right] + \left[\phi_x^{\mathrm{d}*}(f_{\mathrm{s}}(x)) - \varphi_x^{\mathrm{d}*}(f_{\mathrm{s}}(x))\right], \end{aligned} \tag{4.41}$$

where $\phi_x^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ is the discretization of $\phi_x : \mathbb{R}^n \to \mathbb{R}$. For $\phi_x^* - \phi_x^{\mathrm{d}*}$ in (4.41), we can use Lemma 4.2.5, to write

$$\begin{aligned} 0 \le \phi_x^*(f_{\mathrm{s}}(x)) - \phi_x^{\mathrm{d}*}(f_{\mathrm{s}}(x)) &\le \widetilde{e}_1(f_{\mathrm{s}}(x), \phi_x, \mathbb{Y}^{\mathrm{g}}) \\ &= \min_{y \in \partial \phi_x^*(f_{\mathrm{s}}(x))} \left\{ [\, \|f_{\mathrm{s}}(x)\|_2 + \mathrm{L}(\phi_x; \{y\} \cup \mathbb{Y}^{\mathrm{g}})] \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}) \right\} \\ &\le \min_{y \in \partial \mathcal{T} J(x)} \left\{ [\, \|f_{\mathrm{s}}(x)\|_2 + \|f_{\mathrm{i}}(x)\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}}] \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}) \right\}, \end{aligned}$$

where we used the fact that $\phi_x^*(f_{\mathrm{s}}(\cdot)) = \widehat{\mathcal{T}} J(\cdot) = \mathcal{T} J(\cdot)$, and

$$\begin{aligned} \mathrm{L}(\phi_x(\cdot)) &\le \mathrm{L}(C_x^*(-f_{\mathrm{i}}(x)^\top \cdot)) + \mathrm{L}(J^*(\cdot)) \\ &\le \|f_{\mathrm{i}}(x)\|_2 \cdot \mathrm{L}(C_x^*) + \mathrm{L}(J^*) \\ &\le \|f_{\mathrm{i}}(x)\|_2 \cdot \Delta_{\mathrm{dom}(C(x,\cdot))} + \Delta_{\mathrm{dom}(J)} \\ &\le \|f_{\mathrm{i}}(x)\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}}. \end{aligned}$$

Hence,

$$\begin{aligned} 0 \le \phi_x^*(f_{\mathrm{s}}(x)) - \phi_x^{\mathrm{d}*}(f_{\mathrm{s}}(x)) &\le [\, \|f_{\mathrm{s}}(x)\|_2 + \|f_{\mathrm{i}}(x)\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}}] \cdot \min_{y \in \partial \mathcal{T} J(x)} \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}) \\ &= [\, \|f_{\mathrm{s}}(x)\|_2 + \|f_{\mathrm{i}}(x)\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}}] \cdot \mathrm{d}(\partial \mathcal{T} J(x), \mathbb{Y}^{\mathrm{g}}) \\ &= e_{\mathrm{y}}(x) \end{aligned} \tag{4.42}$$

For $\phi_x^{\mathrm{d}*} - \varphi_x^{\mathrm{d}*}$ in (4.41), first observe that for each $y \in \mathbb{Y}^{\mathrm{g}}$, we have (see (4.21a) and (4.22b), and recall that h^{d} is simply a sampled version of h)

$$\phi_x^{\mathrm{d}}(y) - \varphi_x^{\mathrm{d}}(y) = J^{*\mathrm{d}}(y) - J^{\mathrm{d}*\mathrm{d}}(y) = J^*(y) - J^{\mathrm{d}*}(y).$$

Moreover, we can use Lemma 4.2.5, and the fact that $\mathrm{dom}(J) = \mathbb{X}$ is compact, to write

$$\begin{aligned} 0 \le J^*(y) - J^{\mathrm{d}*}(y) &\le [\, \|y\|_2 + \mathrm{L}(J)] \cdot \mathrm{d}_{\mathrm{H}}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}) \\ &\le \max_{y \in \mathbb{Y}^{\mathrm{g}}} [\, \|y\|_2 + \mathrm{L}(J)] \cdot \mathrm{d}_{\mathrm{H}}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}) \\ &\le [\Delta_{\mathbb{Y}^{\mathrm{g}}} + \mathrm{L}(J)] \cdot \mathrm{d}_{\mathrm{H}}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}) = e_{\mathrm{x}}. \end{aligned}$$

That is,

$$0 \le \phi_x^{\mathrm{d}}(y) - \varphi_x^{\mathrm{d}}(y) \le e_{\mathrm{x}}, \quad \forall y \in \mathbb{Y}^{\mathrm{g}}.$$

Then, using the definition of discrete conjugate, we have

$$0 \le \varphi_x^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) - \phi_x^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) \le e_{\mathrm{x}}.$$

Combining the last inequality with the inequality (4.42) completes the proof.

PROOF OF THEOREM 4.5.3

In what follows, we provide the time complexity of each line of the modified d-CDP operation in Algorithm 4. The LLT of line 5 requires $\mathcal{O}(X+Y)$ operations; see Remark 4.2.3. By Assumption 4.5.2, computing φ^{d} in line 6 has a complexity of $\mathcal{O}(Y)$. The LLT of line 7 requires $\mathcal{O}(Y+Z)$ operations. The approximation of line 9 using LERP has a complexity of $\mathcal{O}(\log Z)$; see Remark 4.2.2. Hence, the `for loop` over $x \in \mathbb{X}^{\mathrm{g}}$ requires $\mathcal{O}(X \log Z) = \widetilde{\mathcal{O}}(X)$ operations. The time complexity of the whole operation can then be computed by adding all the aforementioned complexities.

PROOF OF THEOREM 4.5.4

Let $\widehat{\mathscr{T}}^{\mathrm{d}}$ denote the output of the implementation of the d-CDP operator (4.28). Note that the computation of the modified d-CDP operator $\widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}}$ (4.29)differs from that of the d-CDP operator $\widehat{\mathscr{T}}^{\mathrm{d}}$ (4.28) only in the last step. To see this, note that $\widehat{\mathscr{T}}^{\mathrm{d}}$ *exactly* computes $\varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big)$ for $x \in \mathbb{X}^{\mathrm{d}}$ (see Algorithm 3:7). However, in $\widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}}$, the *approximation* $\overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big)$ is used (see Algorithm 4:9), where the approximation uses LERP over the data points $\varphi^{\mathrm{d}*\mathrm{d}} : \mathbb{Z}^{\mathrm{g}} \to \mathbb{R}$. By Corollary 4.2.7, this leads to an over-approximation of $\varphi^{\mathrm{d}*}$, with the upper bound

$$e_{\mathrm{z}} = \Delta_{\mathbb{Y}^{\mathrm{g}}} \cdot \max_{x \in \mathbb{X}^{\mathrm{g}}} \mathrm{d}\big(f_{\mathrm{s}}(x), \mathbb{Z}^{\mathrm{g}}\big) = \Delta_{\mathbb{Y}^{\mathrm{g}}} \cdot \mathrm{d}_{\mathrm{H}}\big(f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\big).$$

Hence, compared to $\widehat{\mathscr{T}}^{\mathrm{d}}$, the operator $\widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}}$ is an over-approximation with the difference bounded by e_{z}, i.e.,

$$0 \le \widehat{\mathscr{T}}_{\mathrm{m}}^{\mathrm{d}} J^{\mathrm{d}}(x) - \widehat{\mathscr{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) \le e_{\mathrm{z}}, \quad \forall x \in \mathbb{X}^{\mathrm{g}}. \tag{4.43}$$

The result then follows from Theorem 4.4.6. Indeed, using the definition of $\widehat{\mathscr{T}}^{\mathrm{d}}$ (4.28), we can define

$$\begin{aligned}
&\widehat{\mathscr{I}}^{\mathrm{d}} J^{\mathrm{d}}(x) \coloneqq \widehat{\mathscr{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) - C_{\mathrm{s}}(x) = \varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big), \quad x \in \mathbb{X}^{\mathrm{g}}, \\
&\varphi^{\mathrm{d}}(y) \coloneqq C_{\mathrm{i}}^*(-B^\top y) + J^{\mathrm{d}*\mathrm{d}}(y), \quad y \in \mathbb{Y}^{\mathrm{g}}.
\end{aligned}$$

Similarly, using the DP operator (4.27), we can also define

$$\mathscr{I} J(x) \coloneqq \mathscr{T} J(x) - C_{\mathrm{s}}(x) = \min_u \big\{C_{\mathrm{i}}(u) + J\big(f(x,u)\big)\big\}.$$

Then, by Theorem 4.4.6, for all $x \in \mathbb{X}^{\mathrm{g}}$, it holds that

$$-e_{\mathrm{x}} \le \mathscr{I} J(x) - \widehat{\mathscr{I}}^{\mathrm{d}} J^{\mathrm{d}}(x) = \mathscr{T} J(x) - \widehat{\mathscr{T}}^{\mathrm{d}} J^{\mathrm{d}}(x) \le e_{\mathrm{y}}^m(x), \tag{4.44}$$

where e_{x} is given in (4.26), and

$$\begin{aligned}
e_{\mathrm{y}}^m(x) &= \big[\,\big\|f_{\mathrm{s}}(x)\big\|_2 + \|B\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}}\big] \cdot \mathrm{d}\big(\partial \mathscr{I} J(x), \mathbb{Y}^{\mathrm{g}}\big) \\
&= \big[\,\big\|f_{\mathrm{s}}(x)\big\|_2 + \|B\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}}\big] \cdot \mathrm{d}\big(\partial\big(\mathscr{T} J - C_{\mathrm{s}}\big)(x), \mathbb{Y}^{\mathrm{g}}\big).
\end{aligned}$$

Combining the inequalities (4.43) and (4.44) completes the proof.

5

INFINITE-HORIZON PROBLEM WITH STOCHASTIC DYNAMICS

Parts of this chapter have been published in Advances in Neural Information Processing Systems **34** (2021) [95].

In this chapter, we discuss the extensions of the proposed conjugate value iteration (ConjVI) algorithm introduced in Chapter 4 for solving the optimal control problem of discrete-time systems, with continuous state-input space. We focus on the extension of the ConjVI Algorithm 4 based on the *modified* d-CDP operator (4.29). However, we note that the same extensions can be applied on the ConjVI Algorithm 3 that utilizes the d-CDP operator (4.22). The extensions are three-fold: We consider *infinite-horizon, discounted cost* problems with *stochastic dynamics*, while *computing the the conjugate of input cost numerically*. We also note that these extensions can be applied independently of one another, corresponding to the problem at hand.

The chapter is organized as follows.[1] We provide the problem statement with the added extensions and its standard solution via the VI algorithm (in primal domain) in Section 5.1. In Section 5.2, we present our main results: We begin with presenting the class of problems that are of interest and then introduce the corresponding ConjVI algorithm and its numerical implementation. The theoretical results on the convergence, complexity, and error of the proposed algorithm along with the guidelines on the construction of dual grids are also provided in this section. In Section 5.3, we compare the performance of the ConjVI algorithm with that of the benchmark VI algorithm through three numerical examples. The technical proofs are provided in Section 5.4. We follow the same notational conventions presented in Section 4.2.1. To facilitate the application of the proposed algorithm, we provide a MATLAB package [96]. In particular, the provided numerical examples in Section 5.3 are also included in the package.

5.1. VI IN PRIMAL DOMAIN

In this chapter, we are concerned with the infinite-horizon, discounted cost, optimal control problems of the form

$$J_\star(x) = \min\ \mathbb{E}_{w_t}\left[\sum_{t=0}^{\infty} \gamma^t C(x_t,u_t)\middle| x_0 = x\right]$$
$$\text{s.t.}\quad x_{t+1} = g(x_t,u_t,w_t),\ x_t \in \mathbb{X},\ u_t \in \mathbb{U},\ w_t \sim \mathbb{P}(\mathbb{W}),\quad \forall t \in \{0,1,\ldots\},$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, and $w_t \in \mathbb{R}^l$ are the state, input and disturbance variables at time t, respectively; $\gamma \in (0,1)$ is the discount factor; $g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^n$ describes the stochastic dynamics; $\mathbb{P}(\cdot)$ is the distribution of the disturbance over the support $\mathbb{W} \subset \mathbb{R}^l$, and $\mathbb{E}_w[\cdot]$ denotes the expectation with respect to the random variable w. $C : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is again the stage cost, and $\mathbb{X} \subset \mathbb{R}^n$ and $\mathbb{U} \subset \mathbb{R}^m$ again describe the state and input constraints, respectively. Throughout this chapter we assume that the problem data satisfy the following properties:

Assumption 5.1.1 (Problem data)**.** [2] *The problem data has the following properties:*

(i) **Disturbance.** *The disturbance w has a finite support $\mathbb{W}^{\mathrm{d}} \subset \mathbb{R}^l$ with a given probability mass function (p.m.f.) $p : \mathbb{W}^{\mathrm{d}} \to [0,1]$.*[3]

[1] For completeness and also for better readability, we try to keep references to Chapter 4 at a minimum and provide all the necessary details in this chapter.

[2] This is an extension of Assumption 4.3.1 for *stochastic* dynamics.

[3] We consider this assumption to simplify the exposition and explicitly include the computational cost of the

(ii) **Dynamics.** *The mapping* $(x,u)\mapsto g(x,u,w)$ *is locally Lipschitz continuous for each* $w\in\mathbb{W}^{\mathrm{d}}$.

(iii) **Constraints.** *The sets* $\mathbb{X}\subset\mathbb{R}^n$ *and* $\mathbb{U}\subset\mathbb{R}^m$ *are compact. Moreover, the set of admissible inputs* $\mathbb{U}(x):=\{u\in\mathbb{U}: g(x,u,w)\in\mathbb{X},\ \forall w\in\mathbb{W}^{\mathrm{d}}\}$ *is nonempty for all* $x\in\mathbb{X}$.

(iv) **Cost function.** $C:\mathbb{X}\times\mathbb{U}\to\mathbb{R}$ *is Lipschitz continuous.*

Assuming the stage cost C is bounded (which is true under Assumption 5.1.1), the optimal value function solves the DP equation $J_\star=\mathcal{T}J_\star$, where $\mathcal{T}$[4] is the DP operator (C and J are extended to infinity outside their effective domains) [97, Prop. 1.2.2]

$$\mathcal{T}J(x):=\min_u\big\{C(x,u)+\gamma\cdot\mathbb{E}_w J\big(g(x,u,w)\big)\big\},\quad\forall x\in\mathbb{X}. \tag{5.1}$$

The operator $\mathcal{T}$ is γ-contractive in the infinity-norm, i.e., $\|\mathcal{T}J_1-\mathcal{T}J_2\|_\infty\le\gamma\|J_1-J_2\|_\infty$ [97, Prop. 1.2.4]. This property means that the VI algorithm $J_{k+1}=\mathcal{T}J_k$ converges to $J_\star$ as $k\to\infty$, for arbitrary initialization J_0. Moreover, assuming that the composition $J\circ g$ (for each w) and the cost C are jointly convex in the state and input variables, $\mathcal{T}$ also preserves convexity [70, Prop. 3.3.1]. Let us note that the properties laid out in Assumption 5.1.1 imply that the set of admissible inputs $\mathbb{U}(x)$ is a compact set for each $x\in\mathbb{X}$. This, in turn, implies that the optimal value in (5.1) is achieved if $J:\mathbb{X}\to\mathbb{R}$ is also assumed to be continuous (which also holds true under Assumption 5.1.1).

Once again, note that the optimization problem (5.1) is infinite-dimensional for the continuous state space $\mathbb{X}$. This renders the exact implementation of VI impossible in most cases. Here, again, we use a sample-based approach, accompanied by a function approximation scheme. To be precise, for a finite grid-like subset $\mathbb{X}^{\mathrm{g}}$ of $\mathbb{X}$, at each iteration $k\ge0$, we take the discrete function $J_k^{\mathrm{d}}:\mathbb{X}^{\mathrm{g}}\to\mathbb{R}$ as the input, and compute the discrete function $J_{k+1}^{\mathrm{d}}=\big[\mathcal{T}\widetilde{J_k^{\mathrm{d}}}\big]^{\mathrm{d}}:\mathbb{X}^{\mathrm{g}}\to\mathbb{R}$, where $\widetilde{J_k^{\mathrm{d}}}:\mathbb{X}\to\mathbb{R}$ is an extension of J_k^{d}; see Section 4.2.2. Moreover, for solving the minimization problem in (5.1) over the control input (for each $x\in\mathbb{X}^{\mathrm{g}}$), we again use the approximation that involves enumeration over a discretization $\mathbb{U}^{\mathrm{d}}\subset\mathbb{U}$ of the inputs space. In the numerical implementation of VI for *stochastic* dynamics, we need to address another issue, namely, computing the expectation in (5.1). Under Assumption 5.1.1-(i), this operation simplifies to

$$\mathbb{E}_w J\big(g(x,u,w)\big)=\sum_{w\in\mathbb{W}^{\mathrm{d}}}p(w)\cdot J\big(g(x,u,w)\big).$$

With these approximations, we end up with the approximate VI algorithm $J_{k+1}^{\mathrm{d}}=\mathcal{T}^{\mathrm{d}}J_k^{\mathrm{d}}$, characterized by the d-DP operator

$$\mathcal{T}^{\mathrm{d}}J^{\mathrm{d}}(x):=\min_{u\in\mathbb{U}^{\mathrm{d}}}\Big\{C(x,u)+\gamma\cdot\sum_{w\in\mathbb{W}^{\mathrm{d}}}p(w)\cdot\widetilde{J^{\mathrm{d}}}\big(g(x,u,w)\big)\Big\},\quad\forall x\in\mathbb{X}^{\mathrm{g}}. \tag{5.2}$$

expectation operation with respect to the disturbance. Indeed, $\mathbb{W}^{\mathrm{d}}$ can be considered as an approximation of the true support $\mathbb{W}$ of the disturbance. Moreover, one can consider other approximation schemes, such as Monte Carlo simulation, for the expectation operation.

[4] We are using the same notations $\mathcal{T}$ for the *discounted, stochastic* DP operation in (5.1), $\mathcal{T}^{\mathrm{d}}$ for its discretization in (5.2), and $\widehat{\mathcal{T}}$ for its dualization in (5.3).

The convergence of approximate VI described above depends on the properties of the extension operation $\widetilde{[\cdot]}$. In particular, if $\widetilde{[\cdot]}$ is non-expansive (in the infinity-norm), then $\mathcal{T}^{\mathrm{d}}$ is also γ-contractive. The error of this approximation ($\lim \left\| J_k^{\mathrm{d}} - J_\star^{\mathrm{d}} \right\|_\infty$) also depends on the extension operation $\widetilde{[\cdot]}$ and its representative power. We refer the interested reader to [69, 97, 98] for detailed discussions on the convergence and error of different approximation schemes for VI.

The d-DP operator (5.2) and the corresponding approximate VI algorithm are again our benchmark for evaluating the performance of the ConjVI algorithm to be introduced. In this regard, we note that the time complexity of the d-DP operation (5.2) is of $\mathcal{O}\left(XUWE\right)$, where E is used to denote the time complexity of a single evaluation of the extension operator $\widetilde{[\cdot]}$; see Remark 4.2.1.

5.2. VI IN CONJUGATE DOMAIN

In this section, we discuss the extension of the ConjVI algorithm for the problem class corresponding to Setting 4.5.1 presented in Chapter 4. In particular, we present the numerical scheme for implementing the proposed algorithm and analyze its convergence, complexity, and error. The problem class of interest is as follows:

Setting 5.2.1. [5] ***(i)*** *The disturbance is additive, i.e., $g(x,u,w) = f(x,u) + w$, where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ describes the deterministic dynamics.* ***(ii)*** *The deterministic dynamics are input-affine with state-independent input dynamics, i.e., $f(x,u) = f_{\mathrm{s}}(x) + B \cdot u$, where $f_{\mathrm{s}} : \mathbb{R}^n \to \mathbb{R}^n$ and $B \in \mathbb{R}^{n\times m}$.* ***(iii)*** *The stage cost is separable in state and input, i.e., $C(x,u) = C_{\mathrm{s}}(x) + C_{\mathrm{i}}(u)$, where $C_{\mathrm{s}} : \mathbb{X} \to \mathbb{R}$ and $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ are the state and input costs, respectively.*

5.2.1. EXTENSION OF CDP OPERATOR

For the problem class of Setting 5.2.1, consider the following reformulation of the optimization problem (5.1) for a fixed $x \in \mathbb{X}$

$$\mathcal{T} J(x) = C_{\mathrm{s}}(x) + \min_{u,z} \left\{ C_{\mathrm{i}}(u) + \gamma \cdot \mathbb{E}_w J(z+w) : z = f(x,u) \right\},$$

where we used additivity of disturbance and separability of stage cost. The corresponding dual problem, i.e., the CDP operator, then reads as

$$\widehat{\mathcal{T}} J(x) := C_{\mathrm{s}}(x) + \max_{y} \min_{u,z} \left\{ C_{\mathrm{i}}(u) + \gamma \cdot \mathbb{E}_w J(z+w) + \left\langle y, f(x,u) - z \right\rangle \right\}, \tag{5.3}$$

where $y \in \mathbb{R}^n$ is the dual variable corresponding to the equality constraint. For input-affine dynamics, we then have (the derivation is similar to the one provided in the proof of Lemma 4.4.2)

$$\epsilon(x) := \gamma \cdot \mathbb{E}_w J(x+w), \qquad x \in \mathbb{X}, \tag{5.4a}$$

$$\phi(y) := C_{\mathrm{i}}^*(-B^\top y) + \epsilon^*(y), \qquad y \in \mathbb{R}^n, \tag{5.4b}$$

$$\widehat{\mathcal{T}} J(x) = C_{\mathrm{s}}(x) + \phi^*\left(f_{\mathrm{s}}(x)\right), \qquad x \in \mathbb{X}. \tag{5.4c}$$

We next provide an alternative representation of the CDP operator that captures the essence of this operation.

[5] This is an extension of Setting 4.5.1 for *stochastic* dynamics.

Proposition 5.2.2 (CDP reformulation). *The CDP operator $\widehat{\mathcal{T}}$ equivalently reads as*

$$\widehat{\mathcal{T}} J(x) = C_{\mathrm{s}}(x) + \min_{u} \left\{ C_{\mathrm{i}}^{**}(u) + \gamma \cdot [\mathbb{E}_w J(\cdot + w)]^{**}\big(f(x,u)\big) \right\}. \tag{5.5}$$

This result implies that the indirect path through the conjugate domain essentially involves substituting the input cost and (expectation of the) value function by their bi-conjugates. In particular, it points to a sufficient condition for zero duality gap.

Corollary 5.2.3 (Equivalence of $\mathcal{T}$ and $\widehat{\mathcal{T}}$). *If the functions $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ and $J : \mathbb{X} \to \mathbb{R}$ are convex, then $\widehat{\mathcal{T}} J = \mathcal{T} J$.*

Hence, $\widehat{\mathcal{T}}$ has the same properties as $\mathcal{T}$ if C_{i} and J are convex. More importantly, if $\mathcal{T}$ and $\widehat{\mathcal{T}}$ preserve convexity, then the *conjugate* VI (ConjVI) algorithm $J_{k+1} = \widehat{\mathcal{T}} J_k$ also converges to the optimal value function $J_\star$, with arbitrary convex initialization J_0. For convexity to be preserved, however, we need more assumptions: First, the state cost $C_{\mathrm{s}} : \mathbb{X} \to \mathbb{R}$ needs to be also convex. Then, for $\widehat{\mathcal{T}} J$ to be convex, a sufficient condition is convexity of $J \circ f$ (jointly in x and u), given that J is convex. The following assumption provides the sufficient conditions for equivalence of the VI and ConjVI algorithms.

Assumption 5.2.4 (Convexity). ***(i)*** *The sets $\mathbb{X} \subset \mathbb{R}^n$ and $\mathbb{U} \subset \mathbb{R}^m$ are convex.* ***(ii)*** *The costs $C_{\mathrm{s}} : \mathbb{X} \to \mathbb{R}$ and $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ are convex.* ***(iii)*** *The deterministic dynamics $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is such that given a convex function $J : \mathbb{X} \to R$, the composition $J \circ f$ is jointly convex in the state and input variables.*

We note that the last condition in the preceding assumption usually does not hold for nonlinear dynamics, however, for $f_{\mathrm{s}}(x) = Ax$ with $A \in \mathbb{R}^{n \times n}$, this is indeed the case for problems satisfying Assumptions 5.1.1 and 5.2.4 and the properties of Setting 5.2.1 [99]. Note that, if convexity is not preserved, then the alternative path suffers from duality gap in the sense that in each iteration it uses the *convex envelope* of (the expectation of) the output of the previous iteration.

5.2.2. EXTENDED d-CDP OPERATOR

In this section, we discuss the numerical implementation of the CDP operator (5.4) that again uses a sample-based approach by solving (5.4) for a finite set $\mathbb{X}^{\mathrm{g}} \subset \mathbb{X}$. The corresponding ConjVI algorithm then involves the consecutive applications of this operator, until some termination condition is satisfied. Algorithm 5 provides the pseudo-code of this procedure, in which the iteration terminates when the difference between two consecutive discrete value functions (in the infinity-norm) is less than a given constant $e_{\mathrm{t}} > 0$; see Algorithm 5:7. The main steps of Algorithm 5 are as follows:

- For the expectation operation in (5.4a), by Assumption 5.1.1-(i), we again have $\mathbb{E}_w J(\cdot + w) = \sum_{w \in \mathbb{W}^{\mathrm{d}}} p(w) \cdot J(\cdot + w)$. Hence, we need to pass the value function $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ through the "scaled expection filter" to obtain $\varepsilon^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \overline{\mathbb{R}}$ in (5.6a) as an approximation of ϵ in (5.4a). Notice that here we are using an extension $\widetilde{J^{\mathrm{d}}} : \mathbb{X} \to \mathbb{R}$ of J^{d} (recall that we only have access to the discrete value function J^{d}).
- To compute ϕ in (5.4b), we need access to two conjugate functions:

– For ϵ^*, we use the approximation $\epsilon^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ in (5.6b), by applying LLT to the data points $\epsilon^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \overline{\mathbb{R}}$, for a properly chosen dual state grid $\mathbb{Y}^{\mathrm{g}} \subset \mathbb{R}^n$.

– Having fixed the dual state grid $\mathbb{Y}^{\mathrm{g}} \subset \mathbb{R}^n$, we need the value of C_{i}^* at $(-B^\top y)$ for $y \in \mathbb{Y}^{\mathrm{g}}$.[6] For that, we use *approximate discrete conjugation*: For a properly chosen dual input grid $\mathbb{V}^{\mathrm{g}} \subset \mathbb{R}^m$, we first employ LLT to compute $C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}} : \mathbb{V}^{\mathrm{g}} \to \mathbb{R}$ in (5.6c) using the data points $C_{\mathrm{i}}^{\mathrm{d}} : \mathbb{U}^{\mathrm{g}} \to \mathbb{R}$, where $\mathbb{U}^{\mathrm{g}}$ is a grid-like finite subset of $\mathbb{U}$.[7] We then use the LERP extension $\overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}$ of $C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}$ to approximate $C_{\mathrm{i}}^{\mathrm{d}*}$ at the required points $(-B^\top y)$ for each $y \in \mathbb{Y}^{\mathrm{g}}$.

With these conjugate functions at hand, we can now compute $\varphi^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ in (5.6d), as an approximation of ϕ in (5.4b).

- To be able to compute the output according to (5.4c), we need to perform another conjugate transform. In particular, we need the value of ϕ^* at $f_{\mathrm{s}}(x)$ for $x \in \mathbb{X}^{\mathrm{g}}$. Here, we again use *approximate discrete conjugation*: We first compute $\varphi^{\mathrm{d}*\mathrm{d}} : \mathbb{Z}^{\mathrm{g}} \to \mathbb{R}$ in (5.6e), by applying LLT to the data points $\varphi^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$, for a properly chosen grid $\mathbb{Z}^{\mathrm{g}} \subset \mathbb{R}^n$. Then, we use the LERP extension $\overline{\varphi^{\mathrm{d}*\mathrm{d}}}$ of $\varphi^{\mathrm{d}*\mathrm{d}}$ to approximate $\varphi^{\mathrm{d}*}$ at the required point $f_{\mathrm{s}}(x)$ for each $x \in \mathbb{X}^{\mathrm{g}}$. Finally, we compute $\widehat{\mathcal{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}}$ in (5.6f) as an approximation of $\widehat{\mathcal{T}} J$ in (5.4c).

With these approximations, we can introduce the *extended* d-CDP operator as follows[8]

$$\epsilon^{\mathrm{d}}(x) := \gamma \cdot \sum_{w \in \mathbb{W}^{\mathrm{d}}} p(w) \cdot \widetilde{J^{\mathrm{d}}}(x + w), \qquad x \in \mathbb{X}^{\mathrm{g}}, \tag{5.6a}$$

$$\epsilon^{\mathrm{d}*\mathrm{d}}(y) = \max_{x \in \mathbb{X}^{\mathrm{g}}} \left\{ \langle x, y \rangle - \epsilon^{\mathrm{d}}(x) \right\}, \qquad y \in \mathbb{Y}^{\mathrm{g}}, \tag{5.6b}$$

$$C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}(v) = \max_{u \in \mathbb{U}^{\mathrm{g}}} \left\{ \langle u, v \rangle - C_{\mathrm{i}}^{\mathrm{d}}(u) \right\}, \qquad v \in \mathbb{V}^{\mathrm{g}}, \tag{5.6c}$$

$$\varphi^{\mathrm{d}}(y) := \overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}(-B^\top y) + \epsilon^{\mathrm{d}*\mathrm{d}}(y), \qquad y \in \mathbb{Y}^{\mathrm{g}}, \tag{5.6d}$$

$$\varphi^{\mathrm{d}*\mathrm{d}}(z) = \max_{y \in \mathbb{Y}^{\mathrm{g}}} \left\{ \langle y, z \rangle - \varphi^{\mathrm{d}}(y) \right\}, \qquad z \in \mathbb{Z}^{\mathrm{g}}, \tag{5.6e}$$

$$\widehat{\mathcal{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}}(x) := C_{\mathrm{s}}(x) + \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big), \qquad x \in \mathbb{X}^{\mathrm{g}}. \tag{5.6f}$$

We will discuss the proper construction of the grids $\mathbb{Y}^{\mathrm{g}}$, $\mathbb{V}^{\mathrm{g}}$, and $\mathbb{Z}^{\mathrm{g}}$ in Section 5.2.4.

5.2.3. ANALYSIS OF CONJVI ALGORITHM

We now provide our main theoretical results concerning the convergence, complexity, and error of the proposed algorithm. Let us begin by presenting the assumptions to be called in this subsection.

[6] Note that, unlike Chapter 4, we are not assuming that C_{i}^* is analytically available; see Assumption 4.5.2.

[7] Note that the set $\mathbb{U}^{\mathrm{g}}$ employed here for discrete conjugation need not be the same as the set $\mathbb{U}^{\mathrm{d}}$ used in the d-DP operator (5.2).

[8] This is the extension of the modified d-CDP operator (4.29) that accounts for discounted cost, additive stochasticity in the dynamics, and numerical approximation of the conjugate of input cost.

Algorithm 5 ConjVI algorithm for Setting 5.2.1

Input: dynamics $f_s : \mathbb{R}^n \to \mathbb{R}^n$, $B \in \mathbb{R}^{n\times m}$; discrete state space $\mathbb{X}^g \subset \mathbb{X}$; discrete input space $\mathbb{U}^g \subset \mathbb{U}$; discrete state cost $C_s^d : \mathbb{X}^g \to \mathbb{R}$; discrete input cost $C_i^d : \mathbb{U}^g \to \mathbb{R}$; discrete disturbance space $\mathbb{W}^d$ and its p.m.f. $p : \mathbb{W}^d \to [0,1]$; discount factor γ; termination bound e_t.
Output: discrete value function $\widehat{J}^d : \mathbb{X}^g \to \mathbb{R}$.

initialization:
construct the grid $\mathbb{V}^g$;
use LLT to compute $C_i^{d*d} : \mathbb{V}^g \to \mathbb{R}$ from $C_i^d : \mathbb{U}^g \to \mathbb{R}$;
construct the grid $\mathbb{Z}^g$;
construct the grid $\mathbb{Y}^g$;
$J^d(x) \leftarrow 0$ for $x \in \mathbb{X}^g$;
$J_+^d(x) \leftarrow C_s^d(x) - \min C_i^d$ for $x \in \mathbb{X}^g$;
iteration:
while $\left\| J_+^d - J^d \right\|_\infty \geq e_t$ **do**
  $J^d \leftarrow J_+^d$;
  d-CDP operation:
  $\varepsilon^d(x) \leftarrow \gamma \cdot \sum_{w\in\mathbb{W}^d} p(w) \cdot \widetilde{J^d}(x+w)$ for $x \in \mathbb{X}^g$;
  use LLT to compute $\varepsilon^{d*d} : \mathbb{Y}^g \to \overline{\mathbb{R}}$ from $\varepsilon^d : \mathbb{X}^g \to \mathbb{R}$;
  for each $y \in \mathbb{Y}^g$ **do**
    use LERP to compute $\overline{C_i^{d*d}}(-B^\top y)$ from $C_i^{d*d} : \mathbb{V}^g \to \mathbb{R}$;
    $\varphi^d(y) \leftarrow \overline{C_i^{d*d}}(-B^\top y) + \varepsilon^{d*d}(y)$;
  end for
  use LLT to compute $\varphi^{d*d} : \mathbb{Z}^g \to \mathbb{R}$ from $\varphi^d : \mathbb{Y}^g \to \mathbb{R}$;
  for each $x \in \mathbb{X}^g$ **do**
    use LERP to compute $\overline{\varphi^{d*d}}\big(f_s(x)\big)$ from $\varphi^{d*d} : \mathbb{Z}^g \to \mathbb{R}$;
    $J_+^d(x) \leftarrow C_s(x) + \overline{\varphi^{d*d}}\big(f_s(x)\big)$;
  end for
end while
output $\widehat{J}^d \leftarrow J_+^d$.

Assumption 5.2.5 (Grids)**.** *Consider the following properties for the grids in Algorithm 5 (consult the Notations in Section 4.2.1):*

(i) The grid $\mathbb{V}^g$ is constructed such that $\mathrm{co}(\mathbb{V}^g_{\mathrm{sub}}) \supseteq \mathbb{L}(C_i^d)$.

(ii) The grid $\mathbb{Z}^g$ is constructed such that $\mathrm{co}(\mathbb{Z}^g) \supseteq f_s\big(\mathbb{X}^g\big)$.

(iii) The construction of $\mathbb{Y}^g$, $\mathbb{V}^g$, and $\mathbb{Z}^g$ requires at most $\mathcal{O}(X+U)$ operations. The cardinality of the grids $\mathbb{Y}^g$ and $\mathbb{Z}^g$ (respectively, $\mathbb{V}^g$) in each dimension is the same as that of $\mathbb{X}^g$ (respectively, $\mathbb{U}^g$) in that dimension so that $Y, Z = X$ and $V = U$.[9]

Assumption 5.2.6 (Extension operator)**.** *Consider the following properties for the extension operator $[\![\cdot]\!]$ in (5.6a):*

[9] The second part of this assumption corresponds to Assumption 4.2.4.

(i) $\widetilde{[\cdot]}$ *is non-expansive with respect to the infinity norm, i.e,* $\|\widetilde{J_1^{\mathrm{d}}} - \widetilde{J_2^{\mathrm{d}}}\|_\infty \le \|J_1^{\mathrm{d}} - J_2^{\mathrm{d}}\|_\infty$ *for two discrete functions* $J_i^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ *and their extensions* $\widetilde{J_i^{\mathrm{d}}} : \mathbb{X} \to \mathbb{R}$ $(i = 1, 2)$.

(ii) Given a function $J : \mathbb{X} \to \mathbb{R}$ *and its discretization* $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$*, the error of* $\widetilde{[\cdot]}$ *is uniformly bounded, i.e.,* $\|J - \widetilde{J^{\mathrm{d}}}\|_\infty \le e_{\mathrm{e}}$ *for some constant* $e_{\mathrm{e}} \ge 0$.

Our first result concerns the contractiveness of the d-CDP operator.

Theorem 5.2.7 (Convergence)**.** *Let Assumptions 5.2.5-(ii) and 5.2.6-(i) hold. Then, the d-CDP operator (5.6) is* γ*-contractive with respect to the infinity-norm.*

The preceding theorem implies that the approximate ConjVI Algorithm 5 is indeed convergent given that the required conditions are satisfied. In particular, for *deterministic* dynamics, $\mathrm{co}(\mathbb{Z}^{\mathrm{g}}) \supseteq f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$ is sufficient for Algorithm 5 to be convergent. We next consider the time complexity of our algorithm.

Theorem 5.2.8 (Complexity)**.** *Let Assumption 5.2.5-(iii) hold. Then, the time complexity of initialization and each iteration in Algorithm 5 are of* $\mathcal{O}(X + U)$ *and* $\widetilde{\mathcal{O}}(XWE)$*, respectively, where* E *denotes the complexity of each evaluation of the operator* $\widetilde{[\cdot]}$ *in (5.6a).*

The requirements of Assumption 5.2.5-(iii) will be discussed in Section 5.2.4. Recall that each iteration of VI (in primal domain) has a complexity of $\mathcal{O}(XUWE)$, where E denotes the complexity of the extension operation in (5.2). That is, ConjVI reduces the quadratic complexity of VI to a linear one by replacing the minimization operation in the primal domain with a simple addition in the conjugate domain. We note however that ConjVI, like our benchmark VI and other approximation schemes that utilize discretization of the continuous state and input spaces, still suffers from the so-called "curse of dimensionality." This is because the sizes X and U of the discretizations increase exponentially with the dimensions n and m of the corresponding spaces. However, for ConjVI, this exponential increase is of rate $\max\{m, n\}$, compared to the rate $m + n$ for standard VI.

Theorem 5.2.9 (Error)**.** *Let Assumptions 5.2.4, 5.2.5-(i)&(ii), and 5.2.6-(i) hold. Consider the true optimal value function* $J_\star = \mathcal{T} J_\star : \mathbb{X} \to \mathbb{R}$ *and its discretization* $J_\star^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$*, and let Assumption 5.2.6-(ii) hold for* $J_\star$*. Also, let* $\widehat{J}^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ *be the output of Algorithm 5 with grids* $\mathbb{Y}^{\mathrm{g}}$*,* $\mathbb{V}^{\mathrm{g}}$*, and* $\mathbb{Z}^{\mathrm{g}}$*. Then,*

$$\|\widehat{J}^{\mathrm{d}} - J_\star^{\mathrm{d}}\|_\infty \le \frac{\gamma(e_{\mathrm{e}} + e_{\mathrm{t}}) + e_{\mathrm{d}}}{1 - \gamma}, \tag{5.7}$$

where $e_{\mathrm{d}} = e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}} + e_{\mathrm{y}} + e_{\mathrm{z}}$*, and*

$$e_{\mathrm{u}} = c_{\mathrm{u}} \cdot \mathrm{d_H}(\mathbb{U}, \mathbb{U}^{\mathrm{g}}), \tag{5.8a}$$
$$e_{\mathrm{v}} = c_{\mathrm{v}} \cdot \mathrm{d_H}\left(\mathrm{co}(\mathbb{V}^{\mathrm{g}}), \mathbb{V}^{\mathrm{g}}\right), \tag{5.8b}$$
$$e_{\mathrm{x}} = c_{\mathrm{x}} \cdot \mathrm{d_H}\left(\mathbb{X}, \mathbb{X}^{\mathrm{g}}\right), \tag{5.8c}$$
$$e_{\mathrm{y}} = c_{\mathrm{y}} \cdot \max_{x \in \mathbb{X}^{\mathrm{g}}} \mathrm{d}\left(\partial(J_\star - C_{\mathrm{s}})(x), \mathbb{Y}^{\mathrm{g}}\right), \tag{5.8d}$$
$$e_{\mathrm{z}} = c_{\mathrm{z}} \cdot \mathrm{d_H}\left(f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\right), \tag{5.8e}$$

with constants $c_{\mathrm{u}}, c_{\mathrm{v}}, c_{\mathrm{x}}, c_{\mathrm{y}}, c_{\mathrm{z}} > 0$ *depending on the problem data.*

Let us first note that Assumption 5.2.4 implies that the DP and CDP operators preserve convexity, and they both have the true optimal value function $J_\star$ as their fixed point (i.e., the duality gap is zero). Otherwise, the proposed scheme can suffer from large errors due to dualization. The remaining sources of error are captured by the three error terms in (5.7): (i) e_{e} is due to the approximation of the value function using the extension operator $\widetilde{[\cdot]}$; (ii) e_{t} corresponds to the termination of the algorithm after a finite number of iterations; (iii) e_{d} captures the error due to the discretization of the primal and dual state and input domains. In particular, Assumptions 5.2.5-(i)&(ii) on the grids $\mathbb{V}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$ are required for bounding the error of approximate discrete conjugations using LERP in (5.6d) and (5.6f); see the proof of Lemmas 5.4.6 and 5.4.8.

5.2.4. CONSTRUCTION OF THE GRIDS

In this subsection, we provide specific guidelines for the construction of the grids $\mathbb{Y}^{\mathrm{g}}$, $\mathbb{V}^{\mathrm{g}}$ and $\mathbb{Z}^{\mathrm{g}}$. The presented guidelines aim to minimize the error terms in (5.8) while taking into account the properties laid out in Assumption 5.2.5. In particular, the schemes described below satisfy the requirements of Assumption 5.2.5-(iii).

CONSTRUCTION OF $\mathbb{V}^{\mathrm{g}}$

Assumption 5.2.5-(i) and the error term e_{v} (5.8b) suggest that we find the smallest dual input grid $\mathbb{V}^{\mathrm{g}}$ such that $\mathrm{co}(\mathbb{V}^{\mathrm{g}}_{\mathrm{sub}}) \supseteq \mathbb{L}(C^{\mathrm{d}}_{\mathrm{i}})$.[10] This latter condition essentially means that $\mathbb{V}^{\mathrm{g}}$ must "more than cover the range of slopes" of the function $C^{\mathrm{d}}_{\mathrm{i}}$; Hence, we need to compute/approximate $\mathrm{L}^{\pm}_j(C^{\mathrm{d}}_{\mathrm{i}})$ for $j = 1,\dots,m$. A conservative approximation is $\mathrm{L}^{-}_j(C_{\mathrm{i}}) = \min \partial C_{\mathrm{i}}/\partial u_j$ and $\mathrm{L}^{+}_j(C_{\mathrm{i}}) = \max \partial C_{\mathrm{i}}/\partial u_j$, assuming C_{i} is differentiable.[11] Having $\mathrm{L}^{\pm}_j(C^{\mathrm{d}}_{\mathrm{i}})$ at our disposal, we can then construct $\mathbb{V}^{\mathrm{g}}_{\mathrm{sub}} = \Pi_{j=1}^{m} \mathbb{V}^{\mathrm{g}}_{\mathrm{sub}\,j}$ such that, in each dimension j, $\mathbb{V}^{\mathrm{g}}_{\mathrm{sub}\,j}$ is uniform with the same cardinality as $\mathbb{U}^{\mathrm{g}}_j$ and $\mathrm{co}(\mathbb{V}^{\mathrm{g}}_{\mathrm{sub}\,j}) = \left[\mathrm{L}^{-}_j(C^{\mathrm{d}}_{\mathrm{i}}), \mathrm{L}^{+}_j(C^{\mathrm{d}}_{\mathrm{i}})\right]$. Finally, we construct $\mathbb{V}^{\mathrm{g}}$ by extending $\mathbb{V}^{\mathrm{g}}_{\mathrm{sub}}$ uniformly in each dimension (by adding a smaller and a larger element to $\mathbb{V}^{\mathrm{g}}_{\mathrm{sub}}$ in each dimension). This numerical scheme has a time complexity of $\mathcal{O}(1)$ assuming, we have access to $\mathrm{L}^{\pm}_j(C^{\mathrm{d}}_{\mathrm{i}})$, $j = 1,\dots,m$.

CONSTRUCTION OF $\mathbb{Z}^{\mathrm{g}}$

According to Assumption 5.2.5-(ii), the grid $\mathbb{Z}^{\mathrm{g}}$ must be constructed such that $\mathrm{co}(\mathbb{Z}^{\mathrm{g}}) \supseteq f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$. This can be simply done by finding the vertices of the smallest box that contains the set $f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})$. Those vertices give the range of $\mathbb{Z}^{\mathrm{g}}$ in each dimension. We can then, for example, take $\mathbb{Z}^{\mathrm{g}}$ to be the uniform grid with the same cardinality as $\mathbb{Y}^{\mathrm{g}}$ in each dimension (so that $Z = Y$). This way, $\mathrm{d_H}\left(f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\right) \leq \mathrm{d_H}\left(\mathrm{co}(\mathbb{Z}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\right)$, and hence e_{z} (5.8e) reduces by using finer grids $\mathbb{Z}^{\mathrm{g}}$. This construction has a time complexity of $\mathcal{O}(X)$.[12]

[10] Recall that $\mathbb{L}(C^{\mathrm{d}}_{\mathrm{i}}) = \Pi_{j=1}^{m}\left[\mathrm{L}^{-}_j(C^{\mathrm{d}}_{\mathrm{i}}), \mathrm{L}^{-}_j(C^{\mathrm{d}}_{\mathrm{i}})\right]$, where $\mathrm{L}^{-}_j(C^{\mathrm{d}}_{\mathrm{i}})$ (respectively, $\mathrm{L}^{+}_j(C^{\mathrm{d}}_{\mathrm{i}})$) is the minimum (respectively, maximum) "slope" of $C^{\mathrm{d}}_{\mathrm{i}}$ along the j-th dimension.

[11] Alternatively, we can directly use the discrete input cost $C^{\mathrm{d}}_{\mathrm{i}} : \mathbb{U}^{\mathrm{g}} \to \mathbb{R}$ for computing $\mathrm{L}^{\pm}_j(C^{\mathrm{d}}_{\mathrm{i}})$. In particular, if C_{i} is convex, we can take $\mathrm{L}^{-}_j(C^{\mathrm{d}}_{\mathrm{i}})$ (respectively, $\mathrm{L}^{+}_j(C^{\mathrm{d}}_{\mathrm{i}})$) to be the minimum first forward difference (respectively, maximum last backward difference) of $C^{\mathrm{d}}_{\mathrm{i}}$ along the j-th dimension. This scheme requires $\mathcal{O}(U)$ operations for computing $\mathrm{L}^{\pm}_j(C^{\mathrm{d}}_{\mathrm{i}})$, $j = 1,\dots,m$.

[12] These guidelines are the same as the ones provided in Remark 4.5.5.

CONSTRUCTION OF $\mathbb{Y}^{\mathrm{g}}$

Construction of the dual state grid $\mathbb{Y}^{\mathrm{g}}$ is more involved. According to Theorem 5.2.9, we need to choose a grid that minimizes e_{y} (5.8d). This can be done by choosing $\mathbb{Y}^{\mathrm{g}}$ such that $\mathbb{Y}^{\mathrm{g}} \cap \partial(J_\star - C_{\mathrm{s}}) \neq \emptyset$ for all $x \in \mathbb{X}^{\mathrm{g}}$ so that $e_{\mathrm{y}} = 0$. Even if we had access to the optimal value function $J_\star$, satisfying such a condition could lead to dual grids $\mathbb{Y}^{\mathrm{g}} \subset \mathbb{R}^n$ of size $\mathcal{O}(X^n)$. Such a large size violates Assumption 5.2.5-(iii) on the size of $\mathbb{Y}^{\mathrm{g}}$, and essentially renders the proposed algorithm impractical for dimensions $n \geq 2$. A more practical condition is $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}) \cap \partial(J_\star - C_{\mathrm{s}}) \neq \emptyset$ for all $x \in \mathbb{X}^{\mathrm{g}}$ so that

$$\max_{x \in \mathbb{X}^{\mathrm{g}}} \mathrm{d}\big(\partial(J_\star - C_{\mathrm{s}})(x), \mathbb{Y}^{\mathrm{g}}\big) \leq \mathrm{d_H}\big(\mathrm{co}(\mathbb{Y}^{\mathrm{g}}), \mathbb{Y}^{\mathrm{g}}\big),$$

and hence e_{y} reduces by using finer grids $\mathbb{Y}^{\mathrm{g}}$. The latter condition is satisfied if $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}) \supseteq \mathbb{L}(J_\star - C_{\mathrm{s}})$, i.e., if $\mathrm{co}(\mathbb{Y}^{\mathrm{g}})$ "covers the range of slops" of $(J_\star - C_{\mathrm{s}})$. Hence, we need to approximate the range of slopes of $(J_\star - C_{\mathrm{s}})$. To this end, we first use the fact that $J_\star$ is the fixed point of DP operator (5.1) to approximate $\mathrm{rng}(J_\star - C_{\mathrm{s}})$ by $R = \frac{\mathrm{rng}(C_{\mathrm{i}}^{\mathrm{d}}) + \gamma \cdot \mathrm{rng}(C_{\mathrm{s}}^{\mathrm{d}})}{1-\gamma}$. We then construct the gird $\mathbb{Y}^{\mathrm{g}} = \Pi_{i=1}^{n} \mathbb{Y}_i^{\mathrm{g}}$ such that, for each dimension i, we have

$$\pm \frac{\alpha R}{\Delta_{\mathbb{X}_i^{\mathrm{g}}}} \in \mathrm{co}(\mathbb{Y}_i^{\mathrm{g}}). \tag{5.9}$$

Here, the coefficient $\alpha > 0$ is a scaling factor mainly depending on the dimension of the state space. In particular, by setting $\alpha = 1$, the value $R/\Delta_{\mathbb{X}_i^{\mathrm{g}}}$ is the slope of a linear function with range R over the domain $\Delta_{\mathbb{X}_i^{\mathrm{g}}}$. This construction has a one-time computational cost of $\mathcal{O}(X + U)$ for computing $\mathrm{rng}(C_{\mathrm{i}}^{\mathrm{d}})$ and $\mathrm{rng}(C_{\mathrm{s}}^{\mathrm{d}})$.

DYNAMIC CONSTRUCTION OF $\mathbb{Y}^{\mathrm{g}}$

Alternatively, we can construct $\mathbb{Y}^{\mathrm{g}}$ *dynamically* at each iteration in order to minimize the corresponding error of *each* application of the d-CDP operator given by[13]

$$e_{\mathrm{y}} = c_{\mathrm{y}} \cdot \max_{x \in \mathbb{X}^{\mathrm{g}}} \mathrm{d}\big(\partial(\mathcal{T}J - C_{\mathrm{s}})(x), \mathbb{Y}^{\mathrm{g}}\big).$$

This means that line 4 in Algorithm 5 is moved inside the iterations, after line 8. Similar to the static scheme described above, the aim here is to construct $\mathbb{Y}^{\mathrm{g}}$ such that $\mathrm{co}(\mathbb{Y}^{\mathrm{g}}) \supseteq \mathbb{L}(\mathcal{T}J - C_{\mathrm{s}})$.[14] Since we do not have access to $\mathcal{T}J$ (it is the output of the current iteration), we can again use the definition of the DP operator (5.2) to approximate $\mathrm{rng}(\mathcal{T}J - C_{\mathrm{s}})$ by $R = \mathrm{rng}(C_{\mathrm{i}}^{\mathrm{d}}) + \gamma \cdot \mathrm{rng}(J^{\mathrm{d}})$, where J^{d} is the output of the previous iteration. We then construct the gird $\mathbb{Y}^{\mathrm{g}} = \Pi_{i=1}^{n} \mathbb{Y}_i^{\mathrm{g}}$ such that, for each dimension i, the inclusion (5.9) holds true. This construction has a one-time computational cost of $\mathcal{O}(U)$ for computing $\mathrm{rng}(C_{\mathrm{i}}^{\mathrm{d}})$ and a per iteration computational cost of $\mathcal{O}(X)$ for computing $\mathrm{rng}(J^{\mathrm{d}})$. Notice, however, that under this dynamic construction, the error bound of Theorem 5.2.9 does not hold true. More importantly, with a dynamic grid $\mathbb{Y}^{\mathrm{g}}$ that varies in each iteration, there is no guarantee for ConjVI to converge. However, as we will see in the numerical examples, the proposed scheme leads to a (possibly non-monotone) convergent behavior.

[13] See Lemma 5.4.7 and Proposition 5.4.3.

[14] These guidelines are a modified version of the ones provided in Section 4.5.3.

5.3. NUMERICAL EXPERIMENTS

In this section, we compare the performance of the proposed ConjVI Algorithm 5 (referred to as ConjVI in this section) with the benchmark VI algorithm that utilizes the d-DP operator (5.2) (referred to as VI in this section) through three numerical examples. In particular, we also consider the *dynamic* scheme for the construction of $\mathbb{Y}^{\mathrm{g}}$ in ConjVI (referred to as ConjVI-d in this section). For the first example, we focus on a synthetic system satisfying the conditions of assumptions considered in this chapter in order to examine our theoretical results. In the second and third examples, we showcase the application of ConjVI in solving the optimal control problem of an inverted pendulum and an unstable batch reactor. We note that the simulations were implemented via MATLAB version R2017b, on a PC with an Intel Xeon 3.60 GHz processor and 16 GB RAM.

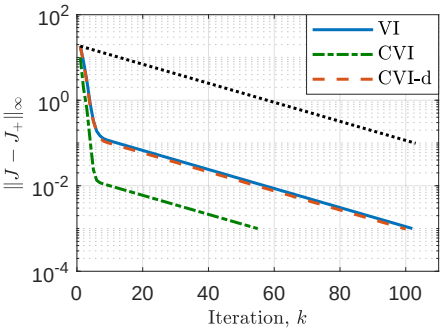
5.3.1. EXAMPLE 1 – SYNTHETIC

We consider the linear system

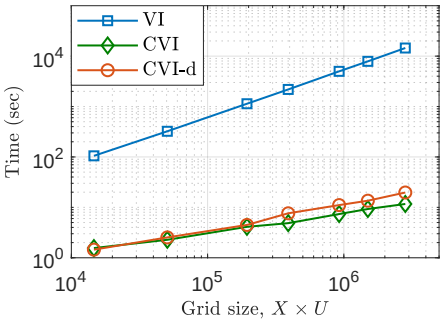
$$x^+ = A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} x + \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} u + w.$$

The problem of interest is the infinite-horizon, optimal control of this system with cost functions $C_{\mathrm{s}}(x) = 10\,\|x\|_2^2$ and $C_{\mathrm{i}}(u) = e^{|u_1|} + e^{|u_2|} - 2$ and discount factor $\gamma = 0.95$. We consider state and input constraint sets $\mathbb{X} = [-1,1]^2 \subset \mathbb{R}^2$ and $\mathbb{U} = [-2,2]^2 \subset \mathbb{R}^2$, respectively. The disturbance is assumed to have a uniform distribution over the finite support $\mathbb{W}^{\mathrm{g}} = \{0, \pm 0.05\} \times \{0\} \subset \mathbb{R}^2$ of size $W = 3$. Notice how the stage cost is a combination of a quadratic term (in state) and an exponential term (in input). Particularly, the control problem at hand does not have a closed-form solution. We use uniform, grid-like discretizations $\mathbb{X}^{\mathrm{g}}$ and $\mathbb{U}^{\mathrm{g}}$ for the state and input spaces such that $\mathrm{co}(\mathbb{X}^{\mathrm{g}}) = \mathbb{X}$ and $\mathrm{co}(\mathbb{U}^{\mathrm{g}}) = \mathbb{U}$. This choice allows us to deploy *multilinear interpolation*, which is non-expansive, as the extension operator $\widetilde{[\cdot]}$ in the d-DP operation (5.2) in VI, and in the d-CDP operation (5.6a) in ConjVI. The grids $\mathbb{V}^{\mathrm{g}}, \mathbb{Z}^{\mathrm{g}} \subset \mathbb{R}^2$ are also constructed uniformly, following the guidelines provided in Section 5.2.2. For the construction of $\mathbb{Y}^{\mathrm{g}} \subset \mathbb{R}^2$, we also follow the guidelines of Section 5.2.2 with $\alpha = 1$. In particular, we also consider the *dynamic* scheme for the construction of $\mathbb{Y}^{\mathrm{g}}$. Moreover, in each implementation of these algorithms, all of the involved grids $(\mathbb{X}^{\mathrm{g}}, \mathbb{U}^{\mathrm{g}}, \mathbb{Y}^{\mathrm{g}}, \mathbb{V}^{\mathrm{g}}, \mathbb{Z}^{\mathrm{g}})$ are chosen to be of the same size N^2 (N points in each dimension). We are particularly interested in the performance of these algorithms, as N increases. We note that the described setup satisfies all of the assumptions in this chapter.

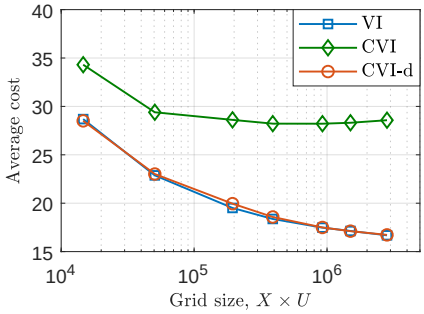
The results of our numerical simulations are shown in Figure 5.1. As shown in Figure 5.1a, both VI and ConjVI are indeed convergent with a rate less than or equal to $\gamma = 0.95$; see Theorem 5.2.7. In particular, ConjVI terminates in $k_{\mathrm{t}} = 55$ iterations, compared to $k_{\mathrm{t}} = 102$ iterations required for VI to reach the termination bound $e_{\mathrm{t}} = 0.001$. Not surprisingly, this faster convergence, combined with the lower time complexity of ConjVI in each iteration, leads to a significant reduction in the running time of this algorithm compared to VI. This effect can be clearly seen in Figure 5.1b, where the runtime of ConjVI for $N = 41^2$ is an order of magnitude less than that of VI for $N = 11^2$. In this regard, we note that the setting of this numerical example leads to $\mathcal{O}(k_{\mathrm{t}} N^4 W)$ and



(a) Convergence rate for $N = 41$.



(b) Running time.



(c) Average cost of 100 random instances of the control problem over $T = 100$ time steps.

Figure 5.1: Performance of VI and ConjVI 5 (CVI) – synthetic example: The black dashed-dotted line in (a) corresponds to exponential convergence with rate $\gamma = 0.95$. CVI-d is the ConjVI algorithm with *dynamic* dual grid $\mathbb{Y}^{\mathrm{g}}$.

$\mathcal{O}(k_{\mathrm{t}} N^2 W)$ time complexities for VI and ConjVI, respectively; see Theorem 5.2.8. Indeed, the running times shown in Figure 5.1b match these complexities.

Since we do not have access to the true optimal value function, we consider evaluating the performance of the greedy policy

$$\mu(x) \in \operatorname*{argmin}_{u \in \mathbb{U}(x) \cap \mathbb{U}^{\mathrm{g}}} \{C(x,u) + \gamma \cdot \mathbb{E}_w \overline{J^{\mathrm{d}}}(g(x,u,w))\},$$

with respect to the discrete value function J^{d} computed using VI and ConjVI (we note that, for finding the greedy control action, we used the same discretization $\mathbb{U}^{\mathrm{g}}$ of the input space and the same extension $\overline{J^{\mathrm{d}}}$ of the value function as the one used in VI and ConjVI algorithms, however, this need not be the case in general). Figure 5.1c reports the average cost of one hundred instances of the optimal control problem with greedy control actions. As can be seen, the reduction in the running time in ConjVI comes with an increase in the cost of the controlled trajectories.

Let us now consider the effect of *dynamic* construction of the dual state grid $\mathbb{Y}^{\mathrm{g}}$. As can be seen in Figure 5.1a, using a dynamic $\mathbb{Y}^{\mathrm{g}}$ leads to a slower convergence (ConjVI-d

terminates in $k_t = 100$ iterations). We note that the relative behavior of the convergence rates in Figures 5.1a was also seen for other grid sizes in the discretization scheme. However, we see a small increase in the running time of ConjVI-d compared to ConjVI since the per iteration complexity for ConjVI-d is again of $\mathcal{O}(k_t N^2 W)$; see Figure 5.1b. More importantly, as depicted in Figure 5.1c, ConjVI-d shows almost the same performance as VI when it comes to the quality of the greedy actions. This is because the dynamic construction of $\mathbb{Y}^g$ in ConjVI-d uses the available computational power (related to the size of the discretization) smartly by finding the smallest grid $\mathbb{Y}^g$ in each iteration, to minimize the error of that same iteration.

5.3.2. EXAMPLE 2 – INVERTED PENDULUM

We now consider the optimal control of an inverted pendulum with quadratic state and input costs $C_s(x) = \|x\|_2^2$ and $C_i(u) = \|u\|_2^2$. The deterministic continuous-time dynamics of the system is described by $\ddot{\theta} = \alpha \sin\theta + \beta\dot{\theta} + \lambda u$, where θ is the angle (with $\theta = 0$ corresponding to upward position), and u is the control input [98, Sec. 4.5.3]. The values of the parameters are $\alpha = 118.6445$, $\beta = -1.599$, and $\lambda = 29.5398$ (corresponding to the values of the physical parameters in [98, Sec. 4.5.3]). Here, we consider the corresponding discrete-time dynamics, by using the forward Euler method with sampling time $\tau = 0.05$. We also introduce stochasticity by considering an additive disturbance in the dynamics. The discrete-time dynamics then reads as

$$x^+ = f_s(x) + Bu + w,$$

where $x = (\theta, \dot{\theta}) \in \mathbb{R}^2$ is the state variable (angle and angular velocity),

$$w \in \mathbb{W}^g = \left\{0, \pm\frac{0.025\pi}{3}, \pm\frac{0.05\pi}{3}\right\} \times \{0, \pm 0.025\pi, \pm 0.05\pi\} \subset \mathbb{R}^2,$$

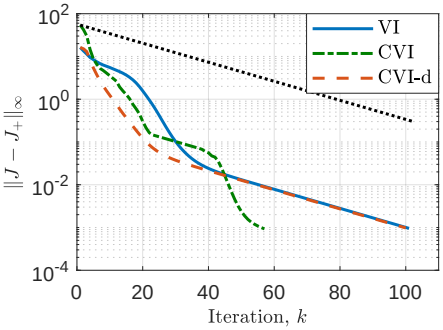
is the disturbance with a uniform distribution, and

$$f_s(x) = x + \tau \cdot \begin{bmatrix} x_2 \\ \alpha \sin x_1 + \beta x_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \lambda \end{bmatrix}.$$

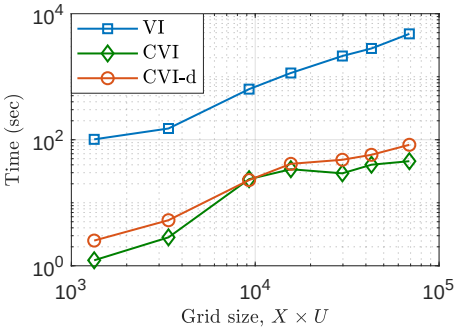
State and input constraints are described by $\mathbb{X} = [-\frac{\pi}{3}, \frac{\pi}{3}] \times [-\pi, \pi] \subset \mathbb{R}^2$ and $\mathbb{U} = [-3, 3] \subset \mathbb{R}$. The discount factor is set to $\gamma = 0.95$ and the termination bound is $e_t = 0.001$.

We use uniform, grid-like discretizations $\mathbb{X}^g$ and $\mathbb{U}^g$ for the state and input spaces such that $\text{co}(\mathbb{X}^g) = [-\frac{\pi}{4}, \frac{\pi}{4}] \times [-\pi, \pi] \subset \mathbb{X}$ and $\text{co}(\mathbb{U}^g) = \mathbb{U}$. With this choice of discrete state space $\mathbb{X}^g$, the feasibility condition of Assumption 5.1.1-(iii) holds for $\text{co}(\mathbb{X}^g)$. (Note that the entire set $\mathbb{X}$ however does not satisfy the feasibility condition). Also, we use *nearest neighbor* extension, which is non-expansive, for the extension operators in (5.2) for VI and in (5.6a) for ConjVI. The grids $\mathbb{V}^g \subset \mathbb{R}$ and $\mathbb{Z}^g, \mathbb{Y}^g \subset \mathbb{R}^2$ are also constructed uniformly, following the guidelines of Section 5.2.4 (with $\alpha = 1$). We again also consider the *dynamic* scheme for the construction of $\mathbb{Y}^g$. Moreover, in each implementation of these algorithms, all the grids are chosen to be of the same size in each dimension, i.e., $X, Y, Z = N^2$ and $U, V = N$.

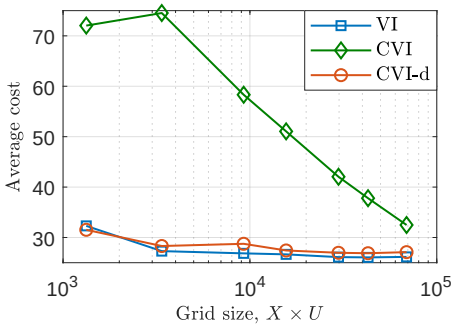
The results of our numerical simulations are shown in Figure 5.2. As reported, we essentially observe the same behaviors as before. In particular, the application of ConjVI, leads to faster convergence and a significant reduction in the running time; see Figures 5.2a and 5.2b. Moreover, as shown in Figures 5.2b and 5.2c, the dynamic scheme for



(a) Convergence rate for $N = 41$.



(b) Running time.



(c) Average cost of 100 random instances of the control problem over $T = 100$ time steps.

Figure 5.2: Performance of VI and ConjVI 5 (CVI) – optimal control of inverted pendulum: The black dashed-dotted line in (a) corresponds to exponential convergence with rate $\gamma = 0.95$. CVI-d is the ConjVI algorithm with *dynamic* dual grid $\mathbb{Y}^g$.

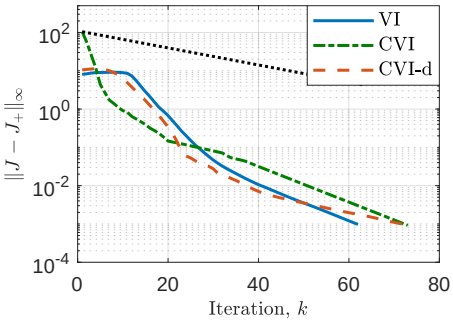
construction of $\mathbb{Y}^g$ leads to a huge improvement in the performance of the corresponding greedy policy at the expense of a small increase in the computational cost.

5.3.3. EXAMPLE 3 – BATCH REACTOR

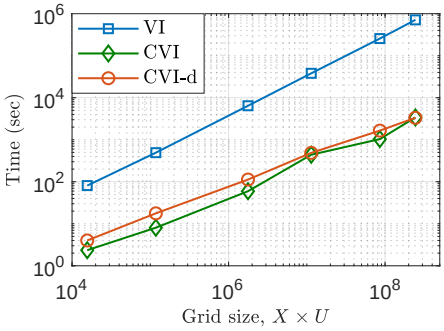
Our last numerical example concerns the optimal control of a system with four states and two input channels, namely, an unstable batch reactor borrowed from [100, Sec. 6]. The discrete-time dynamics are given by

$$x^+ = \begin{bmatrix} 1.08 & -0.05 & 0.29 & -0.24 \\ -0.03 & 0.81 & 0.00 & 0.03 \\ 0.04 & 0.09 & 0.73 & 0.24 \\ 0.00 & 0.19 & 0.05 & 0.91 \end{bmatrix} x + \begin{bmatrix} 0.00 & -0.02 \\ 0.26 & 0.00 \\ 0.08 & -0.13 \\ 0.08 & 0.00 \end{bmatrix} u,$$

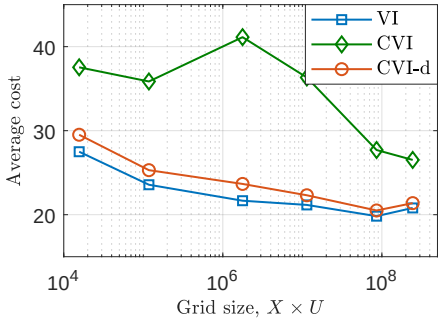
with constraints $x \in \mathbb{X} = [-2,2]^4 \subset \mathbb{R}^4$ and $u \in \mathbb{U} = [-2,2]^2 \subset \mathbb{R}^2$. Note that the dynamics are deterministic. The problem of interest is then the optimal control of the system with quadratic costs $C_s(x) = 2\|x\|_2^2$ and $C_i(u) = \|u\|_2^2$ and discount factor $\gamma = 0.95$. Once again,



(a) Convergence rate for $N = 25$.



(b) Running time.



(c) Average cost of 100 random instances of the control problem over $T = 100$ time steps.

Figure 5.3: Performance of VI and ConjVI 5 (CVI) – optimal control of batch reactor: The black dashed-dotted line in (a) corresponds to exponential convergence with rate $\gamma = 0.95$. CVI-d is the ConjVI algorithm with *dynamic* dual grid $\mathbb{Y}^g$.

we use uniform, grid-like discretizations $\mathbb{X}^g$ and $\mathbb{U}^g$ for the state and input spaces such that $\mathrm{co}(\mathbb{X}^g) = [-1,1]^4 \subset \mathbb{X}$ and $\mathrm{co}(\mathbb{U}^g) = \mathbb{U}$. The grids $\mathbb{V}^g \subset \mathbb{R}$ and $\mathbb{Z}^g, \mathbb{Y}^g \subset \mathbb{R}^2$ are also constructed uniformly, following the guidelines of Section 5.2.4 (with $\alpha = 1$). Moreover, in each implementation of VI and ConjVI(-d), the termination bound is $e_t = 0.001$ and all of the involved grids are chosen to be of the same size in each dimension, i.e., $X, Y, Z = N^4$ and $U, V = N^2$. Finally, we note that we use multi-linear interpolation and *extrapolation* for the extension operator in (5.2) for VI. Due to the extrapolation, the extension operator is no longer non-expansive and hence the convergence of VI is not guaranteed. On the other hand, since the dynamics are deterministic, there is no need for extension in ConjVI (the scaled expectation in (5.6a) in ConjVI reduces to the simple scaling $\varepsilon^d := \gamma \cdot J^d$), and hence the convergence of ConjVI only requires $\mathrm{co}(\mathbb{Z}^g) \supseteq f_s(\mathbb{X}^g)$ and is guaranteed.

The results of our numerical simulations are shown in Figure 5.3. Once again, we see the trade-off between the time complexity and the greedy control performance in VI and ConjVI. On the other hand, ConjVI-d has the same control performance as VI with an insignificant increase in running time compared to ConjVI. In Figure 5.3a, we also observe the non-monotone behavior of ConjVI-d. In this regard, recall that when

the grid $\mathbb{Y}^{\mathrm{g}}$ is constructed dynamically and varies at each iteration, the d-CDP operator is not necessarily contractive, which is here the case for the first six iterations. The VI algorithm also shows a non-monotone behavior, where for the first nine iterations the d-DP operation is actually expansive. As we noted earlier, this is because the extension via multilinear *extrapolation* is expansive.

5.4. TECHNICAL PROOFS

PROOF OF PROPOSITION 5.2.2

We can use the representation (4.21) and the definition of conjugate operation to obtain

$$\begin{aligned}
\widehat{\mathcal{T}} J(x) - C_{\mathrm{s}}(x) &= \max_{y} \{ \langle f_{\mathrm{s}}(x), y \rangle - \phi(y) \} \\
&= \max_{y} \{ \langle f_{\mathrm{s}}(x), y \rangle - C_{\mathrm{i}}^*(-B^\top y) - \epsilon^*(y) \} \\
&= \max_{y} \{ \langle f_{\mathrm{s}}(x), y \rangle - [C_{\mathrm{i}}^*]^{**}(-B^\top y) - \epsilon^*(y) \} \\
&= \max_{y} \left\{ \langle f_{\mathrm{s}}(x), y \rangle - \max_{u \in \mathrm{co}(\mathbb{U})} \left[\langle -B^\top y, u \rangle - C_{\mathrm{i}}^{**}(u) \right] - \epsilon^*(y) \right\} \\
&= \max_{y} \min_{u \in \mathrm{co}(\mathbb{U})} \left\{ C_{\mathrm{i}}^{**}(u) + \langle y, f_{\mathrm{s}}(x) + Bu \rangle - \epsilon^*(y) \right\},
\end{aligned}$$

where we used the fact that $C_{\mathrm{i}}^* : \mathbb{R}^m \to \mathbb{R}$ is proper, closed, and convex, and hence $[C_{\mathrm{i}}^*]^{**} = C_{\mathrm{i}}^*$. This follows from the fact that $\mathrm{dom}(C_{\mathrm{i}}) = \mathbb{U}$ is assumed to be compact (Assumption 5.1.1-(iii)). Hence, the objective function of this maximin problem is convex in u, with $\mathrm{co}(\mathbb{U})$ being compact, which follows from convexity of $C_{\mathrm{i}}^{**} : \mathrm{co}(\mathbb{U}) \to \mathbb{R}$. Also, the objective function is concave in y, which follows from the convexity of ϵ^*. Then, by Sion's Minimax Theorem (see, e.g., [94, Thm. 3]), we have minimax-maximin equality, i.e.,

$$\begin{aligned}
\widehat{\mathcal{T}} J(x) - C_{\mathrm{s}}(x) &= \min_{u} \max_{y} \{ C_{\mathrm{i}}^{**}(u) + \langle y, f(x,u) \rangle - \epsilon^*(y) \} \\
&= \min_{u} \left\{ C_{\mathrm{i}}^{**}(u) + \max_{y} \left[\langle y, f(x,u) \rangle - \epsilon^*(y) \right] \right\} \\
&= \min_{u} \left\{ C_{\mathrm{i}}^{**}(u) + \epsilon^{**}\big(f(x,u)\big) \right\} \\
&= \min_{u} \left\{ C_{\mathrm{i}}^{**}(u) + \gamma \cdot [\mathbb{E}_w J(\cdot + w)]^{**}\big(f(x,u)\big) \right\},
\end{aligned}$$

where for the last equality, we used $[\gamma h]^{**} = \gamma \cdot h^{**}$; see [101, Prop. 13.23–(i)&(iv)].

PROOF OF COROLLARY 5.2.3

By Proposition 5.2.2, we need to show $C_{\mathrm{i}}^{**} = C_{\mathrm{i}}$ and $[\mathbb{E}_w J(\cdot + w)]^{**} = \mathbb{E}_w J(\cdot + w)$ so that

$$\begin{aligned}
C_{\mathrm{i}}^{**}(u) + \gamma \cdot [\mathbb{E}_w J(\cdot + w)]^{**}\big(f(x,u)\big) &= C_{\mathrm{i}}(u) + \gamma \cdot [\mathbb{E}_w J(\cdot + w)]\big(f(x,u)\big) \\
&= C_{\mathrm{i}}(u) + \gamma \cdot \mathbb{E}_w J\big(f(x,u) + w\big) \\
&= C_{\mathrm{i}}(u) + \gamma \cdot \mathbb{E}_w J\big(g(x,u,w)\big).
\end{aligned}$$

This holds if C_{i} and $\mathbb{E}_w J(\cdot + w)$ are proper, closed, and convex. This is indeed the case since $\mathbb{X}$ and $\mathbb{U}$ are compact, and $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ and $J : \mathbb{X} \to \mathbb{R}$ are assumed to be convex.

PROOF OF THEOREM 5.2.7

We begin with two preliminary lemmas on the non-expansiveness of conjugate and multilinear interpolation operations within the d-CDP operation (4.22).

Lemma 5.4.1 (Non-expansiveness of conjugate operator)**.** *Consider two functions h_i ($i = 1,2$), with the same nonempty effective domain $\mathbb{X}$. We have*

$$|h_1^*(y) - h_2^*(y)| \le \|h_1 - h_2\|_\infty, \quad \forall y \in \mathrm{dom}(h_1^*) \cap \mathrm{dom}(h_2^*).$$

Proof. For any $y \in \mathrm{dom}(h_1^*) \cap \mathrm{dom}(h_2^*)$, we have

$$h_1^*(y) = \max_{x\in\mathbb{X}} \langle x, y\rangle - h_1(x) = \max_{x\in\mathbb{X}} \langle x, y\rangle - h_2(x) + h_2(x) - h_1(x).$$

Hence,

$$h_2^*(y) - \|h_1 - h_2\|_\infty \le h_1^*(y) \le h_2^*(y) + \|h_1 - h_2\|_\infty,$$

that is,

$$|h_1^*(y) - h_2^*(y)| \le \|h_1 - h_2\|_\infty.$$

□

Lemma 5.4.2 (Non-expansiveness of interpolative LERP operator)**.** *Consider two discrete functions h_i^{d} ($i = 1,2$) with the same grid-like domain $\mathbb{X}^{\mathrm{g}} \subset \mathbb{R}^n$, and their* interpolative *LERP extensions $\overline{h_i^{\mathrm{d}}} : \mathrm{co}(\mathbb{X}^{\mathrm{g}}) \to \mathbb{R}$. We have*

$$\left\| \overline{h_1^{\mathrm{d}}} - \overline{h_2^{\mathrm{d}}} \right\|_\infty \le \left\| h_1^{\mathrm{d}} - h_2^{\mathrm{d}} \right\|_\infty.$$

Proof. For any $x \in \mathrm{co}(\mathbb{X}^{\mathrm{g}})$, we have ($i = 1,2$)

$$\overline{h_i^{\mathrm{d}}}(x) = \sum_{j=1}^{2^n} \alpha^j \; h_i^{\mathrm{d}}(x^j),$$

where x^j, $j = 1,\dots,2^n$, are the vertices of the hyper-rectangular cell that contains x, and α^j, $j = 1,\dots,2^n$, are convex coefficients (i.e., $\alpha^j \in [0,1]$ and $\sum_j \alpha^j = 1$). Then

$$\left| \overline{h_1^{\mathrm{d}}}(x) - \overline{h_2^{\mathrm{d}}}(x) \right| \le \sum_{j=1}^{2^n} \alpha^j \; \left| h_1^{\mathrm{d}}(x^j) - h_2^{\mathrm{d}}(x^j) \right| \le \left\| h_1^{\mathrm{d}} - h_2^{\mathrm{d}} \right\|_\infty.$$

□

With these preliminary results at hand, we can now show that $\widehat{\mathcal{T}}^{\mathrm{d}}$ is γ-contractive.

Consider two discrete functions $J_i^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$ $(i = 1, 2)$. For any $x \in \mathbb{X}^{\mathrm{d}} \subset \mathbb{R}^n$, we have

$$\begin{aligned}
\left|\widehat{\mathcal{T}}^{\mathrm{d}} J_1^{\mathrm{d}}(x) - \widehat{\mathcal{T}}^{\mathrm{d}} J_2^{\mathrm{d}}(x)\right| &\overset{(5.6f)}{=} \left|\overline{\varphi_1^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big) - \overline{\varphi_2^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big)\right| \overset{\text{Lem. 5.4.2}}{\leq} \left\|\varphi_1^{\mathrm{d}*\mathrm{d}} - \varphi_2^{\mathrm{d}*\mathrm{d}}\right\|_\infty \\
&\overset{\text{Def.}}{\leq} \left\|\varphi_1^{\mathrm{d}*} - \varphi_2^{\mathrm{d}*}\right\|_\infty \overset{\text{Lem. 5.4.1}}{\leq} \left\|\varphi_1^{\mathrm{d}} - \varphi_2^{\mathrm{d}}\right\|_\infty \overset{(5.6d)}{\leq} \left\|\varepsilon_1^{\mathrm{d}*\mathrm{d}} - \varepsilon_2^{\mathrm{d}*\mathrm{d}}\right\|_\infty \\
&\overset{\text{Def.}}{\leq} \left\|\varepsilon_1^{\mathrm{d}*} - \varepsilon_2^{\mathrm{d}*}\right\|_\infty \overset{\text{Lem. 5.4.1}}{\leq} \left\|\varepsilon_1^{\mathrm{d}} - \varepsilon_2^{\mathrm{d}}\right\|_\infty \\
&\overset{(5.6a)}{=} \gamma \cdot \left\| \sum_{w \in \mathbb{W}^{\mathrm{d}}} p(w) \cdot \left(\widetilde{J_1^{\mathrm{d}}}(x+w) - \widetilde{J_2^{\mathrm{d}}}(x+w)\right) \right\|_\infty \\
&\leq \gamma \cdot \left\|\widetilde{J_1^{\mathrm{d}}} - \widetilde{J_2^{\mathrm{d}}}\right\|_\infty \leq \gamma \cdot \left\|J_1^{\mathrm{d}} - J_2^{\mathrm{d}}\right\|_\infty .
\end{aligned}$$

We note that we are using: (i) Assumption 5.2.5-(ii) in the application of Lemma 5.4.2, (ii) $\mathrm{dom}(\varphi_i^{\mathrm{d}*}) = \mathrm{dom}(\varepsilon_i^{\mathrm{d}*}) = \mathbb{R}^n$ for $i = 1, 2$ in the applications of Lemma 5.4.1, and (iii) Assumption 5.2.6-(i) in the last inequality.

5

PROOF OF THEOREM 5.2.8

In what follows, we provide the time complexity of each line of Algorithm 5. In particular, we use the fact that $Y, Z = X$ and $V = U$ by Assumption 5.2.5-(iii). The complexity of construction of $\mathbb{V}^{\mathrm{g}}$ in line 1 is of $\mathcal{O}(X + U)$ by Assumption 5.2.5-(iii). The LLT of line 2 requires $\mathcal{O}(U + V) = \mathcal{O}(U)$ operations [77, Cor. 5]. The complexity of lines 3 and 4 is of $\mathcal{O}(X + U)$ by Assumption 5.2.5-(iii) on the complexity of construction of $\mathbb{Z}^{\mathrm{g}}$ and $\mathbb{Y}^{\mathrm{g}}$. The operation of line 5 also has a complexity of $\mathcal{O}(X)$, and line 6 requires $\mathcal{O}(X+U)$ operations. This leads to the reported $\mathcal{O}(X + U)$ time complexity for initialization.

In each iteration, lines 8 requires $\mathcal{O}(X)$ operations. The complexity of line 9 is of $\mathcal{O}(XWE)$ by the assumption on the complexity of the extension operator $\widetilde{[\cdot]}$. The LLT of line 10 requires $\mathcal{O}(X + Y) = \mathcal{O}(X)$ operations [77, Cor. 5]. The application of LERP in line 12 has a complexity of $\mathcal{O}(\log V)$; see Remark 4.2.3. Hence, the `for loop` over $y \in \mathbb{Y}^{\mathrm{g}}$ requires $\mathcal{O}(Y \log V) = \mathcal{O}(X \log U) = \widetilde{\mathcal{O}}(X)$ operations. The LLT of line 15 requires $\mathcal{O}(Z + Y) = \mathcal{O}(X)$ operations [77, Cor. 5]. The application of LERP in line 17 has a complexity of $\mathcal{O}(\log Z)$; see Remark 4.2.3. Hence, the `for loop` over $x \in \mathbb{X}^{\mathrm{g}}$ requires $\mathcal{O}(X \log Z) = \mathcal{O}(X \log X) = \widetilde{\mathcal{O}}(X)$ operations. The time complexity of each iteration is then of $\widetilde{\mathcal{O}}(XWE)$.

PROOF OF THEOREM 5.2.9

Note that the ConjVI Algorithm 5 involves consecutive applications of the d-CDP operator $\widehat{\mathcal{T}}_{\mathrm{e}}^{\mathrm{d}}$ (4.22), and terminates after a finite number of iterations corresponding to the bound e_{t}. We begin with bounding the difference between the DP and d-CDP operators.

Proposition 5.4.3 (Error of d-CDP operation)**.** [15] *Let $J : \mathbb{X} \to \mathbb{R}$ be a Lipschitz continuous, convex function that satisfies the condition of Assumption 5.2.6-(ii). Assume $C_{\mathrm{i}} : \mathbb{U} \to \mathbb{R}$ is convex. Also, let Assumptions 5.2.5-(i)&(ii) hold. Consider the output of the d-CDP*

[15] This result extends Theorem 4.5.4 on the error of the modified d-CDP operator, by considering the error of extension operation for computing the expectation with respect to to the additive disturbance in (5.6a) and the approximate discrete conjugation of the input cost in (5.6d).

operator $\widehat{\mathcal{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ *and the discretization of the output of the DP operator* $[\mathcal{T} J]^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. *We have*

$$\left\| \widehat{\mathcal{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}} - [\mathcal{T} J]^{\mathrm{d}} \right\|_{\infty} \leq \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{d}}. \tag{5.10}$$

Proof. First note that, by Corollary 5.2.3, the DP and CDP operators are equivalent, i.e., $\mathcal{T} J = \widehat{\mathcal{T}} J$. Hence, it suffices to bound the error of the d-CDP operator $\widehat{\mathcal{T}}_{\mathrm{e}}^{\mathrm{d}}$ with respect to the CDP operator $\widehat{\mathcal{T}}$. We begin with the following preliminary lemma.

Lemma 5.4.4. *The scaled expectation* ϵ *in* (5.4a) *is Lipschitz continuous and convex with a nonempty, compact effective domain. Moreover,* $\mathrm{L}(\epsilon) \leq \gamma \cdot \mathrm{L}(J)$.

Proof. The convexity follows from the fact that expectation preserves convexity and $\gamma > 0$. The effective domain of ϵ is nonempty by the feasibility condition of Assumption 5.1.1-(iii), and is compact since $\mathbb{X}$ is assumed to be compact. Finally, the bound on the Lipschitz constant of ϵ immediately follows from (5.4a). □

We now provide our step-by-step proof. Consider the function ϵ in (5.4a) and its discretization $\epsilon^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \overline{\mathbb{R}}$. Also, consider the discrete function $\varepsilon^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \overline{\mathbb{R}}$ in (5.6a).

Lemma 5.4.5. *We have* $\mathrm{dom}(\epsilon^{\mathrm{d}}) = \mathrm{dom}(\varepsilon^{\mathrm{d}}) \neq \emptyset$. *Moreover,* $\left\| \epsilon^{\mathrm{d}} - \varepsilon^{\mathrm{d}} \right\|_{\infty} \leq \gamma \cdot e_{\mathrm{e}}$.

Proof. The first statement follows from the feasibility condition of Assumption 5.1.1-(iii). For the second statement, note that for every $x \in \mathrm{dom}(\epsilon^{\mathrm{d}}) = \mathrm{dom}(\varepsilon^{\mathrm{d}})$, we can use (5.4a) and (5.6a) to write

$$\begin{aligned}
\left| \epsilon^{\mathrm{d}}(x) - \varepsilon^{\mathrm{d}}(x) \right| &= \gamma \cdot \left| \sum_{w \in \mathbb{W}^{\mathrm{d}}} p(w) \cdot \left(J(x+w) - \widetilde{J^{\mathrm{d}}}(x+w) \right) \right| \\
&\leq \gamma \cdot \sum_{w \in \mathbb{W}^{\mathrm{d}}} p(w) \cdot \left| J(x+w) - \widetilde{J^{\mathrm{d}}}(x+w) \right| \\
&\leq \gamma \cdot \left\| J - \widetilde{J^{\mathrm{d}}} \right\|_{\infty}.
\end{aligned}$$

The result then follows from Assumption 5.2.6-(ii) on J. □

Now, consider the function $\phi : \mathbb{R}^n \to \mathbb{R}$ in (5.4b) and its discretization $\phi^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$. Also, consider the discrete function $\varphi^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ in (5.6d).

Lemma 5.4.6. *We have* $\left\| \phi^{\mathrm{d}} - \varphi^{\mathrm{d}} \right\|_{\infty} \leq \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}}$, *where*

$$\begin{aligned}
e_{\mathrm{u}} &= [\|B\|_2 \cdot \Delta_{\mathbb{Y}^{\mathrm{g}}} + \mathrm{L}(C_{\mathrm{i}})] \cdot \mathrm{d_H}(\mathbb{U}, \mathbb{U}^{\mathrm{d}}), \\
e_{\mathrm{v}} &= \Delta_{\mathbb{U}^{\mathrm{d}}} \cdot \mathrm{d_H}\left(\mathrm{co}(\mathbb{V}^{\mathrm{g}}), \mathbb{V}^{\mathrm{g}} \right), \\
e_{\mathrm{x}} &= \left[\Delta_{\mathbb{Y}^{\mathrm{g}}} + \gamma \cdot \mathrm{L}(J) \right] \cdot \mathrm{d_H}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}).
\end{aligned}$$

Proof. Let $y \in \mathbb{Y}^{\mathrm{g}}$. According to (5.4b) and (5.6d), we have (note that $\varepsilon^{\mathrm{d}*\mathrm{d}}(y) = \varepsilon^{\mathrm{d}*}(y)$)

$$\phi^{\mathrm{d}}(y) - \varphi^{\mathrm{d}}(y) = \phi(y) - \varphi(y) = C_{\mathrm{i}}^*(-B^\top y) - \overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}(-B^\top y) + \epsilon^*(y) - \varepsilon^{\mathrm{d}*}(y). \tag{5.11}$$

First, let us use Lemma 4.2.5 to write

$$0 \le C_{\mathrm{i}}^*(-B^\top y) - C_{\mathrm{i}}^{\mathrm{d}*}(-B^\top y) \le \left[\|-B^\top y\|_2 + \mathrm{L}(C_{\mathrm{i}})\right] \cdot \mathrm{d_H}(\mathbb{U}, \mathbb{U}^{\mathrm{d}})$$
$$\le \left[\|B\|_2 \cdot \Delta_{\mathbb{Y}^{\mathrm{g}}} + \mathrm{L}(C_{\mathrm{i}})\right] \cdot \mathrm{d_H}(\mathbb{U}, \mathbb{U}^{\mathrm{d}}) = e_{\mathrm{u}}. \tag{5.12}$$

Also, Assumption 5.2.5-(i) allows to use Corollary 4.2.7 and write

$$0 \le \overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}(-B^\top y) - C_{\mathrm{i}}^{\mathrm{d}*}(-B^\top y) \le \Delta_{\mathbb{U}^{\mathrm{d}}} \cdot \mathrm{d_H}\left(\mathrm{co}(\mathbb{V}^{\mathrm{g}}), \mathbb{V}^{\mathrm{g}}\right) = e_{\mathrm{v}}. \tag{5.13}$$

Now, by Lemma 5.4.1 (non-expansiveness of conjugation) and Lemma 5.4.5, we have

$$\left|\epsilon^{\mathrm{d}*}(y) - \varepsilon^{\mathrm{d}*}(y)\right| \le \left\|\epsilon^{\mathrm{d}} - \varepsilon^{\mathrm{d}}\right\|_\infty \le \gamma \cdot e_{\mathrm{e}}. \tag{5.14}$$

Moreover, we can use Lemmas 4.2.5 and 5.4.4 to obtain

$$0 \le \epsilon^*(y) - \epsilon^{\mathrm{d}*}(y) \le \left[\|y\|_2 + \mathrm{L}(\epsilon)\right] \cdot \mathrm{d_H}(\mathbb{X}, \mathbb{X}^{\mathrm{g}})$$
$$\le \left[\Delta_{\mathbb{Y}^{\mathrm{g}}} + \gamma \cdot \mathrm{L}(J)\right] \cdot \mathrm{d_H}(\mathbb{X}, \mathbb{X}^{\mathrm{g}}) = e_{\mathrm{x}}. \tag{5.15}$$

5

Combining (5.11)-(5.15), we then have

$$\begin{aligned}\left|\phi^{\mathrm{d}}(y) - \varphi^{\mathrm{d}}(y)\right| &= \left|C_{\mathrm{i}}^*(-B^\top y) - \overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}(-B^\top y) + \epsilon^*(y) - \varepsilon^{\mathrm{d}*}(y)\right| \\ &\le \left|C_{\mathrm{i}}^*(-B^\top y) - C_{\mathrm{i}}^{\mathrm{d}*}(-B^\top y)\right| + \left|C_{\mathrm{i}}^{\mathrm{d}*}(-B^\top y) - \overline{C_{\mathrm{i}}^{\mathrm{d}*\mathrm{d}}}(-B^\top y)\right| \\ &\quad + \left|\epsilon^*(y) - \epsilon^{\mathrm{d}*}(y)\right| + \left|\epsilon^{\mathrm{d}*}(y) - \varepsilon^{\mathrm{d}*}(y)\right| \\ &\le e_{\mathrm{u}} + e_{\mathrm{v}} + \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{x}}.\end{aligned}$$

□

Next, consider the discrete composite functions $[\phi^* \circ f_{\mathrm{s}}]^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ and $[\varphi^{\mathrm{d}*} \circ f_{\mathrm{s}}]^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. In particular, notice that $\phi^* \circ f_{\mathrm{s}}$ appears in (5.4c).

Lemma 5.4.7. *We have* $\left\|[\phi^* \circ f_{\mathrm{s}}]^{\mathrm{d}} - [\varphi^{\mathrm{d}*} \circ f_{\mathrm{s}}]^{\mathrm{d}}\right\|_\infty \le \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}} + e_{\mathrm{y}}$, *where*

$$e_{\mathrm{y}} = \left[\Delta_{f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})} + \Delta_{\mathbb{X}} + \|B\|_2 \cdot \Delta_{\mathbb{U}}\right] \cdot \max_{x \in \mathbb{X}^{\mathrm{g}}} \mathrm{d}\left(\partial(\mathscr{T}J - C_{\mathrm{s}})(x), \mathbb{Y}^{\mathrm{g}}\right).$$

Proof. Let $x \in \mathbb{X}^{\mathrm{g}}$. Also let $\phi^{\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ be the discretization of $\phi : \mathbb{R}^n \to \mathbb{R}$. Since ϕ is convex by construction, we can use Lemma 4.2.5 to obtain (recall that $\mathrm{L}(h; \mathbb{X})$ denotes the Lipschtiz constant of h restricted to the set $\mathbb{X} \subset \mathrm{dom}(h)$)

$$0 \le \phi^*\left(f_{\mathrm{s}}(x)\right) - \phi^{\mathrm{d}*}\left(f_{\mathrm{s}}(x)\right) \le \min_{y \in \partial\phi^*(f_{\mathrm{s}}(x))} \left\{\left[\left\|f_{\mathrm{s}}(x)\right\|_2 + \mathrm{L}\left(\phi; \{y\} \cup \mathbb{Y}^{\mathrm{g}}\right)\right] \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}})\right\} \tag{5.16}$$

By using (5.4c) and the equivalence of DP and CDP operators we have $\phi^* \circ f_{\mathrm{s}} = \widehat{\mathscr{T}}J - C_{\mathrm{s}} = \mathscr{T}J - C_{\mathrm{s}}$. Also, the definition (5.4b) implies that

$$\mathrm{L}(\phi) \le \mathrm{L}\left(C_{\mathrm{i}}^* \circ -B^\top\right) + \mathrm{L}(\epsilon^*) \le \|B\|_2 \cdot \mathrm{L}(C_{\mathrm{i}}^*) + \mathrm{L}(\epsilon^*)$$
$$\le \|B\|_2 \cdot \Delta_{\mathrm{dom}(C_{\mathrm{i}})} + \Delta_{\mathrm{dom}(\epsilon)} \le \|B\|_2 \cdot \Delta_{\mathbb{U}} + \Delta_{\mathbb{X}},$$

where for the last inequality we used the fact that $\mathrm{dom}(\epsilon) \subseteq \mathrm{dom}(J) = \mathbb{X}$. Using these results in (5.16), we have

$$\begin{aligned} 0 \le \phi^*\big(f_{\mathrm{s}}(x)\big) - \phi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) &\le \min_{y \in \partial(\mathscr{T} J - C_{\mathrm{s}})(x)} \Big\{ \big[\|f_{\mathrm{s}}(x)\|_2 + \Delta_{\mathbb{X}} + \|B\|_2 \Delta_{\mathbb{U}} \big] \cdot \mathrm{d}(y, \mathbb{Y}^{\mathrm{g}}) \Big\} \\ &\le \big[\Delta_{f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}})} + \Delta_{\mathbb{X}} + \|B\|_2 \cdot \Delta_{\mathbb{U}} \big] \cdot \max_{x' \in \mathbb{X}^{\mathrm{g}}} \mathrm{d}\big(\partial(\mathscr{T} J - C_{\mathrm{s}})(x'), \mathbb{Y}^{\mathrm{g}}\big) = e_{\mathrm{y}}. \end{aligned} \tag{5.17}$$

Second, by Lemmas 5.4.1 and 5.4.6, we have

$$\Big|\phi^{\mathrm{d}*}(z) - \varphi^{\mathrm{d}*}(z)\Big| \le \Big\|\phi^{\mathrm{d}} - \varphi^{\mathrm{d}}\Big\|_\infty \le \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}}, \tag{5.18}$$

for all $z \in \mathbb{R}^n$, including $z = f_{\mathrm{s}}(x)$. Here, we are using the fact that $\mathrm{dom}(\phi^{\mathrm{d}}) = \mathrm{dom}(\varphi^{\mathrm{d}}) = \mathbb{Y}^{\mathrm{g}}$ and $\mathrm{dom}(\phi^{\mathrm{d}*}) = \mathrm{dom}(\varphi^{\mathrm{d}*}) = \mathbb{R}^n$. Combining inequalities (5.17) and (5.18), we obtain

$$\begin{aligned} \Big|\phi^*\big(f_{\mathrm{s}}(x)\big) - \varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big)\Big| &\le \Big|\phi^*\big(f_{\mathrm{s}}(x)\big) - \phi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big)\Big| + \Big|\phi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) - \varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big)\Big| \\ &\le e_{\mathrm{y}} + \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}}. \end{aligned}$$

This completes the proof. □

Finally, consider the output of the d-CDP operator $\widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. Also, consider the output of the CDP operator $\widehat{\mathscr{T}} J : \mathbb{X} \to \mathbb{R}$ and its discretization $[\widehat{\mathscr{T}} J]^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$.

Lemma 5.4.8. *We have*

$$\Big\|\widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}} - [\widehat{\mathscr{T}} J]^{\mathrm{d}}\Big\|_\infty \le \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}} + e_{\mathrm{y}} + e_{\mathrm{z}} = \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{d}},$$

where

$$e_{\mathrm{z}} = \Delta_{\mathbb{Y}^{\mathrm{g}}} \cdot \mathrm{d}_{\mathrm{H}}\big(f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\big).$$

Proof. Let $x \in \mathbb{X}^{\mathrm{g}}$. According to (5.4c) and (5.6f), we have

$$\widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}}(x) - [\widehat{\mathscr{T}} J]^{\mathrm{d}}(x) = \widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}}(x) - \widehat{\mathscr{T}} J(x) = \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big) - \phi^*\big(f_{\mathrm{s}}(x)\big) \tag{5.19}$$

Now, by Lemma 5.4.7, we have

$$\Big|\phi^*\big(f_{\mathrm{s}}(x)\big) - \varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big)\Big| \le \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}} + e_{\mathrm{y}}. \tag{5.20}$$

Moreover, Assumption 5.2.5-(ii) allows us to use Corollary 4.2.7 and obtain

$$0 \le \overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big) - \varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) \le \Delta_{\mathbb{Y}^{\mathrm{g}}} \cdot \mathrm{d}_{\mathrm{H}}\big(f_{\mathrm{s}}(\mathbb{X}^{\mathrm{g}}), \mathbb{Z}^{\mathrm{g}}\big) = e_{\mathrm{z}}. \tag{5.21}$$

Combining (5.19), (5.20), and (5.21), we then have

$$\begin{aligned} \Big|\widehat{\mathscr{T}}_{\mathrm{e}}^{\mathrm{d}} J^{\mathrm{d}}(x) - [\widehat{\mathscr{T}} J]^{\mathrm{d}}(x)\Big| &= \Big|\overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big) - \phi^*\big(f_{\mathrm{s}}(x)\big)\Big| \\ &\le \Big|\overline{\varphi^{\mathrm{d}*\mathrm{d}}}\big(f_{\mathrm{s}}(x)\big) - \varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big)\Big| + \Big|\varphi^{\mathrm{d}*}\big(f_{\mathrm{s}}(x)\big) - \phi^*\big(f_{\mathrm{s}}(x)\big)\Big| \\ &\le \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{u}} + e_{\mathrm{v}} + e_{\mathrm{x}} + e_{\mathrm{y}} + e_{\mathrm{z}}. \end{aligned}$$

□

The inequality (5.10) then follows from Lemma 5.4.8 by noticing the equivalence of the DP and CDP operators. □

With the preceding result at hand, we can now provide a bound for the difference between the fixed points of the d-CDP and DP operators. To this end, let $\widehat{J}^{\mathrm{d}}_{\star} = \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} \widehat{J}^{\mathrm{d}}_{\star} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ be the fixed point of the d-CDP operator. Recall that $J_{\star} = \mathcal{T} J_{\star} : \mathbb{X} \to \mathbb{R}$ and $J^{\mathrm{d}}_{\star} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ are the true optimal value function and its discretization.

Lemma 5.4.9 (Error of fixed point of d-CDP operator)**.** *We have*

$$\left\| \widehat{J}^{\mathrm{d}}_{\star} - J^{\mathrm{d}}_{\star} \right\|_{\infty} \leq \frac{\gamma \cdot e_{\mathrm{e}} + e_{\mathrm{d}}}{1-\gamma}.$$

Proof. By Assumptions 5.2.5-(ii) and 5.2.6-(i), the operator $\widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}}$ is γ-contractive (Theorem 5.2.7) and hence

$$\left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} \widehat{J}^{\mathrm{d}}_{\star} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} \right\|_{\infty} \leq \gamma \cdot \left\| \widehat{J}^{\mathrm{d}}_{\star} - J^{\mathrm{d}}_{\star} \right\|_{\infty}.$$

Also, notice that Assumptions 5.1.1 and 5.2.4 imply that $J_{\star}$ is Lipschitz continuous and convex. Moreover, $J_{\star}$ is assumed to satisfy the condition of Assumption 5.2.6-(ii). Hence, by Proposition 5.4.3, we have

$$\left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} - [\mathcal{T} J_{\star}]^{\mathrm{d}} \right\|_{\infty} \leq \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{d}}.$$

Using these two inequalities, we can then write

$$\begin{aligned}
\left\| \widehat{J}^{\mathrm{d}}_{\star} - J^{\mathrm{d}}_{\star} \right\|_{\infty} &= \left\| \widehat{J}^{\mathrm{d}}_{\star} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} + \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} - J^{\mathrm{d}}_{\star} \right\|_{\infty} \\
&\leq \left\| \widehat{J}^{\mathrm{d}}_{\star} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} \right\|_{\infty} + \left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} - J^{\mathrm{d}}_{\star} \right\|_{\infty} \\
&= \left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} \widehat{J}^{\mathrm{d}}_{\star} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} \right\|_{\infty} + \left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{\star} - [\mathcal{T} J_{\star}]^{\mathrm{d}} \right\|_{\infty}. \\
&\leq \gamma \cdot \left\| \widehat{J}^{\mathrm{d}}_{\star} - J^{\mathrm{d}}_{\star} \right\|_{\infty} + \gamma \cdot e_{\mathrm{e}} + e_{\mathrm{d}}.
\end{aligned}$$

This completes the proof. □

Finally, we can use the fact that $\widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}}$ is γ-cantractive to provide the following bound on the error due to finite termination of the algorithm. Recall that $\widehat{J}^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$ is the output of Algorithm 5.

Lemma 5.4.10 (Error of finite termination)**.** *We have*

$$\left\| \widehat{J}^{\mathrm{d}} - \widehat{J}^{\mathrm{d}}_{\star} \right\|_{\infty} \leq \frac{\gamma \cdot e_{\mathrm{t}}}{1-\gamma}.$$

Proof. By Assumptions 5.2.5-(ii) and 5.2.6-(i), the operator $\widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}}$ is γ-contractive (Theorem 5.2.7). Let us assume that Algorithm 5 terminates after $k \geq 0$ iterations so that

$\widehat{J}^{\mathrm{d}} = J^{\mathrm{d}}_{k+1}$ and $\left\| J^{\mathrm{d}}_{k+1} - J^{\mathrm{d}}_{k} \right\|_\infty \leq e_{\mathrm{t}}$. Then,

$$\begin{aligned}
\left\| \widehat{J}^{\mathrm{d}} - \widehat{J}^{\mathrm{d}}_\star \right\|_\infty &= \left\| J^{\mathrm{d}}_{k+1} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k+1} + \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k+1} - \widehat{J}^{\mathrm{d}}_\star \right\|_\infty \\
&\leq \left\| J^{\mathrm{d}}_{k+1} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k+1} \right\|_\infty + \left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k+1} - \widehat{J}^{\mathrm{d}}_\star \right\|_\infty \\
&= \left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k+1} \right\|_\infty + \left\| \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} J^{\mathrm{d}}_{k+1} - \widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}} \widehat{J}^{\mathrm{d}}_\star \right\|_\infty \\
&\leq \gamma \cdot \left\| J^{\mathrm{d}}_{k} - J^{\mathrm{d}}_{k+1} \right\|_\infty + \gamma \cdot \left\| J^{\mathrm{d}}_{k+1} - \widehat{J}^{\mathrm{d}}_\star \right\|_\infty \\
&\leq \gamma \cdot e_{\mathrm{t}} + \gamma \left\| \widehat{J}^{\mathrm{d}} - \widehat{J}^{\mathrm{d}}_\star \right\|_\infty ,
\end{aligned}$$

where for the second inequality we used the fact that $\widehat{\mathcal{T}}^{\mathrm{d}}_{\mathrm{e}}$ is a contraction. □

The inequality (5.7) is then derived by combining the results of Lemmas 5.4.9 and 5.4.10.

6

Concluding Remarks

This chapter concludes this thesis by providing some final remarks. In particular, we discuss some of the limitations of the proposed models/approaches and also point to interesting future research directions.

Part One

In the first part of the thesis, we considered a macroscopic model for bounded confidence opinion dynamics with environmental noise. In particular, we studied the effect of exogenous influence by adding a mass of radical (continuum) agents to the original population of the normal agents. The well-posedness of the continuum dynamics expressed as a nonlinear Fokker-Planck equation was established under some assumptions on the initial density of the normal opinions and the density of radical opinions. The long-term behavior of the model was also discussed by considering the corresponding stationary equation. In this regard, we provided a sufficient condition on the noise level that guarantees exponential convergence of the dynamics towards the stationary state that can be made arbitrarily close to the uniform distribution. In the context of opinion dynamics, we derived a theoretical bound on the minimum noise level required to counteract the effect of radical agents and keep the system in a somewhat uniform state.

Exploiting the periodicity of the continuum-agent model, we used Fourier analysis to provide a general framework for characterization of the clustering behavior of the system with the uniform initial distribution. We then applied this framework for a particular distribution of radical opinions, namely, a relatively concentrated triangular distribution. In particular, we studied the effect of the relative mass of the radicals on the critical noise level for order-disorder transition. As expected, the analysis showed that for a larger number of radical agents, the critical noise level increases. We note that this result corresponds to the theoretical result on the global estimate for stationary state. However, comparing the theoretical lower bound on the noise level for the global estimate with its counterpart derived numerically, we find that the theoretical bound is quite conservative, which was expected considering its theoretical nature. We also considered the effect of relative mass and average opinion of radicals on the number, timing, and positioning

of the clusters for noises smaller than the critical noise level. Here, the noise level was shown to be the main factor in determining the number of clusters. Meanwhile, the relative mass of the radicals mainly affects the timing of the clustering behavior, that is, for larger masses of radicals, the clustering behavior is expected to emerge faster. On the other hand, the main effect of the average opinion of the radicals is on the positioning of the clusters; the clusters are positioned in a way that we see a cluster formed around the average opinions of radicals. The simulations of the continuum-agent model and the corresponding discrete-agent model were in agreement with these results.

The most important limitation of the considered macroscopic model for opinion dynamics is in the unwanted effects of the proposed even 2-periodic boundary condition. The proposed boundary condition addresses a disadvantage of simple periodic boundary condition by distinguishing the extreme opinions 0 and 1. However, it reinforces the influence of more extreme neighbors of opinion values in the R-neighborhood of extreme opinions 0 and 1; see Figure 1.1. In a recent paper [102], the authors discussed our even 2-periodic boundary condition and performed a numerical comparison with the "ideal" no-flux boundary condition.

Another restricting aspect of the proposed mean-field model is the underlying assumption that the number of interacting agents is infinite. In some applications such as economics and social sciences, where we are dealing with huge-scale systems with thousands to millions of agents, such an assumption is justifiable. However, in other applications such as multi-agent robotic systems with tens to hundreds of agents, it can be restricting [103]. In this regard, a promising research direction is to look into quantitative bounds on the difference between the population density and its mean-field approximation as a function of the finite size of the population. This seems to be a promising direction considering the results already available in the study of the mean-field limit for interacting particle systems [104].

Another interesting research direction, which can be considered as the next natural step from a control engineering point of view, is the control of the normal population's state/opinion via proper placement of virtual stubborn agents (zealots, leaders). This particular control problem has attracted a lot of attention recently because of the ever-increasing influence of social media in political and commercial campaigns; see, e.g., [105–108]. However, to the best of the author's knowledge, the available literature mainly focuses on employing discrete-agent models for both analysis and synthesis, hence making the application of mean-field models for control synthesis a promising direction. We note that the optimal control of the population density via the corresponding mean-field PDE is a well-established problem in the mathematics and control communities [109, 110], however, the application of mean-field models for proper placement of the virtual stubborn agents has not been explored to the best of our knowledge.

PART TWO

In the second part of the thesis, we considered the approximate implementation of the DP operation arising in the optimal control of discrete-time systems with continuous state and input spaces. The proposed approach involved discretization of the state space and was based on an alternative path that solves the dual problem corresponding to the DP operation by utilizing the LLT algorithm for discrete conjugation. Particularly, we

introduced the d-CDP operator for problems with deterministic input-affine dynamics. There we used the linearity of the dynamics in the input to effectively incorporate the operational duality of addition and infimal convolution, and transformed the minimization in the DP operation to a simple addition at the expense of three conjugate transforms. This, in turn, led to transferring the computational cost from the primal input domain $\mathbb{U}$ to the dual state domain $\mathbb{Y}$. We then modified the proposed d-CDP operator and reduced its time complexity for a subclass of problems with separable data in the state and input variables: For this class, the per-step time complexity was reduced to $\mathcal{O}(X+U)$, compared to the standard complexity of $\mathcal{O}(XU)$. We note that since the proposed scheme essentially solves the discretized dual problem, it is prone to dualization error in non-convex problem. The discretization error, on the other hand, was analyzed and then used to provide concrete guidelines for the construction of a dynamic discrete dual space in the proposed algorithms. We next discussed the extensions of the d-CDP operator for infinite-horizon, discounted cost problems with stochastic dynamics, while computing the conjugate of input cost numerically. In particular, we provided sufficient conditions for the convergence of the corresponding ConjVI algorithm and extended our error and complexity analysis accordingly. Two MATLAB packages were also developed for the implementation of the proposed ConjVI algorithms in this part.

An interesting feature of the conjugate dynamic programming framework proposed in this thesis is that it can be potentially combined with existing tools/techniques for further reduction in time complexity. For example, the proposed framework can be readily combined with sample-based value iteration algorithms that focus on transforming the infinite-dimensional optimization in DP problems into computationally tractable ones (e.g., the common state aggregation technique [69, Sec. 8.1] with piece-wise constant approximation). More interestingly, motivated by the recent quantum speedup for discrete conjugation [79], we envision that the proposed framework paves the way for developing a quantum DP algorithm. Indeed, the proposed algorithms are developed such that any reduction in the complexity of discrete conjugation immediately translates to a reduced computational cost of these algorithms.

The most important drawback of the proposed ConjVI algorithms is their dependence on grid-like discretizations of both primal and dual domains for discrete conjugate operations. Factorized discretizations are particularly suitable for problems with (almost) boxed constraints on the state (and input) spaces. More importantly, with such discretizations, ConjVI, much like the standard VI algorithm, suffers from the curse of dimensionality since the size of the finite representations of the corresponding spaces increases exponentially with the dimension of those spaces. A promising approach to address this issue is to employ adaptive and/or sparse grids for the discretization of the primal space [111, 112]. Moreover, we note that in order to enjoy the linear-time complexity of LLT, we are *only* required to choose a grid-like *dual* grid [77, Rem. 5]; that is, the discretization of the state (and input) space in the primal domain need not be grid-like. However, since the grid-like dual domain is usually chosen to include the same number of points as the primal domain in each dimension, we still face the curse of dimensionality. This, in particular, impairs the performance of the ConjVI Algorithm 3 for problems in which the dimension of the state space is greater than that of the input space. Proper exploitation of the aforementioned property of LLT in such cases calls for a more efficient

construction of the dual grid based on the provided data points in the primal domain.

Consider a discrete function $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$ and its discrete conjugate $h^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ computed using LLT for some finite set $\mathbb{Y}^{\mathrm{g}}$. LLT is, in principle, capable of providing us with the *optimizer mapping*

$$x_\star : \mathbb{Y}^{\mathrm{g}} \to \mathbb{X}^{\mathrm{d}} : y \mapsto \operatorname*{argmax}_{x \in \mathbb{X}^{\mathrm{d}}} \left\{ \langle x, y \rangle - h^{\mathrm{d}}(x) \right\},$$

where for each $y \in \mathbb{Y}^{\mathrm{g}}$, we have $h^{\mathrm{d}*\mathrm{d}}(y) = \langle x_\star(y), y \rangle - h^{\mathrm{d}}\big(x_\star(y)\big)$. Proper exploitation of this capability of LLT is an interesting researh direction that can potentially address one of the drawbacks of the proposed ConjVI Algorithms 4 and 5 by avoiding the approximate discrete conjugation within these algorithms. Let us first recall that by approximate discrete conjugation we mean that we first compute the conjugate function $h^{\mathrm{d}*\mathrm{d}} : \mathbb{Y}^{\mathrm{g}} \to \mathbb{R}$ for some grid $\mathbb{Y}^{\mathrm{g}}$ using the data points $h^{\mathrm{d}} : \mathbb{X}^{\mathrm{d}} \to \mathbb{R}$, and then for any $\tilde{y}$ (not necessarily belonging to $\mathbb{Y}^{\mathrm{g}}$) we use the LERP extension $\overline{h^{\mathrm{d}*\mathrm{d}}}(\tilde{y})$ as an approximation for $h^{\mathrm{d}*}(\tilde{y})$. Indeed, it is possible to avoid this approximation and compute $h^{\mathrm{d}*}(\tilde{y})$ exactly by incorporating a smart search for the corresponding optimizer $\tilde{x} \in \mathbb{X}^{\mathrm{d}}$ for which $h^{\mathrm{d}*}(\tilde{y}) = \langle \tilde{x}, \tilde{y} \rangle - h(\tilde{x})$. To be precise, if $\tilde{y} \in \mathrm{co}(\widetilde{\mathbb{Y}}^{\mathrm{d}})$ for some subset $\widetilde{\mathbb{Y}}^{\mathrm{d}}$ of $\mathbb{Y}^{\mathrm{g}}$, then $\tilde{x} \in \mathrm{co}\left(x_\star(\widetilde{\mathbb{Y}}^{\mathrm{d}})\right)$, where $x_\star : \mathbb{Y}^{\mathrm{g}} \to \mathbb{X}^{\mathrm{d}}$ is the corresponding optimizer mapping. Hence, in order to find the optimizer $\tilde{x} \in \mathbb{X}^{\mathrm{d}}$ corresponding to $\tilde{y}$, it suffices to search in the set $\mathbb{X}^{\mathrm{d}} \cap \mathrm{co}\left(x_\star(\widetilde{\mathbb{Y}}^{\mathrm{d}})\right)$, instead of the entire set $\mathbb{X}^{\mathrm{d}}$. This, in turn, can lead to a lower time complexity for computing the exact discrete conjugate function.[1]

Recall the d-CDP reformulation

$$\widehat{\mathcal{T}}^{\mathrm{d}}[J^{\mathrm{d}}](x) = \min_u \left\{ C(x,u) + J^{\mathrm{d}*\mathrm{d}*}\big(f(x,u)\big) \right\},$$

for deterministic dynamics (Proposition 4.4.5), and note that

$$J^{\mathrm{d}*\mathrm{d}*}(x) = \max_{y \in \mathbb{Y}^{\mathrm{g}}} \left\{ \langle x, y \rangle - J^{\mathrm{d}*\mathrm{d}}(y) \right\},$$

is a max-plus linear combination using the linear basis functions $x \mapsto \langle x, y \rangle$ and coefficients $J^{\mathrm{d}*\mathrm{d}}(y)$, with $y \in \mathbb{Y}^{\mathrm{g}}$ being the slopes for the basis functions. An interesting future research direction is to consider other forms of max-plus linear approximations for the cost functions. In particular, instead of convex, piece-wise affine approximation, one can consider the semi-concave, piece-wise quadratic approximation [86]

$$J^{\mathrm{d}\circledast\mathrm{d}\circledast}(x) = \max_{w \in \mathbb{W}^{\mathrm{g}}} \left\{ c\,\|x - w\|^2 + J^{\mathrm{d}\circledast\mathrm{d}}(w) \right\},$$

for a proper finite set $\mathbb{W}^{\mathrm{g}} \subset \mathbb{X}$ and constant $c > 0$. The important issue then is the fast computation of the coefficients $J^{\mathrm{d}\circledast\mathrm{d}} : \mathbb{W}^{\mathrm{g}} \to \mathbb{R}$ using the data points $J^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$. This

[1] We note that the capability of LLT in providing the optimizer mapping $x_\star : \mathbb{Y}^{\mathrm{g}} \to \mathbb{X}^{\mathrm{d}}$ also allows us to extract the optimal policy within the ConjVI algorithms. In this regard, note that the implementation of the ConjVI algorithms in this thesis *only* provides us with the costs $J_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{R}$, $t = 0, 1, \ldots, T-1$, and *not* the control laws $\mu_t^{\mathrm{d}} : \mathbb{X}^{\mathrm{g}} \to \mathbb{U}^{\mathrm{g}}$, $t = 0, 1, \ldots, T-1$ (in the finite-horizon problem). To address this issue, we have to look at the possibility of extracting the optimal policy within the d-CDP operation by keeping track of the dual pairs in each conjugate transform, i.e., the pairs (x, y) for which $\langle x, y \rangle = h(x) + h^*(y)$. We note that this idea has been implemented in the Master thesis [113].

seems to be possible considering the fact that the operation $[\cdot]^{\circledast}$ closely resembles the "distance transform" [82, 83].

Another interesting research direction is the extension of the proposed conjugate value iteration scheme in a data-driven setup for reinforcement learning problems. This has been the main focus of the author's recent research. In what follows, we provide an overview of the recent developments in this direction. Consider a standard off-line (batch) reinforcement learning problem: We are given a set of size S of samples

$$(x_i, u_i, x_i^+, c_i),\ i = 1, \dots, S,$$

of the agents interaction with the environment. Each sample corresponds to the agent taking the action $u_i \in \mathbb{U} \subset \mathbb{R}^m$ in the state $x_i \in \mathbb{X} \subset \mathbb{R}^n$, which pushes the system to the state $x_i^+ \in \mathbb{X}$, while incurring the cost $c_i \in \mathbb{R}$. The problem is then to use these samples in order to find the Q-function $Q = \mathcal{T}Q : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$. Recall that the DP operator $\mathcal{T}$ is defined as (for deterministic systems)[2]

$$\mathcal{T}Q(x,u) := C(x,u) + \gamma \cdot \min_{u^+} Q(x^+, u^+),$$

A standard solution for this problem is the so-called fitted Q-iteration (FQI) algorithm, in which we use a parametric approximation $\widehat{Q}_\theta : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ for the Q-function, and find the parameter $\theta \in \mathbb{R}^P$ using some form of recursive regressions. Precisely, we solve

$$\theta_{\ell+1} = \text{Regression}\left(\widehat{Q}_{\theta_{\ell+1}}, \left\{\mathbb{S}^{\text{d}}, Q^{\text{d}}_{\ell+1}\right\}\right), \quad \ell = 0, 1, \dots, \tag{6.1}$$

using the data $Q^{\text{d}}_{\ell+1} : \mathbb{S}^{\text{d}} = \{(x_i, u_i)\}_{i=1}^S \to \mathbb{R}$ given by

$$Q^{\text{d}}_{\ell+1}(x_i, u_i) := \mathcal{T}\widehat{Q}_{\theta_\ell}(x_i, u_i) = c_i + \gamma \cdot \min_{u^+} \widehat{Q}_{\theta_\ell}(x_i^+, u^+). \tag{6.2}$$

That is, we find the best $\theta_{\ell+1}$ by fitting $\widehat{Q}_{\theta_{\ell+1}}$ to $\mathcal{T}\widehat{Q}_{\theta_\ell}$ over the samples. Let us now use this basic procedure for developing a similar algorithm in the conjugate domain. To this end, we employ the following *max-plus linear* approximator for the Q-function

$$\widehat{Q}_\theta(x,u) = \max_{(y,v) \in \mathbb{Y}^{\text{d}} \times \mathbb{V}^{\text{d}}} \left\{ \langle y, x \rangle + \langle v, u \rangle - \theta^{\text{d}}(y,v) \right\} = \theta^{\text{d}*}(x,u), \tag{6.3}$$

for proper discrete dual state and input spaces $\mathbb{Y}^{\text{d}}$ and $\mathbb{V}^{\text{d}}$, respectively (we are now treating the vector $\theta \in \mathbb{R}^P$ as a discrete function $\theta^{\text{d}} : \mathbb{Y}^{\text{d}} \times \mathbb{V}^{\text{d}} \to \mathbb{R}$ such that $P = YV$). Note that this is exactly the same type of approximation that is used within the proposed d-CDP operators. Let $\theta_\ell \in \mathbb{R}^P$ be the vector of parameters at the current iteration $\ell \geq 0$. By plugging the approximator (6.3) in (6.2), the values of the Q-function at the sample points in the current iteration can be estimated as follows

$$Q^{\text{d}}_{\ell+1}(x_i, u_i) = c_i + \gamma \cdot \min_{u^+} \theta^{\text{d}*}_\ell(x_i^+, u^+).$$

[2] The Q-function is related to the value function via $J(x) = \min_u Q(x,u)$.

We can then find the updated vector of coefficients $\theta_{\ell+1} \in \mathbb{R}^P$, by fitting $\widehat{Q}_{\theta_{\ell+1}}$ to the estimates computed above, i.e., by setting

$$\widehat{Q}_{\theta_{\ell+1}}(x_i, u_i) = \max_{(y,v)\in \mathbb{Y}^{\mathrm{g}}\times\mathbb{V}^{\mathrm{g}}} \left\{ \langle y, x\rangle + \langle v, u\rangle - \theta^{\mathrm{d}}_{\ell+1}(y, v) \right\} = Q^{\mathrm{d}}_{\ell+1}(x_i, u_i).$$

The proceeding equations form a system of max-plus linear equations. To solve that equation, we can use the largest subsolution [114]

$$\theta^{\mathrm{d}}_{\ell+1}(y, v) = - \max_{(x_i,u_i)\in\mathbb{S}^{\mathrm{d}}} \left\{ \langle y, x_i\rangle + \langle v, u_i\rangle - Q^{\mathrm{d}}_{\ell+1}(x_i, u_i) \right\} = -Q^{\mathrm{d}*}_{\ell+1}(y, v).$$

Hence, the proposed conjugate FQI involves the following steps at each iteration:

1. Compute $Q^{\mathrm{d}}_{\ell+1}(x_i, u_i) = c_i + \gamma \cdot \min_{u^+} \theta^{\mathrm{d}*}_{\ell}(x^+_i, u^+)$ for $(x_i, u_i) \in \mathbb{S}^{\mathrm{d}}$;
2. Compute $\theta^{\mathrm{d}}_{\ell+1}(y, v) = -Q^{\mathrm{d}*}_{\ell+1}(y, v)$ for $(y, v) \in \mathbb{Y}^{\mathrm{d}} \times \mathbb{V}^{\mathrm{d}}$.

Now, the important observation is that if the sets $\mathbb{S}^{\mathrm{d}}$, $\mathbb{Y}^{\mathrm{d}}$, and $\mathbb{V}^{\mathrm{d}}$ are grid-like, then we can use the LLT algorithm for computing the involved discrete conjugate operation. This can reduce the time complexity of each iteration of the FQI algorithm to $\mathcal{O}(S + P)$ from to the standard $\mathcal{O}(SP)$ (disreg the complexity of solving the minimization over u^+ in the first step).

REFERENCES

[1] M. A. S. Kolarijani, A. V. Proskurnikov, and P. Mohajerin Esfahani, "Macroscopic noisy bounded confidence models with distributed radical opinions," *IEEE Transactions on Automatic Control*, vol. 66, no. 3, pp. 1174–1189, 2021.

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwanga, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4–5, pp. 175–308, 2006.

[3] Y.-Y. Liu and A.-L. Barabási, "Control principles of complex systems," *Reviews of Modern Physics*, vol. 88, p. 035006, 2016.

[4] S. H. Strogatz, *Sync: The Emerging Science of Spontaneous Order*. New York: Hyperion Press, 2003.

[5] C. W. Wu, *Synchronization in Complex Networks of Nonlinear Dynamical Systems*. Singapore: World Scientific, 2007.

[6] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton and Oxford: Princeton University Press, 2010.

[7] A. Pogromsky, G. Santoboni, and H. Nijmeijer, "Partial synchronization: from symmetry towards stability," *Physica D: Nonlinear Phenomena*, vol. 172, no. 1, pp. 65 – 87, 2002.

[8] W. Wu, W. Zhou, and T. Chen, "Cluster synchronization of linearly coupled complex networks under pinning control," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 4, pp. 829–839, 2009.

[9] W. Xia and M. Cao, "Clustering in diffusively coupled networks," *Automatica*, vol. 47, no. 11, pp. 2395–2405, 2011.

[10] L. Pecora, F. Sorrentino, A. Hagerstrom, T. Murphy, and R. Roy, "Cluster synchronization and isolated desynchronization in complex networks with symmetries," *Nature Communications*, vol. 5, p. 4079, 2014.

[11] L. Lu, C. Li, S. Bai, L. Gao, L. Ge, and C. Han, "Cluster synchronization between uncertain networks with different dynamics," *Physica A*, vol. 469, pp. 429 – 437, 2017.

[12] R. Abelson, "Mathematical models of the distribution of attitudes under controversy," in *Contributions to Mathematical Psychology*, N. Frederiksen and H. Gulliksen, Eds. New York: Holt, Rinehart & Winston, Inc, 1964, pp. 142–160.

[13] T. Kurahashi-Nakamura, M. Mäs, and J. Lorenz, "Robust clustering in generalized bounded confidence models," *Journal of Artificial Societies and Social Simulation*, vol. 19, no. 4, p. 7, 2016.

[14] N. Friedkin, "The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem," *IEEE Control Systems Magazine*, vol. 35, no. 3, pp. 40–51, 2015.

[15] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Reviews of Modern Physics*, vol. 81, pp. 591–646, 2009.

[16] H. Xia, H. Wang, and Z. Xuan, "Opinion dynamics: A multidisciplinary review and perspective on future research," *International Journal of Knowledge and Systems Science*, vol. 2, no. 4, pp. 72–91, 2011.

[17] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, vol. 1, pp. 3–49, 2011.

[18] A. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annual Reviews in Control*, vol. 43, 2017, 65-79.

[19] ——, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annual Reviews in Control*, vol. 45, 2018, 166–190.

[20] A. V. Proskurnikov, C. Ravazzi, and F. Dabbene, "Dynamics and structure of social networks from a systems and control viewpoint: A survey of Roberto Tempo's contributions," *Online Social Networks and Media*, vol. 7, pp. 45 – 59, 2018.

[21] L. Mastroeni, P. Vellucci, and M. Naldi, "Agent-based models for opinion formation: A bibliographic survey," *IEEE Access*, vol. 7, pp. 58 836–58 848, 2019.

[22] A. Aydogdu, S. T. McQuade, and N. P. Duteil, "Opinion dynamics on a general compact Riemannian manifold," *Networks & Heterogeneous Media*, vol. 12, p. 489, 2017.

[23] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 1–27, 2013.

[24] J. Liu, X. Chen, T. Basar, and M. Belabbas, "Exponential convergence of the discrete- and continuous-time Altafini models," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6168–6182, 2017.

[25] M. McPherson, L. Smith-Lovin, and J. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, pp. 415–444, 2001.

[26] C. G. Lord and L. Ross, "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2098–2109, 1979.

[27] U. Krause, "A discrete nonlinear and non-autonomous model of consensus formation," *Communications in Difference Equations*, vol. 2000, pp. 227–236, 2000.

[28] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems*, vol. 3, no. 01–04, pp. 87–98, 2000.

[29] S. Motsch and E. Tadmor, "Heterophilious dynamics enhances consensus," *SIAM Review*, vol. 56, no. 4, pp. 577–621, 2013.

[30] S. R. Etesami, "A simple framework for stability analysis of state-dependent networks of heterogeneous agents," *SIAM Journal on Control and Optimization*, vol. 57, no. 3, pp. 1757–1782, 2019.

[31] B. Chazelle and C. Wang, "Inertial Hegselmann-Krause systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3905–3913, 2017.

[32] A. Mirtabatabaei and F. Bullo, "Opinion dynamics in heterogeneous networks: Convergence conjectures and theorems," *SIAM Journal on Control and Optimization*, vol. 50, no. 5, pp. 2763–2785, 2012.

[33] S. R. Etesami and T. Başar, "Game-theoretic analysis of the Hegselmann-Krause model for opinion dynamics in finite dimensions," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1886–1897, 2015.

[34] R. Hegselmann and U. Krause, "Truth and cognitive division of labour: First steps towards a computer aided social epistemology," *Journal of Artificial Societies and Social Simulation*, vol. 9, no. 3, p. 1, 2006.

[35] ——, "Opinion dynamics under the influence of radical groups, charismatic leaders and other constant signals: a simple unifying model," *Networks & Heterogeneous Media*, vol. 10, no. 3, pp. 477–509, 2015.

[36] Y. Zhao, L. Zhang, M. Tang, and G. Kou, "Bounded confidence opinion dynamics with opinion leaders and environmental noises," *Computers & Operations Research*, vol. 74, pp. 205–213, 2016.

[37] M. Porfiri, E. M. Bollt, and D. J. Stilwell, "Decline of minorities in stubborn societies," *The European Physical Journal B*, vol. 57, no. 4, pp. 481–486, 2007.

[38] M. Pineda, R. Toral, and E. Hernández-García, "Noisy continuous-opinion dynamics," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 08, p. P08001, 2009.

[39] E. Ben-Naim, P. Krapivsky, and S. Redner, "Bifurcations and patterns in compromise processes," *Physica D*, vol. 183, no. 3, pp. 190–204, 2003.

[40] S. Grauwin and P. Jensen, "Opinion group formation and dynamics: Structures that last from nonlasting entities," *Physical Review E*, vol. 85, p. 066113, 2012.

[41] M. Pineda, R. Toral, and E. Hernández-García, "The noisy Hegselmann-Krause model for opinion dynamics," *The European Physical Journal B*, vol. 86, no. 12, p. 490, 2013.

[42] A. Carro, R. Toral, and M. San Miguel, "The role of noise and initial conditions in the asymptotic solution of a bounded confidence, continuous-opinion model," *Journal of Statistical Physics*, vol. 151, pp. 131–149, 2013.

[43] V. D. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "On the 2r conjecture for multi-agent systems," in *2007 European Control Conference (ECC)*, 2007, pp. 874–881.

[44] C. Wang, Q. Li, W. E, and B. Chazelle, "Noisy Hegselmann-Krause systems: Phase transition and the 2R-conjecture," *Journal of Statistical Physics*, vol. 166, no. 5, pp. 1209–1225, 2017.

[45] M. Huang and J. H. Manton, "Opinion dynamics with noisy information," in *52nd IEEE Conference on Decision and Control*, 2013, pp. 3445–3450.

[46] W. Su, G. Chen, and Y. Hong, "Noise leads to quasi-consensus of Hegselmann–Krause opinion dynamics," *Automatica*, vol. 85, pp. 448–454, 2017.

[47] J. Lorenz, "Continuous opinion dynamics under bounded confidence: a survey," *International Journal of Modern Physics C*, vol. 18, no. 12, pp. 1819–1838, 2007.

[48] V. Blondel, J. Hendrickx, and J. Tsitsiklis, "Continuous-time average-preserving opinion dynamics with opinion-dependent communications," *SIAM Journal on Control and Optimization*, vol. 48, pp. 5214–5240, 2010.

[49] J. Hendrickx and A. Olshevsky, "On symmetric continuum opinion dynamics," *SIAM Journal on Control and Optimization*, vol. 54, no. 5, pp. 2893–2918, 2016.

[50] A. Mirtabatabaei, P. Jia, and F. Bullo, "Eulerian opinion dynamics with bounded confidence and exogenous inputs," *SIAM Journal on Applied Dynamical Systems*, vol. 13, no. 1, pp. 425–446, 2014.

[51] C. Canuto, F. Fagnani, and P. Tilli, "An Eulerian approach to the analysis of Krause's consensus models," *SIAM Journal on Control and Optimization*, vol. 50, pp. 243–265, 2012.

[52] L. Boudin and F. Salvarani, "Opinion dynamics: Kinetic modelling with mass media, application to the Scottish independence referendum," *Physica A*, vol. 444, pp. 448–457, 2016.

[53] A. Nordio, A. Tarable, C.-F. Chiasserini, and E. Leonardi, "Belief dynamics in social networks: A fluid-based analysis," *IEEE Transactions on Network Science and Engineering*, vol. 5, no. 4, pp. 276–287, 2018.

[54] J. Garnier, G. Papanicolaou, and T.-W. Yang, "Consensus convergence with stochastic effects," *Vietnam Journal of Mathematics*, vol. 45, no. 1, pp. 51–75, 2017.

[55] B. Chazelle, Q. Jiu, Q. Li, and C. Wang, "Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics," *Journal of Differential Equations*, vol. 263, no. 1, pp. 365 – 397, 2017.

[56] L. Evans, *Partial Differential Equations*. American Mathematical Society, 2010.

[57] D. A. Dawson, "Critical dynamics and fluctuations for a mean-field model of cooperative behavior," *Journal of Statistical Physics*, vol. 31, no. 1, pp. 29–85, 1983.

[58] K. Oelschlager, "A martingale approach to the law of large numbers for weakly interacting stochastic processes," *The Annals of Probability*, vol. 12, no. 2, pp. 458–479, 1984.

[59] J. Gärtner, "On the McKean-Vlasov limit for interacting diffusions," *Mathematische Nachrichten*, vol. 137, no. 1, pp. 197–248, 1988.

[60] J. A. Carrillo, R. S. Gvalani, G. A. Pavliotis, and A. Schlichting, "Long-time behaviour and phase transitions for the McKean–Vlasov equation on the torus," *Archive for Rational Mechanics and Analysis*, Jul 2019.

[61] R. A. Adams and J. J. Fournier, *Sobolev Spaces*. Elsevier, 2003, vol. 140.

[62] M. A. S. Kolarijani, A. V. Proskurnikov, and P. Mohajerin Esfahani, "Long-term behavior of mean-field noisy bounded confidence models with distributed radicals," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 6158–6163.

[63] M. Pineda, R. Toral, and E. Hernández-García, "Diffusing opinions in bounded confidence processes," *The European Physical Journal D*, vol. 62, no. 1, pp. 109–117, 2011.

[64] M. A. S. Kolarijani and P. Mohajerin Esfahani, "Discrete conjugate dynamic programming (d-CDP) MATLAB package," Available online at https://github.com/AminKolarijani/d-CDP, 2021.

[65] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*. Belmont, MA: Athena Scientific, 2019.

[66] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[67] N. Balaji, S. Kiefer, P. Novotnỳ, G. A. Pérez, and M. Shirmohammadi, "On the complexity of value iteration," *preprint arXiv:1807.04920*, 2018.

[68] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*, 3rd ed. Belmont, MA: Athena Scientific, 2005.

[69] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2011.

[70] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, MA: Athena Scientific, 2009.

[71] R. Bellman and W. Karush, "Mathematical programming and the maximum transform," *Journal of the Society for Industrial and Applied Mathematics*, vol. 10, no. 3, pp. 550–567, 1962.

[72] A. O. Esogbue and C. W. Ahn, "Computational experiments with a class of dynamic programming algorithms of higher dimensions," *Computers & Mathematics with Applications*, vol. 19, no. 11, pp. 3 – 23, 1990.

[73] C. M. Klein and T. L. Morin, "Conjugate duality and the curse of dimensionality," *European Journal of Operational Research*, vol. 50, no. 2, pp. 220 – 228, 1991.

[74] R. Rockafellar, *Conjugate Duality and Optimization*. Philadelphia: Society for Industrial and Applied Mathematics, 1974.

[75] L. Corrias, "Fast Legendre-Fenchel transform and applications to Hamilton-Jacobi equations and conservation laws," *SIAM Journal on Numerical Analysis*, vol. 33, no. 4, pp. 1534–1558, 1996.

[76] Y. Lucet, "A fast computational algorithm for the Legendre-Fenchel transform," *Computational Optimization and Applications*, vol. 6, no. 1, pp. 27–57, 1996.

[77] ——, "Faster than the fast Legendre transform, the linear-time Legendre transform," *Numerical Algorithms*, vol. 16, no. 2, pp. 171–185, 1997.

[78] ——, "What shape is your conjugate? A survey of computational convex analysis and its applications," *SIAM Review*, vol. 52, no. 3, pp. 505–542, 2010.

[79] D. Sutter, G. Nannicini, T. Sutter, and S. Woerner, "Quantum Legendre-Fenchel transform," *preprint arXiv:2006.04823*, 2020.

[80] Y. Achdou, F. Camilli, and L. Corrias, "On numerical approximation of the Hamilton-Jacobi-transport system arising in high frequency approximations," *Discrete & Continuous Dynamical Systems-Series B*, vol. 19, no. 3, 2014.

[81] G. Costeseque and J.-P. Lebacque, "A variational formulation for higher order macroscopic traffic flow models: Numerical investigation," *Transportation Research Part B: Methodological*, vol. 70, pp. 112 – 133, 2014.

[82] P. F. Felzenszwalb and D. P. Huttenlocher, "Distance transforms of sampled functions," *Theory of computing*, vol. 8, no. 1, pp. 415–428, 2012.

[83] Y. Lucet, "New sequential exact Euclidean distance transform algorithms based on convex analysis," *Image and Vision Computing*, vol. 27, no. 1, pp. 37 – 44, 2009.

[84] M. Jacobs and F. Léger, "A fast approach to optimal transport: the back-and-forth method," *arXiv preprint arXiv:1905.12154*, 2019.

[85] R. Carpio and T. Kamihigashi, "Fast value iteration: an application of Legendre-Fenchel duality to a class of deterministic dynamic programming problems in discrete time," *Journal of Difference Equations and Applications*, vol. 26, no. 2, pp. 209–222, 2020.

[86] W. M. McEneaney, *Max-plus methods for nonlinear control and estimation*. Springer Science & Business Media, 2006.

[87] W. M. McEneaney, "Max-plus eigenvector representations for solution of nonlinear H_∞ problems: basic concepts," *IEEE Transactions on Automatic Control*, vol. 48, no. 7, pp. 1150–1163, 2003.

[88] M. Akian, S. Gaubert, and A. Lakhoua, "The max-plus finite element method for solving deterministic optimal control problems: Basic properties and convergence analysis," *SIAM Journal on Control and Optimization*, vol. 47, no. 2, pp. 817–848, 2008.

[89] F. Bach, "Max-plus matching pursuit for deterministic Markov decision processes," *arXiv preprint arXiv:1906.08524*, 2019.

[90] E. Berthier and F. Bach, "Max-plus linear approximations for deterministic continuous-state Markov decision processes," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 767–772, 2020.

[91] K. Murota, *Discrete Convex Analysis*. Society for Industrial and Applied Mathematics, 2003.

[92] A. Sidford, M. Wang, X. Wu, and Y. Ye, "Variance reduced value iteration and faster algorithms for solving Markov decision processes," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2018, pp. 770–787.

[93] I. Joó and L. L. Stachó, "A note on Ky Fan's minimax theorem," *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 39, no. 4, pp. 401–407, 1982.

[94] S. Simons, "Minimax theorems and their proofs," in *Minimax and Applications*, D.-Z. Du and P. M. Pardalos, Eds. Boston, MA: Springer US, 1995, pp. 1–23.

[95] M. A. S. Kolarijani, G. F. Max, and P. Mohajerin Esfahani, "Fast approximate dynamic programming for infinite-horizon Markov decision processes," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021.

[96] M. A. S. Kolarijani and P. Mohajerin Esfahani, "Conjugate value iteration (ConjVI) MATLAB package," Available online at https://github.com/AminKolarijani/ConjVI, 2021.

[97] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. II*, 3rd ed. Belmont, MA: Athena Scientific, 2007.

[98] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*. CRC press, 2017.

[99] D. Bertsekas, "Linear convex stochastic control problems over an infinite horizon," *IEEE Transactions on Automatic Control*, vol. 18, no. 3, pp. 314–315, 1973.

[100] A. S. Kolarijani, S. C. Bregman, P. Mohajerin Esfahani, and T. Keviczky, "A decentralized event-based approach for robust model predictive control," *IEEE Transactions on Automatic Control*, vol. 65, no. 8, pp. 3517–3529, 2020.

[101] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, 2nd ed. Springer, New York, NY, 2017.

[102] B. D. Goddard, B. Gooding, G. A. Pavliotis, and H. Short, "Noisy bounded confidence models for opinion dynamics: the effect of boundary conditions on phase transitions," *preprint arXiv:2009.03131*, 2021.

[103] K. Elamvazhuthi and S. Berman, "Mean-field models in swarm robotics: A survey," *Bioinspiration & Biomimetics*, vol. 15, p. 015001, 2020.

[104] P.-E. Jabin and Z. Wang, *Mean Field Limit for Stochastic Particle Systems*. Cham: Springer International Publishing, 2017, pp. 379–402.

[105] R. Hegselmann, S. Konig, S. Kurz, C. Niemann, and J. Rambau, "Optimal opinion control : The campaign problem," *Journal of Artificial Societies and Social Simulation*, vol. 18, pp. 1–18, 2014.

[106] N. Masuda, "Opinion control in complex networks," *New Journal of Physics*, vol. 17, p. 033031, 2015.

[107] F. Dietrich, S. Martin, and M. Jungers, "Control via leadership of opinion dynamics with state and time-dependent interactions," *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1200–1207, 2018.

[108] I. Douven and R. Hegselmann, "Mis- and disinformation in a bounded confidence model," *Artificial Intelligence*, vol. 291, p. 103415, 2021.

[109] M. Annunziato and A. Borzì, "A Fokker-Planck control framework for stochastic systems," *EMS Surveys in Mathematical Sciences*, vol. 5, no. 1, pp. 65–98, 2018.

[110] A. Bensoussan, J. Frehse, P. Yam *et al.*, *Mean field games and mean field type control theory*. Springer, 2013, vol. 101.

[111] H.-J. Bungartz and M. Griebel, "Sparse grids," *Acta Numerica*, vol. 13, p. 147–269, 2004.

[112] D. Pflüger, B. Peherstorfer, and H.-J. Bungartz, "Spatially adaptive sparse grids for high-dimensional data-driven problems," *Journal of Complexity*, vol. 26, no. 5, pp. 508–522, 2010.

[113] C. Rodopoulos, "Conjugate dynamic programming," Master thesis, Deflt University of Technology, 2021.

[114] B. De Schutter, T. van den Boom, J. Xu, and S. S. Farahani, "Analysis and control of max-plus linear discrete-event systems: An introduction," *Discrete Event Dynamic Systems*, vol. 30, p. 25–54, 2020.