Optimal Agent Cooperation with Local Information

by

Jan De Mot

Burgerlijk Werktuigkundig Ingenieur
Katholieke Universiteit Leuven, 2000

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2005

Author . . . .
Department of Mechanical Engineering
May 12, 2005

Certified by . . . .
Eric Feron
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Certified by . . . .
Daniela Pucci de Farias
Assistant Professor of Mechanical Engineering
Thesis Advisor

Certified by . . . .
John N. Tsitsiklis
Professor of Electrical Engineering and Computer Science
Thesis Advisor

Accepted by . . . .
Lallit Anand
Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Students

Optimal Agent Cooperation with Local Information

by

Jan De Mot

Submitted to the Department of Mechanical Engineering
on May 12, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Multi-agent systems are in general believed to be more efficient, robust, and versatile than their single-agent equivalents. However, it is not an easy task to design strategies that fully exploit the multi-agent benefits, and with this in mind we address several multi-agent system design issues. Specifically, it is of central importance to determine the optimal agent group composition, which involves a trade-off between the cost and performance increase per additional agent. Further, truly autonomous agents solely rely on on-board environment measurements, the design of which requires quantifying the multi-agent performance as a function of the locally observed environment areas. In this thesis, we focus on the collaborative search for individually rewarding resources, *i.e.* it is possible for multiple agents to incur the same reward. The system objective is to maximize the aggregate rewards incurred.

Motivated by a cooperative surveillance context, we formulate a graph traversal problem on an unbounded structured graph, and restrain the agent motion spatially so that only the lateral agent separation is controlled. We model the problem mathematically as a discrete, infinite state, infinite horizon Dynamic Program and convert it using standard techniques to an equivalent Linear Program (LP) with infinitely many constraints. The graph spatial invariance allows to decompose the LP into a set of infinitely many coupled LPs, each with finitely many constraints. We establish that the unique bounded function that simultaneously satisfies the latter LPs is the problem optimal value function. Based on this, we compute the two-agent optimal value function explicitly as the solution of an LP with finitely many constraints for small agent separations, and implicitly in the form of a recursion for large agent separations, satisfying adequate connection constraints. Finally, we propose a similar method to compute the state probability distribution in steady state under an optimal policy, summarizing the agent behavior at large separations in a set of connection constraints, which is sufficient to compute the probability distribution at small separations.

We analyze and compare the optimal performance of various problem instances. We confirm and quantify the intuition that the performance increases with the group size. Some results stand out: for cone-shaped local observation, two agents incur 25% less cost than a single agent in a mine field type environment (scarce though high costs); further, for some environment specifics, a third agent provides little to no performance increase. Then, we compare various local observation zones, and quantify their effect on the overall group performance. Finally, we study the agent spatial distribution under an optimal policy, and observe that as rewards are scarcer, the agents tend to spread in order to gather information on a larger environment part.

Thesis Supervisor: Eric Feron
Title: Associate Professor of Aeronautics and Astronautics

A careless word may kindle strife,
a cruel word may wreck a life,
a timely word may level stress,
a loving word may heal and bless.

anonymous

Acknowledgments

At the top of my thank-you list appears without a doubt my advisor, Professor Eric Feron. I can safely claim that a set of timely, pointed remarks propelled this work forward at the crucial moments where the frustration level was high. But more importantly, I found in him a constant, excellent source of enthusiasm and support. On one or two occasions, a late-night desperate email was answered promptly with a comforting appointment at 9am the next day. On a more personal note, I am very grateful for his style of advising, which forced me to be independent, and allowed me to meet my other half, Cielo, through various trips to Mexico. Thank you for your confidence.

It is an honor for me to thank Professor John Tsitsiklis. He impressed me with his impeccable teaching style: fast, clean, error-free, and very clear. His academic rigor was a continuous example and his mere presence on my thesis committee was enough to provide me with extra nerves at our meetings, but also made me push my limits even further. I couldn't have hoped for a better benchmark in academics. On a technical note, I thank him and his former post-doc Dr Shie Mannor for suggesting the proof of the Lemmas 15 and 16.

It is a pleasure to thank Professor Daniela Pucci de Farias for the last-minute chairing of my thesis committee. Words of appreciation go to Professor Sanjay Sarma who, in spite of his busy schedule, continued to be very interested and enthusiastic about my work. I benefited a lot from interactions with Professor Nicola Elia who helped me through the more explorative, not always easy period of my effort. I thank Professor John Hansman, who contributed the cooperative soaring application, allowing for a less militarily inspired motivation of my work, and Professor Pablo Parrilo, for his helpful suggestions regarding positivity of polynomials. Dr Vishwesh Kulkarni deserves credit for his initial steps towards improving my writing skills and for exposing me to the more entrepreneurial aspect of engineering.

I thank the Belgian American Educational Foundation for their generous financial support during my first year, making this adventure possible.

The Feron-lab is ... huge and its dynamics have changed several times during my stay. So have the office mates. Undeniably, the healthy and friendly work atmosphere in the office pulled me through the more difficult parts of my stay. Emilio Frazzoli, Vlad Gavrilets, David Dugail, Lucia Pallottina and John Harper helped create that environment during the first year, mixing a good dose of humor and social interaction with hard work. I owe a great deal to Rodin Lyasoff, the de facto network administrator, saving me a tremendous amount of time with his ingenuous solutions to just about anything. I am very grateful to Masha Ishutkina and Tom Schouwenaars for the many discussions, and the collective, therapeutical venting of frustrations. I owe the quality of my thesis defense to Masha and to Chris Dever, who took the time to sit through a dry run of my presentation providing extremely useful comments. Thank you, Bernard Mettler, Ji-Hyun Yang, Zhi-Hong Mao, Greg Mark, Farmey Joseph, Kara Sprague, Sommer Gentry, Animesh Chakravarthy, Emily Craparo and the rest of the controls group. Of course, nothing in the lab would function smoothly without the care of the administrative staff.

I have the greatest respect for Ioannis Martinos, friend and former office mate. His style to bypass the consequences of his background have not stopped amazing me; I have learned more from him than he probably realizes. I will miss the regular, relaxing Saturday lunches in the Miracle of Science (where?) with Franck Billarant, who I am delighted to count among my best friends. A special word of thanks goes to Phil Root, it was an enormous pleasure to work with him, but above all, I appreciate his warm, caring friendship and the excitement of being around his son Andrew. Thomas Gervais was my roommate for four years. I believe that sharing this experience with him, with ups and downs, has created a lasting friendship. I learned an enormous amount from him, and his interest in politics, languages, literature, and cooking provided an excellent escape from the otherwise dominating MIT.

I have the utmost respect for my parents. Through the years, I have realized the value of the principles they instilled in me. I thank them for their unconditional support, for the smooth, pressureless communication and for being an eternal source of advice and wisdom. To them I dedicate this thesis.

Finally, many words of love go to Cielo Díez, my companion in good and bad moments. Her patience in this endeavor is unequalled; her tenderness, enthusiasm, energy and love are amazing. Cielo is the connection between my period at MIT and the future, and I am thrilled about what is to come.

Contents

List of Figures

Chapter 1

Introduction

Cooperative multi-agent systems exhibit significant advantages over single-agent systems. Enhanced robustness, efficiency, flexibility and execution speed are only some of the typically claimed advantages in many applications. Fully exploiting the benefits of multi-agent systems is not an easy task. One of the main design issues is related to choosing the right agent group composition. To this end, one needs to have a fundamental understanding of agent cooperation. This thesis builds insight in the origin of cooperation and focuses on a set of important multi-agent design issues, including optimal group sizes.

1.1 The Collaborative Search for Individually Rewarding Resources

In recent years, multi-agent navigation problems have become of major interest to many researchers, mainly driven by the potential benefits multi-agent systems pose compared to their single-agent equivalents. The main idea is that a group of multiple, possibly cheap and diverse agents executes tasks faster, more reliably, and more efficiently than a single agent.

To adequately situate the range of multi-agent problems this thesis addresses, we distinguish two general problem classes: the collaborative search for *globally* rewarding resources and the collaborative search for *individually* rewarding resources. This thesis focuses on a set of problems in the latter class. We discuss example applications and past and current work in the literature regarding both problem types.

A reward is denominated global if, once incurred by one member of the agent group, it is no longer a possible reward for any other agent. In other words, the group objective is obtained if all rewards are collected, regardless of whether one agent incurs all rewards, or each agent incurs some rewards. For example, in a reconnaissance mission [69, 70], or map building mission [60, 73, 78], if an agent observed and assessed the situation in one part of the environment, there is no reward for a second agent observing or mapping the same part of the environment. We find a second example in formation flight [3, 27, 28, 68]. One of the motivations behind formation flight stems from the energy savings multiple agents enjoy by reducing the drag in formation. The first agent in the formation is not able to reduce its drag, only the following agents are. Since the total energy spent is minimized, this problem indeed qualifies as part of the search for global rewards. Here, 'search' is interpreted as the process of finding the optimal agent configuration. A third example is the search and rescue problem [4, 54]. Similarly as in the reconnaissance problem, there is no reward to be collected in areas where previously an agent made observations.

A reward is denominated local if, once incurred by one member of the agent group, it is still a reward to be collected by any other member of the agent group. In other words, it is potentially beneficial for the agent group if more than one agent incurs the same reward, usually discovered by the first agent to incur the reward. A simple example occurs when a group of people visit a huge museum, with hidden treasures. The group can decide to spread out, while reporting discovery of such treasures by phone to the other group members. The latter then decide whether the reward is high enough for them to make the detour and incur the same reward. All people potentially incur the same reward by visually enjoying the treasure. The same reasoning applies to the search of particularly enjoyable landscapes while traveling with multiple cars along different routes, or even with multiple people in a single car, observing different directions. Whale watching is especially suited as example, since usually a whale is only visible for a limited amount of time and potentially reappears in completely different locations around the vessel. Agents can be interpreted as different passengers on one vessel, or, on a larger scale, as a set of vessels looking for whales.

A more technical example appears in competitive gliding [67]. The objective of a single glider in competition is to reach a known target in minimum time. To this end, the glider searches for thermals, cylinders of rising hot air, circling in which allows it to gain altitude or potential energy. Upon reaching sufficient height, the glider descends and gains speed, converting potential energy into kinetic energy and advancing towards the target, while searching for the next thermal if necessary [34, 51]. In some gliding competitions, cooperation is allowed, and in [67], the author reports cases where multiple cooperating gliders reach the target in shorter time than single gliders. Indeed, the probability of detecting a sufficiently powerful thermal increases when multiple gliders advance in parallel, covering a larger area. Each glider shares the position and the strength of the encountered thermals and based on this information, the other gliders possibly decide to deviate from their current trajectory, joining a thermal they would otherwise not have detected, thus incurring the same reward.

Another application arises in traffic management [44, 50]. Here, we consider a group of vehicles, part of a taxi or bus company, or a group of friends in different cars, entering a traffic-plagued city in as short a time as possible. Rather than relying on the traffic forecast, which is not always available, delayed or inaccurate, the vehicles spread out, each gathering information on the traffic situation of the route of choice. This information is shared, allowing each vehicle to choose an adequate (traffic-free) route in a better-informed manner. Multiple vehicles incur the same reward if each transfers to the same route.

In the military domain, various missions require a group of agents to traverse a danger ridden environment. Examples include troop transportation in urban warfare settings, or the traversal of a vehicle platoon through an unsafe area towards a search and rescue mission [30]. In these applications, it is clear that if one agent discovers a safe area, it is potentially beneficial for the other agents to traverse the same safe area on the way to the target. Again, more than one agent possibly incurs the same reward.

1.2 Multi-Agent Systems

Single agents are the building block of multi-agent systems; hence, it is worth presenting a brief overview of the state of art in this area. In fact, single-agent navigation has been addressed extensively and rather successfully in the past two decades. The single-agent motion planning problem consists of finding the control input to a vehicle such that it moves from an initial location to a final location, optimizing a certain objective function. The problem of basic motion planning in uncluttered environments for vehicles with general non-linear dynamics is treated effectively using techniques of optimal control [2, 13] and, related, dynamic programming [6]. More specific techniques for non-linear dynamics are described extensively in [71]. On a higher level, algorithms exist for navigation in obstacle cluttered environments in, based on various principles. We refer to [45, 46] for an overview. Recently, extensions for real-time solutions have been developed. For example, in [32] the authors present a motion planning algorithm which accounts for the computation time based on dynamic programming, while the work in [72] uses a mixed integer linear programming receding horizon approach to the same effect. We can safely conclude that the single-agent motion planning problem is reasonably well understood.

Multi-agent problems, however, pose a set of interesting new challenges (see [16] for an overview). As the union of single agents, we assume that once the individual tasks, target sequences, are determined, possibly in real-time, existing single-agent motion planning algorithms suffice for the lower level execution. Hence, we can focus on the higher level problems related to multi-agent systems; we distinguish: determining the optimal group size and composition, the adequate sensing and communication infrastructure, the decision structure and the corresponding (optimal) policies. In the following paragraphs, we give a brief overview of the state of the literature for each and summarize the aspects we address in this thesis, focusing on the collaborive search for individually rewarding resources.

Group Size

An essential question in the design of multi-agent systems is how many agents are required for satisfactory task completion. This includes the possibility of a heterogeneous agent group where specific subtasks are assigned in hardware to agent subsets of the group. For example, the group can consist of a set of small cheap observers gathering threat information in the environment, and thus providing necessary information to design safe strategies for expensive larger agents with core task capabilities. A specific military example is the transportation of personnel through a threat cluttered area. The optimal number of each agent type is the result of a trade-off between the increase in group performance and the increase in cost per additional agent, the latter including the agents fixed cost, and costs related to the complexity increase.

Several researchers address this problem, mostly based on simulation or experiments with some ad-hoc multi-agent strategy, confirming that the group performance improves with an increasing group size. For example, consider the problem of exploration of an unknown environment. One expects that n agents reduce the time to completion by a factor $1/n$, at least. In [15, 73], the authors propose an ad-hoc strategy where the utility of already explored areas is reduced. They show experimentally that the exploration time decreases sub-linearly with the number of agents for an example environment. In [5], the authors consider a multi-agent multi-target acquisition problem in an unsafe environment, and for a strategy based on a suboptimal hierarchical approach, they present performance results as a function of the number of agents assigned to each target for an example problem. The simulation results seem to indicate that assigning an equal number of agents to each target yields a better performance. Further, in [17], a multi-agent system is used for land-mine detection. Local, suboptimal strategies based on a potential field type approach are designed, and their performance is compared for different agent group sizes. In elevator control, multiple elevators reduce the average waiting time significantly. Results on optimal off-peak elevator positions and high traffic strategies can be found in [12] and [53]. The authors use an exact dynamic programming method with some simplifying assumptions. Simulations show average waiting time reductions of up to 70% with two to eight elevators.

Energy efficient formation flying is another area exhibiting a lot of research activity. Much attention goes to control strategies that stabilize formation flight, yet not many results are available that quantify the total energy saved. In [68] the authors consider a set of aircraft whereby each aircraft has a different initial and final position. They devise suboptimal strategies that allow subsets of the aircraft to join and travel in formation towards their respective targets. Simulation results for a ten agent case indicate a significant decrease of the energy spent per agent.

In each of these results, it is clear that multiple agents perform better than single agents, but rigorous statements quantifying this improvement lack, also recognized in [1]. However, some bounds on the optimal agent group performance exist. In particular, in [66], the authors present a proof that two and three agents explore a convex polygonal environment at least two and three times faster, respectively, than a single agent, under some fairly restrictive technical conditions. A similar result is described in [9, 10] where the authors treat a version of the dynamic traveling repairman problem. A group of m agents, moving at constant speed in the Euclidean plane is required to serve demands whose time of arrival and location are stochastic in nature. For infinite capacity per agent, they establish a lower bound to the system time (average wait plus service time) that decreases by a factor of

$1/m^2$ as compared to the single-agent case. An optimal policy for the light traffic case is provided, several heavy traffic suboptimal policies are analyzed. The decentralized version of this problem is treated in [31], with similar results.

In this thesis, we rigorously quantify the group performance as a function of the group size, for the problem of collaborative search for individually rewarding resources, with limited local environment information.

Local Observation

An essential property of truly autonomous agents is their ability to face unknown environments merely relying on observed or measured information. Several parameters of local observation zones have an influence on the multi-agent performance: the direction (angle over which observations are made), the range (the distance over which observations are made), and the timing. It seems clear that not all possibly gathered information is of equal importance. For example, in the target acquisition problem in an unknown environment, where agents observe a local environment area visually, it is unnecessary to first map the environment completely and then navigate towards the target in a known environment. It suffices to selectively make observations on those parts of the environment relevant for the particular target acquisition problem at hand. This idea appears in [37, 38, 36] where the authors treat a pursuit-evader problem in an unknown environment. The pursuers observe locally, mapping relevant parts of the environment, while focusing on acquiring the evader. A suboptimal greedy policy is presented whereby at each decision step, the agents move to the position with highest likelihood of encountering the evader. Further, in [60], the authors present an algorithm for a single agent navigating to a target at a known position, through an unknown environment. The agent repeatedly observes, focusing on the area in the target direction, and then navigates in the observed and thus known part of the environment towards the target. Extensions to multi-agent systems seem fairly straightforward.

Usually, researchers choose a particular local observation zone and study the resulting multi-agent system. A extensive part of the literature deals with visibility graphs as a model for local observation. That is, they assume each agent can observe what is not blocked by an obstacle. Other researchers assume a circle shaped local observation around each agent's current position. We find many examples in map building [66, 69, 73, 75] and search problems [65]. In [52], the authors use a scheme whereby each agent observes the state of the agents located within a certain radius R, and prove stability in the sense that all agents reach a consensus on the heading to adopt. Little to no effort goes in determining the relevance of the observations in the different observation directions.

In [4], we find an example where the observation range is at play. In particular, the authors consider a multi-agent navigation problem whereby the agents are required to reach a set of targets, while avoiding threats. The environment structure is known, threat locations are unknown. Each agent observes the environment over a distance equivalent to what can be reached in L time steps, and the authors study the agent performance under suboptimal policies, for one specific L. The extension of this work, whereby one investigates the effect of choosing different values for L on the agent performance, could provide insight in the value of local observations as a function of their distance to the current agent position.

Finally, in [58, 59], we find an interesting extension, whereby two agents use flashlights to identify each other as fast as possible. In this work, optimal policies are devised that indicate when it is opportune to make an observation, studying the observation timing.

In this thesis, we study different shapes and sizes of local observation zones, in order to determine and quantify the value of the observation directions and range towards the optimal multi-agent performance.

Communication

In multiple agent systems where each agent gathers local information, relative to its current location, communication is essential to devise optimal agent strategies based on the union of the individually gathered observations. In a centralized decision making scheme, one central computation unit, possibly one of the participating agents, gathers the observations, computes and communicates the optimal decisions to each agent. However, sometimes a decentralized decision structure is sufficient for the agents to take optimal coordinated decisions. Here, only the information that is relevant for each individual agent's decision process is required to be communicated. Indeed, in a two-agent example and for the case where the inter-agent distance is large, local information the first agent observes is irrelevant to the second agent, rendering the communication of observations of little use. This contrasts with the situation where two agents are at a small inter-agent distance; here, each agent can exploit the observations of the other agents, rendering communication necessary for optimality. Hence, there seems to be a quantifiable inter-agent distance beyond which communication is unnecessary.

In the literature, communication requirements and the effects on the agent performance due to communication limitations have been treated for specific problems. In particular, in [29], the authors consider the problem of formation flight, and study the minimum information required to be communicated between agents for formation stability. They employ graphs with varying topology representing the structure of the communication network and relate the formation stability to properties of the graph. A decentralized information exchange between vehicles is devised. In [55, 56] an agent swarm is interpreted as a reconfigurable sensor array, in search for increasing concentrations of the measured quantity. The agents communicate findings to a central unit, which computes the location of a set of virtual leaders, and communicates sufficient information to each agent so that it can stabilize with respect to the latter virtual leaders. A problem of a slightly different type can be found in [14, 19, 20], where an agent group is required to cover an area as efficiently as possible. Each agent has a sensing capability whose accuracy decreases as a function of the sensing distance. Further, a given function indicates environment areas of higher importance where accurate sensing is necessary. The authors propose a decentralized strategy where each agent gathers sufficient information of the agent positions around its own location to compute a moving direction. The agents are guaranteed to converge to a local optimum, *i.e.* a configuration which locally maximizes the integrated sensing quality, weighed with the location importance. In [47], the authors present a variation where the energy required for communication is directly included in the objective function.

In this thesis, we intend to study the communication requirements for the particular problem of cooperative search of individually rewarding resources. We intend to quantify the inter-agent distance beyond which locally observed information is irrelevant to other agents, thus rendering its communication of no use. In other words, we study the minimal information to be shared among the agents so that each is able to determine an optimal decision, and this as a function of the inter-agent distance.

Decision Structure and Agent Strategies

Finally, there is the need to devise an algorithm that, based on (part of) the current system state, devises a decision for each agent that is at best optimal according to a well-defined objective. In general, we distinguish two decision making schemes: a centralized and a decentralized one. In the former scheme, it is usually possible, at high computation cost, to compute the optimal strategy for each agent, while in the latter, one sacrifices optimality for a more computationally efficient approach, while maintaining essential system properties, for example stability. The choice of the decision making scheme is closely related to the system communication requirements, since centralized, optimal decision making in general requires full system state information.

Decentralized methods are popular in systems with a large number of agents, since its centralized equivalent is computationally too complex. For the problem where a group of agents exchange their values asynchronously and form weighted averages with values possessed by their neighbors, we refer to [79], where under weak conditions, the agents are guaranteed to converge to a common value. Some extensions are presented in [11]. We find more recent example applications of this work in [57, 76, 77] where each agent gathers information on the state of a limited number of agents around its current location (fixed and variable topology versions are considered). The authors prove stability in the sense that under this scheme and some technical conditions, the agent reach consensus on the heading. An approach validated by experiments can be found in [33], where fixed formations are controlled in a distributed fashion. In particular, each agent maintains a fixed relative position with respect to an *a priori* chosen second agent, using visual sensing. Other examples appear in [14,19,20], where a decentralized scheme is used to cover an area with a set of agents, in [80], where the cooperative search problem is handled in a distributed fashion, and in [28], where formation stability is established using local strategies.

In centralized methods, the full system state is available for strategy design. One can use an optimization tool of choice; dynamic programming is very common. We repeat that centralized method suffer from computational complexity issues, but yield optimal policies.

In this thesis, we study properties of optimal centralized agent strategies, detecting subspaces of the system state space where simple decentralized, or almost (*quasi*-) decentralized strategies are optimal.

1.3 Thesis Objectives

Throughout this thesis, we consider the problem of collaborative search for individually rewarding resources. On a general level, the main issue we intend to address, is the determining of the optimal (with respect to a well-defined objective function) agent group composition and size for this class of problems as a function of the task and the environment specifics. In particular, our objective is to give specific answers to the multi-agent system designer on how many agents of which capabilities are optimally required for successful task execution. To this end, we compute the group performance as a function of the group composition and size. Information on the agent fixed costs and the additional costs as a result of the system complexity then suffices to determine the optimal trade-off between performance and cost.

A second general objective of this thesis is to gain a deeper understanding and a more trustworthy intuition regarding multi-agent cooperation. In other words, we aim to understand the mechanics of cooperation in its simplest form, hoping that this insight will lend itself to more adequate multi-agent system design. We specifically stress that we refrain from using approximate methods to determine the optimal agent strategies, thus creating unbiased results which otherwise arise from suboptimal problem solutions. To this end, we analyze optimal policy properties of multi-agent groups operating in different environments. We specifically focus on discovering particular patterns in the optimal agent trajectories and on studying spatial agent distribution properties, which shed light on possible underlying simple cooperation dynamics.

Another objective of this thesis is to address issues related to the design of the local observation and measurement infrastructure of the multi-agent system. As mentioned earlier, autonomous agents are truly autonomous if they are capable of performing tasks in partially or completely unknown environments, merely relying on locally observed and possibly shared information. Depending on the task at hand, it seems clear that not all local information that can possibly be observed, is of equal importance to the performance of the agent group. For example, in a target acquisition problem, the information in the general direction of the agent movement is more important than environment information in other directions. We aim to study the influence of the different types of observation zones on the group performance for the problem involving individually rewarding resources.

Finally, a last objective of this thesis concerns the study of the communication requirements of multi-agent groups. In particular, we study the optimal agent strategies and analyze the communication requirements that are sufficient for the agents to compute optimal decisions at each system state. Behind this idea lies the reasoning that as the distance between two agents is large, local information one agent observes is of no immediate relevance to the other agent, rendering its communication of no use. Therefore, in this thesis, we focus on quantifying the inter-agent distance beyond which communication is unnecessary, for different environment specifics, hence simplifying the multi-agent system design and reducing the communication infrastructure requirements.

One of the main challenges in multi-agent centralized optimal policy design, with locally observed environment information, is the computational complexity. In [21], the authors prove NP-hardness [74] of a simplified version of the search problem whereby the searcher only has local environment information and minimizes the time to detection. The multi-agent extension is at least as complex, also recognized by the authors in [35, 64], who take a more general approach and present complexity results for different multi-agent problem types. We recognize the complexity issues related to multi-agent systems. However, we also recognize the importance of exact optimal solutions to possibly simple problems for insight building. In this thesis, we present an algorithm which deals with the inherent computational complexity sufficiently to solve non-trivial instances of the collaborative search for individually rewarding resources. The thus provided intuition serves as a great means to develop solid heuristics for the inevitable approximate approaches, required for large-scale systems [7].

1.4 Statement of Contributions

We summarize the thesis contributions as follows. First, we describe the results regarding the development of policy design and analysis tools; then we present the analysis results regarding the performance and properties of some multi-agent navigation problems. In particular:

- We model the collaborative search for individually rewarding resources as a traversal problem of a structured graph. The model simplicity allows to focus exclusively on agent cooperation, and provides a tool to test the optimal performance of relatively small multi-agent systems, creating benchmarks.
- We present and prove correctness of an algorithm which exploits the graph spatial invariance due to its infinite dimensions, to compute the optimal agent policies. The algorithm is sufficiently efficient to compute the exact solution for the navigation

problem with limited size agent groups. Further, the unboundedness of the navigation graph provides the additional advantage that the analysis results are independent of unpleasant effects due to an otherwise bounded graph.

- Finally, for analysis purposes, we present an algorithm which computes the agent state probability distribution in steady state given an optimal policy, for the graph traversal problem. We exploit the problem spatial invariance to deal with the infinite size problem state space.

The previous allows for the computation of the optimal policy for small multi-agent graph traversal problems. In what follows we cite the thesis contributions regarding the analysis results for different types of multi-agent designs. Specifically, we study the single, two- and three-agent navigation problems, with local observation zones ranging from wide angle cone-shaped views to narrow tunnel-shaped observation in the agent movement direction. In particular:

- We quantify the optimal agent performance as a function of the agent group size for a set of environment parameters and local observation zones. Two conclusions stand out:
  - In an environment with a cost structure similar to a mine-field, two agents with cone-shaped local observations provide a 25% performance increase as compared to a single agent.
  - The optimal performance of a three-agent vs. a two-agent group is not significantly better for short tunnel-shaped local observation zones and for a mine-field type environment. This is one instance where we can conclude that two agents suffice.

- For a fixed group size, we compare the optimal agent performance for several local observation zones, indicating which local areas of the environment contain more value towards the agent performance. We confirm and quantify the intuition that the more is observed, the better the performance. We also conclude that cone-shaped local information leads to a dramatic performance increase as compared to tunnel-shaped local information.

- We present characteristics of the optimal agent policies, to further our understanding of cooperative agent behavior. For example, we confirm and quantify the intuition that if rewards are scarce, the agents spread out in expected sense, decreasing the intersection of the respective local observation zones, thus gathering more environment information and increasing the probability of encountering a reward. Additionally, we observe that for any environment cost structure, and for a short tunnel-shaped local observation zone, the lateral distance between the members of a two-agent group is upper bounded in steady state under an optimal policy.

- Finally, from the analysis of optimal policies, we have that beyond a problem specific lateral inter-agent distance, the sharing of local observations is unnecessary for the local computation of an optimal policy. The optimal policy can therefore be considered as decentralized at the latter inter-agent distances, but remains necessarily centralized otherwise.

1.5 Thesis Outline

The thesis is organized as follows. In Chapter 2, we present some asymptotic properties of the multi-agent navigation problem, which we exploit in the development of an algorithm for the computation of optimal agent policies. We give a conceptual overview of the algorithm and introduce the 'vehicle' for our analysis: the multi-agent graph traversal problem. Then, in Chapter 3, we formulate the problem mathematically as a Dynamic Program (DP) and convert it using standard techniques into an equivalent Linear Program (LP) with infinitely many constraints. Exploiting the problem structure, we decompose the LP into an infinite set of coupled LPs with finitely many constraints and show that the unique bounded vector that solves the latter LPs represents the optimal value function, formally characterized as the unique bounded fixed point of a properly defined function F. In Chapter 4, we present and prove correctness of an algorithm to compute the latter fixed point of F for various multi-agent problems. It is computed explicitly and described implicitly in the form a recursion at small and large lateral inter-agent distances, respectively. In Chapter 5, we exploit the problem structure again to compute the system state probability distribution in steady state and under an optimal policy, for the same multi-agent problems and in Chapter 6, we present the analysis results. Finally, in Chapter 7, we summarize the thesis and suggest directions for future work.

Chapter 2

Asymptotic Problem Structure and Algorithm Concept

In this chapter, we introduce specific instances of the problem of collaborative search for individually rewarding resources in a target acquisition setting, in Section 2.1. Then, we study the asymptotic properties of the optimal agent strategies as the inter-agent distance is large and approaches infinity, in Section 2.2. For a time-invariant environment whereby the agents are constrained to advance towards a target spatially synchronously, the considerations regarding the agent behavior at large agent separations allow for the development of a sufficiently efficient algorithm to compute optimal agent policies. In Section 2.3, we provide an overview of the basic concepts that underlie the latter algorithm. Finally, in Section 2.4, we present the navigation problems on a structured graph that constitute the vehicle of our analysis.

2.1 Cooperative Target Acquisition

Many multi-agent problems exhibit the property that several agents can incur the same reward, yielding the individual reward structure introduced in Section 1.1. The main component of multi-agent tasks involves the navigation of an agent group towards a target, as efficiently as possible. Here, we associate a cost to different environment areas, reflecting its adversity to agent traversal. The notion of adversity envelops anything from the energy or time required to traverse, possible exposure to threats, visibility, etc, with the constraint that a reward (or negative cost) can be incurred by several agents. We define the local

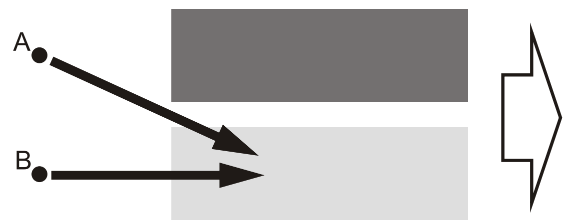

Figure 2-1: Agents A and B observe the dark and light area, respectively, and share this information. The dark area represents a high-cost region, while a low cost is associated with the light area. Agent A joins B, and both incur the lower cost.

area an agent observes as its *local observation zone*, while the *reachable zone* indicates the environment area an agent can reach given its current position and motion constraints. Given that each agent locally observes the environment state and shares this information, the group objective is to acquire the target by minimizing the sum of the incurred costs.

In this thesis, we constrain the agent movement further so that at all times the agents are positioned at an equal number of steps from the target, leaving their lateral separation as the unique degree of freedom. In other words, at each time step, the agents find an optimal decision, and move in *spatial synchrony*, towards the target. This requirement significantly reduces the problem complexity, but still allows for very rich, insight providing problems. Relaxing the spatial synchrony constraint is a challenging avenue for future work (see Section 7.2.4).

We give an example of a situation illustrating the possible benefits of cooperation. Consider a two-agent group positioned in spatial synchrony some number of steps away from the target. The two agents (A and B) are identical, and are required to advance in a general movement towards the target. Let the agent separation be such that the reachable zone of agent A overlaps with the local observation zone of the agent B. Due to the agent symmetry, the opposite holds as well. If agent B observes a particularly friendly environment area, only known to agent A by communication, then A can choose to join B in case its own observation indicates a particularly adverse local environment. Fig. 2-1 illustrates this example. The large arrow indicates the general target direction and each agent can deviate from that direction by some specific maximum angle. Two agents observe a larger part of the environment than a single agent could, and in this case the extra information possibly leads to a more cost-beneficial strategy.

In the general N-agent situation, two tendencies are key to the agent cooperative behav-

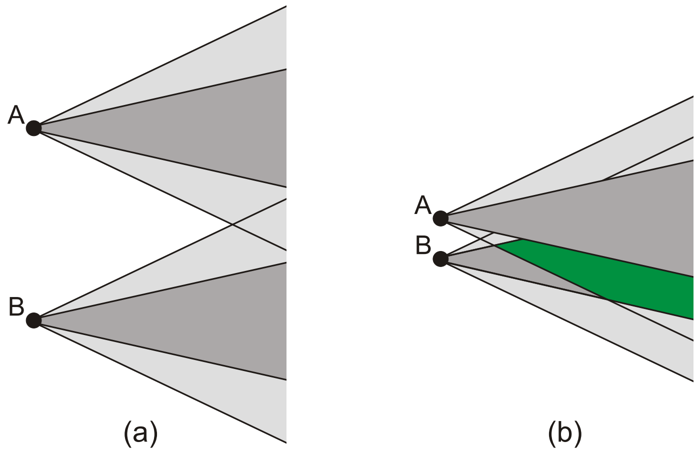

Figure 2-2: Dark and light gray areas represent the local observation zones and reachable zones, respectively, associated with the different agents. (a) Large agent separation: no observation zone overlap, but A cannot reach what B observes and vice versa. (b) Small agent separation: large information overlap, but A can reach what B (and not A) observes (green area).

ior, illustrated in Fig. 2-2 for a two-agent group. First, since more environment information yields an enhanced efficiency, the agents tend to spread, decreasing the size of the intersection of the local observation zones to an empty set (Fig. 2-2(a)). On the other hand, the agents tend to converge so that overlap occurs between the reachable zone of one agent and the local observation zone of another agent (Fig. 2-2(b)), at the cost of a local observation zone overlap and less total environment information. Only in a situation of this type can an agent take advantage of opportunities in the form of potentially cheap environment areas the other agent observes (green area in Fig. 2-2(b)) and exploit the benefit multiple agents have over single agents.

2.2 Optimal Policy at Large Agent Separations

In this section, we present insight regarding the agent behavior in the limiting case where one or more inter-agent separations approach infinity, for spatially synchronous problems. The insight is merely based on intuition and is presented without rigorous proof at this point. It serves as a tool to understand the reasoning at the basis of the algorithm to compute multi-agent optimal policies.

We start with the smallest multi-agent problem and consider a two-agent group ($N = 2$, where N denotes the group size) with a particular reachable and local observation zone associated with each agent. At large agent separation, the local observation zone of one agent does not overlap with the reachable zone of the other agent and vice versa. Hence, what one agent observes is of no relevance to the other agent, since the latter cannot reach the area the former observes. The opposite occurs at small separations. Indeed, the reachable and local observation zones of the respective agents overlap, and hence, each agent can exploit the extra relevant environment information, rendering the small separations particularly attractive. If the agents are at infinite separation, they cannot reach a small separation in finite time. Hence, the agents best behave as two single agents, both employing a single-agent optimal policy. At very large, but not infinite separation, it seems clear that the agents still adopt a single-agent optimal policy, but choose the particular optimal policy that minimizes the resulting separation at each decision step. In this manner, the agents 'hope' to reach the attractive small separations at some future time and incur the benefits of cooperation.

A similar reasoning holds for $N > 2$. There are $N - 1$ agent separations, one between each pair of neighboring agents. Consider the case where one separation approaches infinity, splitting the agent group in two smaller groups of sizes N_1 and N_2, where $N_1 + N_2 = N$. At infinite separation, the two groups behave independently. That is, the groups adopt policies which are optimal for groups of size N_1 and N_2, respectively. The combination of the two policies constitutes the global N-agent optimal policy for the case the group is split into groups of sizes N_1 and N_2. As in the two-agent case, at very large but not infinite separations, the two groups seem to best adopt an optimal N_1- and N_2-agent policy with the understanding that degrees of freedom in the optimal policies are used to minimize the resulting group separation at each decision step.

From the previous reasoning, we have that the optimal policy of a group of size N split in any possible way by a large separation, should consist of the combination of two policies, optimal for the respective group sizes in the split. It is intuitively clear that for each possible way to split a group of N agents, there exists a specific separation beyond which the latter two policies are adopted. Indeed, at large separations, the two groups behave independently, adopting the respective subgroup optimal policies; at small separations, the groups cooperate and a different policy is optimal. This observation is exploited to the fullest in the next chapters, where we present an algorithm to compute optimal agent policies. In the next section, we present the algorithm conceptually.

We make two further remarks. First, in the case more than two agent separations are infinitely large, it is clear that the resulting subgroups behave independently, adopting a policy optimal for each particular subgroup size. For very large, yet not infinitely large separations, it appears that each group still adopts such policy. However, it is not clear what happens in the case more than one policy is optimal for a particular subgroup. We give an example. Let the agent group be split in three subgroups of sizes N_1, N_2 and N_3, respectively, and such that $N = N_1 + N_2 + N_3$. Further, consider a system state where the middle subgroup (of size N_2) has multiple optimal decision available. It can use this decision freedom to decrease its separation with one of the two adjacent subgroups (of sizes N_1 and N_3). Rigorous analysis is necessary to assess this issue.

A second remark concerns the relaxation of spatial synchrony. In a time independent environment, the multi-agent strategy in the case longitudinal separation is allowed, is for the agents to sequentially traverse the environment, whereby each agent uses all previously observed information. That is, the first agent traverses the environment till reaching the target with only its own local observations as environment information. Then, the second agent advances to the target, using both its own observations and the observations of the first agent. This is continued till the last agent reaches the target. However, the problem becomes more interesting in a time-dependent environment, whereby area costs vary following a particular model, for example a Markov chain. Then, the equilibrium longitudinal agent separation reflects a trade-off between the size of the area on which information is observed, and the reliability of the observations. We leave the relaxation of spatial synchrony to future work (see Section 7.2.4) and limit ourselves to the spatially synchronous problem.

2.3 Algorithm Concept

In this section, we present the concept of an algorithm that computes the optimal agent policy, without entering into specific details. We consider a two-agent group. From the previous section, we have that only at separations where an overlap exists between the local observation zone of one agent and the reachable zone of the second agent, there is a possible opportunity for the two agents to incur a aggregate lower cost than in the non-cooperative case. The latter situation occurs at small agent separations. Further, in the previous section, we present intuition that at large agent separations the agents adopt single-agent optimal policies. Therefore, we assume the existence of a specific separation $\bar{s}$ such at all separations $s \geq \bar{s}$, a combination of single-agent optimal policies constitutes the two-agent optimal policy. The set of states for which $s < \bar{s}$ and for which $s \geq \bar{s}$, are connected and inter-dependent. That is, one cannot compute the optimal policy at the states for which $s < \bar{s}$ independently from the optimal policy computation at states for which $s \geq \bar{s}$ and vice versa.

In the algorithm design, we assume that the optimal two-agent policy for $s \geq \bar{s}$ consists of a combination of two single-agent optimal policies. Then, we solve for the optimal agent policy at $s < \bar{s}$, ensuring that the solution method contains the correct *connection constraints* between the sets of states for which $s < \bar{s}$ and for which $s \geq \bar{s}$, respectively. The connection constraints are extracted from the agent behavior at $s \geq \bar{s}$, based on the initial assumption, and are sufficient to determine the two-agent optimal policy at $s < \bar{s}$. In other words, the necessary information on the part of the state space at separations $s \geq \bar{s}$ that is required to render computing the optimal policy at $s < s$ a stand-alone problem, is summarized in the connection constraints. Finally, we need to check the correctness of the initial assumption on the optimal agent policy at the states for which $s \geq \bar{s}$.

For the N-agent case, we envision a hierarchical approach, whereby we compute the N-agent optimal policy given the optimal policies for groups of size $N-1$ and smaller. Let s_i, for $i = 1, 2, \ldots, N-1$, denote the separation between the ith and $(i+1)$th agent. Then, we solve for the optimal group policy for the agent separations $s_i < \bar{s}_i$, where $\bar{s}_i$ is defined similarly as in the previous paragraph and account for the various ways the agent group can split with the appropriate connection constraints. For further details, we refer to our remarks for future work in Section 7.2.3.

2.4 Notation and Problem Formulation: Navigation on a Structured Graph

In this section, we introduce the notation and formulate the N-agent navigation problem as a graph traversal problem with partial edge cost information. In particular, N agents are required to traverse a graph while minimizing the sum of the discounted costs of the traversed edges. To each agent is associated a set of edges whose costs the agents observe and share, which is the incentive for cooperation.

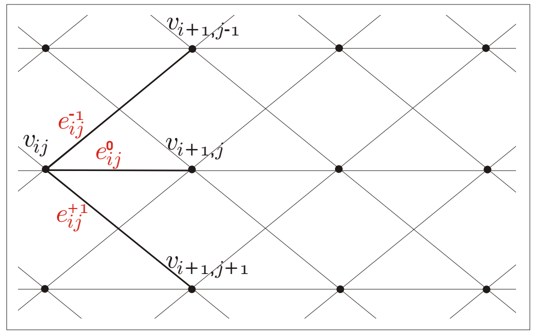

Figure 2-3: Notation related to graph G. Black dots represent vertices, lines are edges.

We grid the navigation terrain into sectors and associate a vertex $v \in V$ to each sector. Edges $e \in E$ connect pairs of vertices, which reduces the navigation problem into a graph traversal problem on the graph $G(V, E)$. We refer the reader to the literature for descriptions of algorithms to convert generic navigation terrains into a graph (see for example [45] and references therein). One typical example is the Voronoi graph for obstacle cluttered environments, where edges consist of points at equal distance from its closest obstacles (see for example [18]).

The structured transition graph we consider (see Fig. 2-3) consists of an infinite number of vertical lines of vertices, having an infinite number of vertices each. A horizontal vertex array is referred to as a *lane*, while a *stage* refers to a vertical line of vertices. Vertex v_{ij} denotes the vertex at stage i and lane j where the first stage and lane are chosen arbitrarily. Three edges e_{ij}^k, $k \in \{-1, 0, 1\}$, connect each vertex v_{ij} to the set of three vertices $v_{i+1,j+k}$ ($k \in \{-1, 0, 1\}$).

To each edge e_{ij}^k is associated a cost c_{ij}^k reflecting the edge traversal cost and the reward an agent collects upon reaching $v_{i+1,j+k}$. For example, if *a priori* information suggests that the environment sector associated with vertex $v_{i+1,j+k}$ is a particularly unsafe sector, then c_{ij}^k is relatively high. The traversal cost in practical applications relates, for instance, to the energy an agent requires to traverse an edge. In this work, we assume that the edge costs are independent identically distributed (i.i.d.) random variables picked from a finite set $\mathcal{L} = \{0, 1\}$. We denote p as the probability of encountering a zero edge cost. *A priori*, only the edge cost statistics are available and the edge costs are constant, *i.e.* independent of time.

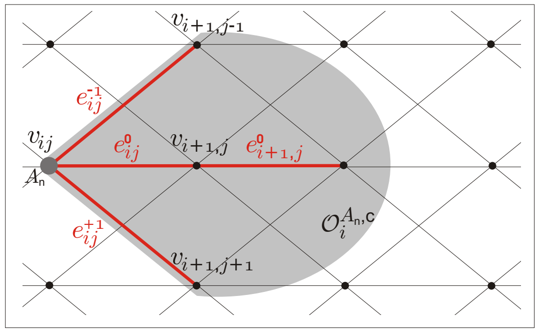

Figure 2-4: The gray area represents the cone-shaped local observation zone for agent A_n located at vertex v_{ij}. Formally, $\mathcal{O}_i^{A_n,c}$ is the set of red edges.

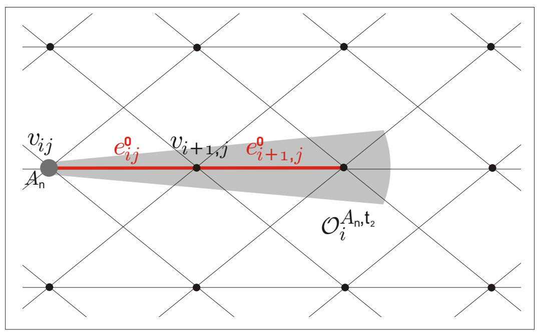

Figure 2-5: The gray area represents the short tunnel-shaped local observation zone for agent A_n located at vertex v_{ij}. Formally, $\mathcal{O}_i^{A_n,t_2}$ is the set of red edges.

In this thesis, we consider the graph traversal problem for a set of N agents, referred to as agents A_n, for $n = 1, 2, \ldots, N$. The graph traversal problem is modeled as a discrete time decision making problem where at time zero, agent A_n for $n = 1, 2, \ldots, N$ is positioned at the vertex at stage zero and on lane $l_0^{A_n}$ ($\in \mathbb{N}$ and with $l_0^{A_n} \leq l_0^{A_{n+1}}$, for $n = 1, 2, \ldots, N-1$). The subscript of lane l denotes the current time; the agents start at time zero. At each time step, each agent chooses to traverse one of the three available links to the next stage. Therefore, we let the time index coincide with the stage index and at time i the agents reach stage i at lanes $l_i^{A_n}$, for $n = 1, 2, \ldots, N$. Here, The agents are relabeled at each time step such that $l_i^{A_n} \leq l_i^{A_{n+1}}$, for $i \geq 0$, and for $n = 1, 2, \ldots, N$. Since at each time step, the agents are positioned at the same stage, we refer to this as *spatially synchronous* motion.

At each stage i, we associate a local observation zone to agent A_n, for $n = 1, 2, \ldots, N$, namely $\mathcal{O}_i^{A_n}$, a set of edges whose cost is observed and known. In this thesis, we define

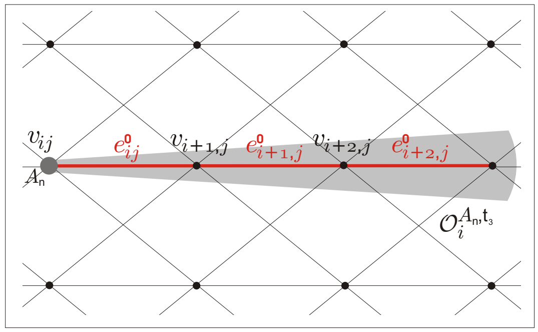

Figure 2-6: The gray area represents the extended tunnel-shaped local observation zone for agent A_n located at vertex v_{ij}. Formally, $\mathcal{O}_i^{A_n,t_3}$ is the set of red edges.

three types of local observation zones for agent A_n positioned at vertex $l_i^{A_n} = j$ (hence, A_n is located at vertex v_{ij}). First, there is the *cone-shaped* local observation zone (see Fig. 2-4), for which

$$\mathcal{O}_i^{A_n,c} = \{e^0_{ij}, e^0_{i+1,j}, e^{-1}_{ij}, e^1_{ij}\}.$$

Then, there is the *short tunnel-shaped* local observation zone (see Fig. 2-5), for which

$$\mathcal{O}_i^{A_n,t_2} = \{e^0_{ij}, e^0_{i+1,j}\}.$$

Finally, there is the *extended tunnel-shaped* local observation zone (see Fig. 2-6), for which

$$\mathcal{O}_i^{A_n,t_3} = \{e^0_{ij}, e^0_{i+1,j}, e^0_{i+2,j}\}.$$

The additional superscripts c, t_2, and t_3, to the general local observation zone symbol $\mathcal{O}_i^{A_n}$ for agent A_n at time/stage i, indicate the cone-shaped, short and extended tunnel-shaped local observation zones, respectively. In this thesis, we only consider problems where to each agent is associated the same type of local observation zone. Further, of unobserved edges, only the *a priori* available information, *i.e.* the edge cost statistics, is known. The term *local* stems from the fact that only edges in the neighborhood of the current agent position are included in $\mathcal{O}_i^{(\cdot)}$. At stage i, the set of agents incurs the costs associated with the edges traversed to reach stage $i+1$. Upon arrival at a vertex, each agent observes the costs of until then unknown edges and communicates these to the other agent. This provides an incentive for cooperation since observations by one agent possibly are of use to another agent.

The consideration of an infinitely wide and long graph involves an idealization with respect to practical path planning problems which have bounded environments. However, this idealization is analogous to the approach taken in control of dynamical systems, where a long but finite time horizon is typically modeled with infinite horizon tools. Limiting the graph width and length imposes boundary conditions on the problem which often do not lead to elegant and insightful solutions.

We formulate the following main problem.

Problem 1 ***[Main]*** *Let a set of N agents be located at positions $l_0^{A_n}$ on graph $G(V, E)$, for $n = 1, 2, \dots, N$. The edge costs are time invariant and i.i.d. random variables over $\mathcal{L}$ with probability p for a zero edge cost. The agents navigate spatially synchronously through the graph in the direction of increasing stage indices, infinitely far. Furthermore, the agents share the costs of the edges in their respective local observation zones perfectly and instantaneously upon reaching a vertex.*

Then, find the navigation strategy for the N agents so that the expected discounted sum of costs incurred at each stage is minimized. The expected value is taken over all initially unobserved edge costs; costs incurred at stage i are discounted by factor α^i, where $0 < \alpha < 1$.
$\square$

In this thesis, we solve Problem 1 for $N = 1$ and $N = 2$ three times, once for each type of local observation zone. Further, we solve Problem 1 for $N = 3$, where to each agent is associated the short tunnel-shaped local observation zone.

2.5 Chapter Summary

In this chapter, we introduce the multi-agent target acquisition problem, the focus of this thesis. The problem is of the collaborative search for individually rewarding resources type. In particular, the agent group objective is to reach a known target incurring the minimum aggregate environment cost, whereby more than one agent possibly incurs the same (low) cost. Each agent observes environment information in a local area around its current position.

We restrict the agents to spatially synchronous motion and present the intuitive notion that at infinite separations, the resulting agent subgroups behave independently, adopting a policy optimal for each particular subgroup size. We argue that at small separations, when there exists an overlap between the reachable zone and local observation zone of two neighboring agents, effective cooperation possibly takes place. Indeed, in such situation, each agent obtains relevant environment information from its neighbor, which it can exploit.

We present the concept of an algorithm to compute the optimal agent policy for a two-agent problem, exploiting the problem structure. In particular, we solve for the optimal agent policy at the states with separation $s < \bar{s}$ accounting for the agent behavior at states with separation $s \geq \bar{s}$ by means of a set of connection constraints.

Finally, we model the multi-agent target acquisition problem as a traversal problem of an unbounded, structured graph.

Chapter 3

The Dynamic Program as an Infinite Set of Linear Programs

In Chapter 2, we introduced the formal problem statement for a multi-agent target acquisition problem on a structured graph. In this chapter, we take the first steps towards solving this problem. In particular, in Section 3.1.1, we formulate the two-agent problem mathematically as a *Dynamic Program* (DP) with an infinitely large state space, for the cone-shaped, the short and the extended tunnel-shaped local observation zones. Then, we present an equivalent *Linear Program* (LP) with an infinite number of constraints in Section 3.1.2. Finally, in Section 3.2, we decompose the latter LP into an infinite set of coupled LPs with a finite number of constraints, exploiting the constraint structure. We show that the unique bounded simultaneous solution of the LPs represents the DP optimal value function. The results in this section are previously described in [24] and more extensively in [23].

3.1 Mathematical Problem Formulation

In Section 3.1.1, we present a DP formulation of the two-agent graph traversal problem, for any of the three local observation zones. Then, in Section 3.1.2, we transform the DP into an equivalent LP, whose structure we analyze in subsequent sections.

3.1.1 DP Formulation

We cast the two-agent navigation problem as a discounted cost, infinite horizon DP problem as follows. Since graph G exhibits spatial invariance properties in horizontal and vertical

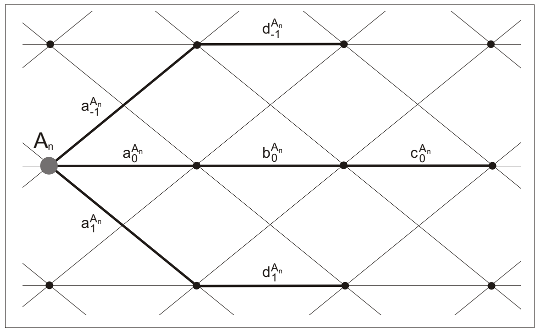

Figure 3-1: Definition of relevant edge costs relative to the current position of agent A_n.

direction, and since we have spatially synchronous motion, we choose the system state x, an element of set $\mathcal{S}_2$, to be independent from the current stage and from the absolute positions of the agent pair. For an N-agent problem, we denote the state space as $\mathcal{S}_N$, though we omit the subscript N when it is clear from the context which group size the state space is associated with. For ease of exposition, we refer to the cost of the edges in the local observation zone associated with an agent A_n (for $n = 1, 2$) relative to the current position of A_n, rather than in absolute manner. In particular, let

$$\begin{aligned} a_k^{A_n} &= c_i(l_i^{A_n}, k), \text{ for } k \in \{-1, 0, 1\}, \\ b_0^{A_n} &= c_{i+1}(l_i^{A_n}, 0), \\ c_0^{A_n} &= c_{i+2}(l_i^{A_n}, 0), \end{aligned}$$

for $n = 1, 2$ and where i is the current stage. Further, we define

$$d_m^{A_n} = c_{i+1}(l_i^{A_n} + m, 0), \text{ for } m = -1, 1,$$

for $n = 1, 2$, which are four edges that are not an element of $\mathcal{O}^{A_n}$ (for $n = 1, 2$), but whose cost is possibly observed at the previous time step in the case of the extended tunnel-shaped local observation zones. We refer to Fig. 3-1, for the edge cost definitions relative to the current observer location. This set of edge costs is sufficient to cover the at the current stage known and reachable edges for the three local observation zones considered. Let $\mathcal{C}$ denote the vector with as entries the cost of the latter edges. In other words, $\mathcal{C}$ contains the costs of the edges in $\mathcal{O}^{A_n}$, for $n = 1, 2$ and a subset of $\{d_m^{A_n}\}$, for $n = 1, 2$, and $m = -1, 1$.

We refer to Chapter 4 for the specific definitions of $\mathcal{C}$ for the three local observation zones considered. Then, $x = (s, \mathcal{C})$, where $s \in \mathbb{N}^+$ denotes the agent *separation*,

$$s = l^{A_2} - l^{A_1}.$$

Let $\mathcal{S}(s) \subset \mathcal{S}$ denote the set of states associated with separation s. Let $\mathbf{u} = (u^{A_1}, u^{A_2}) \in U$ denote the decision vector where $u^{(\cdot)} \in \{-1, 0, 1\}$ represents the three possible decisions available to each agent (*i.e.* traversing edges with cost $a_k^{(\cdot)}$, with $k \in \{-1, 0, 1\}$). The set $U = \{-1, 0, 1\}^2$ comprises the nine possible decision combinations for the agents and is invariant. Given x_i and $\mathbf{u}_i$ at time i, the agent cluster moves into the new state

$$x_{i+1} = f(x_i, \mathbf{u}_i), \tag{3.1}$$

where $f : \mathcal{S} \times U \to \mathcal{S}$ is the state transition function. The cost incurred in the transition from state x_i to x_{i+1} is

$$g(x_i, \mathbf{u}_i) = a_{u_{A_1}}^{A_1} + a_{u_{A_2}}^{A_2},$$

which is the sum of the edge costs of the edges agents A_1 and A_2 choose to traverse.

Let policy $\mu : \mathcal{S} \to U$ be a particular agent policy. Then, the expected discounted cost $J_\mu(x_0)$ the agents incur in advancing for an infinite number of time steps, given initial state x_0 and under policy μ is

$$J_\mu(x_0) = \lim_{N \to \infty} E\left[\sum_{i=0}^{N} \alpha^i g(x_i, \mu(x_i))\right],$$

subject to the state transition function (3.1), and where $0 < \alpha < 1$ is the discount factor. The cost function $J_\mu(x)$, for all $x \in \mathcal{S}$, satisfies

$$J_\mu(x) = E\left[g(x, \mu(x)) + \alpha J_\mu(f(x, \mu(x)))\right], \tag{3.2}$$

where we drop the time index i since we have time invariance and an infinite horizon. The discount factor α ensures that J_μ is well defined since with a bounded local cost g, the total cost incurred is bounded. Using the principle of optimality, the optimal value function

$J^*(x)$ satisfies the corresponding Bellman equation:

$$J^*(x) = \min_{\mathbf{u} \in U} E\left[g(x, \mathbf{u}) + \alpha J^*(f(x, \mathbf{u}))\right]. \tag{3.3}$$

A policy that minimizes the right hand side (RHS) of Eq. (3.3) is referred to as an optimal policy μ^*. Since the problem at hand has bounded local cost, we have that J^* is the unique bounded solution of the Bellman equation (see [6]). The optimal policy μ^* is not necessarily unique.

3.1.2 LP Formulation

We reformulate the DP defined in Section 3.1.1 as an equivalent LP. For any function $J : \mathcal{S} \to \mathbb{R}$, we denote the function obtained after applying the DP mapping as

$$(TJ)(x) = \min_{\mathbf{u} \in U} E[g(x, \mathbf{u}) + \alpha J(f(x, \mathbf{u}))], \qquad x \in \mathcal{S}.$$

Value iteration creates a sequence of functions $(T^k J) : \mathcal{S} \to \mathbb{R}$, for $k = 0, 1, \ldots$ and $x \in \mathcal{S}$, such that

$$\lim_{k \to \infty} (T^k J)(x) = J^*(x), \qquad x \in \mathcal{S}, \tag{3.4}$$

where T^k indicates the application of operator T with itself, k times, and for any bounded function $J : \mathcal{S} \to \mathbb{R}$. If J is chosen such that $J(x) \leq (TJ)(x)$, for all $x \in \mathcal{S}$, then, using the monotonicity property of DP (see for example [6]), we have that $(T^k J)(x) \leq (T^{k+1} J)(x)$, for $x \in \mathcal{S}$, and for $k = 0, 1, \ldots$. With the convergence property of value iteration [see Eq. (3.4)], it follows that $J(x) \leq J^*(x)$, for all $x \in \mathcal{S}$. Since we have a monotonously increasing sequence of functions $T^k J \leq J^*$ and since $J^* = TJ^*$, we have that J^* is the "largest" function J that satisfies $J \leq TJ$, which is equivalent to the set of constraints

$$J(x) \leq E\left[g(x, \mathbf{u}) + \alpha J(f(x, \mathbf{u}))\right], \qquad x \in \mathcal{S}, \mathbf{u} \in U. \tag{3.5}$$

In particular, J^* is the "upper right corner" of the polyhedron formed by the set of equality constraints (3.5). With the purpose of formulating the LP cost function and of rewriting the previous set of constraints in typical LP form, we define new variables. Let the vector

$\mathbf{x}^s \in \mathbb{R}^{n_s}$ contain the value function $J(x)$ for $x \in \mathcal{S}(s)$ as its entries, for $s = 0, 1, \ldots$. The dimension n_s, for $s = 0, 1, \ldots$, represents the number of different sets of known and reachable edges the agents can encounter at separation s. For clarity of the exposition, keeping the respective DP and LP notation conventions, we write in short that

$$J(x) \sim \mathbf{x}^s, \qquad x \in \mathcal{S}(s). \tag{3.6}$$

Let $\mathbf{e}$ denote a vector whose entries are all ones. Let $\mathbf{T}$ denote a linear operator so that, when applied to a vector $\mathbf{x}$ of size n, the vector $\mathbf{Tx}$, of size $n|U|$, has as its first $|U|$ entries the first entry of $\mathbf{x}$, the next $|U|$ entries are the second entry of $\mathbf{x}$, etc. In what follows, we assume the dimensions of $\mathbf{T}$ to be clear from the context. Then, we have the following equivalent LP:

$$\begin{array}{ll} \text{maximize} & \mathbf{e}^T \mathbf{x}^{\tilde{s}} \\ \text{subject to} & \mathbf{T}\mathbf{x}^s \leq \displaystyle\sum_{\sigma=\max\{-2,-s\}}^{2} \mathbf{A}_s^{\sigma} \mathbf{x}^{s+\sigma} + \mathbf{b}^s, \quad s = 0, 1, \ldots, \end{array} \tag{3.7}$$

where $(\cdot)^T$ denotes the transpose of a vector and for any $0 \leq \tilde{s} < \infty$. Since the optimal value function is the maximum element of the lattice [22] formed by the constraints in Eq. (3.7), the objective function can be any linear combination with positive coefficients of any set of entries of the vectors $\mathbf{x}^s$, $s = 0, 1, \ldots$, where the sum of the coefficients is finite. In the remainder of this paper, we adopt different versions according to our needs. We denote this LP by $\mathcal{LP}_g$. The matrices $\mathbf{A}_s^{\sigma}$, for $s = 0, 1, \ldots$, reflect the dependence of the constraint set associated with $\mathbf{x}^s$ [Eq. (3.7)] on $\mathbf{x}^{s+\sigma}$ and are deducted from Eq. (3.5) in a strenuous but straightforward manner. For $s \geq 2$, we have that $\mathbf{A}^{\sigma} = \mathbf{A}_s^{\sigma}$, for $\sigma = -2, \ldots, 2$.

In order to illustrate the LP problem structure adequately, we define the vectors

$$\begin{aligned} \mathbf{X}_g &= \begin{bmatrix} (\mathbf{x}^0)^T & (\mathbf{x}^1)^T & \ldots \end{bmatrix}^T, \\ \mathbf{B}_g &= \begin{bmatrix} (\mathbf{b}^0)^T & (\mathbf{b}^1)^T & \ldots \end{bmatrix}^T, \end{aligned}$$

and the matrix $\mathbf{A}_g$ such that Eq. (3.7) can be rewritten in typical LP-format as $(\mathbf{T} - \mathbf{A}_g)\mathbf{X}_g \leq \mathbf{B}_g$. Note that $\mathbf{X}_g$, $\mathbf{B}_g$, and $\mathbf{A}_g$ have infinite dimensions. Fig. 3-2 shows the block structure of $\mathbf{A}_g$. Empty squares are zero block matrices, the grey squares denote the

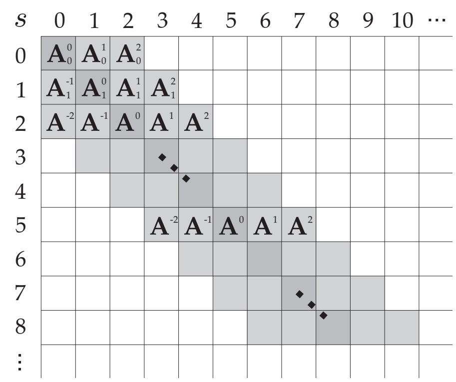

Figure 3-2: Matrix $\mathbf{A}_g$. Empty squares are zero-matrices, the grey squares denote the relevant matrices in Eq. (3.7).

relevant matrices that appear in Eq. (3.7). Since $\mathbf{T}$ is block-diagonal, the matrix $\mathbf{T} - \mathbf{A}_g$ has the same structure as $\mathbf{A}_g$. The LP has a general block multi-diagonal structure, but is not of the *staircase* type since non-zero blocks appear above and under the main diagonal. For staircase LPs, efficient distributed algorithms based on nested decomposition have been developed (see [40, 41] and references therein). In our problem, the optimal value function at separation $s \geq 2$ depends on the optimal value function at lower *and* greater separations (namely, at separations $s + \sigma$, for $\sigma = -2, \ldots, 2$), which renders the problem significantly different from staircase LPs where, in equivalent terms, there is dependence uniquely on the optimal value function at lower separations.

3.2 LP Decomposition and Solution as a Fixed Point

In this section we define a set of LPs $\mathcal{Q} = \{\mathcal{LP}(s)\}$, $s = 0, 1, \ldots$, where $\mathcal{LP}(s)$ refers to the LP associated with separation s. We show that $\mathcal{Q}$ is equivalent to $\mathcal{LP}_g$.

We have the following definition of $\mathcal{LP}(s)$ (for all $s \geq 0$).

Definition 1 *Let $\hat{\mathbf{x}}^{s-2}$, $\hat{\mathbf{x}}^{s-1}$, $\hat{\mathbf{x}}^{s+1}$ and $\hat{\mathbf{x}}^{s+2}$ be given vectors of appropriate dimensions. Then, we define $\mathcal{LP}(s)$ for $s = 0, 1, \ldots$ as*

$$\begin{aligned} &\textit{maximize} && \mathbf{e}^T \mathbf{x}^s \\ &\textit{subject to} && (\mathbf{T} - \mathbf{A}_s^0)\mathbf{x}^s \leq \sum_{\sigma \in S^s} \mathbf{A}_s^\sigma \hat{\mathbf{x}}^{s+\sigma} + \mathbf{b}^s, \end{aligned} \tag{3.8}$$

where $S^0 = \{1, 2\}$, $S^1 = \{-1, 1, 2\}$ and $S^s = \{-2, -1, 1, 2\}$, for $s \geq 2$. $\square$

The matrices $\mathbf{e}$, $\mathbf{T}$, $\mathbf{A}_s^\sigma$ and $\mathbf{b}^s$ of the appropriate dimensions are defined as in Section 3.1.2. Note that the RHS of the constraint set (3.8) is known.

In the rest of this section, we establish the equivalence between $\mathcal{LP}_g$ and the set $\mathcal{Q}$ of LPs. In particular, we show that the vectors $\mathbf{x}^s$, $s = 0, 1, \ldots$ of suitable dimension solve $\mathcal{LP}_g$ if and only if $\mathbf{x}^s$ solves $\mathcal{LP}(s)$ where $\hat{\mathbf{x}}^{\sigma+s}$ in Eq. (3.8) is equal to $\mathbf{x}^{\sigma+s}$ for $\sigma \in S^s$ and for all $s \geq 0$.

We first introduce some notation. Let $\mathcal{K} = \mathbb{R}^{\mathbb{N}}$. Define the vector $\mathbf{X} \in \mathcal{K}$ as the concatenation of the set of vectors $\{\mathbf{x}^s\}$ $(s = 0, 1, \ldots)$. In particular,

$$\mathbf{X}^T \doteq \begin{bmatrix} (\mathbf{x}^0)^T & (\mathbf{x}^1)^T & \ldots \end{bmatrix}^T, \tag{3.9}$$

where $\mathbf{x}^s \in \mathbb{R}^{n_s}$, for $s \geq 0$. The vector $\mathbf{X}^* \in \mathcal{K}$ is defined similarly from the set $\{\mathbf{x}^{s,*}\}$ (for $s = 0, 1, \ldots$), *i.e.*, the optimal value function and solution of $\mathcal{LP}_g$. The following function definition is key to the development in this section.

Definition 2 *We define the function $F : \mathcal{K} \to \mathcal{K} : F(\mathbf{X}) = \mathbf{X}'$, where $\mathbf{X}$ and $\mathbf{X}'$ are the concatenation of respective sets $\{\mathbf{x}^s\}$ and $\{(\mathbf{x}^s)'\}$ $(s = 0, 1, \ldots)$ [see Eq. (3.9)], and where $(\mathbf{x}^s)'$ solves $\mathcal{LP}(s)$ with constraint parameters [see Eq. (3.8)] $\hat{\mathbf{x}}^{s+\sigma} = \mathbf{x}^{s+\sigma}$, for $\sigma \in S^s$.* $\square$

We show that $\mathcal{Q}$ is equivalent to $\mathcal{LP}_g$ by establishing that $\mathbf{X}^*$ is the unique bounded fixed point of F, presented in the remaining part of this section. We essentially prove that F is a contraction.

We need the following set of auxiliary problems $\mathcal{P}(s)$, one for each separation $s \geq 0$. Let the set $\mathcal{M}_s = \bigcup_{\sigma \in S^s} \mathcal{S}(s + \sigma)$. Then, each auxiliary problem is an optimal stopping version of the Main Problem 1, as follows:

Problem 2 ***[Auxiliary]*** *Let two agents be positioned on an arbitrary stage of the graph $G(V, E)$ at separation $s \geq 0$. The edge costs are time invariant and i.i.d. random variables over $\mathcal{L}$ with probability p for a zero edge cost. The agents navigate spatially synchronously through the graph in the direction of increasing stage indices. Furthermore, the agents share the costs of the edges in their respective local observation zones perfectly and instantaneously upon reaching a vertex. Let $g(x, \mathbf{u})$ be defined as before as the sum of the edge costs the two agents incur at state x and choosing decision $\mathbf{u}$, and let $f(x, \mathbf{u})$ be the state transition function. Let $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$ be any bounded function. Two situations occur:*

- *The current agent separation equals s. Then, the local cost incurred in advancing to the next stage equals $g(x, \mathbf{u})$ if $f(x, \mathbf{u}) \in \mathcal{S}(s)$ and $g(x, \mathbf{u}) + \alpha \hat{J}(f(x, \mathbf{u}))$ otherwise.*

- *The current agent separation differs from s. Then, the process terminates and the agents do not incur further costs.*

Then, find the navigation strategy for both agents so that the expected discounted sum of costs incurred at each stage is minimized (discount factor $0 < \alpha < 1$). $\square$

In other words, in Problem 2, the agents stick to separation s incurring the associated aggregate edge costs, or move to a separation differing from s incurring a one-time cost and stop.

Before we relate the solution of problem $\mathcal{P}(s)$ to the solution of $\mathcal{LP}(s)$, we define operator $T_{s,\hat{J}_s}$ as follows.

Definition 3 *Given an arbitrary function $J_s : \mathcal{S}(s) \to \mathbb{R}$, and an associated bounded function $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$ then, for all $x \in \mathcal{S}(s)$,*

$$(T_{s,\hat{J}_s} J_s)(x) \quad = \quad \min_{\mathbf{u} \in U} E\left[v(x, \mathbf{u}) + \alpha V_s(f(x, \mathbf{u}))\right], \tag{3.10}$$

where

$$v(x, \mathbf{u}) = \begin{cases} g(x, \mathbf{u}), & f(x, \mathbf{u}) \in \mathcal{S}(s) \\ g(x, \mathbf{u}) + \alpha \hat{J}_s(f(x, \mathbf{u})), & f(x, \mathbf{u}) \in \mathcal{M}_s, \end{cases} \tag{3.11}$$

and where

$$V_s(x) = \begin{cases} J_s(x), & x \in \mathcal{S}(s) \\ 0, & x \in \mathcal{M}_s. \end{cases} \tag{3.12}$$

$\square$

In words, $T_{s,\hat{J}_s}$ executes one value iteration step for the DP associated with Problem 2, given $\hat{J}_s$ and J_s. From the traditional value iteration results, we have that for an arbitrary bounded function $J_s : \mathcal{S}(s) \to \mathbb{R}$ and a given bounded associated function $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$

$$\tilde{J}_s(x) = \lim_{k \to \infty} \left(T^k_{s,\hat{J}_s} J_s\right)(x), \qquad x \in \mathcal{S}(s), \tag{3.13}$$

where $\tilde{J}_s(x)$ ($x \in \mathcal{S}(s)$) denotes the optimal value function for Auxiliary Problem 2, where in the notation we do not explicitly mention the dependence of $\tilde{J}_s$ on the problem parameter $\hat{J}_s$.

The following lemma relates $\tilde{J}_s$ to the solution of $\mathcal{LP}(s)$.

Lemma 1 *Let $\tilde{\mathbf{x}}^s \sim \tilde{J}_s(x)$, for $x \in \mathcal{S}(s)$, where $\tilde{J}_s$ is computed as in Eq. (3.13), given any bounded function $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$. Then $\tilde{\mathbf{x}}^s$ solves $\mathcal{LP}(s)$, with $\hat{\mathbf{x}}^{s+\sigma} \sim \hat{J}_s(x)$, for $x \in \mathcal{S}(s+\sigma)$, and $\sigma \in S^s$.* $\square$

Proof: With the LP formulation of the DP associated with the Auxiliary Problem 2 (see Section 3.1.2 for details), we have that $\tilde{J}_s$ is the solution to the following LP:

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{x \in \mathcal{S}(s)} J_s(x) \\ \text{subject to} & J_s(x) \le E[v(x,\mathbf{u}) + \alpha V_s(f(x,\mathbf{u}))], \qquad x \in \mathcal{S}(s), \mathbf{u} \in U, \end{array}$$

with $v(x,\mathbf{u})$ and $V_s(x,\mathbf{u})$ as defined in Eqs. (3.11,3.12). This LP is equivalent to an LP with the same cost function but with constraints $J_s(x) \le E[g(x,\mathbf{u}) + \alpha \hat{V}_s(f(x,\mathbf{u}))]$, for all $x \in \mathcal{S}(s)$, and $\mathbf{u} \in U$, and where

$$\hat{V}_s(x) = \begin{cases} J_s(x), & x \in \mathcal{S}(s) \\ \hat{J}_s(x), & x \in \mathcal{M}_s. \end{cases}$$

With $\mathbf{x}^s \sim J_s(x)$, for $x \in \mathcal{S}(s)$, and for $\hat{\mathbf{x}}^{s+\sigma} \sim \hat{J}_s(x)$, for $x \in \mathcal{S}(s)$, $\sigma \in S^s$, the equivalence of the last LP with $\mathcal{LP}(s)$ follows easily. Hence, $\tilde{\mathbf{x}}^s \sim \tilde{J}_s$, for $x \in \mathcal{S}(s)$ is the solution to $\mathcal{LP}(s)$ and the proof is complete. $\blacksquare$

The following lemma establishes an essential monotonicity property of $T_{s,\hat{J}_s}$.

Lemma 2 *For any two functions $J_s : \mathcal{S}(s) \to \mathbb{R}$ and $J'_s : \mathcal{S}(s) \to \mathbb{R}$ and for any two bounded functions $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$ and $\hat{J}'_s : \mathcal{M}_s \to \mathbb{R}$ such that*

$$\begin{aligned} J_s(x) &\le J'_s(x), \qquad x \in \mathcal{S}(s), \\ \hat{J}_s(x) &\le \hat{J}'_s(x), \qquad x \in \mathcal{M}_s, \end{aligned} \tag{3.14}$$

we have

$$(T^k_{s,\hat{J}_s} J_s)(x) \le (T^k_{s,\hat{J}'_s} J'_s)(x), \qquad x \in \mathcal{S}(s),\ k = 1, 2, \ldots \tag{3.15}$$

$\square$

Proof: We use induction. For $k = 0$, Eq. (3.15) holds trivially by Eq. (3.14). For the induction hypothesis, assume

$$(T^k_{s,\hat{J}_s} J_s)(x) \leq (T^k_{s,\hat{J}'_s} J'_s)(x), \qquad x \in \mathcal{S}(s), \tag{3.16}$$

for some k, then we have, for $x \in \mathcal{S}(s)$,

$$\begin{aligned}(T^{k+1}_{s,\hat{J}_s} J_s)(x) &= \min_{\mathbf{u}\in U} E\left[v(x,\mathbf{u}) + \alpha(V^k_s)(f(x,\mathbf{u}))\right] \\ &\leq \min_{\mathbf{u}\in U} E\left[v'(x,\mathbf{u}) + \alpha(V'^k_s)(f(x,\mathbf{u}))\right] \\ &= (T^{k+1}_{s,\hat{J}'_s} J'_s)(x).\end{aligned} \tag{3.17}$$

Here, $v(x,\mathbf{u})$ and $v'(x,\mathbf{u})$ are as defined in Eq. (3.11), with $\hat{J}_s(x)$ and $\hat{J}'_s(x)$, respectively. Further,

$$V^k_s(x) = \begin{cases} (T^k_{s,\hat{J}_s} J_s)(x), & x \in \mathcal{S}(s), \\ 0, & x \in \mathcal{M}_s. \end{cases}$$

The function V'^k_s is defined similarly. In Eq. (3.17), we use induction hypothesis (3.16) for the inequality relation and twice the definition of $T_{s,\hat{J}_s}$ [Eq. (3.10)]. This concludes the induction and hence the proof. ■

Let $e : \mathcal{S} \to \mathbb{R}$ denote the unity function such that $e(x) = 1$, for all $x \in \mathcal{S}$. Then, we have the following property of $T_{s,\hat{J}_s}$.

Lemma 3 *For $k = 1, 2, \ldots$, for any function $J_s : \mathcal{S}(s) \to \mathbb{R}$, for any bounded function $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$ and for any scalar $r \geq 0$, we have*

$$(T^k_{s,\hat{J}_s+re}(J_s + re))(x) \leq (T^k_{s,\hat{J}_s} J_s)(x) + \alpha r, \qquad x \in \mathcal{S}(s), \tag{3.18}$$

$$(T^k_{s,\hat{J}_s-re}(J_s - re))(x) \geq (T^k_{s,\hat{J}_s} J_s)(x) - \alpha r, \qquad x \in \mathcal{S}(s). \tag{3.19}$$

□

Proof: We give the proof of Eq. (3.18) by induction (the proof of Eq. (3.19) is similar). From the definition of $T_{s,\hat{J}_s}$ [Eq. (3.10)], we have

$$(T_{s,\hat{J}_s+re}(J_s \pm re))(x) = (T_{s,\hat{J}_s} J_s)(x) \pm \alpha r, \qquad x \in \mathcal{S}(s). \tag{3.20}$$

For $k = 1$, Eq. (3.18) holds, by Eq. (3.20). Assume, as induction hypothesis that

$$(T^k_{s,\hat{J}_s+re}(J_s + re))(x) \leq (T^k_{s,\hat{J}_s} J_s)(x) + \alpha r, \quad x \in \mathcal{S}(s).$$

Then, applying operator $T_{s,\hat{J}_s+re}$ to both sides of the previous equation, and using Lemma 2, yields

$$\left[T^{k+1}_{s,\hat{J}_s+re}(J_s + re)\right](x) \leq \left[T_{s,\hat{J}_s+re}\left((T^k_{s,\hat{J}_s} J_s) + \alpha re\right)\right](x), \quad x \in \mathcal{S}(s). \tag{3.21}$$

Two cases occur:

- First, consider the set of pairs of $x \in \mathcal{S}(s)$ and $\mathbf{u} \in U$ such that $f(x, \mathbf{u}) \in \mathcal{S}(s)$. Then, for the RHS of Eq. (3.21), we have the following inequality:

$$\begin{aligned}
\left[T_{s,\hat{J}_s+re}\left((T^k_{s,\hat{J}_s} J_s) + \alpha re\right)\right](x) &\leq \min_{\mathbf{u}} E\left[g(x, \mathbf{u}) + \alpha\left((T^k_{s,\hat{J}_s} J_s)\,(f(x, \mathbf{u})) + \alpha r\right)\right], \\
&= \min_{\mathbf{u}} E\left[g(x, \mathbf{u}) + \alpha(T^k_{s,\hat{J}_s} J_s)(f(x, \mathbf{u}))\right] + \alpha^2 r, \\
&\leq \min_{\mathbf{u}} E\left[g(x, \mathbf{u}) + \alpha(T^k_{s,\hat{J}_s} J_s)(f(x, \mathbf{u}))\right] + \alpha r,
\end{aligned} \tag{3.22}$$

  where, for all $x \in \mathcal{S}(s)$, the minimization is over all $\mathbf{u} \in U$ such that $f(x, \mathbf{u}) \in \mathcal{S}(s)$.

- Second, consider the set of pairs of $x \in \mathcal{S}(s)$ and $\mathbf{u} \in U$ such that $f(x, \mathbf{u}) \in \mathcal{M}_s$. Then, for the RHS of Eq. (3.21), we have the following inequality:

$$\begin{aligned}
\left[T_{s,\hat{J}_s+re}\left((T^k_{s,\hat{J}_s} J_s) + \alpha re\right)\right](x) &\leq \min_{\mathbf{u}} \left[g(x, \mathbf{u}) + \alpha(\hat{J}_s(f(x, \mathbf{u})) + r)\right], \\
&= \min_{\mathbf{u}} \left[g(x, \mathbf{u}) + \alpha\hat{J}_s(f(x, \mathbf{u}))\right] + \alpha r,
\end{aligned} \tag{3.23}$$

  where, for all $x \in \mathcal{S}(s)$, the minimization is over all $\mathbf{u} \in U$ such that $f(x, \mathbf{u}) \in \mathcal{M}_s$.

Combining Eqs. (3.22) and (3.23), yields

$$\left[T_{s,\hat{J}_s+re}\left((T^k_{s,\hat{J}_s} J_s) + \alpha re\right)\right](x) \quad \leq \quad \left[T^{k+1}_{s,\hat{J}_s} J_s\right](x) + \alpha r, \quad x \in \mathcal{S}(s),$$

and with Eq. (3.21), we have

$$\left[T^{k+1}_{s,\hat{J}_s+re}(J_s + re)\right](x) \leq \left[T^{k+1}_{s,\hat{J}_s} J_s\right](x) + \alpha r, \quad x \in \mathcal{S}(s),$$

which concludes the induction and the proof. ■

We now establish two properties of F, based on which we show that F is a contraction. We first introduce some notation. For an arbitrary $\mathbf{X}^0 \in \mathcal{K}$, define $\mathbf{X}^k \in \mathcal{K}$, for $k = 1, 2, \ldots$ as

$$\begin{aligned} \mathbf{X}^k &= F(\mathbf{X}^{k-1}), \\ &= \left[\begin{array}{ccc} (\mathbf{x}^{0,k})^T & (\mathbf{x}^{1,k})^T & \ldots \end{array} \right]^T. \end{aligned} \tag{3.24}$$

Vector $(\mathbf{X}')^k \in \mathcal{K}$ is defined similarly. We establish the first property of F, monotonicity.

Lemma 4 *For any two bounded vectors $\mathbf{X}^0, (\mathbf{X}')^0 \in \mathcal{K}$ such that*

$$\mathbf{X}^0 \leq (\mathbf{X}')^0, \tag{3.25}$$

we have

$$F^k(\mathbf{X}^0) \leq F^k((\mathbf{X}')^0), \quad k = 1, 2, \ldots \tag{3.26}$$

□

Proof: We use induction. For $k = 0$, Eq. (3.26) holds trivially by Eq. (3.25). For the induction hypothesis, assume that $F^k(\mathbf{X}^0) \leq F^k((\mathbf{X}')^0)$, or using the notation in Eq. (3.24), we have that

$$\mathbf{X}^k \leq (\mathbf{X}')^k, \tag{3.27}$$

for some k. From Eq. (3.24), we have that $F(\mathbf{X}^k) = \mathbf{X}^{k+1}$, and with Lemma 1, this yields for all $s = 0, 1, \ldots$,

$$\mathbf{x}^{s,k+1} \sim \left(\lim_{l \to \infty} T^l_{s, \hat{J}^k_s} J^k_s \right)(x), \quad x \in \mathcal{S}(s),$$

where $J^k_s(x) \sim \mathbf{x}^{s,k}$, for $x \in \mathcal{S}(s)$ and where $\hat{J}^k_s(x) \sim \mathbf{x}^{s+\sigma,k}$, for $x \in \mathcal{S}(s+\sigma)$, for all $\sigma \in S^s$. A similar statement holds for $(\mathbf{x}')^{s,k+1}$. From Lemma 2, and with the induction hypothesis (3.27), we have that for $s = 0, 1, \ldots$

$$\left(T^l_{s, \hat{J}^k_s} J^k_s \right)(x) \leq \left(T^l_{s, (\hat{J}'_s)^k} (J'_s)^k \right)(x), \quad x \in \mathcal{S}(s), \tag{3.28}$$

for $l = 0, 1, \ldots$. Therefore, taking the limit for $l \to \infty$ in Eq. (3.28), we have that $\mathbf{x}^{s,k+1} \leq (\mathbf{x}')^{s,k+1}$, for $s = 0, 1, \ldots$, from which it follows that $\mathbf{X}^{k+1} \leq (\mathbf{X}')^{k+1}$, concluding the induction and hence the proof. ■

The following lemma provides the second property of F required to show that F is a contraction.

Lemma 5 *For $k = 1, 2, \ldots$, for any bounded vector $\mathbf{X} \in \mathcal{K}$, and for any scalar $r \geq 0$, we have*

$$F^k(\mathbf{X} + r\mathbf{e}) \leq F^k(\mathbf{X}) + \alpha^k r\mathbf{e}, \tag{3.29}$$

$$F^k(\mathbf{X} - r\mathbf{e}) \geq F^k(\mathbf{X}) - \alpha^k r\mathbf{e}, \tag{3.30}$$

where $\mathbf{e} \in \mathcal{K}$ contains ones as its entries. □

Proof: We prove Eq. (3.29). The proof of Eq. (3.30) is similar. We define the functions $J_s : \mathcal{S}(s) \to \mathbb{R}$, for $s = 0, 1, \ldots$, such that

$$J_s(x) \sim \mathbf{x}^s, \quad x \in \mathcal{S}(s), \tag{3.31}$$

with

$$\mathbf{X} = \begin{bmatrix} (\mathbf{x}^0)^T & (\mathbf{x}^1)^T & \ldots \end{bmatrix}^T.$$

Similarly, we define the functions $\hat{J}_s : \mathcal{M}_s \to \mathbb{R}$ such that

$$\hat{J}_s(x) \sim \{\mathbf{x}^{s+\sigma}\}, \quad x \in \mathcal{M}_s, \sigma \in S^s. \tag{3.32}$$

From Lemma 3, we have

$$\left(T^l_{s,\hat{J}_s+re}(J_s + re)\right)(x) \leq (T^l_{s,\hat{J}_s} J_s)(x) + \alpha r, \quad x \in \mathcal{S}(s),$$

and taking the limit for $l \to \infty$, with Eqs. (3.31,3.32) and with Lemma 1, this yields

$$F(\mathbf{X} + r\mathbf{e}) \leq F(\mathbf{X}) + \alpha r\mathbf{e}. \tag{3.33}$$

We use induction. For $k = 1$, Eq. (3.29) holds by Eq. (3.33). For the induction hypothesis, suppose $F^k(\mathbf{X} + r\mathbf{e}) \leq F^k(\mathbf{X}) + \alpha^k r\mathbf{e}$. Applying F to both sides of the previous equation, and using Lemma 4, yields

$$\begin{aligned} F^{k+1}(\mathbf{X} + r\mathbf{e}) &\leq F(F^k(\mathbf{X}) + \alpha^k r\mathbf{e}), \\ &\leq F^{k+1}(\mathbf{X}) + \alpha(\alpha^k r\mathbf{e}), \\ &= F^{k+1}(\mathbf{X}) + \alpha^{k+1} r\mathbf{e}, \end{aligned}$$

where the second inequality follows from Eq. (3.33). This concludes the induction, and hence the proof. ■

The following lemma provides the basis for main result of this section.

Lemma 6 *The function F is a contraction mapping. In particular, given any two bounded vectors $\mathbf{X}, \mathbf{X}' \in \mathcal{K}$, we have that*

$$\max |F^k(\mathbf{X}) - F^k(\mathbf{X}')| \leq \alpha^k \max |\mathbf{X} - \mathbf{X}'|,$$

where the $\max(\cdot)$ *is taken over all components of its argument.* □

Proof: The proof, based on Lemmas 4 and 5, is identical to the proof of the fact that operator T is a contraction as can be found in [6]. We give it here for completeness. Let $m = \max |\mathbf{X} - \mathbf{X}'|$. It follows that

$$\mathbf{X} - m\mathbf{e} \leq \mathbf{X}' \leq \mathbf{X} + m\mathbf{e}. \tag{3.34}$$

We apply F^k to both sides of Eq. (3.34) and use Lemmas 4 and 5 to obtain

$$F^k(\mathbf{X}) - \alpha^k m\mathbf{e} \leq F^k(\mathbf{X}') \leq F^k(\mathbf{X}) + \alpha^k m\mathbf{e},$$

which yields

$$\max |F^k(\mathbf{X}) - F^k(\mathbf{X}')| \leq \alpha^k m,$$

which concludes the proof. ■

The following lemma provides an essential property of contraction mappings (see [6]).

Lemma 7 *If the function $F : \mathcal{K} \rightarrow \mathcal{K}$ is a contraction mapping, then there exists a unique fixed point of F, within the class of bounded functions. In particular, there exists a unique bounded vector $\mathbf{X}^* \in \mathcal{K}$ such that $F(\mathbf{X}^*) = \mathbf{X}^*$. Furthermore, if $\mathbf{X}$ is any bounded vector in $\mathcal{K}$, then*

$$\lim_{k\rightarrow\infty} \|F^k(\mathbf{X}) - \mathbf{X}^*\| = 0.$$

$\square$

Proof: For a proof, see [48] and [49]. $\blacksquare$

We have now left to show that the solution of $\mathcal{LP}_g$ is the unique fixed point of F. The following theorem provides the main result.

Theorem 1 *The solution $\mathbf{X}^*$ of $\mathcal{LP}_g$ is the unique bounded fixed point of F.* $\square$

Proof: From Lemma 6, we have that F is a contraction and from Lemma 7, F has a unique bounded fixed point. Therefore, we only need to show that the solution of $\mathcal{LP}_g$, *i.e.*

$$\mathbf{X}^* = \begin{bmatrix} (\mathbf{x}^{0,*})^T & (\mathbf{x}^{1,*})^T & \dots \end{bmatrix}^T,$$

is a fixed point of F. For a all $s \geq 0$, let

$$\hat{\mathbf{x}}^{s+\sigma} = \mathbf{x}^{s+\sigma,*}, \qquad \sigma \in S^s. \tag{3.35}$$

Furthermore, let the solution of $\mathcal{LP}(s)$, with $\hat{\mathbf{x}}^{s+\sigma}$ ($\sigma \in S^s$) as in Eq. (3.35), be denoted as $\tilde{\mathbf{x}}^s$. Then, for $\mathbf{X}^*$ to be a fixed point of F, we need to show that for all $s = 0, 1, 2, \dots$ we have that

$$\tilde{\mathbf{x}}^s = \mathbf{x}^{s,*}. \tag{3.36}$$

Since $\mathbf{X}^*$ solves $\mathcal{LP}_g$, $\mathbf{x}^{s,*}$ is also a feasible solution of $\mathcal{LP}(s)$, with $\hat{\mathbf{x}}^{s+\sigma}$ as in Eq. (3.35). It follows that $\tilde{\mathbf{x}}^s \geq \mathbf{x}^{s,*}$. To establish Eq. (3.36), we show that $\tilde{\mathbf{x}}^s \leq \mathbf{x}^{s,*}$, by contradiction. Assume that

$$\tilde{\mathbf{x}}^s \geq \mathbf{x}^{s,*}, \tag{3.37}$$

with for at least one i

$$\tilde{x}_i^s > x_i^{s,*}, \tag{3.38}$$

where x_i^s denotes the ith component of $\mathbf{x}^s$. We show that

$$\tilde{\mathbf{X}}^s = \left[\begin{array}{cccccccc} (\mathbf{x}^{0,*})^T & (\mathbf{x}^{1,*})^T & \ldots & (\mathbf{x}^{s-1,*})^T & (\tilde{\mathbf{x}}^s)^T & (\mathbf{x}^{s+1,*})^T & \ldots \end{array} \right]^T,$$

is a feasible solution for $\mathcal{LP}_g$, as follows. The constraints associated with separation s [see Eq. (3.7)] are satisfied, since $\tilde{\mathbf{x}}^s$ satisfies the constraints of $\mathcal{LP}(s)$ with $\hat{\mathbf{x}}^{s+\sigma}$ as in Eq. (3.35). The constraints associated with all other separations also hold since $\tilde{\mathbf{x}}^s$ appears in the RHS with nonnegative coefficient matrices $\mathbf{A}_s^{(\cdot)}$. From assumption (3.37) it follows that $\tilde{\mathbf{X}}^s \geq \mathbf{X}^*$, where, under assumption (3.38) and for at least one $l \geq 1$ we have that $\tilde{X}_l^s > X_l^*$, where X_l^s denotes the lth element of $\mathbf{X}^s$. Since $\mathbf{X}^*$ is optimal, we reached a contradiction and it follows that Eq. (3.36) holds for all $s \geq 0$. This concludes the proof. ∎

We have shown that it suffices to compute the unique fixed point of F to find the optimal problem value function.

3.3 Extension to Larger Agent Groups

So far in this chapter, we formulated the two-agent problem as a DP, for the three types of local observation zones considered. We then used a standard procedure to convert the DP into an equivalent LP with infinitely many constraints, and decomposed the latter into a set of infinitely many LPs with a finite number of constraints. Finally, we proved that the unique bounded solution of the LP set corresponds to the optimal value function. It is clear that the DP and LP formulations, and the LP decomposition result hold for any finite-size local observation zone, although we omitted this fact earlier in the chapter, focusing on the three specific local observation zones considered.

The second remark concerns N-agent groups, with $N > 2$. The DP and equivalent LP formulations hold for the general N-agent problem. In this thesis, we focus on two-agent problems and on one three-agent problem (with short tunnel-shaped local observation zones) which can be interpreted as a two-agent problem since we assume an upper bound on the separation between the closest two agents (see Section 4.5 for further details). Hence, we refrained from presenting the DP and LP formulations for the general N-agent problem, simplifying significantly the notation. However, it seems clear that even the LP decomposition into a set of infinitely many LPs with a finite number of constraints (one such LP per agent separation set) holds true, with minor (mostly notational) modification in the proof.

3.4 Chapter Summary

In this chapter, we formulate the two-agent navigation problem on a structured graph as a Dynamic Program (DP), focusing on the three types of local observation zones considered. We then use a standard procedure to convert the DP into an equivalent Linear Program (LP), with infinitely many constraints. Then follows the main result in this chapter: we decompose the LP into a set $\mathcal{Q}$ of infinitely many LPs with a finite number of constraints, one per agent separation, and establish that the unique bounded solution to $\mathcal{Q}$ is the problem optimal value function. To this end, we define a function F which takes a candidate optimal value function represented properly by the vector $\mathbf{X}$, and converts it into $F(\mathbf{X})$, a new candidate value function computed as the union of the solutions to the LPs in $\mathcal{Q}$ with $\mathbf{X}$ as parameter. Equivalently, we prove that the unique bounded fixed point of the function F corresponds to the problem optimal value function. In the next chapter, we present a method to compute this fixed point of F for a set of navigation problems involving two and three agents, and the three types of local observation zones.

Chapter 4

The Computation of the Linear Program Set Fixed Point

In Chapter 3, we formulated the two-agent problem as a DP and used a standard procedure to convert the DP into an equivalent LP ($\mathcal{LP}_g$) with infinitely many constraints. We then decomposed $\mathcal{LP}_g$ into a set $\mathcal{Q}$ of infinitely many LPs (one for each agent separation) with a finite number of constraints and established the equivalence between $\mathcal{Q}$ and $\mathcal{LP}_g$ in the sense that the unique bounded solution to the LPs in $\mathcal{Q}$, formulated as the bounded fixed point of a function F, represents the problem optimal value function. In this chapter, we present an algorithm to compute the fixed point of F for a set of navigation problems with two and three agents, and for the three types of local observation zones. We compute the optimal value function explicitly at the separations $s < \bar{s}$ (for a specific well-defined $\bar{s}$) and provide an implicit description in the form of a recursion for the optimal value function at the separations $s \geq \bar{s}$.

4.1 Algorithm Overview

For large s, it is intuitively clear that the optimal agent policy converges to a "stationary" policy in the sense that as s increases, the optimal agent decision at each state becomes independent of the agent separation s. In fact, in Section 2.2, we argue that intuitively at large s, both agents adopt a single-agent optimal policy. Hereby, each agent's decision depends uniquely on its own local observation zone, and ties are broken by choosing the particular decision that yields the smallest agent separation at the next stage. On the other hand, for small s, the optimal policy is expected to be significantly different from the stationary policy valid at large s, since opportunities are exploited at small separations. The algorithm to compute the optimal value function over the whole state space relies on the two mentioned areas in the state space: for large s, the optimal policy is stationary (independent of separation s), and at small s, the actual cooperation takes place. The algorithm ensures the correct *connection* between the two state space areas.

The problem we face is computing the solution of an infinite set of LPs, each with a finite number of constraints. To this end, we devise from $\mathcal{Q}$ a causal *linear time invariant* (LTI) system with state $\mathbf{x}^s$, and with s as the equivalent of "time", for $s \geq \bar{s}$, with $\bar{s}$ a particular small separation. We carefully choose the output so that its positivity indicates optimality of $\mathbf{x}^s$ in the corresponding LP of $\mathcal{Q}$, for $s \geq \bar{s}$.

To determine the initial conditions for this LTI system, we formulate another LP with a finite number of constraints, denoted $\mathcal{LP}_{in}$, with variables $\mathbf{x}^s$ for $s < \bar{s}$ and with an extra set of equality constraints so that the solution to $\mathcal{LP}_{in}$ is the particular value function that exclusively excites the stable modes in the LTI system for $s \geq \bar{s}$. The added equality constraints are the connection between the system behavior at small and large separations. We establish that this provides the unique bounded fixed point of F and hence, the optimal value function.

In Section 4.2, we present the derivation in depth for the two-agent problem with cone-shaped local observation zones. Further, in Section 4.3, we show by a coupling argument that the agent separation of the two-agent problem with short tunnel-shaped local observation zones is upper bounded by three. Hence, for this case, it is sufficient to solve a simple finite state DP. Then, in Section 4.4, we treat the two-agent problem with extended tunnel-shaped local observation zones, focusing on the aspects that are significantly different from the in-depth exposition in Section 4.2, on the two-agent problem with cone-shaped local observation zones. Finally, in Section 4.5, we present the fixed point computation for the three-agent problem with short tunnel-shaped local observation zones.

4.2 Two Agents with Cone-Shaped Local Observation

In this section, we treat the two-agent problem with cone-shaped local observation zones in depth. In Section 4.2.1, we introduce some notation. In Section 4.2.2, we study the single-agent problem, and compute the optimal single-agent policies. Then, in Section 4.2.3, we assume that there exists a particular $\bar{s} \geq 0$ such that for all $s \geq \bar{s}$ a particular policy based on set of optimal single-agent policies, is optimal. We derive an LTI system with as state, the optimal value function at one separation, where "separation" is the equivalent of "time" in a traditional LTI system, in Sections 4.2.4 and 4.2.5. In Section 4.2.6, we formulate a LP whose solution is the optimal value function at separations $s < \bar{s}$, satisfying the fixed point condition. Furthermore, the initial condition for the LTI system derived from this LP solution, is such that the simulation of the LTI system yields a sequence of value functions for $s \geq \bar{s}$ that also satisfy the fixed point condition. Finally, in Section 4.2.7, we present a fix of the previous algorithm for a specific value of parameter p at which an otherwise invertible matrix becomes singular.

4.2.1 Notation

In this section, we provide information on the organization of the optimal value function in the vectors $\mathbf{x}^s$ of size n_s, for $s \geq 0$. For the two-agent problem at hand, $n_0 = 16$ and $n_s = 136$, for $s \geq 1$. Let

$$\mathcal{C}^{A_n} = \begin{bmatrix} a_0^{A_n} & b_0^{A_n} & a_{-1}^{A_n} & a_1^{A_n} \end{bmatrix}^T \tag{4.1}$$

denote the vector with as entries the costs of the edges in $\mathcal{O}^{A_n}$, for $n = 1, 2$ (see Fig. 3-1). For this problem, the set of known edges $\mathcal{C}$ at each stage is defined as the pair $(\mathcal{C}^{A_1}, \mathcal{C}^{A_2})$. The first entry of $\mathbf{x}^0$ represents the value of the value function for two agents at separation zero, and with local observation zones

$$\mathcal{C}^{A_1} = \mathcal{C}^{A_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T,$$

the second entry of $\mathbf{x}^0$ represents the value of the value function at local observation zone

$$\mathcal{C}^{A_1} = \mathcal{C}^{A_2} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T,$$

and so forth. Similarly, the ith entry of $\mathbf{x}^s$, for $i = 1, \ldots, n$, represents the value of the value function for two agents at separation $s > 0$, and at the ith local observation zone pair $(\mathcal{C}^{A_1}, \mathcal{C}^{A_2})$. The list of local observation zones is constructed as in the case where $s = 0$, but omitting local observation zone pairs for which

$$\mathcal{C}^{A_1} >_b \mathcal{C}^{A_2}, \tag{4.2}$$

due to the problem symmetry. In particular, the local observation zone pair where $\mathcal{C}^{A_1} = \mathcal{C}_1$ and $\mathcal{C}^{A_2} = \mathcal{C}_2$ is equivalent to the local observation zone pair where $\mathcal{C}^{A_1} = \mathcal{C}_2$ and $\mathcal{C}^{A_2} = \mathcal{C}_1$ and we omit one of the two equivalent pairs following the rule in Eq. (4.2). In the latter rule, the inequality is to be interpreted in a binary manner as if the vectors $\mathcal{C}^{A_1}$ and $\mathcal{C}^{A_2}$ were binary numbers (remember that the entries of $\mathcal{C}^{A_1}$ and $\mathcal{C}^{A_2}$ are in $\mathcal{L} = \{0, 1\}$). For example,

$$\mathcal{C}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} >_b \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} = \mathcal{C}_2,$$

and the 25th entry of $\mathbf{x}^s$ represents the value of the value function at local observation zone pair $(\mathcal{C}_2, \mathcal{C}_1)$.

4.2.2 Single-Agent Problem

We formulate the single-agent problem and find the optimal single-agent policies, a building block for the two-agent problem.

Problem 3 ***[Single Agent]*** *Let one agent be positioned on an arbitrary stage of graph $G(V, E)$. The edge costs are time invariant and i.i.d. random variables over $\mathcal{L}$ with probability p for a zero edge cost. The agent observes the costs of the links in its local observation zone and navigates through the graph in the direction of increasing stage indices. Then, find the navigation strategy so that the expected discounted sum of costs incurred at each stage is minimized (discount factor $0 < \alpha < 1$).* □

The state $x = \mathcal{C} \in \mathcal{S}_1$ is the set of costs of the edges in the local observation zone associated with the agent, summarized in vector $\mathcal{C}$ [see Eq. (4.1)]. We distinguish two types of edge costs: a controllable edge cost is determined at least in part by decision $u \in U_1 = \{-1, 0, 1\}$, whereas an uncontrollable edge cost is independent of u. Edge cost a_0 is controllable, as opposed to a_{-1}, a_1 and b_0, which are uncontrollable. In particular, if at the current state the agent chooses the edge straight ahead ($u = 0$), then $a_0' = b_0$, where $(\cdot)'$ denotes the value of $(\cdot)$ at the next time step. On the other hand, for any $u \in U_1$, the edge costs a'_{-1}, a'_1 and b'_0 are only known probabilistically, according to the *a priori* information available on unobserved edge costs, and do not depend on the agent decision. Hence, a_0 is (partially) controllable. Since a_0 is the only controllable edge cost, from Bellman's equation, it suffices to compute the expected value of the optimal value function for $a_0 = 0$ and for $a_0 = 1$ to determine the optimal value function at all states.

Let a policy $\mu_1 : \mathcal{S}_1 \to U_1$ be determined by the following two rules:

1. Traverse the edge with the smallest cost,
2. Break ties by traversing the edge that leads to the smallest cost a'_0, given the information at the current stage. An unknown a'_0 is considered better than $a'_0 = 1$ but worse than $a'_0 = 0$.

Then, we can compute $J_{\mu_1} : \mathcal{S}_1 \to \mathbb{R}^+$, the value function associated with policy μ_1. We have that

$$E[J_{\mu_1}(x)|a_0 = 0] = \frac{(1-p)^4\left(1+p-p^2\right)\alpha}{(1-\alpha)\left(1+p\alpha-5p^2\alpha+6p^3\alpha-2p^4\alpha\right)}, \tag{4.3}$$

$$E[J_{\mu_1}(x)|a_0 = 1] = \frac{(1-p)^2\left(1-p\alpha-2p^2\alpha+3p^3\alpha-p^4\alpha\right)}{(1-\alpha)\left(1+p\alpha-5p^2\alpha+6p^3\alpha-2p^4\alpha\right)}, \tag{4.4}$$

where $0 \leq p \leq 1$ and $0 < \alpha < 1$. The function $J_{\mu_1}(x)$ at the system states $x = x_i$, for $i = 1, \ldots, 16$ then becomes

$$\begin{aligned}
J_{\mu_1}(x_{i_1}) &= \alpha E[J_{\mu_1}(x)|a_0 = 0], \text{ for } i_1 \in \{1, 2, 3, 4\}, \\
J_{\mu_1}(x_{i_2}) &= \alpha E[J_{\mu_1}(x)], \text{ for } i_2 \in \{5, 6, 7, 9, 10, 11, 13, 14, 15\}, \\
J_{\mu_1}(x_8) &= \alpha E[J_{\mu_1}(x)|a_0 = 1], \\
J_{\mu_1}(x_{12}) &= 1 + \alpha E[J_{\mu_1}(x)|a_0 = 0], \\
J_{\mu_1}(x_{16}) &= 1 + \alpha E[J_{\mu_1}(x)].
\end{aligned}$$

We have the following lemma.

Lemma 8 *The policy μ_1 is optimal for the single-agent problem 3, for $0 \leq p \leq 1$ and $0 \leq \alpha < 1$.* □

Proof: We show that J_{μ_1} solves Bellman's equation and hence represents the optimal value function J_1^*, where the subscript indicates we treat a one-agent problem. It suffices to show that at x_i, for $i = 1, \ldots, 16$,

$$J_{\mu_1}(x_i) \leq g(x_i, u) + E[J_{\mu_1}(f(x_i, u))], u \in U_1,$$

where $g(x, u)$ denotes the cost of the edge traversed when taking decision u in state x, and where $f(x, u)$ denotes the state reached upon taking decision u in state x. Establishing the previous equation amounts to establishing that

$$\begin{aligned} E[J_{\mu_1}(x)|a_0 = 0] &\leq E[J_{\mu_1}(x)|a_0 = 1], && (4.5) \\ \alpha E[J_{\mu_1}(x)|a_0 = 1] &\leq 1 + \alpha E[J_{\mu_1}(x)], && (4.6) \\ \alpha E[J_{\mu_1}(x)] &\leq 1 + \alpha E[J_{\mu_1}(x)|a_0 = 0]. && (4.7) \end{aligned}$$

Substituting the Eqs (4.3,4.4) in the Eqs (4.5-4.7) yields the conditions

$$\begin{aligned} 0 &\leq \frac{(1-p)^2}{P_d(p,\alpha)}, && (4.8) \\ 0 &\leq \frac{Pn,1(p,\alpha)}{P_d(p,\alpha)}, && (4.9) \\ 0 &\leq \frac{P_{n,2}(p,\alpha)}{P_d(p,\alpha)}, && (4.10) \end{aligned}$$

for $0 \leq p \leq 1$ and $0 < \alpha < 1$, where

$$\begin{aligned} P_{n1}(p,\alpha) &= 1 - 3p^2\alpha + 5p^3\alpha - 2p^4\alpha, \\ P_{n2}(p,\alpha) &= 1 - \alpha + 4p\alpha - 8p^2\alpha + 7p^3\alpha - 2p^4\alpha, \\ P_d(p,\alpha) &= 1 + p\alpha - 5p^2\alpha + 6p^3\alpha - 2p^4\alpha. \end{aligned}$$

Since the three polynomials are linear in α, it suffices to verify nonnegativity of $P_{n1}(p,\alpha)$ and $P_{n2}(p,\alpha)$ and positivity of $P_d(p,\alpha)$ for $\alpha = 0$ and $\alpha = 1$, for all $0 \leq p \leq 1$. We have that

$$P_{n1}(p,0) = P_{n2}(p,0) = P_d(p,0) = 1,$$

so that we have left to establish nonnegativity of $P_{n1}(p,1)$ and $P_{n2}(p,1)$ and positivity of $P_d(p,1)$, for $0 \le p \le 1$. For each polynomial, we formulate a *Linear Matrix Inequality* (LMI) constraint. We give details for $P_d(p,1)$; similar procedures lead to the proof of nonnegativity of $P_{n1}(p,1)$ and $P_{n2}(p,1)$. For an in-depth treatment, see [62]. In particular, we find an auxiliary polynomial $P_d^a(p)$ such that

$$P'_d(p) = P_d(p,1) + P_d^a(p)\left((p-1/2)^2 - 1/4\right) > 0, \tag{4.11}$$

where

$$P_d^a(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix},$$

with parameters a_1, a_2 and a_3 such that $P_d^a(p)$ is *Sum of Squares* (SOS), or equivalently, its coefficient matrix is positive semi-definite. Indeed, if such a polynomial $P_d^a(p)$ can be found, then for $0 \le p \le 1$ we have that the second term of $P'_d(p)$ is non-positive for $0 \le p \le 1$ and therefore $P_d(p,1)$ is positive for $0 \le p \le 1$. Requiring the $P'_d(p)$ to be SOS guarantees nonnegativity and transforms Eq. (4.11) into an LMI constraint, although it is a more stringent requirement. Here, we require positive definiteness of the coefficient matrix of $P'_d(p)$, rather than positive semi-definiteness, since we want to prove strict inequality in Eq. (4.11). The SOSTOOLS Sum of Squares Optimization Toolbox for Matlab [63] uses semi-definite programming to find a satisfying $P_d^a(p)$, which yields

$$P_d^a(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 3 & -\frac{7}{5} \\ -\frac{7}{5} & \frac{41}{10} \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix}$$

as a possible positive definite quadratic form for the auxiliary polynomial (although only positive semi-definiteness is required). The corresponding quadratic form for $P'_d(p)$ is

$$P'_d(p) = \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}^T \begin{bmatrix} 1 & -1 & -\frac{8}{15} \\ -1 & \frac{28}{15} & -\frac{9}{20} \\ -\frac{8}{15} & -\frac{9}{20} & \frac{21}{10} \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix},$$

whose coefficient matrix can easily be verified to be positive definite. We conclude that $P_d(p,\alpha) > 0$ for $0 \leq p, \alpha \leq 1$.

Similarly, we can compute auxiliary polynomials $P^a_{n1}(p)$ and $P^a_{n2}(p)$. In particular, we have that

$$\begin{aligned} P^a_{n1}(p) &= \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix}, \\ P^a_{n2}(p) &= \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 4 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix}, \end{aligned}$$

where the coefficient matrices are positive definite. The corresponding quadratic forms of $P'_{n1}(p,1)$ and $P'_{n2}(p,1)$ are

$$\begin{aligned} P'_{n1}(p) &= \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}^T \begin{bmatrix} 1 & -1 & -\frac{1}{3} \\ -1 & \frac{5}{3} & -\frac{1}{2} \\ -\frac{1}{3} & -\frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}, \\ P'_{n2}(p) &= \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix}, \end{aligned}$$

where the coefficient matrices are positive definite. Again, the last four coefficient matrices need only be positive semi-definite. Therefore, the proof is complete. ■

Hence, for any relevant p and α ($0 \leq p \leq 1$, $0 < \alpha < 1$), the same single-agent policy is optimal.

4.2.3 Mixed Forward-Backward Recursion

In this section, we extract from $\mathcal{LP}(s)$ a recursion writing the optimal value function at separation s as a function of the optimal value function at separations $s+\sigma$, for $\sigma = \max\{-2,-s\},\ldots,2$. This recursion is neither forward, nor backward since at separation s, the RHS includes terms at lower and greater separations. Let $\mu^s_\infty : \mathcal{S}(s) \to U$ denote the policy, whereby at separation s each agent follows the single-agent optimal policy and where ties are broken by choosing the pair of decisions that minimizes the resulting agent separation. Then, we make the following assumption:

Assumption 1 For any $0 \leq p \leq 1$ and $0 < \alpha < 1$, we look for an optimal policy $\mu^* : \mathcal{S} \to U$ in the set of policies $\{\mu \mid \mu(x) = \mu_\infty^s(x)$ for $x \in \mathcal{S}(s), s \geq \bar{s}\}$ for a particular $\bar{s} \geq 0$.

This assumption is motivated in part by the development in Section 4.2.2, where we show that for the set of relevant p and α, the same set of policies is optimal. Note that under Assumption 1, μ_∞^s is independent of s for $s \geq \bar{s}$ and therefore, for $s \geq \bar{s}$ we omit the superscript s. One sanity check is that the optimal two-agent value function should converge to twice the optimal single-agent value function for s going to infinity, since for infinitely large s the two-agent problem is equivalent to the combination of two single-agent problems. Assumption 1 satisfies this check, since for infinitely large s, each agent indeed uses a single-agent optimal policy.

We write $\mathcal{LP}(s)$ in standard form as follows:

$$\text{maximize} \quad \mathbf{c}^T \begin{bmatrix} \mathbf{x}^s \\ \mathbf{y}^s \end{bmatrix} \tag{4.12}$$

$$\text{subject to} \quad (\mathbf{T} - \mathbf{A}^0)\mathbf{x}^s + \mathbf{y}^s = \sum_{\sigma \in S^s} \mathbf{A}^\sigma \hat{\mathbf{x}}^{s+\sigma,*} + \mathbf{b}^s, \tag{4.13}$$

$$\mathbf{x}^s = \mathbf{x}_+^s - \mathbf{x}_-^s \tag{4.14}$$

$$\mathbf{x}_+^s, \mathbf{x}_-^s, \mathbf{y}^s \geq 0,$$

denoted $\mathcal{LP}_f(s)$. Here, $\mathbf{c}^T = \begin{bmatrix} \mathbf{e}^T & \mathbf{0}^T \end{bmatrix}^T$ with the vectors $\mathbf{e}$ and $\mathbf{0}$ of the same dimension as $\mathbf{x}^s$ and $\mathbf{y}^s$, respectively. Writing $\mathbf{x}^s = \mathbf{x}_+^s - \mathbf{x}_-^s$, with $0 \leq \mathbf{x}_+^s, \mathbf{x}_-^s \in \mathbb{R}^{n'}$, where $n' = n_0$ for $s = 0$, and $n' = n$ for $s \geq 1$, is a common procedure for converting a free variable to the standard form format. Similary, introducing slack variables $0 \leq \mathbf{y}^s \in \mathbb{R}^{n'|U|}$ is a standard method to transform inequalities into equalities. We refer the reader to [8] for details on converting a general LP into standard form. The version of $\mathcal{LP}_f(s)$ where $\hat{\mathbf{x}}^{s+\sigma} = \mathbf{x}^{s+\sigma,*}$, for $\sigma \in S^s$, is denoted $\mathcal{LP}_f^*(s)$. Since $\mathbf{x}^{s+\sigma,*} \geq 0$, for any $s \geq 0$ and $\sigma \in S^s$, we can set $\mathbf{x}_-^s = 0$ in $\mathcal{LP}_f^*(s)$, a helpful simplification.

We now derive a matrix equation relating $\mathbf{x}^{s,*}$ to $\mathbf{x}^{s+\sigma,*}$ ($\sigma \in S^s$) for $s \geq 0$. To this end, we show a property of the optimal basic feasible solution, which holds for all $s \geq 0$. In particular, consider the optimal basic feasible solution pair $(\mathbf{x}^{s,*}, \mathbf{y}^{s,*})$ of $\mathcal{LP}_f^*(s)$ for any $s \geq 0$. At $(\mathbf{x}^{s,*}, \mathbf{y}^{s,*})$, the $n|U|$ linear equality constraints (4.13) are satisfied, where n needs to be replaced by n_0 for $s = 0$ (from here on, we omit this fact). Furthermore, from the definition of a basic solution, we have that, out of all the constraints that are active,

$n+n|U|$ (*i.e.* the number of decision variables) are linearly independent. Therefore, at least n of the inequality constraints $\mathbf{y}^s \geq 0$ are active. To each particular $x_i^{s,*}$ $(0 \leq i \leq n)$, *i.e.* at separation s and for the pair of observation zones corresponding to index i, is associated a set of $|U|$ equality constraints (see the definition of $\mathbf{T}$ in Section 3.1.2). Let $\mathbf{y}_i^s \in \mathbb{R}^{|U|}$ contain the slack variables associated with the equality constraint set related to x_i^s. We have the following property of $\mathbf{y}_i^{s,*}$.

Lemma 9 *At least one of the components of* $\mathbf{y}_i^{s,*}$ *equals zero, for all* $i = 1, \ldots, n$, *and for all* $s = 0, 1, \ldots$ □

Proof: One way to see this is by contradiction. Suppose the opposite is true for some i, *i.e.* $\mathbf{y}_i^{s,*} > 0$, then we can increase $x_i^{s,*}$, adapting the elements of $\mathbf{y}_i^{s,*}$ accordingly so that the set of $|U|$ equality constraints (4.13) associated with x_i^s remains satisfied, until some element of $\mathbf{y}_i^s$ equals zero. Increasing x_i^s does not jeopardize equality constraints related to other components of $\mathbf{x}^s$, since x_i^s appears in the RHS of those constraints with non-negative coefficients.

This procedure yields a feasible solution $(\bar{\mathbf{x}}^s, \bar{\mathbf{y}}^s)$ such that $\bar{\mathbf{x}}^s \geq \mathbf{x}^{s,*}$, where strict inequality holds for at least one component. Since $(\mathbf{x}^{s,*}, \mathbf{y}^{s,*})$ is assumed optimal, this is a contradiction and the proof is complete. ■

We now derive a linear equation in $\mathbf{x}^{s,*}$, and $\mathbf{x}^{s+\sigma,*}$ for $\sigma \in S^s$ for all $s \geq 0$. Since for all $i = 1, \ldots, n$, at least one element of $\mathbf{y}_i^s$ equals zero (see Lemma 9), we can construct a set of n equations with x_i^s, *i.e.* the components of $\mathbf{x}^s$, as the n unknowns. Specifically, for each $i = 1, \ldots, n$ we can pick an index j_i^s $(1 \leq j_i^s \leq |U|)$, such that

$$(\mathbf{y}_i^{s,*})_{j_i^s} = 0, \tag{4.15}$$

where the LHS denotes the (j_i^s)th component of $\mathbf{y}_i^{s,*}$. Then, for each $i = 1, \ldots, n$, and out of the $|U|$ equality constraints associated with x_i^s, take the equality constraint with index j_i^s as row i of matrix $\mathbf{I} - \mathbf{D}_{s,0}$, *i.e.*, $\mathbf{I}$ and $\mathbf{D}_{s,0}$ contain the relevant rows of $\mathbf{T}$ and $\mathbf{A}_0$, respectively. Then, for all $s \geq 0$, we have that

$$(\mathbf{I} - \mathbf{D}_{s,0})\mathbf{x}^{s,*} = \sum_{\sigma \in S^s} \mathbf{D}_{s,\sigma}\mathbf{x}^{s+\sigma,*} + \mathbf{d}_s, \tag{4.16}$$

where $\mathbf{I}-\mathbf{D}_{s,0}, \mathbf{D}_{s,\sigma} \in \mathbb{R}^{n\times n}$ (for all $\sigma \in S^s$), where $\mathbf{D}_{s,\sigma}$ and $\mathbf{d}_s \in \mathbb{R}^n$ contain the relevant rows of $\mathbf{A}_s^\sigma$ (for all $\sigma \in S^s$) and $\mathbf{b}^s$, respectively. Note that in general the optimal basic feasible solution need not be the same for all s, hence the addition of subscript s to the composing matrices of Eq. (4.16).

We now restrict our attention to separations $s \geq \bar{s}$, where under Assumption 1, the same basic feasible solution is optimal, and therefore we can write $\mathbf{D}_0 \doteq \mathbf{D}_{s',0}$, for all $s' \geq \bar{s}$. The matrices $\mathbf{D}_\sigma$, for $\sigma \in S^s$ and the vector $\mathbf{d}$ are defined similarly. We have then, for $s \geq \bar{s}$, that

$$(\mathbf{I}-\mathbf{D}_0)\mathbf{x}^{s,*} = \sum_{\sigma\in S^s} \mathbf{D}_\sigma \mathbf{x}^{s+\sigma,*} + \mathbf{d}. \tag{4.17}$$

The structure of Main Problem 1 allows us to reduce the dimension of the matrices and vectors in Eq. (4.17). Recall that in this problem, state $x = (s, \mathcal{C}^A, \mathcal{C}^B)$ is a triple containing the agent separation s and the costs of the edges in the local observation zones associated with both agents, summarized in the vectors $\mathcal{C}^A$ and $\mathcal{C}^B$ [see Eq. (4.1)]. The edge costs a_0^A and a_0^B are controllable (see Section 4.2.2 for the definition of a controllable edge cost); a_1^A, a_{-1}^A, b_0^A, a_1^B, a_{-1}^B, and b_0^B, are uncontrollable.

We introduce some notation. Let

$$E\mathbf{x}_{qr}^{s,*} = E[J^*(x)|a_0^A = q, a_0^B = r], \tag{4.18}$$

where $q, r \in \mathcal{L}\cup\{\times\}$, and where the expected value is taken over the uncontrollable edge costs. Here, $q = \times$ (resp. $r = \times$) indicates that a_0^A (resp. a_0^B) is unobserved. From symmetry, we have that $E\mathbf{x}_{qr}^{s,*} = E\mathbf{x}_{rq}^{s,*}$, for $q, r \in \mathcal{L}\cup\{\times\}$. Furthermore, we have the relation

$$E\mathbf{x}_{q\times}^{s,*} = pE\mathbf{x}_{q0}^{s,*} + (1-p)E\mathbf{x}_{q1}^{s,*}, \qquad \text{for } q \in \mathcal{L}\cup\{\times\}, \tag{4.19}$$

which follows immediately from the definition (4.18) of $E\mathbf{x}_{qr}^{s,*}$.

We now use $E\mathbf{x}_{qr}^{s,*}$ and the fact that only a_0^A and a_0^B are controllable to simplify Eq. (4.17), for $s \geq \bar{s}$. We investigate two cases. First, if at a particular separation $s \geq \bar{s}$, the optimal two-agent decision $\mathbf{u}_i \in U$ associated with some i ($0 \leq i \leq n$) is such that $s' = s+2$ or $s' = s-2$, then $x_i^{s,*} = d_i + \alpha E\mathbf{x}_{\times\times}^{s',*}$. Therefore, in Eq. (4.17), we have that the ith row of the matrices $\mathbf{D}_0$, $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ equals zero. Furthermore, $\mathbf{D}_2$ and $\mathbf{D}_{-2}$ have at most rank one. In particular, we have

$$\mathbf{D}_2\mathbf{x}^{s+2,*} = \mathbf{D}_2' E\mathbf{x}_{\times\times}^{s+2,*} \tag{4.20}$$

where $\mathbf{D}_2' \in \mathbb{R}^{n\times 1}$. A similar equation holds in the case $s' = s-2$.

Second, for $\mathbf{u}_i$ such that $s' = s+1$ or $s' = s-1$, we have that $x_i^{s,*} = d_i + \alpha E\mathbf{x}_{q\times}^{s',*}$, with $q = b_0^A$ if $u_i^A = 0$ and $q = b_0^B$ if $u_i^B = 0$. Hence, the matrices $\mathbf{D}_1$ and $\mathbf{D}_{-1}$ are at most of rank two since at most two linearly independent linear combinations of $\mathbf{x}^{s+1,*}$ (resp. $\mathbf{x}^{s-1,*}$), namely $E\mathbf{x}_{0\times}^{s+1,*}$ and $E\mathbf{x}_{1\times}^{s+1,*}$ (resp. $E\mathbf{x}_{0\times}^{s-1,*}$ and $E\mathbf{x}_{1\times}^{s-1,*}$), suffice in Eq. (4.17). We have

$$\mathbf{D}_1\mathbf{x}^{s+1,*} = \mathbf{D}_1' \left[\begin{array}{cc} E\mathbf{x}_{0\times}^{s+1,*} & E\mathbf{x}_{1\times}^{s+1,*} \end{array}\right]^T, \tag{4.21}$$

where $\mathbf{D}_1' \in \mathbb{R}^{n\times 2}$. A similar equation holds in the case $s' = s-1$. For policy μ_∞, we have $\mathrm{rank}(\mathbf{D}_2') = \mathrm{rank}(\mathbf{D}_{-2}') = 1$ and $\mathrm{rank}(\mathbf{D}_1') = \mathrm{rank}(\mathbf{D}_{-1}') = 2$.

Finally, if $\mathbf{u}_i$ is such that $s' = s$, we have that, in similar fashion, matrix $\mathbf{D}_0$ is of rank three and hence

$$\mathbf{D}_0\mathbf{x}^{s,*} = \mathbf{D}_0' \left[\begin{array}{ccc} E\mathbf{x}_{00}^{s,*} & E\mathbf{x}_{01}^{s,*} & E\mathbf{x}_{11}^{s,*} \end{array}\right]^T, \tag{4.22}$$

where $\mathbf{D}_0' \in \mathbb{R}^{n\times 3}$, with the appropriate components. With Eqs (4.20-4.22), we write Eq. (4.17) as

$$\begin{aligned} (\mathbf{I} - \mathbf{D}_0'') \left[\begin{array}{ccc} E\mathbf{x}_{00}^{s,*} & E\mathbf{x}_{01}^{s,*} & E\mathbf{x}_{11}^{s,*} \end{array}\right]^T &= \dots \\ \mathbf{D}_{-2}'' E\mathbf{x}_{\times\times}^{s-2,*} + \mathbf{D}_{-1}'' \left[\begin{array}{c} E\mathbf{x}_{0\times}^{s-1,*} \\ E\mathbf{x}_{1\times}^{s-1,*} \end{array}\right] &+ \mathbf{D}_1'' \left[\begin{array}{c} E\mathbf{x}_{0\times}^{s+1,*} \\ E\mathbf{x}_{1\times}^{s+1,*} \end{array}\right] + \mathbf{D}_2'' E\mathbf{x}_{\times\times}^{s+2,*} + \mathbf{d}'', \end{aligned} \tag{4.23}$$

where $\mathbf{D}_\sigma'' = \mathbf{K}\mathbf{D}_\sigma'$, for $\sigma \in S^s$, and where $\mathbf{d}'' = \mathbf{K}\mathbf{d}$. Here, the matrix $\mathbf{K} \in \mathbb{R}^{3\times n}$ is such that

$$\left[\begin{array}{ccc} E\mathbf{x}_{00}^{s} & E\mathbf{x}_{01}^{s} & E\mathbf{x}_{11}^{s} \end{array}\right]^T = \mathbf{K}\mathbf{x}^s, \tag{4.24}$$

where the entries of $\mathbf{K}$ are known polynomials in p. In words, linear operator $\mathbf{K}$ takes the appropriate linear combination of the elements of $\mathbf{x}^s$ to compute the expected values in the

LHS of Eq. (4.24). The explicit expressions for the matrices $\mathbf{D}''_\sigma$ for $\sigma = -2, \ldots, 2$ and for the vector $\mathbf{d}''$ are

$$
\begin{aligned}
\mathbf{D}''_{-2} &= \alpha \begin{bmatrix} p^2(1-p)^2 \\ p(1-3p+5p^2-4p^3+p^4) \\ (-1+2p-3p^2+p^3)^2 \end{bmatrix}, \\
\mathbf{D}''_{-1} &= \alpha \begin{bmatrix} 2p^2(1-p) & 2p(1-p)^4 \\ p(1-p+2p^3-p^4) & (1-p)^3(1-2p+3p^2-p^3) \\ 2p(1-p)^2(1-2p+3p^2-p^3) & 0 \end{bmatrix}, \\
\mathbf{D}''_{0,1} &= \alpha p^2 \begin{bmatrix} (1+2p^2-6p^3+6p^4-2p^5) \\ (1-p)^2(1+p-p^2+3p^3-p^4) \\ (1-2p+6p^3-7p^4+2p^5) \end{bmatrix}, \\
\mathbf{D}''_{0,2} &= \alpha p(1-p)^2 \begin{bmatrix} 2(1-p)(1+2p^2-2p^3) \\ (1-p)(1-p^2+6p^3-2p^4) \\ 4p(1-2p+3p^2-p^3) \end{bmatrix}, \\
\mathbf{D}''_{0,3} &= \alpha(1-p)^3 \begin{bmatrix} (1-p)^2(1-p+2p^2) \\ (1-p)p(1-p+3p^2-p^3) \\ 2p(1-2p+3p^2-p^3) \end{bmatrix}, \\
\mathbf{D}''_{1} &= \alpha p \begin{bmatrix} 2(1-p)^2 p & 2(1-p)^5 \\ p(2-5p+6p^2-4p^3+p^4) & (1-p)^4 \\ 2(1-p)^3 p & 0 \end{bmatrix}, \\
\mathbf{D}''_{2} &= \alpha(1-p)^2p^2 \begin{bmatrix} (1-p)^2 \\ (1-p) \\ 1 \end{bmatrix}, \\
\mathbf{d}'' &= (1-p)^2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix},
\end{aligned}
$$

where $\mathbf{D}''_{0,i}$ denotes the ith column of $\mathbf{D}''_0$, for $i = 1, 2, 3$. The recursion in Eq. (4.23) is neither forward nor backward in the sense that the optimal value function at separation s depends (linearly) on the optimal value function at separations both lower and greater than s. In the following section, we convert Eq. (4.23) to a forward equation, so that given the optimal value function up till separation s, we can compute the optimal value function at larger separations.

4.2.4 Forward Recursion

In this section, we transform Eq. (4.23) into a forward recursion which allows us, under Assumption 1, to compute the optimal value function for $s \geq \bar{s}$, given the appropriate initial condition, involving $E\mathbf{x}_{\times\times}^{\bar{s}+1,*}$, $E\mathbf{x}_{q\times}^{\bar{s},*}$ $(q = 1, 2)$ and all information on the optimal value function at separations $s < \bar{s}$.

We first perform a change of coordinates in Eq. (4.23) and rearrange the terms in an appropriate way. In particular, we express the expected values of the optimal value function at a particular separation s as a function of $E\mathbf{x}_{\times\times}^{s,*}$, $E\mathbf{x}_{1\times}^{s,*}$ and $E\mathbf{x}_{11}^{s,*}$. Specifically, we define $\mathbf{P}_1 \in \mathbb{R}^{3\times 3}$ and $\mathbf{P}_2 \in \mathbb{R}^{2\times 2}$ as

$$\begin{aligned}
\mathbf{P}_1 &= \begin{bmatrix} p^2 & 2p(1-p) & (1-p)^2 \\ 0 & p & 1-p \\ 0 & 0 & 1 \end{bmatrix}, \\
\mathbf{P}_2 &= \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix},
\end{aligned}$$

so that

$$\begin{bmatrix} E\mathbf{x}_{\times\times}^{s,*} & E\mathbf{x}_{1\times}^{s,*} & E\mathbf{x}_{11}^{s,*} \end{bmatrix}^T = \mathbf{P}_1 \begin{bmatrix} E\mathbf{x}_{00}^{s,*} & E\mathbf{x}_{01}^{s,*} & E\mathbf{x}_{11}^{s,*} \end{bmatrix}^T, \tag{4.25}$$

$$\begin{bmatrix} E\mathbf{x}_{\times\times}^{s,*} & E\mathbf{x}_{1\times}^{s,*} \end{bmatrix}^T = \mathbf{P}_2 \begin{bmatrix} E\mathbf{x}_{0\times}^{s,*} & E\mathbf{x}_{1\times}^{s,*} \end{bmatrix}^T. \tag{4.26}$$

Both $\mathbf{P}_1$ and $\mathbf{P}_2$ are invertible for $0 < p < 1$. The cases where $p = 0$ or $p = 1$, yield trivial navigation problems and or not of interest. Using Eqs (4.25,4.26), Eq. (4.23) becomes

$$\begin{aligned}
\mathbf{R}_0 \begin{bmatrix} E\mathbf{x}_{\times\times}^{s,*} & E\mathbf{x}_{1\times}^{s,*} & E\mathbf{x}_{11}^{s,*} \end{bmatrix}^T &= \dots \\
\mathbf{R}_{-2} E\mathbf{x}_{\times\times}^{s-2,*} + \mathbf{R}_{-1} \begin{bmatrix} E\mathbf{x}_{\times\times}^{s-1,*} \\ E\mathbf{x}_{1\times}^{s-1,*} \end{bmatrix} &+ \mathbf{R}_1 \begin{bmatrix} E\mathbf{x}_{\times\times}^{s+1,*} \\ E\mathbf{x}_{1\times}^{s+1,*} \end{bmatrix} + \mathbf{R}_2 E\mathbf{x}_{\times\times}^{s+2,*} + \mathbf{r},
\end{aligned} \tag{4.27}$$

where $\mathbf{R}_0 = (\mathbf{I} - \mathbf{D}_0'')\mathbf{P}_1^{-1}$, where $\mathbf{R}_\sigma = \mathbf{D}_\sigma''$, for $\sigma \in \{-2, 2\}$, where $\mathbf{R}_\sigma = \mathbf{D}_\sigma''\mathbf{P}_2^{-1}$, for $\sigma \in \{-1, 1\}$, and where $\mathbf{r} = \mathbf{d}''$. We regroup the terms in Eq. (4.27) to obtain the forward recursion

$$\mathbf{R}'_2 \begin{bmatrix} E\mathbf{x}_{\times\times}^{s+2,*} \\ E\mathbf{x}_{1\times}^{s+1,*} \\ E\mathbf{x}_{11}^{s,*} \end{bmatrix} = \sum_{\sigma=-2}^{1} \mathbf{R}'_\sigma \begin{bmatrix} E\mathbf{x}_{\times\times}^{s+\sigma,*} \\ E\mathbf{x}_{1\times}^{s+\sigma-1,*} \\ E\mathbf{x}_{11}^{s+\sigma-2,*} \end{bmatrix} + \mathbf{r}, \tag{4.28}$$

where

$$\begin{aligned} \mathbf{R}'_2 &= \begin{bmatrix} -\mathbf{R}_2 & -\mathbf{R}_{1,2} & \mathbf{R}_{0,3} \end{bmatrix}, \\ \mathbf{R}'_1 &= \begin{bmatrix} \mathbf{R}_{1,1} & -\mathbf{R}_{0,2} & \mathbf{0} \end{bmatrix}, \\ \mathbf{R}'_0 &= \begin{bmatrix} -\mathbf{R}_{0,1} & \mathbf{R}_{-1,2} & \mathbf{0} \end{bmatrix}, \\ \mathbf{R}'_{-1} &= \begin{bmatrix} \mathbf{R}_{-1,1} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \\ \mathbf{R}'_{-2} &= \begin{bmatrix} \mathbf{R}_{-2} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \end{aligned} \tag{4.29, 4.30}$$

where $\mathbf{R}_{k,l}$ denotes the lth column of $\mathbf{R}_k$. Eq. (4.28) is only then a well-defined forward recursion if $\mathbf{R}'_2$ is invertible. We have the following lemma.

Lemma 10 *The matrix $\mathbf{R}'_2 \in \mathbb{R}^{3\times 3}$ is invertible for all p and α such that $0 < \alpha < 1$ and $0 < p < 1$, $p \neq p'$, where*

$$p' = \frac{1}{3}\left(4 - \left(\frac{2}{25 - 3\sqrt{69}}\right)^{\frac{1}{3}} - \left(\frac{25 - 3\sqrt{69}}{2}\right)^{\frac{1}{3}}\right) \simeq 0.245.$$

□

Proof: Define

$$\begin{aligned} P_1(p) &= -1 + 5p - 4p^2 + p^3, \\ P_2(p,\alpha) &= -1 - 2p - p^2\,(1-\alpha) - 10p^3\,\alpha + 33p^4\,\alpha - 42p^5\,\alpha + 26p^6\alpha - 8p^7\alpha + p^8\alpha. \end{aligned}$$

Then, we have

$$\det(\mathbf{R}'_2) = -(1-p)^6 p\alpha^2\, P_1(p)\, P_2(p,\alpha). \tag{4.31}$$

We now analyze the polynomial in Eq. (4.31) with some straightforward manipulations. In particular, one can verify that p' is the only real-valued root of $P_1(p)$, and therefore, $\det(\mathbf{R}_2') = 0$ for $p = p'$ and any $0 < \alpha < 1$. Then, we establish that $P_2(p, \alpha) \neq 0$, for $0 < p < 1$ and $0 < \alpha < 1$. Specifically, we show that $P_2(p, \alpha) < 0$ for $0 < p < 1$ and $0 < \alpha < 1$. Since $P_2(p, \alpha)$ is linear in α, it suffices to show that both $P_2(p, 0) < 0$ and $P_2(p, 1) < 0$ for $0 < p < 1$. The former can easily be verified to hold by noticing that $P_2(p, 0) = -1 - 2p - p^2$. Further, we have that

$$P_2(p, 1) = -1 + p\,(-2 + p^2(1-p)^2\, P_2'(p))$$

where

$$P_2'(p) = -10 + 13p - 6p^2 + p^3.$$

From this, we can see that the inequality $P_2(p, 1) < 0$ holds for $0 < p < 1$ if $P_2'(p) < 0$ for $0 < p < 1$. Since $P_2'(0) < P_2'(1) < 0$, we have as sufficient condition that

$$\frac{d}{dp} P_2'(p) = 13 - 12p + 3p^2 > 0.$$

This condition can easily been verified analytically to hold, which completes the proof. ■

Lemma 10 indicates that for all relevant $p \neq p'$ and α, Eq. (4.28) is a well-defined forward recursion. For $p = p'$, rearranging terms in Eq. (4.28) yields a different, well-defined forward recursion, for which the algorithm presented in the remainder of this thesis is valid. We omit the details regarding this special case and assume from here on that $p \neq p'$.

4.2.5 LTI System

In this section, we analyze the recursion in Eq. (4.28) with control-theoretic tools. Rearranging the terms in Eq. (4.28) and adding appropriate variables yields

$$E\mathbf{x}^{s+1,*} = \mathbf{A}(E\mathbf{x}^{s,*}) + \mathbf{B}, \tag{4.32}$$

where

$$E\mathbf{x}^{s,*} = \Big[\begin{array}{ccccccc} E\mathbf{x}_{\times\times}^{s+2,*} & E\mathbf{x}_{1\times}^{s+1,*} & E\mathbf{x}_{11}^{s,*} & E\mathbf{x}_{\times\times}^{s-2,*} & E\mathbf{x}_{\times\times}^{s-1,*} & E\mathbf{x}_{1\times}^{s-1,*} & \dots \end{array}$$
$$\begin{array}{ccc} E\mathbf{x}_{\times\times}^{s,*} & E\mathbf{x}_{\times\times}^{s+1,*} & E\mathbf{x}_{1\times}^{s,*} \end{array} \Big]^T, \quad (4.33)$$

and with

$$\mathbf{A} = \left[\begin{array}{ccc|c|c|c|c|c} & & & 0 & & 0 & & \\ & (\mathbf{R}_2')^{-1}\mathbf{R}_1' & & 0 & (\mathbf{R}_2')^{-1}\mathbf{R}_{-2,1}' & 0 & (\mathbf{R}_2')^{-1}\mathbf{R}_{-1,1}' & (\mathbf{R}_2')^{-1}\mathbf{R}_{0,(1,2)}' \\ & & & 0 & & 0 & & \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \quad 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \quad 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \quad 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \quad 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \quad 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \quad 0 \end{array} \right],$$

and

$$\mathbf{B} = \Big[\begin{array}{ccccccc} ((\mathbf{R}_2')^{-1}\mathbf{r})^T & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \Big]^T.$$

Eq. (4.32) is a necessary condition on the optimal value function, for $s \geq \bar{s}$. The superscript s of the state vector indicates that at "time" s, the optimal value function is completely known at separations up until and including s. Note that the entries $E\mathbf{x}_{\times\times}^{s-2,*}$ and $E\mathbf{x}_{1\times}^{s-1,*}$ of $E\mathbf{x}^{s,*}$ are unnecessary in the state transition equation, but are required in the state vector to define the system output adequately.

We now consider any state sequence $E\mathbf{x}_c^s$, for $s = \bar{s}, \bar{s}+1, \dots$, and formulate the state transition equation of an LTI system with system matrix $\mathbf{A}$ as follows:

$$E\mathbf{x}_c^{s+1} = \mathbf{A}(E\mathbf{x}_c^s) + \mathbf{B}, \quad (4.34)$$

where $E\mathbf{x}_c^s$ is defined as in Eq. (4.33), but omitting the asterisk; the subscript c indicates that the value function associated with $E\mathbf{x}_c^s$, for $s = \bar{s}, \bar{s}+1, \dots$, is considered a candidate optimal value function. We determine the output $\mathbf{z}^s \in \mathbb{R}^{n|U|}$ (for $s \geq \bar{s}$) so that non-negativity of $\mathbf{z}^s$

verifies that $\mathbf{x}_c^s$ is indeed the optimal solution of $\mathcal{LP}_f(s)$, where $\hat{\mathbf{x}}^{s+\sigma} = \mathbf{x}_c^{s+\sigma}$, for $\sigma \in S^s$, an LP which we denote as $\mathcal{LP}_f^c(s)$. Here, $\mathbf{x}_c^s$ and $\mathbf{x}_c^{s+\sigma}$ (for $s \geq \bar{s}$ and $\sigma \in S^s$) are computed from the state vectors $E\mathbf{x}_c^s$ and $E\mathbf{x}_c^{s+\sigma}$ using Eq. (3.2) with μ_∞ as policy [recall the notation convention in Eq. (3.6)]. To this end, we rewrite the constraints of $\mathcal{LP}_f^c(s)$ [see Eq. (4.13)] as

$$\mathbf{A}_{LP} \begin{bmatrix} \mathbf{x}^s \\ \mathbf{y}^s \end{bmatrix} = \mathbf{b}_{LP}^s,$$

with

$$\begin{aligned} \mathbf{A}_{LP} &= \begin{bmatrix} \mathbf{T} - \mathbf{A}^0 & \mathbf{I} \end{bmatrix}, \\ \mathbf{b}_{LP}^s &= \sum_{\sigma \in S^s} \mathbf{A}^\sigma \mathbf{x}_c^{s+\sigma} + \mathbf{b}^s, \end{aligned}$$

where $\mathbf{I}$ has the appropriate dimensions. Note that we only consider non-negative candidate optimal value functions $\mathbf{x}_c^s$, and therefore, in $\mathcal{LP}_f^c(s)$, we can set $\mathbf{x}_-^s = 0$ [see Eq. (4.14)]. We need the following property of matrix $\mathbf{A}_{LP}$.

Lemma 11 *The first n columns of the matrix $\mathbf{A}_{LP}$ are linearly independent.* □

Proof: Equivalently, we need to establish that the matrix $\mathbf{T} - \mathbf{A}^0$ is full rank. Recall that there are $|U|$ rows of $\mathbf{T} - \mathbf{A}^0$ associated with $\mathbf{x}_i^s$ $(1 \leq i \leq n)$, one row for each agent decision pair (choose any $s \geq \bar{s}$). For any i $(1 \leq i \leq n)$, the row of $\mathbf{A}^0$ associated with the decision $\mathbf{u}$ where the two agents diverge, has all zero entries, since $s' = s+2$, where s' is the separation at the next stage. Therefore, for each i $(1 \leq i \leq n)$, there is an r $(1 \leq r \leq n|U|)$ such that $\mathbf{T}_{ri} - \mathbf{A}_{ri}^0 = 1$ and $\mathbf{T}_{ri'} - \mathbf{A}_{ri'}^0 = 0$ for $i' \neq i$, where $(\cdot)_{ri}$ is the entry on the rth row and ith column of $(\cdot)$. Therefore, the first n columns of A_{LP} are linearly independent, which concludes the proof. ■

Let $\mu^s : \mathcal{S}(s) \to U$ be any two-agent policy for the Auxiliary Problem $\mathcal{P}(s)$, where $\hat{J}_s(x) \sim \mathbf{x}_c^{s+\sigma}$, for $\sigma \in S^s$, $s \in \mathcal{M}_s$. Let $\mathbf{x}_{\mu^s}^s \sim J_{\mu^s}(x)$, for $x \in \mathcal{S}(s)$, denote the value function associated with stationary policy μ^s, *i.e.* the cost when the agents apply policy μ^s at all stages. We show a property of $\mathbf{x}_{\mu^s}^s$ in the following lemma.

Lemma 12 *For any μ^s, there exists a vector $\mathbf{y} \in \mathbb{R}^{n|U|}$ such that $(\mathbf{x}_{\mu^s}^s, \mathbf{y})$ is a basic solution of $\mathcal{LP}_f^c(s)$.* □

Proof: At the state associated with index i ($1 \leq i \leq n$), the agents take decision $\mathbf{u}_i \in U$ under policy μ^s. For each i, let index j_i ($1 \leq j_i^s \leq |U|$) correspond to the decision $\mathbf{u}_i$. We choose

$$(\mathbf{y}_i)_{j_i} = 0 \qquad \text{for all } i = 1, \ldots, n. \tag{4.35}$$

We now define the matrix $\mathbf{B}_{LP}^{\mu^s} \in \mathbb{R}^{n|U| \times n|U|}$ as consisting of a copy of the matrix $\mathbf{A}_{LP}$, where the columns associated with $(\mathbf{y}_i)_{j_i}$ are removed. We show that $\mathbf{B}_{LP}^{\mu^s}$ is full rank. From Lemma 11, we have that the first n columns of $\mathbf{B}_{LP}^{\mu^s}$ are linearly independent. The last $(|U|-1)n$ columns of $\mathbf{B}_{LP}^{\mu^s}$ are linearly independent since they consist of different columns of $\mathbf{I}$. Lastly, each of the first n columns is linearly independent of the last $(|U|-1)n$ columns, since the ith column of $\mathbf{B}_{LP}^{\mu^s}$ ($1 \leq i \leq n$), has a nonzero element on the row corresponding to $(\mathbf{y}_i)_{j_i}$, and that the same row only contains zeros on its last $(|U|-1)n$ columns. Therefore, the columns of $\mathbf{B}_{LP}^{\mu^s}$ are linearly independent and $\mathbf{B}_{LP}^{\mu^s}$ is an invertible matrix.

Let $\mathbf{w} \in \mathbb{R}^{(|U|-1)n}$ contain all entries of $\mathbf{y}$ except for the entries $(\mathbf{y}_i)_{j_i}$. With $\mathbf{B}_{LP}^{\mu^s}$, we have

$$\begin{bmatrix} \mathbf{x}_{\mu^s} \\ \mathbf{w} \end{bmatrix} = (\mathbf{B}_{LP}^{\mu^s})^{-1} \mathbf{b}_{LP}^s. \tag{4.36}$$

The LHS of Eq. (4.36), with Eq. (4.35), satisfies the equality constraints of $\mathcal{LP}_f^c(s)$. Since $\mathbf{A}_{LP}$ has linearly independent rows, and given the definition of $\mathbf{B}_{LP}^{\mu^s}$, we apply Theorem 2.4 in [8] to conclude that $\mathbf{x}_{\mu^s}$ is a basic solution of $\mathcal{LP}_f^c(s)$, with corresponding matrix $\mathbf{B}_{LP}^{\mu^s}$. ■

From Lemma 12, it follows that to policy μ_∞ is associated a basic solution of $\mathcal{LP}_f^c(s)$ with basis matrix $\mathbf{B}_{LP} \in \mathbb{R}^{n|U| \times n|U|}$. Two conditions need to hold for matrix $\mathbf{B}_{LP}$ to be optimal (see [8] for a detailed treatment). First, the basic solution associated with $\mathbf{B}_{LP}$ is required to be dual feasible. In particular, we have that

$$\bar{\mathbf{c}}^T = \mathbf{c}^T - \mathbf{c}_B^T \mathbf{B}_{LP}^{-1} \mathbf{A}_{LP} \geq 0, \tag{4.37}$$

where $\bar{\mathbf{c}}$ is the *reduced cost* vector, $\mathbf{c}$ is the cost vector of $\mathcal{LP}_f^c(s)$ and where $\mathbf{c}_B$ is the cost vector corresponding to the basis $\mathbf{B}_{LP}$. This condition is independent of $\mathbf{b}_{LP}^s$ and thus needs to be checked only once for each specific α and p, and not for all separations $s \geq \bar{s}$. We have the following lemma.

Lemma 13 *The vector $\bar{\mathbf{c}}^T = \mathbf{c}^T - \mathbf{c}_B^T \mathbf{B}_{LP}^{-1} \mathbf{A}_{LP} \geq 0$, for all p and α such that $0 < p < 1$ and $0 < \alpha < 1$.* □

Proof: The basis matrix $\mathbf{B}_{LP}$ corresponds to the policy μ_∞, a two-agent policy that is a combination of two optimal single-agent policies (see Section 4.2.3). Further, in Lemma 8, we establish that for any $0 < \alpha < 1$ and $0 \leq p \leq 1$, the same set of single-agent policies is optimal. Therefore, in the limit for $s \to \infty$, the policy μ_∞ is an optimal two-agent policy for all $0 < \alpha < 1$ and $0 \leq p \leq 1$.

The vector of reduced costs $\bar{\mathbf{c}}$ is independent of separation s, for $s \geq \bar{s}$. Hence, since in the limit for $s \to \infty$ the basis matrix $\mathbf{B}_{LP}$ is optimal, we have that $\bar{\mathbf{c}}^T \geq 0$. ■

Lemma 13 allows us to conclude that dual feasibility is satisfied for $s \geq \bar{s}$, under Assumption 1.

The second condition is primal feasibility. In particular, we need to check that

$$\mathbf{B}_{LP}^{-1} \mathbf{b}_{LP}^s \geq 0, \tag{4.38}$$

for all $s \geq \bar{s}$. Therefore, we define the output $\mathbf{z}^s$ of the LTI system as

$$\mathbf{z}^s = \mathbf{B}_{LP}^{-1} \left(\mathbf{M}_1 E \mathbf{x}_c^s + \mathbf{M}_2 \right), \tag{4.39}$$

where $\mathbf{M}_1 \in \mathbb{R}^{n|U| \times 9}$ and $\mathbf{M}_2 \in \mathbb{R}^{n|U|}$ are such that $\mathbf{b}_{LP}^s = \mathbf{M}_1 E \mathbf{x}_c^s + \mathbf{M}_2$. If dual feasibility holds for a particular p and α, then $\mathbf{z}^s \geq 0$ for all $s \geq \bar{s}$, indicates optimality of $\mathbf{B}_{LP}$ for all $s \geq \bar{s}$.

Therefore, given a state and output sequence of linear system (4.34,4.39) such that the output is non-negative at all separations, then the state sequence represents the optimal value function at separations $s \geq \bar{s}$. The next section deals with the only remaining issue of computing the adequate system initial condition $E\mathbf{x}_c^{\bar{s}-1}$.

4.2.6 Computation of the Optimal Value Function at Small Separations

We first establish a necessary condition on the initial state of system (4.34). Then, we formulate a LP whose solution is the optimal value function for separations $s < \bar{s}$. The necessary condition on the initial system state is one of the LP constraints, linking the optimal value function at small separations to the optimal value function for $s \to \infty$.

Under Assumption 1, $s = \bar{s}$ is the smallest separation for which the complete optimal value function can be computed with forward recursion (4.28). Therefore, $E\mathbf{x}_c^{s'}$, for any $s' \geq \bar{s} - 1$, is a valid initial state of system (4.34). We choose $E\mathbf{x}_c^{\bar{s}-1}$. Since the optimal value function for the two-agent navigation problem is upper bounded by twice the optimal value function for the single-agent navigation problem, $E\mathbf{x}_c^{\bar{s}-1}$ can only excite the stable modes of system (4.34), leading to a necessary condition on $E\mathbf{x}_c^{\bar{s}-1}$ as follows.

First we have the following lemma, regarding the spectrum $\lambda(\mathbf{A})$ of the system matrix $\mathbf{A}$ [see Eq. (4.34)]. The subscripts u and i indicate eigenvalues corresponding to unstable and stable system modes, respectively.

Lemma 14 *The following holds for all α and p such that $0 < \alpha < 1$ and $0 < p < 1$, $p \neq p'$. The spectrum $\lambda(\mathbf{A})$ of matrix $\mathbf{A}$ contains exactly three eigenvalues $\lambda_{u,i} \in \lambda(\mathbf{A})$, for $i = 1, 2, 3$, such that $|\lambda_{u,i}| > 1$. Furthermore, for all other eigenvalues $\lambda_{s,j} \in \lambda(\mathbf{A})$, for $j = 1, \dots, 6$, we have that $|\lambda_{s,j}| < 1$.* □

Proof: The proof is as follows. We first determine the characteristic equation of $\mathbf{A}$ as

$$C_{\mathbf{A}}(p, \alpha, Y) = \frac{Y^3}{-p^3(\det \mathbf{R}_2')} \sum_{i=0}^{6} P_i(p, \alpha) Y^i,$$

with $Y \in \mathbb{C}$ and where $P_i(p, \alpha)$ for $i = 0, \dots, 6$ are polynomials in p and α. Note that for a particular pair of parameters p and α, the values of Y such that $C_{\mathbf{A}}(p, \alpha, Y) = 0$ are the eigenvalues of matrix $\mathbf{A}$. For $0 < p \neq p' < 1$ and for $0 < \alpha < 1$, the factor $-p^3(\det \mathbf{R}_2') \neq 0$ (see Lemma 10) and can therefore be omitted. We assume from here in that $p \neq p'$, in this proof. The polynomials $P_i(p, \alpha)$ are continuous functions of p and α, for $i = 0, \dots, 6$. Therefore, the number of eigenvalues λ with $|\lambda| < 1$ can only be different at the parameter pairs (p_1, α_1) and (p_2, α_2), for $0 < p_1, p_2, \alpha_1, \alpha_2 < 1$, if there exists a parameter pair (p, α) and a $\lambda \in \mathbb{C}$, with $|\lambda| = 1$, for which $C_{\mathbf{A}}(p, \alpha, \lambda) = 0$, with $p_1 \leq p \leq p_2$ and $\alpha_1 \leq \alpha \leq \alpha_2$. Hence, if we have that $C_{\mathbf{A}}(p, \alpha, Y) \neq 0$ for any $0 < p < 1$ and $0 < \alpha < 1$ and for any $Y \in \mathbb{C}$ with $|Y| = 1$, it follows that the number of eigenvalues λ such that $|\lambda| < 1$, is constant for all parameter pairs (p, α). In other words, with the continuous variation of p and α, no eigenvalue crosses the unit circle in the complex plane. Then, it remains to be verified numerically for some parameter pair $(\bar{p}, \bar{\alpha})$, that there exist exactly three λ_i, for $i = 1, 2, 3$, such that $C_{\mathbf{A}}(\bar{p}, \bar{\alpha}, \lambda_i) = 0$, and with $|\lambda_i| < 1$.

Let

$$C'_{\mathbf{A}}(p, \alpha, Y) = Y^3 \sum_{i=0}^{6} P_i(p, \alpha) Y^i. \tag{4.40}$$

We now show there is no parameter pair (p, α) and a $Y \in \mathbb{C}$ with $|Y| = 1$ for which $C'_{\mathbf{A}}(p, \alpha, Y) = 0$. Since $Y^3 \neq 0$ for $|Y| = 1$, we can omit the factor Y^3 from Eq. (4.40) which yields

$$C''_{\mathbf{A}}(p, \alpha, Y) = \sum_{i=0}^{6} P_i(p, \alpha) Y^i,$$

and use $C''_{\mathbf{A}}(p, \alpha)$ in stead of $C'_{\mathbf{A}}(p, \alpha)$. Next, we perform a change of variables that maps the unit circle in the complex plane into the imaginary axis. In particular, let

$$Y = \frac{1 + W}{1 - W},$$

where $W \in \mathbb{C}$. The equation $C''_{\mathbf{A}}(p, \alpha, Y)$ is transformed into equation $D_{\mathbf{A}}(p, \alpha, W)$ and proving that no root of $C''_{\mathbf{A}}(p, \alpha, Y)$ crosses the unit circle in the complex plane is equivalent to proving that no root of $D_{\mathbf{A}}(p, \alpha, W)$ crosses the imaginary axis, $W = j\omega$, for $\omega \in \mathbb{R}$. Then, we define

$$D_{\mathbf{A}}(p, \alpha, W) = \frac{1}{(1 - W)^6} \sum_{i=0}^{6} P'(p, \alpha) W^i,$$

where $W \in \mathbb{C}$, and where $P'_i(p, \alpha)$ for $i = 0, \ldots, 6$ are polynomials in p and α. The factor $(1 - W)^6 \neq 0$ for $W = j\omega$, with $\omega \in \mathbb{R}$, and can be omitted. Hence, equivalently, we show there is no parameter pair $0 < p, \alpha < 1$, and $\omega \in \mathbb{R}$ such that $D'_{\mathbf{A}}(p, \alpha, j\omega) = 0$, with

$$D'_{\mathbf{A}}(p, \alpha, W) = \sum_{i=0}^{6} P'_i(p, \alpha) W^i.$$

Substituting $W = j\omega$ and separating the real and imaginary parts of the resulting equation, yields

$$D'_{\mathbf{A}}(p,\alpha,j\omega) = D'_{\mathbf{A},1}(p,\alpha,\omega) + jD'_{\mathbf{A},2}(p,\alpha,\omega),$$

with

$$\begin{aligned} D'_{\mathbf{A},1}(p,\alpha,\omega) &= P'_0 - P'_2\omega^2 + P'_4\omega^4 - P'_6\omega^6 \\ D'_{\mathbf{A},2}(p,\alpha,\omega) &= P'_1\omega - P'_3\omega^3 + P'_5\omega^5, \end{aligned} \tag{4.41}$$

where the dependence of P'_i on p and α, for $i = 0,\ldots,6$, is omitted for notational simplicity. The proof has been reduced to establishing that $D'_{\mathbf{A},1}(p,\alpha,\omega) = 0$ and $D'_{\mathbf{A},2}(p,\alpha,\omega) = 0$ cannot both hold for any $0 < p,\alpha < 1$ and any $\omega \in \mathbb{R}$. We focus on Eq. (4.41), and show that $D'_{\mathbf{A},1}(p,\alpha,\omega) > 0$ for $0 < p,\alpha < 1$ and $\omega^2 \in \mathbb{R}^+$. We first perform a change of variables, with $\bar{\omega} = \omega^2$. This yields

$$D''_{\mathbf{A},1}(p,\alpha,\bar{\omega}) = P'_0 - P'_2\bar{\omega} + P'_4\bar{\omega}^2 - P'_6\bar{\omega}^3. \tag{4.42}$$

We apply the Routh-Hurwitz Stability Criterion [43] on Eq. (4.42) and show that if $\bar{\omega}$ is such that $D''_{\mathbf{A},1}(p,\alpha,\bar{\omega}) = 0$ for some $0 < p,\alpha,1$, then $\Re(\bar{\omega}) < 0$. Hence, the corresponding $\omega = \sqrt{\bar{\omega}}$ is complex with a non-zero imaginary part and therefore there exists no $\omega \in \mathbb{R}$ such that $D'_{\mathbf{A},1}(p,\alpha,\omega) = 0$.

We have left to show that the requirements for the Routh-Hurwitz Stability Criterion to hold, are satisfied. Specifically, we need to show that

$$\frac{-P'_4}{P'_6} > 0, \tag{4.43}$$

$$\frac{-P'_4 P'_2}{(P'_6)^2} - \frac{-P'_0}{P'_6} > 0, \tag{4.44}$$

for $0 < p,\alpha < 1$. Condition (4.43) is equivalent to

$$-P'_6(p,\alpha) > 0, \tag{4.45}$$

$$P'_4(p,\alpha) > 0, \tag{4.46}$$

for $0 < p,\alpha < 1$. If Eqs (4.45,4.46) hold, then it can easily be seen that it is sufficient to show that

$$
\begin{aligned}
P_4'(p,\alpha) &> -P_6'(p,\alpha), && (4.47)\\
-P_2'(p,\alpha) &> P_0'(p,\alpha), && (4.48)
\end{aligned}
$$

for condition (4.44) to hold. We refer to the Lemmas 19-22, in Appendix A, where we establish that the conditions in Eqs (4.45-4.48) hold for $0 < p, \alpha < 1$, and hence the proof is complete. ∎

Note that from Lemma 10, we have that for $p = p'$ and any $0 < \alpha < 1$, the forward recursion in Eq. (4.28) is not well-defined. Hence, $\mathbf{A}$ is not well-defined at that parameter value.

With

$$\tilde{E}\mathbf{x}_c^s = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} - E\mathbf{x}_c^s, \tag{4.49}$$

where $\mathbf{I} - \mathbf{A}$ is invertible from Lemma 14, the state transition equation (4.34) becomes

$$\tilde{E}\mathbf{x}_c^{s+1} = \mathbf{A}\tilde{E}\mathbf{x}_c^s. \tag{4.50}$$

We use Schur's unitary triangularization theorem [42], to write Eq. (4.50) as

$$\mathbf{V}(\tilde{E}\mathbf{x}_c^{s+1}) \;=\; \mathbf{U}\mathbf{V}(\tilde{E}\mathbf{x}_c^s), \tag{4.51}$$

where $\mathbf{V}$ is a unitary matrix and where $\mathbf{U} = \mathrm{diag}(\mathbf{U}_s, \mathbf{U}_u)$ with $\mathbf{U}_s \in \mathbb{R}^{6\times 6}$ and $\mathbf{U}_u \in \mathbb{R}^{3\times 3}$ upper triangular matrices containing as diagonal elements $\lambda_{s,j}$ for $j = 1, \ldots, 6$ and $\lambda_{u,i}$ for $i = 1, 2, 3$, respectively (see Lemma 14), and where $\mathbf{VA} = \mathbf{UV}$. The initial state $E\mathbf{x}_c^{\bar{s}-1}$ excites only the stable modes if and only if $\mathbf{V}_u(\tilde{E}\mathbf{x}_c^{\bar{s}-1}) = \mathbf{0}$, where $\mathbf{V}_u$ consists of the rows of $\mathbf{V}$ associated with the unstable system modes. With Eq. (4.49), this yields $\mathbf{V}_u\left((\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} - E\mathbf{x}_c^{\bar{s}-1}\right) = \mathbf{0}$, or

$$\mathbf{V}_u E\mathbf{x}_c^{\bar{s}-1} = \mathbf{V}_u(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}. \tag{4.52}$$

Let $\mathbf{V}_1 \in \mathbb{R}^{3\times 3}$ contain the first two columns and the last column of $\mathbf{V}_u$, *i.e.* the columns corresponding to the entries $E\mathbf{x}^{s+2}_{\times\times,c}$, $E\mathbf{x}^{s+1}_{1\times,c}$ and $E\mathbf{x}^{s+1}_{\times\times,c}$.

Eq. (4.52) allows us to formulate linear program $\mathcal{LP}_{in}$, whose solution is the optimal value function for separations $s < \bar{s}$. In particular, the structure of $\mathcal{LP}_{in}$ is similar to the structure of $\mathcal{LP}_g$ for separations $s < \bar{s}$, with minor changes. To compute the optimal value function at separation $\bar{s}-1$, information is required on the optimal value function at separations $s = \bar{s}, \bar{s}+1$. This information is provided implicitly by adding Eq. (4.52), the necessary condition on $E\mathbf{x}^{\bar{s}-1}_c$, as one of the $\mathcal{LP}_{in}$ constraints. This yields the following LP:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{s=0}^{\bar{s}-1} \mathbf{e}^T\mathbf{x}^s \\
\text{subject to} \quad & (\mathbf{T}-\mathbf{A}^0_s)\mathbf{x}^s \leq \sum_{\sigma\in S^s} \mathbf{A}^\sigma_s\mathbf{x}^{s+\sigma} + \mathbf{b}^s, \quad s = 0,1,\ldots,\bar{s}-3, && (4.53)\\
& (\mathbf{T}-\mathbf{A}^0_{\bar{s}-2})\mathbf{x}^{\bar{s}-2} \leq \sum_{\sigma\in\{-2,-1,1\}} \mathbf{A}^\sigma_{\bar{s}-2}\mathbf{x}^{\bar{s}-2+\sigma} + \bar{\mathbf{A}}^2_{\bar{s}-2}E\mathbf{x}^{\bar{s}}_{\times\times} + \mathbf{b}^{\bar{s}-2}, && (4.54)\\
& (\mathbf{T}-\mathbf{A}^0_{\bar{s}-1})\mathbf{x}^{\bar{s}-1} \leq \sum_{\sigma\in\{-2,-1\}} \mathbf{A}^\sigma_{\bar{s}-1}\mathbf{x}^{\bar{s}-1+\sigma} + \bar{\mathbf{A}}^1_{\bar{s}-1}\begin{bmatrix} E\mathbf{x}^{\bar{s}}_{1\times} \\ E\mathbf{x}^{\bar{s}}_{\times\times}\end{bmatrix} + \ldots \\
& \qquad\qquad \bar{\mathbf{A}}^2_{\bar{s}-1}E\mathbf{x}^{\bar{s}+1}_{\times\times} + \mathbf{b}^{\bar{s}-1}, && (4.55)\\
& \mathbf{V}_u(E\mathbf{x}^{\bar{s}-1}) = \mathbf{V}_u(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}, && (4.56)
\end{aligned}
$$

where $\bar{\mathbf{A}}^1_s$ and $\bar{\mathbf{A}}^2_s$ are such that

$$
\begin{aligned}
\bar{\mathbf{A}}^1_s \begin{bmatrix} E\mathbf{x}^{s+1}_{1\times} \\ E\mathbf{x}^{s+1}_{\times\times}\end{bmatrix} &= \mathbf{A}^1_s\mathbf{x}^{s+1}, && (4.57)\\
\bar{\mathbf{A}}^2_s E\mathbf{x}^{s+2}_{\times\times} &= \mathbf{A}^2_s\mathbf{x}^{s+2}. && (4.58)
\end{aligned}
$$

A reasoning identical to the one that led to the derivation of Eqs. (4.20) and (4.21) leads to Eqs. (4.57) and (4.58), where we use Eq. (4.19) with q taken to be $\times$ to yield Eq. (4.57).

The following theorem links the solution of $\mathcal{LP}_{in}$ and the stable state trajectory of system (4.34) to the fixed point of F, and therefore to the optimal value function, under some technical conditions. Let $\mathbf{x}^s_c$, for $s = 0,1,\ldots,\bar{s}-1$ and $E\mathbf{x}^{\bar{s}}_{1\times,c}, E\mathbf{x}^{\bar{s}}_{\times\times,c}$ and $E\mathbf{x}^{\bar{s}+1}_{\times\times,c}$ denote the solution of $\mathcal{LP}_{in}$. Let the vector $E\mathbf{x}^{\bar{s}-1}_c$ [see Eq. (4.33)] be determined, based on $\mathbf{x}^s_c$, for $s = \bar{s}-3, \bar{s}-2, \bar{s}-1$ and on $E\mathbf{x}^{\bar{s}}_{1\times,c}, E\mathbf{x}^{\bar{s}}_{\times\times,c}$ and $E\mathbf{x}^{\bar{s}+1}_{\times\times,c}$. Lastly, let $E\mathbf{x}^s_c$, for $s = \bar{s}, \bar{s}+1, \ldots$ be the state of system (4.34) at separation s, with $E\mathbf{x}^{\bar{s}-1}_c$ as initial condition.

For $s \geq \bar{s}$, let $\mathbf{x}_c^s$ be defined as

$$\mathbf{x}_c^s \;\sim\; J_c(x) \;=\; g(x, \mu_\infty(x)) + \alpha E[J_c(f(x, \mu_\infty(x)))], \qquad x \in \mathcal{S}(s)$$

where the expected value is taken over the unknown edge costs. Note that the required expected values of $J_c(x)$, for $x \in \mathcal{S}(s)$ and $s \geq \bar{s}$, are determined from $E\mathbf{x}_c^s$ with Eq. (4.18).

Theorem 2 *For any α and p such that $0 < \alpha < 1$ and $0 < p < 1$, $p \neq p'$, the vectors $\mathbf{x}_c^s$ for $s = 0, 1, \ldots$, represent the Main Problem 1 optimal value function for some $\bar{s} \geq 0$ if $\mathbf{V}_1$ is rank three and if for the output of system (4.34), we have that*

$$\mathbf{z}^s \geq 0, \qquad s = \bar{s}, \bar{s}+1, \ldots. \tag{4.59}$$

Furthermore, we have that the optimal two-agent policy for separation $s \geq \bar{s}$ is μ_∞. $\square$

Proof: We show that

$$\mathbf{X}_c = \left[\begin{array}{cccc} (\mathbf{x}_c^0)^T & (\mathbf{x}_c^1)^T & (\mathbf{x}_c^2)^T \cdots \end{array} \right]^T,$$

is the unique fixed point of the function F, so that with Theorem 1, we have that $\mathbf{X}_c$ represents the optimal value function and hence is equal to $\mathbf{X}^*$.

We can see that the constraints in Eqs (4.53-4.55) of $\mathcal{LP}_{in}$ are identical to the constraints associated with the separations $s = 0, \ldots, \bar{s}-1$ in $\mathcal{LP}_g$. Without any other constraint, the presence of $E\mathbf{x}_{\times\times}^{\bar{s}}$, $E\mathbf{x}_{1\times}^{\bar{s}}$ and $E\mathbf{x}_{\times\times}^{\bar{s}+1}$ in the RHS of Eqs (4.54-4.55) yields an unbounded $\mathcal{LP}_{in}$ solution. However, with the necessary condition in Eq. (4.56), where the matrix $\mathbf{V}_1$ is of rank three, we can express $E\mathbf{x}_{1\times}^{\bar{s}}, E\mathbf{x}_{\times\times}^{\bar{s}}$ and $E\mathbf{x}_{\times\times}^{\bar{s}+1}$ as unique affine functions of $\mathbf{x}^s$, for $s = 0, \ldots, \bar{s}$. Since, $\mathbf{x}_c^s$ for $s = 0, \ldots, \bar{s}-1$, solve $\mathcal{LP}_{in}$, we have that $\mathbf{x}_c^s$ is the solution of $\mathcal{LP}(s)$ with $\hat{\mathbf{x}}^{s+\sigma} = \mathbf{x}_c^{s+\sigma}$ ($\sigma \in S^s$). The latter follows from the definition of $\mathcal{LP}(s)$.

For the given parameter ranges, the matrices $\mathbf{R}_2'$ and $\mathbf{I}-\mathbf{A}$ are invertible (see Lemmas 10 and 14), and therefore the LTI system (4.34) and the constraint in Eq. (4.56) are well defined. Then, from the latter constraint, we have that $E\mathbf{x}_c^{\bar{s}-1}$ as initial state for system (4.34) is such that only the stable modes are excited. Furthermore, the vector $\mathbf{x}_c^s$ solves $\mathcal{LP}(s)$ with $\hat{\mathbf{x}}^{s+\sigma} = \mathbf{x}_c^{s+\sigma}$ ($\sigma \in S^s$), for $s = \bar{s}, \bar{s}+1, \ldots$ since both the primal [see Eq. (4.59)] and dual [see Lemma 13] feasibility conditions are satisfied for all $s \geq \bar{s}$.

Hence, $\mathbf{x}_c^s$ solves $\mathcal{LP}(s)$ with $\hat{\mathbf{x}}_c^{s+\sigma} = \mathbf{x}_c^{s+\sigma}$ ($\sigma \in S^s$), for $s = 0, 1, \ldots$ and therefore $\mathbf{X}_c$ is the unique fixed point of F and represents the optimal value function. Lastly, from Assumption 1, we have that for $s \geq \bar{s}$, policy μ_∞ is optimal, which concludes the proof. ■

We now have a computationally simple algorithm that computes the optimal value function for $s < \bar{s}$ as the solution of $\mathcal{LP}_{in}$, which is relatively small-sized given that $\bar{s}$ is not too large. The optimal value function for $s \geq \bar{s}$ is implicitly described via a LTI system. In Section 6.5, we give $\bar{s}$ for different probabilities p; for most p, separation $\bar{s} = 3$. Furthermore, numerical experiments indicate that the technical condition on the rank of $\mathbf{V}_1$ holds in all cases computed.

Brief Algorithm Summary

Finally, we summarize the iterative use of the presented algorithm. In particular, we start by constructing the LTI system using policy μ_∞, the combination of two single-agent optimal policies. We extract the connection constraints from the LTI system, necessary conditions which ensure that the initial condition only excites stable LTI modes. Then, we start an iterative process and choose a particular, small $\bar{s}$. We solve $\mathcal{LP}_{in}$, including the connection constraints, with the chosen value for $\bar{s}$, extract the LTI initial condition and simulate the LTI system. If its output is non-negative, the algorithm terminates successfully. Otherwise, start a new iteration by solving $\mathcal{LP}_{in}$ with $\bar{s}$ incremented by one, and so forth.

4.2.7 Ill-Defined Forward Recursion: a Fix

For the sake of completeness, we present a solution for when $p = p'$. In this case, the forward recursion as developed in Section 4.2.4 is not well-defined. In this section, we present an alternative forward recursion and the corresponding LTI system which allows for the computation of the necessary extra constraints in $\mathcal{LP}_{in}$, along the lines of the method presented in the sections 4.2.4-4.2.6.

From Lemma 10, we have that for $p = p'$ the matrix $\mathbf{R}_2'$ is singular. Therefore, the transition from the mixed forward-backward recursion [Eq. (4.27)] to a purely forward recursion [Eq. (4.28)] does not longer hold, and an alternative forward recursion is necessary. We start with the mixed forward-backward recursion from Eq. (4.27), repeated here for convenience:

$$\begin{aligned}\mathbf{R}_0 \begin{bmatrix} E\mathbf{x}_{\times\times}^{s,*} & E\mathbf{x}_{1\times}^{s,*} & E\mathbf{x}_{11}^{s,*} \end{bmatrix}^T &= \dots \\ \mathbf{R}_{-2} E\mathbf{x}_{\times\times}^{s-2,*} + \mathbf{R}_{-1} \begin{bmatrix} E\mathbf{x}_{\times\times}^{s-1,*} \\ E\mathbf{x}_{1\times}^{s-1,*} \end{bmatrix} &+ \mathbf{R}_1 \begin{bmatrix} E\mathbf{x}_{\times\times}^{s+1,*} \\ E\mathbf{x}_{1\times}^{s+1,*} \end{bmatrix} + \mathbf{R}_2 E\mathbf{x}_{\times\times}^{s+2,*} + \mathbf{r}. \end{aligned} \tag{4.60}$$

Recall the definition of matrix $\mathbf{R}'_2$ from Eq. (4.29):

$$\mathbf{R}'_2 = \begin{bmatrix} -\mathbf{R}_2 & -\mathbf{R}_{1,2} & \mathbf{R}_{0,3} \end{bmatrix}.$$

For $p = p'$, the matrix $\mathbf{R}'_2 \in \mathbb{R}^{3\times3}$ is singular; more specifically, $\mathrm{rank}(\mathbf{R}'_2) = 2$ and we have that the first and second column of $\mathbf{R}'_2$ are linearly dependent, or $f_1\,\mathbf{R}'_{2,1} = \mathbf{R}'_{2,2}$, where

$$f_1 = -\frac{2(1-p')^2}{p'}.$$

Equivalently, and from the definition of $\mathbf{R}'_2$, we have that $f_1\,\mathbf{R}_2 = \mathbf{R}_{1,2}$. We use the latter relation to rearrange the terms in Eq. (4.60) which yields

$$\mathbf{R}''_2 \begin{bmatrix} E\mathbf{x}_{\times\times}^{s+2} - f_1 E\mathbf{x}_{1\times}^{s+1} \\ E\mathbf{x}_{\times\times}^{s+1} \\ E\mathbf{x}_{11}^{s} \end{bmatrix} = \sum_{\sigma=-1}^{1} \mathbf{R}''_\sigma \begin{bmatrix} E\mathbf{x}_{\times\times}^{s+\sigma} - f_1 E\mathbf{x}_{1\times}^{s+\sigma-1} \\ E\mathbf{x}_{\times\times}^{s+\sigma-1} \\ E\mathbf{x}_{11}^{s+\sigma-2} \end{bmatrix} + \mathbf{r}, \tag{4.61}$$

with

$$\begin{aligned} \mathbf{R}''_2 &= \begin{bmatrix} -\mathbf{R}_2 & -\mathbf{R}_{1,1} + \frac{1}{f_1}\mathbf{R}_{0,2} & \mathbf{R}_{0,3} \end{bmatrix}, \\ \mathbf{R}''_1 &= \begin{bmatrix} \frac{1}{f_1}\mathbf{R}_{0,2} & -\mathbf{R}_{0,1} + \frac{1}{f_1}\mathbf{R}_{-1,2} & \mathbf{0} \end{bmatrix}, \\ \mathbf{R}''_0 &= \begin{bmatrix} -\frac{1}{f_1}\mathbf{R}_{-1,2} & \mathbf{R}_{-1,1} & \mathbf{0} \end{bmatrix}, \\ \mathbf{R}''_{-1} &= \begin{bmatrix} \mathbf{0} & \mathbf{R}_{-2} & \mathbf{0} \end{bmatrix}, \end{aligned}$$

where all matrices are computed for $p = p'$ and $0 < \alpha < 1$. One can easily check by inspection that

$$\det(\mathbf{R}''_2) = (0.4424 + 0.0058\alpha)\alpha > 0,$$

for $0 < \alpha < 1$. Hence, $\mathbf{R}_2''$ is invertible and the recursion in Eq. (4.61) constitutes a well-defined forward recursion, which we now transform in a straightforward manner to an equivalent LTI system. Here, for simplicity, we only retain states necessary for the system equation, omitting states required for the system output. In particular, we have that

$$E\mathbf{x}^{s+1,*} = \mathbf{A}' E\mathbf{x}^{s,*} + \mathbf{B}', \tag{4.62}$$

with state vector

$$\begin{aligned} E\mathbf{x}^{s,*} \quad = \quad \Big[\; (E\mathbf{x}_{\times\times}^{s+2,*} - f_1 E\mathbf{x}_{1\times}^{s+1,*}) \quad E\mathbf{x}_{\times\times}^{s+1,*} \quad E\mathbf{x}_{11}^{s,*} \quad (E\mathbf{x}_{\times\times}^{s+1,*} - f_1 E\mathbf{x}_{1\times}^{s,*}) \quad \ldots \\ E\mathbf{x}_{\times\times}^{s,*} \quad E\mathbf{x}_{\times\times}^{s-1,*} \; \Big]^T, \end{aligned}$$

with system matrix

$$\mathbf{A}' = \left[\begin{array}{cc|c|cc|c} \multicolumn{2}{c|}{} & 0 & \multicolumn{2}{c|}{} & \\ \multicolumn{2}{c|}{(\mathbf{R}_2'')^{-1}\mathbf{R}_{1,(1,2)}''} & 0 & \multicolumn{2}{c|}{(\mathbf{R}_2'')^{-1}\mathbf{R}_{0,(1,2)}''} & (\mathbf{R}_2'')^{-1}\mathbf{R}_{-1,2}'' \\ \multicolumn{2}{c|}{} & 0 & \multicolumn{2}{c|}{} & \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right],$$

and with

$$\mathbf{B}' = \Big[\; ((\mathbf{R}_2'')^{-1}\mathbf{r})^T \quad 0 \quad 0 \quad 0 \; \Big]^T.$$

The remainder of the development for $p = p'$ is identical as in Section 4.2.6 where the eigenvectors corresponding to the eigenvalues associated with unstable system modes (gathered as rows in matrix $\mathbf{V}_u'$) are extracted from system matrix $\mathbf{A}'$ yielding a necessary condition for the optimal value function, here repeated from Section 4.2.6,

$$\mathbf{V}_u'(E\mathbf{x}^{\bar{s}-1}) = \mathbf{V}_u'(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}'.$$

This condition, added as a constraint in $\mathcal{LP}_{in}$ ensures that the initial condition for the LTI system, as computed in $\mathcal{LP}_{in}$, is such that only the stable system modes are excited.

However, without a proof, we mention that for all values for $0 < \alpha < 1$ tested, the LTI system in Eq. (4.62) has at least two unstable modes, rather than the previously established three. To explain this difference, we repeat the $\mathcal{LP}_{in}$ constraints in Eq. (4.54) and Eq. (4.55) adapted for the case at hand where $p = p'$, which yields

$$\begin{aligned}
(\mathbf{T} - \mathbf{A}^{0}_{\bar{s}-2})\mathbf{x}^{\bar{s}-2} &\leq \sum_{\sigma \in \{-2,-1,1\}} \mathbf{A}^{\sigma}_{\bar{s}-2}\mathbf{x}^{\bar{s}-2+\sigma} + \bar{\mathbf{A}}^{1}_{\bar{s}-2}E\mathbf{x}^{\bar{s}}_{\times\times} + \mathbf{b}^{\bar{s}-2}, \\
(\mathbf{T} - \mathbf{A}^{0}_{\bar{s}-1})\mathbf{x}^{\bar{s}-1} &\leq \sum_{\sigma \in \{-2,-1\}} \mathbf{A}^{\sigma}_{\bar{s}-1}\mathbf{x}^{\bar{s}-1+\sigma} + \dots \\
&\qquad \bar{\mathbf{A}}^{2}_{\bar{s}-1}E\mathbf{x}^{\bar{s}}_{\times\times} + \bar{\mathbf{A}}^{3}_{\bar{s}-1}(E\mathbf{x}^{\bar{s}+1}_{\times\times} - f_1 E\mathbf{x}^{\bar{s}}_{1\times}) + \mathbf{b}^{\bar{s}-1},
\end{aligned}$$

where $\bar{\mathbf{A}}^{1}_{\bar{s}-2}$, $\bar{\mathbf{A}}^{2}_{\bar{s}-1}$ and $\bar{\mathbf{A}}^{3}_{\bar{s}-1}$ are such that

$$\begin{aligned}
\bar{\mathbf{A}}^{1}_{\bar{s}-2}E\mathbf{x}^{\bar{s}}_{\times\times} &= \mathbf{A}^{2}_{\bar{s}-2}\mathbf{x}^{\bar{s}} \\
\bar{\mathbf{A}}^{2}_{\bar{s}-1}E\mathbf{x}^{\bar{s}}_{\times\times} + \bar{\mathbf{A}}^{3}_{\bar{s}-1}(E\mathbf{x}^{\bar{s}+1}_{\times\times} - f_1 E\mathbf{x}^{\bar{s}}_{1\times}) &= \mathbf{A}^{1}_{\bar{s}-1}\mathbf{x}^{\bar{s}} + \mathbf{A}^{2}_{\bar{s}-1}\mathbf{x}^{\bar{s}+1}.
\end{aligned}$$

The latter relations are derived with a similar reasoning that lead to the Eqs. (4.57-4.58), additionally taking into consideration the linear dependence of the first two columns of $\mathbf{R}'_2$.

Notice that the constraints of $\mathcal{LP}_{in}$ as adapted for $p = p'$ only depend on two linearly independent functions of the optimal value function at separations $\bar{s}$ and $\bar{s}+1$, namely $E\mathbf{x}^{\bar{s}}_{\times\times}$ and $(E\mathbf{x}^{\bar{s}+1}_{\times\times} - f_1 E\mathbf{x}^{\bar{s}}_{1\times})$. Let $\mathbf{V}'_1$ contain as its columns the first and second columns of $\mathbf{V}'_u$. The matrix $\mathbf{V}'_1$ has at least two rows. Hence, for the proof of Theorem 2 to hold in the case where $p = p'$, the matrix $\mathbf{V}_1$ is required to be of rank two. Numerical experiments indicate that this is the case for different $0 < \alpha < 1$.

Therefore, the alternative method presented in this section for the particular case when $p = p'$ and the original forward recursion is ill-defined, is correct.

4.3 Two Agents with Short Tunnel-Shaped Local Observation

For the two-agent problem with short tunnel-shaped local observation zones, we show that the agent separation is upper bounded by three in steady state under an optimal policy.

Further, we establish that if the initial agent separation is larger than three, the optimal policy is such that the separation sequence is non-increasing until reaching a separation of three or less.

Let an optimal navigation policy for the two-agent cluster minimize

$$\lim_{K\to\infty} E\left[\sum_{i=0}^{K} \alpha^i g(x_i, \mathbf{u}_i)\right],$$

where $g(x_i, \mathbf{u}_i)$ denotes the edge cost sum of the two traversed edges at stage i, where the two-agent system is at state x_i, and takes decision $\mathbf{u}_i$. Given an initial agent separation s_0 at stage 0 on $\mathcal{G}$, let $\{s_i\}$, $i = 1, 2, \ldots$ denote the agent separation sequence with s_i the separation at stage i. Let stage $d \geq 0$ be the first stage where $s_i \leq 3$. The following lemma states that the maximum agent separation of a two-agent cluster is bounded above by three:

Lemma 15 *Given $s_0 \leq 3$, then for a two-agent group under an optimal policy, the separation sequence $\{s_i\}$, $i \geq 0$, satisfies:* $\max_i\{s_i\} \leq 3$. □

Proof: We use a coupling argument. We consider two arbitrary trajectories $t_n = \{l_{i,n}\}$, $i = 0, 1 \ldots$, one for each agent A_n ($n = 1, 2$) where A_n is positioned on lane $l_{i,n}$ at stage i (see Fig. 4-1). Furthermore, let a_i^n denote the cost of the edge straight ahead of agent A_n at stage i. Let $u_{i,n}$ indicate the decision agent A_n takes at stage i. Note that $l_{i,n} - l_{i-1,n} = u_{i-1,n}$.

We construct two trajectories, $t'_n = \{l'_{i,n}\}$, for $n = 1, 2$, one for each agent. Along the new trajectory set, the agents incur each cost with the same probability at each stage as along the trajectories t_n, guaranteeing that the expected value of the discounted sum of incurred costs is equal along both trajectory sets. However, the new trajectory set is constructed such that at each stage the agent separation does not exceed three (see Fig. 4-1). In other words, for any trajectory set (with $s_0 \leq 3$), we can find a trajectory set that is at least as good and with the property that the agents do not exceed separation three. In particular, t'_n is such that at each stage i, the agents observe edge costs a_i^n with the same probability as in t_n. Furthermore, $|u_{i,n}| = |u_{i,n}|$, or, if at stage i, agent A_n chooses to traverse edge a_i^n in the trajectory t_n, then A_n does the same in trajectory t'_n; if at stage i, agent A_n chooses to take a diagonal edge in trajectory t_n, it chooses a diagonal edge in the

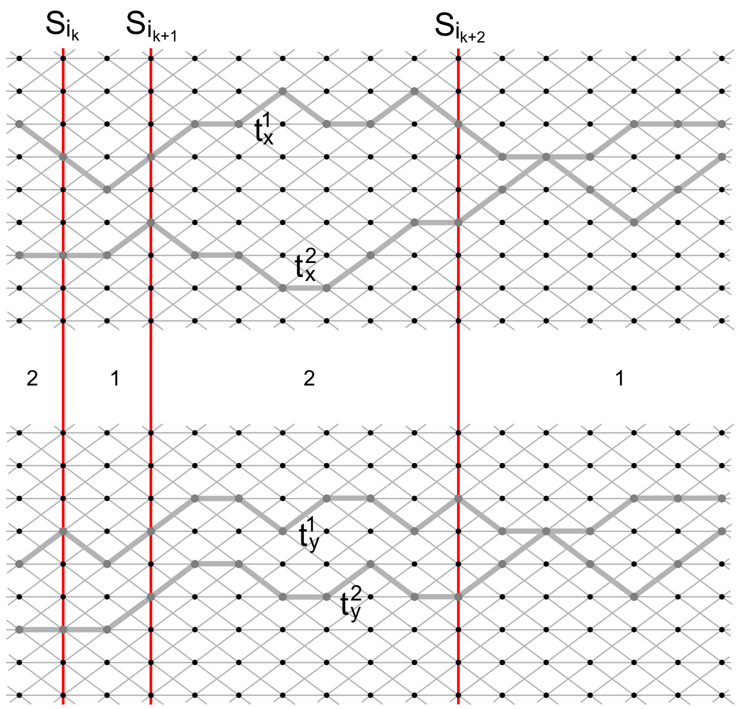

Figure 4-1: Principle of the proof. The bold grey lines denote the respective agent trajectories, while the red vertical lines denote initial/final segment stages. Note that the maximum separation of the bottom trajectory set is three.

trajectory t'_n. This guarantees that the probability by which particular costs are incurred along t_n and t'_n is equal, since unknown edge costs are i.i.d. random variables.

Now, we show how to construct t'_n for $n = 1, 2$ such that $|l'_{i,1} - l'_{i,2}| \leq 3$, for $i = 0, 1, \ldots$. We split trajectories t_n, for $n = 1, 2$ in an alternating sequence of two types of segments, determined by the agent separation in the following way. Let $\{i_k\}$, for $k \geq 0$, denote the set of stage indices where consecutive segments start ($i_0 = 0$), then:

- Let a σ_1-type segment start at stage i_k and end at stage i_{k+1}. Then, $s_i \leq 3$, for $i_k \leq i \leq i_{k+1}$. In other words, the maximum agent separation of the segment is three (see Fig. 4-1). Note that $s_{i_k - 1} > 3$, unless $k = 0$, and $s_{i_{(k+1)}+1} > 3$. Since $s_0 \leq 3$, σ_1-segments start at stages with even index k.

- Every σ_1-segment is followed by a σ_2-type segment, e.g. starting at stage i_{k+1} and ending at stage i_{k+2}. Then, $s_i > 3$, for $i_{(k+1)} + 1 \leq i \leq i_{(k+2)} - 1$. In other words, the minimum agent separation over all stages in between the initial and final segment stage, is at least four (See Fig. 4-1).

We construct t'_n, for $n = 1, 2$, so that at the start of all σ_1-segments, i.e. at states i_k for k

even, we have that

$$s_{i_k} = s'_{i_k}. \tag{4.63}$$

Choosing $u'_{i,n} = u_{i,n}$ for $n = 1, 2$ ensures that s and s' are equal in σ_1-segments.

In σ_2-segments, we construct t'_n for $n = 1, 2$ so that at all stages the agent separation equals two or three and so that Eq. (4.63) holds. Let a σ_2-segment start at stage i_{k+1}, and end at stage i_{k+2}. Since i_{k+1} is the last stage of a σ_1-segment, we have that $s'_{i_{k+1}} = s_{i_{k+1}} \in \{2, 3\}$. Furthermore, $u'_{i,n}$ for $n = 1, 2$, $i_{k+1} \le i < i_{k+2}$, is chosen such that $s'_i \in \{2, 3\}$ in the following way. Three distinct cases can occur at a particular stage i, for $i_{k+1} \le i < i_{k+2}$ (see Fig. 4-1 for examples):

- $u_{i,n} = 0$ for $n = 1, 2$ in which case we choose $u'_{i,n} = 0$, for $n = 1, 2$, i.e. A_1 and A_2 choose the straight ahead edge leaving their separation unchanged.

- $u_{i,1} = 0$ and $|u_{i,2}| = 1$, in which case we choose $u'_{i,1} = 0$ and $|u'_{i,2}| = 1$ so that $s_{i+1} \in \{2, 3\}$, i.e. the agent separation changes by one, from two to three, or from three to two. The symmetric case is similar.

- $|u_{i,n}| = 1$ for $n = 1, 2$ in which case we choose $u'_{i,n} = 1$ for $n = 1, 2$, or $u'_{i,n} = -1$ for $n = 1, 2$, i.e. A_1 and A_2 move parallel leaving their separation unchanged.

The construction of t'_n, $n = 1, 2$, implies that $s'_{i_{k+2}} = s_{i_{k+2}}$. In particular, if $s_{i_{k+1}} = s_{i_{k+2}}$, then along this σ_2-segment the separation changes an even number of times by one. The construction of t'_n, $n = 1, 2$, implies an equal number of separation changes by one as in t_n, $n = 1, 2$. Therefore, if $s_{i_{k+1}} = s'_{i_{k+1}}$, it follows that $s_{i_{k+2}} = s'_{i_{k+2}}$. A similar reasoning holds when $s_{i_{k+1}} \neq s_{i_{k+2}}$. This concludes the proof. ■

Lemma 15 shows that once the agent separation is within three, it remains within three under an optimal policy. For $s_0 > 3$, the following result shows that under an optimal policy the agent separation does not increase until the agents reach separation three or smaller.

Lemma 16 *Given $s_0 > 3$. Then, for a two-agent group under an optimal policy, at all stages i, for $1 \le i < d$, where $s_d \le 3$, the separation $s_i \ge s_{i+1} > 3$,* □

Proof: We use a coupling argument. Take two arbitrary agent trajectories t_n, for $n = 1, 2$. Refer to the proof of Lemma 15 for the notation. We construct two trajectories, t'_n, $n = 1, 2$, along which the agents incur the costs with the same probability as along the trajectories t_n, for $n = 1, 2$ respectively and which have the property of Lemma 16: $s_i \geq s_{i+1}$, for $1 \leq i < d$. At each stage i, let the agents observe a_i^n along t'_n. The probability of observing a_i^n on both t_n and t'_n, $n = 1, 2$ is equal, since unknown edge costs are i.i.d. random variables and since we choose $|u_{i,n}| = |u_{i,n}|$ for $n = 1, 2$. That is, if at stage i, agent A_n takes edge a_i^n on t_n, then it takes edge a_i^n on t'_n, and if agent A_n takes a diagonal edge on t_n, it takes a diagonal edge on t'_n. This guarantees that the probabilities by which costs are incurred along t_n and t'_n are equal.

Next, we show how to construct t'_n, for $n = 1, 2$. Let $s'_0 = s_0$. Let stage $r \geq 0$, be such that $s_r \leq 3$, where $s_i > 3$, for $0 \leq i \leq r - 1$. We construct the two trajectories t'_n, for $n = 1, 2$, such that $s_r = s'_r$. At stage r we can then apply Lemma 15, guaranteeing a maximum agent separation of three from stage r on. If no such r exists, then omit the previous step and consider r infinitely large in what follows. From stages 0 to $r - 1$, do the following until the agent separation along the trajectories t'_n, for $n = 1, 2$, equals two or three. If $u_{i,n} = 0$, take $u'_{i,n} = 0$; if $|u_{i,n}| = 1$, take $u'_{i,n}$ such that the $s_i > s_{i+1}$. Then, follow a similar strategy as in the proof of Lemma 15 until stage r, guaranteeing that the maximum agent separation does not exceed three. Similarly as in the proof of Lemma 15, we can establish that the trajectories t'_n, for $n = 1, 2$ constructed with the previous procedure, are such that $s_r = s'_r$. This concludes the proof. ■

Given an initial separation $s_0 > 3$, we have that from Lemma 16, the agents eventually reach a separation of three under an optimal policy. Hence, from Lemma 15, we conclude that in steady state, the agent separation does not exceed three. We compute the optimal policy by solving a DP where decisions leading to an agent separation larger than three are not allowed. This yields a standard finite state space DP, which can be solved with any method of choice.

4.4 Two Agents with Extended Tunnel-Shaped Local Observation

In this section, we consider the two-agent problem with extended tunnel-shaped local observation zones. We mainly focus on the technical aspects which differ from the two-agent case with cone-shaped local observation zones. In Section 4.4.1, we introduce some notation and

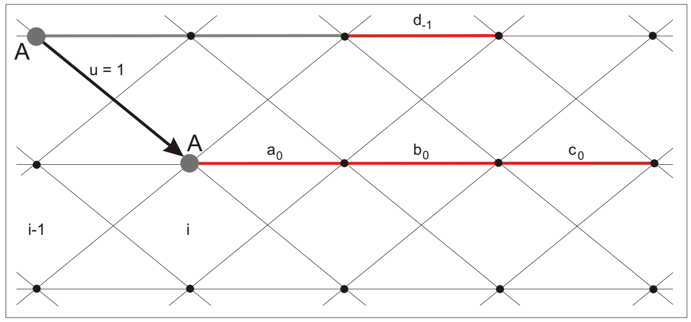

Figure 4-2: Extended tunnel-shaped local observation zone: Example of an at the current stage i unobserved, though known edge cost (edge d_{-1}). Indeed, agent A observes edge d_{-1} at stage $i-1$. This edge is reachable and unobserved for A when positioned at stage i, after taking decision $u = 1$. Note that the edges a_0, b_0, and c_0 constitute A's local observation zone at stage i, the set of red edges denotes the set of known reachable edge costs at stage i.

present the single-agent optimal policies in a specific interval for p. Then, in Section 4.4.2, we present the recursion which implicitly describes the optimal value function evolution for $s \geq \bar{s}$ and indicate how to proceed with its rather cumbersome transformation into a forward recursion. Finally, in Section 4.4.3, we present the connection constraints which suffice for the problem instance at hand.

4.4.1 Single Agent Problem

In this section, we define the system state for the single-agent problem with extended tunnel-shaped local observation zone and provide the optimal policies, computed by solving a finite state, infinite horizon discounted cost DP, with any method of choice. We keep our habit of referring to a particular edge by the symbol representing its cost.

We now define the system state, with corresponding state space $\mathcal{S}_1$. The subscript indicates we deal with a single-agent problem. For the extended tunnel-shaped local observation zone, it is possible for an edge cost to be known at the current agent position without being an element of the agent's current local observation zone. Fig. 4-2 shows an example where edge cost d_{-1} is known at the current agent position (at stage i) through observation at the previous agent position (at stage $i-1$). Therefore, at each stage the system state denotes the set of at that stage known costs of the reachable edges. Two types of states are possible. First, there is the set of states that occur in case at the previous stage, agent A takes decision $u = 0$ ($u \in U_1 = \{-1, 0, 1\}$, the decision space), *i.e.* chooses to traverse the edge a_0, the edge straight ahead, not deviating from its current lane. In this case, $x = \mathcal{C}$ where

$\mathcal{C} = \begin{bmatrix} a_0 & b_0 & c_0 \end{bmatrix}^T$, that is, the local observation zone associated with the agent at its current position. Second, there is the set of states that occur in case at the previous stage $|u| = 1$, that is, A chooses to traverse a diagonal edge. Then, the state $x = (\mathcal{C}, d)$. That is, the state consists of the local observation zone $\mathcal{C}$ associated with the current agent position, and edge cost $d = d_u$, known through observation at the previous stage and reachable from the current stage (see Fig. 4-2). Due to symmetry, it is not necessary to make a distinction between the cases where $u = 1$ and $u = -1$, at the previous stage.

Other aspects of the mathematical problem formulation are similar to the cone-shaped local observation zone case, and a the solution of a finite state, infinite horizon discounted cost DP can be computed using any standard method, for example by solving its LP equivalent with CPLEX. Numerical computation of the optimal value function indicates that the same set of policies is optimal for the values of $p = 0.42, 0.43, \ldots, 0.99$. A change in the optimal policy set occurs for some value of $0.41 \leq p \leq 0.42$. The latter can be determined in a straightforward manner by analysis of the LP, equivalent to the DP, but is outside the scope of this thesis. We focus on the single-agent optimal policy $\mu_1^* : \mathcal{S}_1 \rightarrow U_1$ at the presented values for p, easily described in words. In particular, if the system state x is such that $a_0 = 0$, we have that $\mu^*(x) = 0$. Otherwise, there are three possibilities. First, d is unknown, in which case $\mu^*(x) \in \{-1, 1\}$. Second, $d = 0$, in which case $\mu^*(x)$ is such that A traverses the diagonal edge leading to the lane edge d is part of. Third, $d = 1$, in which case $\mu^*(x)$ is such that A traverses the diagonal edge leading to the lane edge d is not part of. We now proceed with the two-agent case with extended tunnel-shaped local observation zones.

4.4.2 Value Function Recursion at Large Separations

In this section, we use the optimal single-agent policy to define the recursion describing the evolution of the optimal value function at separations $s \geq \bar{s}$, where $\bar{s}$ is defined similarly as before as the smallest separation where a combination of two single agent optimal policies constitutes the two-agent optimal policy. Again, in case the single agent optimal policy exhibits a degree of freedom, that pair of decisions is chosen that results in the smallest agent separation at the next stage. As before, $\bar{s}$ is determined during the execution of the algorithm that computes the two-agent optimal value function. With a slight abuse of notation, we use the same symbols as in the two-agent case with cone-shaped local observation zones, omitting the explicit indication that in this section we treat its equivalent with extended tunnel-shaped local observation zones. Further note that for the values of $p = 0.42, 0.43, \ldots, 0.99$, the same single-agent policy is optimal, and hence the development that follows for the two-agent version of the problem only holds for the latter values of p. For the values of p where the single agent optimal policy differs from the policy at $p = 0.42, 0.43, \ldots, 0.99$, a different analysis is necessary for the two-agent behavior at large separations ($s \geq \bar{s}$).

We first define the system state $x \in \mathcal{S}$, where $\mathcal{S}$ is the system state space, and the vector $\mathbf{x}^{s,*}$ which, as before, contains as its entries the optimal value function at the states associated with separation s, *i.e.*, element of $\mathcal{S}(s)$. The system state $x = (s, \mathcal{C}^{A_1}, \mathcal{C}^{A_2}, \mathcal{D})$, where s is the agent separation, the vectors $\mathcal{C}^{A_1}$ and $\mathcal{C}^{A_2}$ have as entries the costs of the edges in the local observation zones of the agents A_1 and A_2, respectively, and $\mathcal{D}$ is the set of at the current stage unobserved, though reachable and known edges, observed by either agent at the previous stage. The set $\mathcal{D}$ can contain no, one or two edge costs depending on the agents decisions at the previous stage, in similar fashion as for the single agent version of this problem.

We now define the entries of the vector $\mathbf{x}^{s,*}$ for large s. The number of system states per separation is constant for $s \geq 3$, and hence the following definition of $\mathbf{x}^{s,*}$ holds for $s \geq 3$. Since the length of $\mathbf{x}^{s,*}$, for $s \geq 3$ equals 820, we describe which system state is associated to which vector entry in words. The first 36 elements of $\mathbf{x}^{s,*}$ concern the states associated with separation s with known edge costs as in Fig. 4-3(a). In this case, both agents chose to stay on the same lane at the previous stage, or $u_-^{A_1} = u_-^{A_2} = 0$, where the subscript indicates a decision taken at the previous stage. In the same manner as for the two-agent cone-shaped local observation zone instance of this problem, we can exploit symmetry to reduce the number of states associated with this case from 64 to 36. The ordering of the states in vector $\mathbf{x}^{s,*}$ is similar to the cone-shaped local observation zone version of this problem, described in Section 4.2.1, but with the vectors $\mathcal{C}^{A_1}$ and $\mathcal{C}^{A_2}$.

The following 128 entries in vector $\mathbf{x}^{s,*}$ denote the optimal value function associated with separation s and in the case the known edges at the current stage are as in Fig. 4-3(b). In this case, we have that $u_-^{A_1} = 1$ and $u_-^{A_2} = 0$. There is no symmetry to exploit, and hence, there are 128 possible states $(s, \mathcal{C}^{A_1}, \mathcal{C}^{A_2}, d_{-1}^{A_1})$ for this case, counting 'binary' from $(0, 0, 0, 0, 0, 0, 0)$ to $(1, 1, 1, 1, 1, 1, 1)$, as described in Section 4.2.1. Similarly, the next 128

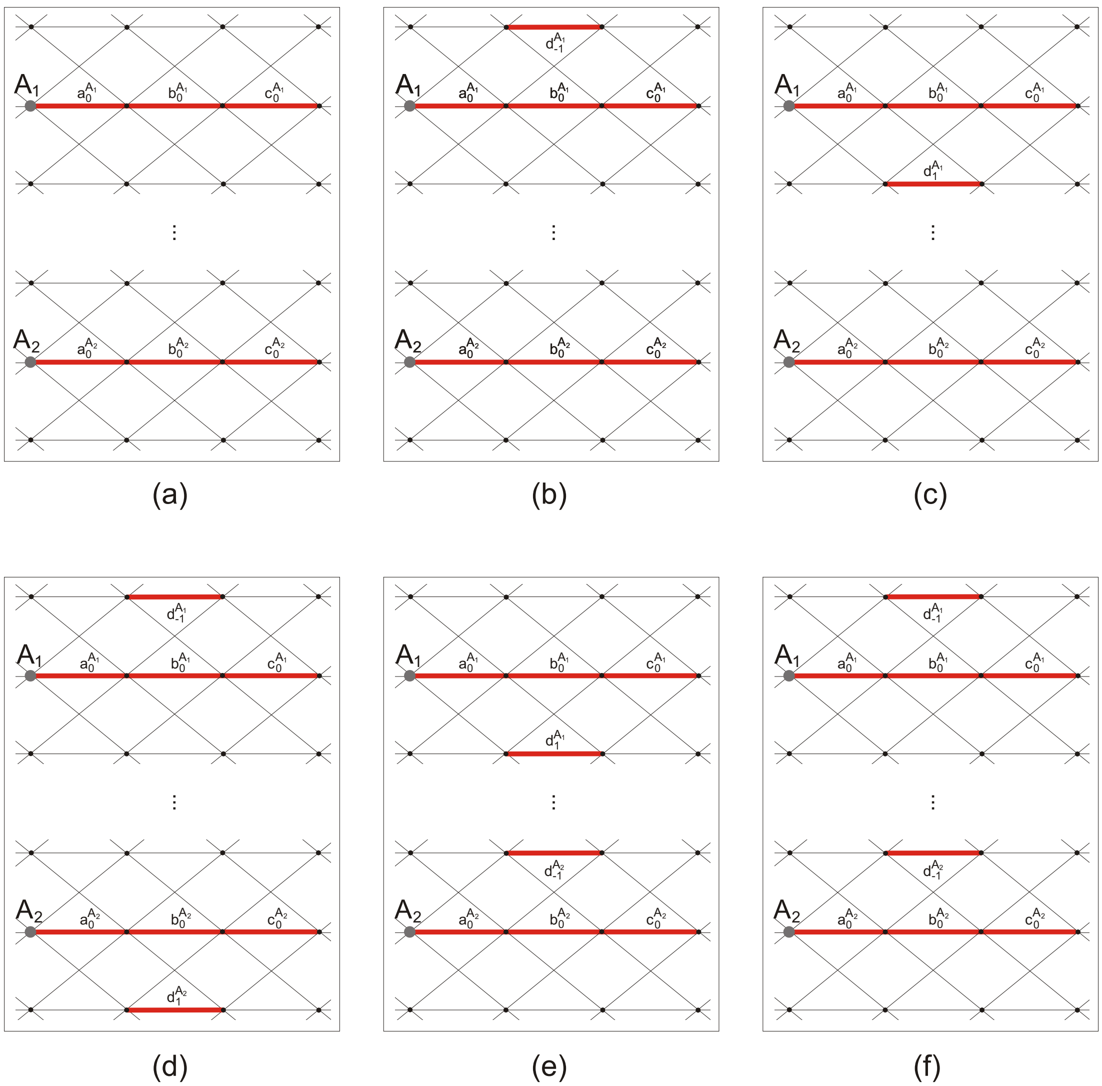

Figure 4-3: Two agents with extended tunnel-shaped local observation: six possible sets of edges (in red) with known cost at the current stage, for large separation s. The edges are part of a local observation zone or are observed at the previous stage.

entries in vector $\mathbf{x}^{s,*}$ denote the optimal value function associated with separation s and in the case the known edges at the current stage are as in Fig. 4-3(c), which is when $u_-^{A_1} = -1$ and $u_-^{A_2} = 0$.

The next 136 entries of $\mathbf{x}^{s,*}$ are associated with the case in Fig. 4-3(d) at separation s, where $u_-^{A_1} = 1$ and $u_-^{A_2} = -1$. This case exhibits symmetry, exploited similarly as in the case of Fig. 4-3(a), with the vectors $(\mathcal{C}^{A_1}, d_{-1}^{A_1})$ and $(\mathcal{C}_1^{A_2}, d_1^{A_2})$, leading to 136 possible states. Similarly, the following 136 entries are associated with the situation in Fig. 4-3(e), where $u_-^{A_1} = -1$ and $u_-^{A_2} = 1$.

Finally, the last 256 entries of $\mathbf{x}^{s,*}$ are associated with the situation in Fig. 4-3(f) at separation s, where $u_-^{A_1} = u_-^{A_2} = 1$. There is no symmetry, and hence the 256 entries are associated with the states produced by the binary sequence from $(0, 0, 0, 0, 0, 0, 0, 0)$ to $(1, 1, 1, 1, 1, 1, 1, 1)$ of the vector $(\mathcal{C}^{A_1}, d_{-1}^{A_1}, \mathcal{C}^{A_2}, d_{-1}^{A_2})$.

For the description of the optimal value function evolution at large agent separations, we can, as before in Section 4.2.2, reduce the size of the set of states by taking the expected value of the optimal value function over the uncontrollable edge costs, thus constructing the vector $\tilde{E}\mathbf{x}^{s,*}$ with 78 entries. In particular, for the situation in Fig. 4-3(a), the edges $a_0^{A_1}$, $b_0^{A_1}$, $a_0^{A_2}$ and $b_0^{A_2}$ are controllable. We can reduce the number of states to 10, by taking the expected value over the uncontrollable edges $c_0^{A_1}$ and $c_0^{A_2}$ and exploiting the symmetry of this case similarly as before. Specifically, the 10 first entries of $\tilde{E}\mathbf{x}^{s,*}$ are associated to the expected value of the optimal value function, given the edge costs $(a_0^{A_1}, b_0^{A_1})$ and $(a_0^{A_2}, b_0^{A_2})$. The next 16 entries of $\tilde{E}\mathbf{x}^{s,*}$ are associated with the expected value of the value function at separation s given the controllable edge costs $(a_0^{A_1}, a_0^{A_2}, b_0^{A_2}, d_{-1}^{A_1})$ in Fig. 4-3(b), ordered as before, with no symmetry. Similarly, the following 16 entries of $\tilde{E}\mathbf{x}^{s,*}$ are associated with the expected value of the optimal value function, given the controllable edge costs $(a_0^{A_1}, a_0^{A_2}, b_0^{A_2}, d_1^{A_1})$ in Fig. 4-3(c). The next 10 elements of $\tilde{E}\mathbf{x}^{s,*}$ are associated with the expected value of the optimal value function, given the controllable edge costs $(a_0^{A_1}, d_{-1}^{A_1})$ and $(a_0^{A_2}, d_1^{A_2})$. We exploit the symmetry and order the states in $\tilde{E}\mathbf{x}^{s,*}$ as before. Similarly, the following 10 elements of $\tilde{E}\mathbf{x}^{s,*}$ are associated with Fig. 4-3(e), with controllable edge costs $(a_0^{A_1}, d_1^{A_1})$ and $(a_0^{A_2}, d_{-1}^{A_2})$. Finally, the last 16 elements of $\tilde{E}\mathbf{x}^{s,*}$ are associated with the case in Fig. 4-3(f) at separation s, with controllable edge costs $(a_0^{A_1}, d_{-1}^{A_1}, a_0^{A_2}, d_{-1}^{A_2})$ and no symmetry.

As in the two-agent cone-shaped local observation zone instance of this problem, we assume the existence of a particular separation $\bar{s}$ such that for all $s \geq \bar{s}$, the optimal two-agent policy consists of a combination of two single-agent optimal policies where ties are broken by choosing the particular pair of decisions at each state that leads to the smallest separation at the next stage. Similarly as before, from $\mathcal{LP}(s)$, the LP with finitely many constraints associated with separation s and constructed by decomposing $\mathcal{LP}_g$, we can derive the following equation, connecting the optimal value function at separation s with the optimal value function at the separations $s-2$, $s-1$, $s+1$ and $s+2$, for $s \geq \bar{s}$. Specifically, we have that

$$\tilde{E}\mathbf{x}^{s,*} = \sum_{\sigma=-2}^{2} \mathbf{D}_\sigma \tilde{E}\mathbf{x}^{s+\sigma,*} + \mathbf{d},$$

where the matrices $\mathbf{D}_\sigma$ (78×78) and the vector $\mathbf{d}$ are extracted from the constraints of $\mathcal{LP}(s)$, similarly as in Section 4.2.3.

We now describe the conversion from the latter mixed forward-backward recursion into a forward recursion and a corresponding LTI system, a technically not difficult, though rather tedious task. Here, we note that theoretically, the z-transform allows to compute an explicit solution to the recursion at hand, given an adequate initial condition. However, it requires the symbolic inversion of the fairly large matrix, which is practically infeasible in this case. We take an approach that is more closely related to the problem structure and start by gathering the terms in $\tilde{E}\mathbf{x}^{s,*}$, which yields,

$$(\mathbf{I} - \mathbf{D}_0)\tilde{E}\mathbf{x}^{s,*} = \mathbf{D}_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}_{-1}\tilde{E}\mathbf{x}^{s-1,*} + \mathbf{D}_1\tilde{E}\mathbf{x}^{s+1,*} + \mathbf{D}_2\tilde{E}\mathbf{x}^{s+2,*} + \mathbf{d}.$$

Left-multiplying by $(\mathbf{I} - \mathbf{D}_0)^{-1}$, an inverse which numerical experiments indicate exists for the values of p considered, yields

$$\tilde{E}\mathbf{x}^{s,*} = \mathbf{D}'_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}'_{-1}\tilde{E}\mathbf{x}^{s-1,*} + \mathbf{D}'_1\tilde{E}\mathbf{x}^{s+1,*} + \mathbf{D}'_2\tilde{E}\mathbf{x}^{s+2,*} + \mathbf{d}', \tag{4.64}$$

where the primed matrices equal there unprimed equivalent, pre-multiplied by $(\mathbf{I} - \mathbf{D}_0)^{-1}$.

As a next step, we simplify the latter recursion by exploiting the fact that the matrices $\mathbf{D}'_1$ and $\mathbf{D}'_2$ have four and three linearly independent columns, respectively. In fact, we can write that $\mathbf{D}'_1 = \mathbf{C}_1\mathbf{L}_1$, and $\mathbf{D}'_2 = \mathbf{C}_2\mathbf{L}_2$, where $\mathbf{C}_1$ and $\mathbf{C}_2$ have four columns (the ones with indices 27, 31, 35, and 39 of $\mathbf{D}'_1$) and three columns (the ones with indices 53, 55, and

60 of $\mathbf{D}'_2$), respectively, and are of full rank. The matrices $\mathbf{L}_1$ and $\mathbf{L}_2$ are of appropriate dimensions. The previous matrices can be found in a fairly straightforward manner by computing the null space of both $\mathbf{D}'_1$ and $\mathbf{D}'_2$. Substituting in Eq. (4.64) yields

$$\tilde{E}\mathbf{x}^{s,*} = \mathbf{D}'_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}'_{-1}\tilde{E}\mathbf{x}^{s-1,*} + \mathbf{C}_1\tilde{E}_1\mathbf{x}^{s+1,*} + \mathbf{C}_2\tilde{E}_2\mathbf{x}^{s+2,*}, \tag{4.65}$$

where $\tilde{E}_1\mathbf{x}^{s+1,*}$ and $\tilde{E}_2\mathbf{x}^{s+2,*}$ denote the vectors $\mathbf{L}_1\tilde{E}\mathbf{x}^{s+1,*}$ and $\mathbf{L}_2\tilde{E}\mathbf{x}^{s+2,*}$, respectively, and represent the appropriate expected value of the optimal value function at the respective separations. Note that the vectors $\tilde{E}_1\mathbf{x}^{s+1,*}$ and $\tilde{E}_2\mathbf{x}^{s+2,*}$ have four and three entries respectively. We now write the recursion in Eq. (4.65) as follows:

$$\begin{bmatrix} \mathbf{I} & -\mathbf{C}_1 & -\mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \tilde{E}\mathbf{x}^{s,*} \\ \tilde{E}_1\mathbf{x}^{s+1,*} \\ \tilde{E}_2\mathbf{x}^{s+2,*} \end{bmatrix} = \mathbf{D}'_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}'_{-1}\tilde{E}\mathbf{x}^{s-1,*} + \mathbf{d}'.$$

Considering the vectors $\tilde{E}\mathbf{x}^{s,*}$, $\tilde{E}_1\mathbf{x}^{s+1,*}$ and $\tilde{E}_2\mathbf{x}^{s+2,*}$ as variables, and the RHS as known, the previous equation corresponds to a system of 78 equations in 85 unknowns. Note that in the previous scheme, it is assumed that the optimal value function at separations smaller than or equal to $s-1$ is known, and that we can compute the optimal value function at separation s. This starts approaching a forward recursion. However, in its current form, the system of equations has no unique solution, we need more equations. Indeed, at the two previous, hypothetical, recursion iterations, we compute $\tilde{E}_1\mathbf{x}^{s,*}$, and $\tilde{E}_2\mathbf{x}^{s,*}$, information that needs to be included in the current set of equations. This yields,

$$\mathbf{M} \begin{bmatrix} \tilde{E}\mathbf{x}^{s,*} \\ \tilde{E}_1\mathbf{x}^{s+1,*} \\ \tilde{E}_2\mathbf{x}^{s+2,*} \end{bmatrix} = \begin{bmatrix} \mathbf{D}'_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}'_{-1}\tilde{E}\mathbf{x}^{s-1,*} + \mathbf{d}' \\ \tilde{E}_1\mathbf{x}^{s,*} \\ \tilde{E}_2\mathbf{x}^{s,*} \end{bmatrix},$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} & -\mathbf{C}_1 & -\mathbf{C}_2 \\ \mathbf{L}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_2 & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

We now have a system with 85 equations in 85 unknowns. However the matrix $\mathbf{M}$ only is of rank 81, and thus only has 81 linearly independent columns. The previous holds for $p = 0.42, 0.43, \ldots, 0.99$, the values of p for which we compute the optimal two-agent policy of the problem at hand.

In the next step, we remove four of the first 78 columns of $\mathbf{M}$ so that the resulting matrix is of full rank. Determining the null space of $\mathbf{M}$ helps in choosing the appropriate columns to be removed. We have that $\mathbf{M} = \mathbf{M}'\mathbf{L}_3$, where $\mathbf{M}'$ has 81 columns and is of full rank. Here, we have removed the columns with indices 19, 21, 36, and 50. We now need to add four variables. Since we just removed the 19th, 21st, 36th and 50th element of $\tilde{E}\mathbf{x}^{s,*}$ from the set of variables, we need to compute their values at one of the next iterations. Hence, we choose to add the variables with the same indices, but at the separation $s-1$. This yields

$$\begin{bmatrix} \mathbf{M}' & \mathbf{D}^c_{-1} \end{bmatrix} \begin{bmatrix} \mathbf{L}_3 \begin{bmatrix} \tilde{E}\mathbf{x}^{s,*} \\ \tilde{E}_1\mathbf{x}^{s+1,*} \\ \tilde{E}_2\mathbf{x}^{s+2,*} \end{bmatrix} \\ \tilde{E}\mathbf{x}_c^{s-1,*} \end{bmatrix} = \begin{bmatrix} \mathbf{D}'_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}^r_{-1}\tilde{E}\mathbf{x}_r^{s-1,*} + \mathbf{d}' \\ \tilde{E}_1\mathbf{x}^{s,*} \\ \tilde{E}_2\mathbf{x}^{s,*} \end{bmatrix}, \tag{4.66}$$

where $\mathbf{D}^c_{-1}$ contains the columns 19, 21, 36, and 50 of the matrix $\mathbf{D}'_{-1}$, where $\mathbf{D}^r_{-1}$ contains all but the previous columns of $\mathbf{D}'_{-1}$. Further, $\tilde{E}\mathbf{x}_r^{s-1,*}$ contains the elements of $\tilde{E}\mathbf{x}^{s-1,*}$ with indices 19, 21, 36, and 50, and $\tilde{E}\mathbf{x}_r^{s-1,*}$ contains all elements of $\tilde{E}\mathbf{x}^{s-1,*}$, but the ones with the previous indices. Inspection of the resulting recursion tells us that at the current iteration, certain elements of the RHS cannot be known as is. Indeed, not $\tilde{E}_1\mathbf{x}^{s,*}$ is known, though a linear combination of the latter vector with elements of $\tilde{E}\mathbf{x}^{s-1,*}$ and $\tilde{E}_2\mathbf{x}^{s+1,*}$, computed at the previous iteration. Since at the current iteration, no information is computed at separation $s-2$ or lower, we can assume for now that the optimal value function at the latter separations is known, computed from previous iterations. Hence, we assume that $\tilde{E}_2\mathbf{x}^{s,*}$ in the RHS of Eq. (4.66) can be computed from the results of the previous iterations. Finally, not $\tilde{E}\mathbf{x}_r^{s-1,*}$ is known, though a linear combination of its entries and the entries in the vectors $\tilde{E}_1\mathbf{x}^{s,*}$ and $\tilde{E}_2\mathbf{x}^{s+1,*}$. In particular, we have that

$$\mathbf{M}'' \begin{bmatrix} \mathbf{L}_3 \begin{bmatrix} \tilde{E}\mathbf{x}^{s,*} \\ \tilde{E}_1\mathbf{x}^{s+1,*} \\ \tilde{E}_2\mathbf{x}^{s+2,*} \end{bmatrix} \\ \tilde{E}\mathbf{x}_c^{s-1,*} \end{bmatrix} = \begin{bmatrix} \mathbf{D}'_{-2}\tilde{E}\mathbf{x}^{s-2,*} + \mathbf{D}^r_{-1}\mathbf{L}_{3,1} \begin{bmatrix} \tilde{E}\mathbf{x}^{s-1,*} \\ \tilde{E}_1\mathbf{x}^{s,*} \\ \tilde{E}_2\mathbf{x}^{s+1,*} \end{bmatrix} + \mathbf{d}' \\ \mathbf{L}_{3,2} \begin{bmatrix} \tilde{E}\mathbf{x}^{s-1,*} \\ \tilde{E}_1\mathbf{x}^{s,*} \\ \tilde{E}_2\mathbf{x}^{s+1,*} \end{bmatrix} \\ \tilde{E}_2\mathbf{x}^{s,*} \end{bmatrix},$$

where $\mathbf{L}_{3,1}$ denotes the sub-matrix of $\mathbf{L}_3$ that contains its first 74 rows, and where $\mathbf{L}_{3,2}$ is the sub-matrix of $\mathbf{L}_3$ containing the rows with indices $75, \ldots, 78$. The matrix $\mathbf{M}''$ is adapted from its predecessor in Eq. (4.66) to accommodate the changes in the RHS, and preserve the equality. Note that the last seven columns of $\mathbf{L}_{3,1}$ are zero and that only its columns corresponding to the elements in $\tilde{E}\mathbf{x}_c^{s-1,*}$ call for a change in the corresponding columns of the predecessor of $\mathbf{M}''$. The same holds true for the matrix $\mathbf{L}_{3,2}$, where the last three columns are zero, and changes to the columns associated with $\tilde{E}\mathbf{x}_c^{s-1,*}$ in the predecessor of $\mathbf{M}''$ suffice.

We now have 85 equations in 85 variables with coefficient matrix $\mathbf{M}''$. However, the latter is of rank 84, and thus only has 84 linearly independent columns, yielding a set of equations with no unique solution, yet. The last step is to remove the last column of $\mathbf{M}''$, which contains only zeros as entries, and replace the variable in the fourth entry of $\tilde{E}\mathbf{x}_c^{s-1,*}$ by its equivalent, though at separation $s-2$. As before, the RHS and the coefficient matrix need to be adapted accordingly. This yields

$$\mathbf{M}_f \begin{bmatrix} \mathbf{L}_3 \begin{bmatrix} \tilde{E}\mathbf{x}^{s,*} \\ \tilde{E}_1\mathbf{x}^{s+1,*} \\ \tilde{E}_2\mathbf{x}^{s+2,*} \end{bmatrix} \\ \tilde{E}\mathbf{x}_{c,1,2,3}^{s-1,*} \\ \tilde{E}\mathbf{x}_{c,4}^{s-2,*} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{-2}^{t}\tilde{E}\mathbf{x}_t^{s-2,*} + \mathbf{D}_{-1}^{r}\mathbf{L}_{3,1} \begin{bmatrix} \tilde{E}\mathbf{x}^{s-1,*} \\ \tilde{E}_1\mathbf{x}^{s,*} \\ \tilde{E}_2\mathbf{x}^{s+1,*} \end{bmatrix} + \mathbf{d}' \\ \mathbf{L}_{3,2} \begin{bmatrix} \tilde{E}\mathbf{x}^{s-1,*} \\ \tilde{E}_1\mathbf{x}^{s,*} \\ \tilde{E}_2\mathbf{x}^{s+1,*} \end{bmatrix} \\ \mathbf{L}_{3,3} \begin{bmatrix} \tilde{E}\mathbf{x}^{s-2,*} \\ \tilde{E}_1\mathbf{x}^{s-1,*} \\ \tilde{E}_2\mathbf{x}^{s,*} \end{bmatrix} - \mathbf{L}_{3,4}\tilde{E}\mathbf{x}_{c,1,2}^{s-2,*} \end{bmatrix}, \tag{4.67}$$

where $\tilde{E}\mathbf{x}_{c,1,2,3}^{s-1,*}$ denotes the vector with the first three entries of $\tilde{E}\mathbf{x}_c^{s-1,*}$, while $\tilde{E}\mathbf{x}_{c,4}^{s-2,*}$ is the fourth element of $\tilde{E}\mathbf{x}_{c,4}^{s-2,*}$. Further, $\mathbf{D}_{-2}^{t}$ equals matrix $\mathbf{D}'_{-2}$, but without the 50th column of the latter. Similarly, $\tilde{E}\mathbf{x}_t^{s-2,*}$ equals the vector $\tilde{E}\mathbf{x}^{s-2,*}$, but without its 50th entry. Then, $\mathbf{L}_{3,3}$ is the sub-matrix of $\mathbf{L}_3$ containing the last three rows of $\mathbf{L}_3$, and $\mathbf{L}_{3,4}$ is the sub-matrix of $\mathbf{L}_3$ containing the three last rows of the 19th and the 20th columns of $\mathbf{L}_3$. The columns with indices $79, \ldots, 82$, corresponding to the entries in $\tilde{E}_1\mathbf{x}^{s-1,*}$ contain zeros, and hence no compensation is required in $\mathbf{M}''$ regarding these terms. Further, only the columns with indices 19, 21 and 50, associated with the vector $\tilde{E}\mathbf{x}^{s-2,*}$ in $\mathbf{L}_{3,3}$ contain non-zero elements. The extra term with coefficient matrix $\mathbf{L}_{3,4}$ compensates for the first two, while changes in the last column of $\mathbf{M}''$ compensate for the non-zero column in $\mathbf{L}_{3,3}$ with index 50.

The transition of the mixed forward-backward original recursion in Eq. (4.64) to the forward recursion in Eq. (4.67) is now complete. Indeed, each iteration provides sufficient information to compute the RHS for the next iteration, and the matrix $\mathbf{M}_f$ is invertible for the values of p for which computed ($p = 0.42, 0.43, \ldots, 0.99$). The conversion to a LTI system is now straightforward and in the next section, we address the connection constraints required for this problem.

4.4.3 Sufficient Connection Constraints

In this section, we present the connection constraints sufficient for $\mathcal{LP}_{in}$ (see Section 4.2.6 for the two-agent cone-shaped local observation zone version of the problem), to produce the optimal value function for $s < \bar{s}$. In particular, using a development similar to the one used for the two-agent cone-shaped local observation zone, we require as additional necessary condition that the optimal value function at separations $s < \bar{s}$ is such that only the stable system modes of the system

$$E\mathbf{x}^{s+1,*} = E\mathbf{x}^{s,*} + \mathbf{B},$$

are excited. The latter LTI system is derived based on the forward recursion developed in the previous section using a method similar to one presented in Section 4.2.5 for the two-agent cone-shaped local observation zone problem. The system exhibits five unstable modes, providing five additional equality constraints for $\mathcal{LP}_{in}$:

$$\mathbf{V}_u(E\mathbf{x}^{\bar{s}-1}) = \mathbf{V}_u(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}, \tag{4.68}$$

defined similarly as in Section 4.2.6 and where $(\mathbf{I} - \mathbf{A})^{-1}$ is invertible for the values of p for which computed. From the development in the previous section, we notice that the number of linearly independent linear combinations of the optimal value function at the separations $\bar{s}$ and $\bar{s}+1$ appearing in the constraints associated with the separations $\bar{s}-2$ and $\bar{s}-1$, equals ten (the four entries of $\tilde{E}_1\mathbf{x}^{\bar{s},*}$, the three entries of $\tilde{E}_2\mathbf{x}^{\bar{s}+1,*}$, and the three entries of $\tilde{E}_2\mathbf{x}^{\bar{s},*}$). Hence, the number of necessary equality constraints in Eq. (4.68) is insufficient for an equivalent version of Theorem 2 to hold. Additional connection constraints are required.

In fact, adding as extra constraint one iteration of the LTI system, yields the required set of connection constraints. For this, the optimal value function at separation $\bar{s}$ needs to be included in $\mathcal{LP}_{in}$, so that one such iteration can be enforced. Therefore, the complete set of connection constraints becomes:

$$
\begin{aligned}
\mathbf{V}_u(E\mathbf{x}^{\bar{s}}) &= \mathbf{V}_u(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}, \\
E\mathbf{x}^{\bar{s}} &= \mathbf{A}E\mathbf{x}^{\bar{s}-1} + \mathbf{B}.
\end{aligned}
$$

The linear program in $\mathcal{LP}_{in}$ then contains the constraints associated with the separations $s = 0, 1, \ldots, \bar{s}$, rather than at the separations $s = 0, 1, \ldots, \bar{s}-1$ as is the case for the cone-shaped local observation zone version of the problem at hand.

More rigorously, we have that an equivalent version of Lemma 13 holds, establishing the dual feasibility of $\mathcal{LP}(s)$ for $s \geq \bar{s}$ when adopting a pair of single-agent optimal policies (policy μ_∞) as two-agent policy for $s \geq \bar{s}$. Further, we define the system output $\mathbf{z}^s$, for $s = \bar{s}, \ldots$ as in Section 4.2.6. Then, we can verify optimality of the thus computed value function (solving $\mathcal{LP}_{in}$ for $s \leq \bar{s}$ and simulating the LTI system for $s = \bar{s}, \bar{s}+1 \ldots$) for each relevant p by verifying that the value function satisfies Bellman's equation for $s \leq \bar{s}$ and by verifying non-negativity of $\mathbf{z}^s$, for $s = \bar{s}, \bar{s}+1, \ldots$. This assures that we obtain the unique bounded fixed point of the function F.

Finally, we remark that, although not investigated further, the combination of connection constraints employed most likely ensures that the linear combinations of the optimal value function at the separations $\bar{s}+1$ and $\bar{s}+2$ that appear in the RHS of the constraints associated to the separations $\bar{s}-1$ and $\bar{s}$, can be expressed in a unique manner as a linear function of the value function at smaller separations, allowing for a stronger result, similar to the one in Theorem 2. As for the cone-shaped local observation version of the problem, we solve for the optimal value function with increasing $\bar{s}$ until we find the optimal value function. We present values for $\bar{s}$ associated with the extended tunnel-shaped local observation zone in Section 6.5.

4.5 Three Agents with Short Tunnel-Shaped Local Observation

In this section, we treat the three-agent problem with short tunnel-shaped local observation zones. In particular, we present those aspects related to the problem solution that differ from the solution method of the two-agent cone-shaped and extended tunnel-shaped local observation zone cases.

We start by defining the system state. Let three agents be located at the lanes l^{A_1}, l^{A_2} and l^{A_3}, at some stage of the graph G. At each stage, the agents are (re)labeled such that $l^{A_1} \leq l^{A_2} \leq l^{A_3}$. We define $s_1 = l^{A_2} - l^{A_1}$ and $s_2 = l^{A_3} - l^{A_2}$ as the two separations between three agents. We can assume without loss of generality that $s_1 \leq s_2$, due to the problem symmetry. Then, at each stage, we define the system state as $x = (s_1, s_2, \mathcal{C}^{A_1}, \mathcal{C}^{A_2}, \mathcal{C}^{A_3})$, containing the two agent separations, and the costs of the edges in the local observation zones associated with each agent.

In Section 4.3, we establish an upper bound on the agent separation in steady state under an optimal policy for the two-agent version of the problem at hand. This provides us with the intuition that if s_2 is large, then $s_1 \leq 3$, in steady state under an optimal policy. We formulate the following assumption:

Assumption 2 Let $\{s_1^i\}$ and $\{s_2^i\}$, for $i = 0, 1, \ldots$ denote two sequences of agent separations under an optimal policy in steady state, for the three-agent problem with short tunnel-shaped local observation zones. Note that $s_1^i \leq s_2^i$ for $i = 0, 1, \ldots$. Then, $s_1^i \leq 3$ for $i = 0, 1, \ldots$.

In words, the previous assumption states that two out of three agents remain within a separation of three in steady state under an optimal policy.

Under Assumption 2, the three-agent problem can be treated effectively as a two-agent problem. Indeed, the two agents whose separation remains under three in the three-agent problem can be treated as a single agent. The separation s_2 plays the role of the separation s in the two-agent cases presented thus far. The state vector $\mathbf{x}^{s_2}$ contains the states associated with separation s_2 and with the separation $s_1 = 0, \ldots, \min\{s_2, 3\}$. We give details for the vector $\mathbf{x}^{s_2}$ for $s_2 \geq \bar{s}_2$, where $\bar{s}_2$ is defined as the smallest separation s_2 where the optimal three-agent policy consists of a combination of an optimal two-agent policy (adopted by the two agents whose separation is smaller than three) and an optimal single-agent policy (adopted by the third agent). In particular, the vector $\mathbf{x}^{s_2}$ contains four sets of states. The first set is associated with the situations where $s_1 = 0$, and contains 16 states, one for each possible set of two local observation zones, and ordered similarly as in the two-agent cone-shaped local observation zone instance of this problem. Then follow the sets of states associated with separation $s_1 = 1, 2, 3$, each containing 64 elements and ordered similarly. The problem exhibits no useful symmetry for large s_2.

Further, we mention that the rest of the solution development follows a similar line as outlined for the two-agent extended tunnel-shaped local observation zone case, in the previous section. For $s_2 \geq \bar{s}_2$, one can formulate the mixed forward-backward recursion, transformed to a forward recursion and corresponding LTI system with a similar method as outlined in the previous section. Again, the z-transform method is practically of no use here, due to its complexity. Further, the connection constraints contain an additional iteration of the LTI system, as before for the two-agent extended tunnel-shaped local observation zone problem instance. Note that the single- and the two-agent optimal policies used here, are optimal for all the values of $p \geq 0.384$ for which computed (see Section subsec:tunnel2 for further details). For some $0.383 < p' < 0.384$, a change in the optimal policy occurs and for $p < p'$ a different set of policies is optimal than for $p > p'$. Hence, we have a different recursion for $s_2 \geq \bar{s}_2$ in the problem solution. As for the two-agent extended tunnel-shaped local observation zone case, the development of the latter recursion is a time-intensive and tedious task, omitted in this thesis. We believe that it is possible to automate this task, but consider it outside the scope of this thesis. However, to obtain reliable results for some values of p such that $p < p'$, we compute an upper and lower bound to the problem optimal value function considering a bounded (in width) and a cylindrically shaped navigation graph, respectively. A large graph width (for the upper bound) and a large cylinder diameter (for the lower bound) allow for determining bounds that are as tight as the accuracy of CPLEX, at the cost of a large computation time. The problem complexity of the two-agent problem with extended tunnel-shaped local observation zones did not allow the computation of similar upper and lower bounds to the problem optimal value function.

Finally, rigorous confirmation of Assumption 2 is not provided in this thesis, and part of future work (see Section 7.2.2). However, we note that numerical experiments further confirm the intuition of a separation upper bound for the two closest agents. In particular, one can compute the optimal value function for $s_1 > 3$, for a bounded though fairly large environment, and study optimal policy properties.

4.6 Chapter Summary

In this chapter, we present an algorithm that computes the optimal value function for various problem instances as the fixed point of the function F, defined in the previous chapter. We develop the algorithm in detail for the two-agent cone-shaped local observation zone problem instance. In particular, we compute the optimal value function explicitly for the separations $s < \bar{s}$ by solving an LP with finitely many constraints. Here, $\bar{s}$ denotes the smallest separation for which the optimal two-agent policy consists of a combination of two single-agent policies. The separation $\bar{s}$ is determined through the algorithm. Further, the set of LP-constraints contains connection constraints which summarize the information on the optimal value function for $s \geq \bar{s}$ necessary for the LP-solution to represent the optimal value function for $s < \bar{s}$. The optimal value function at $s \geq \bar{s}$ is described implicitly in the form of a discrete linear time invariant system with as state vector the optimal value function at separation s in some form; the LTI system is connected to the LP by way of the connection constraints, which ensure that the LP-solution provides the adequate initial condition for the LTI system. We establish that, with a proper definition of the system output, its non-negativity guarantees that the computed value function is a fixed point of F, and therefore optimal. For the case a particular choice of $\bar{s}$ leads to a negative system output for some $s \geq \bar{s}$, we increment $\bar{s}$ by one and restart the algorithm.

For the two-agent problem with short tunnel-shaped local observation zones, we establish an upper bound on the separation under an optimal policy in steady state. Therefore, an analysis as in the cone-shaped local observation zone version of this problem is unnecessary. For the two-agent problem with extended local observation zones, and for the three-agent problem with short tunnel-shaped local observation zones, we describe the technical aspects of the algorithm that differ. For the latter case, we assume that two out of three agents remain within separation two, allowing to treat the problem successfully with the presented two-agent algorithm.

Chapter 5

The Steady State Separation Probability Distribution

In Chapter 4, we compute the optimal policy for various instances of the target acquisition problem. In this chapter, we provide a method to compute the state probability distribution in steady state and under an optimal policy. We treat the two-agent problem with cone-shaped local observation zones in Section 5.1, and some differing aspects on the two-agent problem with short and extended tunnel-shaped local observation zone and on the three-agent problem with short tunnel-shaped local observation zones in Section 5.2.

5.1 Two Agents with Cone-Shaped Local Observation Zones

In this section, we study the steady state agent behavior under an optimal policy. In particular, we present a method to compute the probability distribution of the agent separation after initial transition effects have died out. The results in this section are also presented in [26].

With the optimal policy μ^*, the state evolution can be described by means of an infinite state Markov chain. Specifically, let $\delta_{qr}^s \in \mathcal{D}$ ($s \geq 0$, $q, r \in \mathcal{L}$) denote the state whereby the agent separation is s, and whereby the observation zone pair is any pair for which $a_0^A = q$ and $a_0^B = r$. Set $\mathcal{D}$ is the set of Markov chain states. For the remainder of this section, we denote by *state* the Markov chain state δ_{qr}^s. Fig. 5-1 shows the Markov chain structure conceptually, only depicting outgoing arcs for the three states associated with separation $s \geq 2$. This element is repeated for all separations $s \geq 2$ with natural extensions

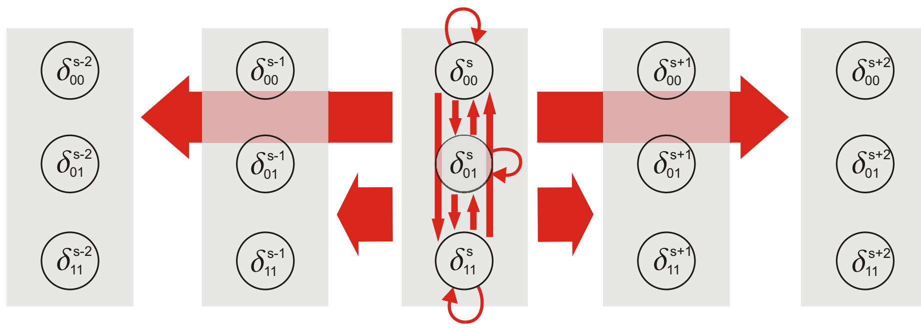

Figure 5-1: Element of the two-agent Markov chain structure. For clarity, only the probabilities of the transitions out of the states associated with separation s are shown. Large arrows indicate possible transitions from any of the three states associated with separation s, to any of the three states associated with separation $s' \in \{s-2, s-1, s, s+1, s+2\}$.

at separations $s = 0$ and $s = 1$. Let $p^k(\delta)$ denote the probability for being at state δ after k state transitions. Let

$$\pi_k = \left[\begin{array}{cccc} (\mathbf{p}_0^k)^T & (\mathbf{p}_1^k)^T & (\mathbf{p}_2^k)^T & \ldots \end{array} \right]^T,$$

where

$$\mathbf{p}_s^k = \left[\begin{array}{ccc} p^k(\delta_{00}^s) & p^k(\delta_{01}^s) & p^k(\delta_{11}^s) \end{array} \right]^T.$$

In words, π_k contains the state probability distribution after k state transitions. We define $q(\delta'|\delta)$ as the probability the next state is δ', given the current state δ. Given the edge cost statistics and μ^*, we can compute $q(\delta'|\delta)$ for all $\delta, \delta' \in \mathcal{D}$. Let matrix $\mathbf{Q}$ denote the state transition matrix. Since from separation s, only the separations $s + \sigma$ (for $\sigma = -2, \ldots, 2$) can be reached, the matrix $\mathbf{Q}$ has a block multi-diagonal structure (see Fig. 5-2). In $\mathbf{Q}$, the entry on the row and column corresponding to δ' and δ, respectively, contains $q(\delta'|\delta)$. Sub-matrix $\mathbf{Q}_\sigma^s \in \mathbb{R}^{3\times3}$ ($\sigma = \max\{-2, -s\}, \ldots, 2$) describes the transitions from the states associated with separation s to the states associated with separation $s + \sigma$.

We have that policy μ_∞ is optimal for all $s \geq \bar{s}$. Therefore, the matrix $\mathbf{Q}$ has a recurring structure for all $s \geq \bar{s}$. In particular, let $\mathbf{Q}_s$ contain the rows of $\mathbf{Q}$ corresponding to separation s, for $s \geq \bar{s}$. Then,

$$\mathbf{Q}_s = \left[\begin{array}{cccccccccc} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}_{-2} & \mathbf{Q}_{-1} & \mathbf{Q}_0 & \mathbf{Q}_1 & \mathbf{Q}_2 & \mathbf{0} & \cdots \end{array} \right]^T,$$

where the zero-matrices are of the appropriate dimensions. Specific expressions for the matrices $\mathbf{Q}_\sigma$, for $\sigma = -2, \ldots, 2$ as a function of parameter p are

$$\mathbf{Q}_{-2} = p^2(1-p)^2 \begin{bmatrix} (1-p)^2p^2 & (1-p)p^2 & p^2 \\ 2(1-p)^3p & 2(1-p)^2p & 2(1-p)p \\ (1-p)^4 & (1-p)^3 & (1-p)^2 \end{bmatrix},$$

$$\mathbf{Q}_{-1} = p(1-p) \begin{bmatrix} 2(1-p)p^2 & p^2(2-p)(1-p+p^2) & 2(1-p)^2p^2 \\ 2(1-p)^2p(2-2p+p^2) & (1-p)p(3-5p+4p^2-p^3) & 2(1-p)^3p \\ 2(1-p)^5 & (1-p)^4 & 0 \end{bmatrix},$$

$$\mathbf{Q}_0 = \begin{bmatrix} p^2(1+2p^2-6p^3+6p^4-2p^5) & p^2(1-p)^2\,P_1(p) & p^2(1-p)\,P_2(p) \\ 2(1-p)^3p(1+2p-2p^3) & (1-p)^3p\,P_3(p) & 4(1-p)^2p^2\,P_4(p) \\ (1-p)^5(1-p+2p^2) & (1-p)^4p\,P_5(p) & 2(1-p)^3p\,P_4(p) \end{bmatrix},$$

$$\mathbf{Q}_1 = \begin{bmatrix} 2(1-p)p^3 & p^2(1-p+2p^3-p^4) & 2(1-p)^2p^2\,P_4(p) \\ 2(1-p)^2p^2(2-2p+p^2) & p(2-p)(1-p)(1-p+p^2)^2 & 2(1-p)^3p\,P_4(p) \\ 2(1-p)^5p & (1-p)^4(1-2p+3p^2-p^3) & 0 \end{bmatrix},$$

$$\mathbf{Q}_2 = \begin{bmatrix} (1-p)^2p^4 & (1-p)p^3P_4(p) & p^2(P_4(p))^2 \\ 2(1-p)^3p^3 & 2(1-p)^2p^2P_4(p) & 2(1-p)p(P_4(p))^2 \\ (1-p)^4p^2 & (1-p)^3pP_4(p) & (1-p)^2(P_4(p))^2 \end{bmatrix},$$

where

$$\begin{aligned} P_1(p) &= 1+p-p^2+3p^3-p^4, \\ P_2(p) &= 1-p-p^2+5p^3-2p^4, \\ P_3(p) &= 1-p^2+6p^3-2p^4, \\ P_4(p) &= 1-2p+3p^2-p^3, \\ P_5(p) &= 1-p+3p^2-p^3. \end{aligned}$$

Fig. 5-2 shows the recurring structure of $\mathbf{Q}$ for $s \geq \bar{s}$.

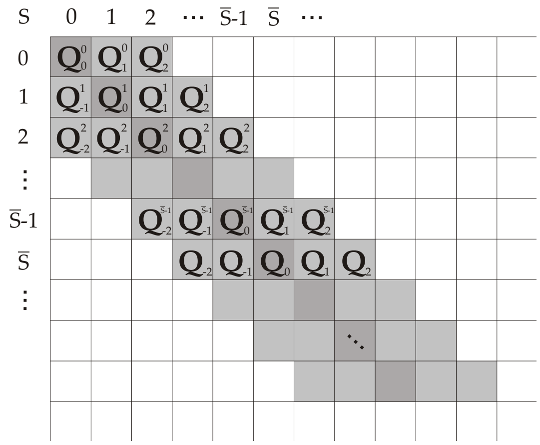

Figure 5-2: Conceptual representation of the block multi-diagonal structure of Markov chain transition matrix $\mathbf{Q}$.

We wish to compute

$$\pi^* = \lim_{k \to \infty} \pi^k,$$

the steady state probability distribution. The vector π^* is the eigenvector corresponding to the unique unit eigenvalue of $\mathbf{Q}$. We exploit the structure of $\mathbf{Q}$ to obtain π^*, using a principle similar to the one underlying the algorithm to compute the optimal value function presented earlier. Specifically, we formulate an autonomous LTI system whose state trajectories are the steady state probabilities for $s \geq \bar{s}$, where the separation s plays the role of "time" in a classical LTI system. We then formulate a set of linear equalities whose solution are the steady state probabilities for $s < \bar{s}$. The set includes linear equalities that ensure the computation of initial conditions so that only the stable LTI modes are excited. Lastly, the condition that the probabilities are required to sum up to one ensures the set of equalities yields a unique solution. The latter condition is equivalent to the maximization in $\mathcal{LP}_{in}$. We mention here that one finds a similar problem in the treatment of queues [39], where the steady state probability of a queue length of L depends on the steady state probability of queue lengths of both $L-1$ and $L+1$. The separation s which appears in the two-agent graph traversal problem at hand, takes on the role of the queue length L in the equivalence. In [39], one computes the steady state probability distribution easily by imposing the condition that the sum of all probabilities equals one, not sufficient for the problem at hand.

We have that

$$\mathbf{Q}\pi^* = \pi^*. \tag{5.1}$$

Therefore, for $s \geq \bar{s}$, we have that

$$\mathbf{p}_s^* = \sum_{\sigma=-2}^{2} \mathbf{Q}_\sigma \mathbf{p}_{s+\sigma}^*, \tag{5.2}$$

where $\mathbf{p}_s^*$ is the steady state version of $\mathbf{p}_s^k$. We transform this equation into a forward recursive equation. In particular, we have that $\text{rank}(\mathbf{Q}_2) = 1$. Therefore,

$$\mathbf{Q}_2 = \mathbf{q}_2(\mathbf{r}_1)^T, \tag{5.3}$$

where

$$\begin{aligned} \mathbf{q}_2 &= p^2(1-p)^2 \begin{bmatrix} p^2 \\ 2p(1-p) \\ (1-p)^2 \end{bmatrix}, \\ \mathbf{r}_1^T &= \begin{bmatrix} 1 & P_r(p) & (P_r(p))^2 \end{bmatrix}, \end{aligned}$$

with

$$P_r(p) = \frac{P_4(p)}{p(1-p)}.$$

Since $\text{rank}(\mathbf{Q}_1) = 2$ and

$$\text{rank}\left(\begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix}\right) = 2,$$

we have that

$$\mathbf{Q}_1 = \mathbf{q}_{11}(\mathbf{r}_2)^T + \mathbf{q}_{12}(\mathbf{r}_3)^T, \tag{5.4}$$

where

$$\begin{aligned}
\mathbf{q}_{11} &= \begin{bmatrix} \frac{2p^3(1-p)^3}{P_r(p)} \\ \frac{2p^2(1-p)^4}{P_r(p)} \\ 0 \end{bmatrix}, \\
\mathbf{q}_{12} &= \begin{bmatrix} p^2(1-3p+6p^2-4p^3+p^4) \\ p(1-p)(2-7p+14p^2-13p^3+6p^4-p^5) \\ (1-p)^4(1-2p+3p^2-p^3) \end{bmatrix}, \\
\mathbf{r}_2^T &= \begin{bmatrix} \frac{2}{P_r(p)} & 1 & 0 \end{bmatrix}, \\
\mathbf{r}_3^T &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}.
\end{aligned}$$

Further, $\mathrm{rank}(\mathbf{I}-\mathbf{Q}_0) = 3$. Let $\rho_s^{i,*} = (\mathbf{r}_i)^T \mathbf{p}_s^*$, for $i = 1, 2, 3$, or more compact,

$$\rho_s^* = \mathbf{R}_\rho \mathbf{p}_s^*, \tag{5.5}$$

where

$$\rho_s^* = \begin{bmatrix} \rho_s^{1,*} & \rho_s^{2,*} & \rho_s^{3,*} \end{bmatrix}^T,$$

and where

$$\mathbf{R}_\rho = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}^T.$$

Since $\det(\mathbf{R}_\rho) = -1$, the matrix $\mathbf{R}_\rho$ is invertible. Then, we can write Eq. (5.2) as

$$\begin{aligned}
(\mathbf{I}-\mathbf{Q}_0)\mathbf{R}_\rho^{-1}\rho_s^* \quad = \quad & \mathbf{Q}_{-2}\mathbf{R}_\rho^{-1}\rho_{s-2}^* + \mathbf{Q}_{-1}\mathbf{R}_\rho^{-1}\rho_{s-1}^* + \\
& \begin{bmatrix} \mathbf{q}_{11} & \mathbf{q}_{12} \end{bmatrix} \begin{bmatrix} \rho_{s+1}^{1,*} \\ \rho_{s+1}^{2,*} \end{bmatrix} + \mathbf{q}_2 \rho_{s+2}^{1,*}.
\end{aligned} \tag{5.6}$$

Let the vector $\mathbf{q}_{0i} \in \mathbb{R}^{3\times 1}$ denote the ith column of the matrix $(\mathbf{I}-\mathbf{Q}_0)\mathbf{R}_\rho^{-1}$, for $i = 1, 2, 3$.

We now write recursion (5.6) as

$$\mathbf{Q}_2' \rho_{s+2}'^* = \sum_{\sigma\in\{1,0,-1,-2\}} \mathbf{Q}_\sigma' \rho_{s+\sigma}'^*, \tag{5.7}$$

where

$$\begin{aligned}
\mathbf{Q}_2' &= \begin{bmatrix} -\mathbf{q}_2 & -\mathbf{q}_{12} & \mathbf{q}_{03} \end{bmatrix}, \\
\rho_s'^* &= \begin{bmatrix} \rho_s^{1,*} & \rho_{s-1}^{2,*} & \rho_{s-2}^{3,*} \end{bmatrix}^T,
\end{aligned}$$

and where $\mathbf{Q}_\sigma'$ (for $\sigma \in \{1, 0, -1, -2\}$) can be determined from Eq. (5.6). The following lemma establishes the condition for Eq. (5.7) to be a well-defined forward recursion, namely the invertibility of the matrix $\mathbf{Q}_2'$ for $0 < p < 1$.

Lemma 17 *The matrix $\mathbf{Q}_2'$ is invertible for $0 < p < 1$.* □

Proof: We show that the determinant of the matrix $\mathbf{Q}_2'$ differs from zero for $0 < p < 1$. In particular, we have that

$$\det(\mathbf{Q}_2') = (1-p)^3 p^2 \, P_{\mathbf{Q},1}(p) \, P_{\mathbf{Q},2}(p) \, P_{\mathbf{Q},2}(p),$$

with

$$\begin{aligned}
P_{\mathbf{Q},1}(p) &= 1 - 2p + 5p^2 - 4p^3 + p^4, \\
P_{\mathbf{Q},2}(p) &= 1 - 2p + 5p^2 - 3p^3 + 5p^4 - 4p^5 + p^6, \\
P_{\mathbf{Q},3}(p) &= 1 - 2p + 3p^2 + p^3 - 5p^4 + 4p^5 - p^6.
\end{aligned}$$

We establish that the polynomials $P_{\mathbf{Q},i}(p) > 0$ for $0 < p < 1$ and for $i = 1, 2, 3$. We follow a procedure similar to the one used in the proof of Lemma 8. Specifically, for $P_{\mathbf{Q},i}(p)$, for $i = 1, 2, 3$, find an SOS auxiliary polynomial $P_{\mathbf{Q},i}^a(p)$ such that

$$P_{\mathbf{Q},i}^t(p) = P_{\mathbf{Q},i}(p) + P_{\mathbf{Q},i}^a(p) \left((p - 1/2)^2 - 1/4 \right) \geq 0, \tag{5.8}$$

for $p \in \mathbb{R}$. Indeed, if such a polynomial can be found, then the second term of $P_{\mathbf{Q},i}^t(p)$ is nonpositive for $0 < p < 1$, and therefore its first term, the polynomial $P_{\mathbf{Q},i}(p)$, must be strictly positive. In stead of requiring positivity of $P_{\mathbf{Q},i}^t(p)$, we show that it is SOS with a positive definite coefficient matrix, a more stringent condition, but practically easier to handle. We transform Eq. (5.8) into an LMI constraint and find a feasible solution using semi-definite programming [62], by way of the SOSTOOLS Sum of Squares Optimization Toolbox for Matlab [63]. Specifically, we have that

$$
\begin{aligned}
P^a_{\mathbf{Q},1} &= \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix}, \\
P^t_{\mathbf{Q},1} &= \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{3}{2} & \frac{1}{3} \\ -\frac{3}{2} & \frac{16}{3} & -3 \\ \frac{1}{3} & -3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix},
\end{aligned}
$$

with positive definite coefficient matrices. Further, we have that

$$
\begin{aligned}
P^a_{\mathbf{Q},2} &= 2, \\
P^t_{\mathbf{Q},2} &= \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix}^T \begin{bmatrix} 1 & -2 & \frac{1}{3} & 0 \\ -2 & \frac{19}{3} & -\frac{3}{2} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{3}{2} & \frac{16}{3} & -2 \\ 0 & -\frac{1}{6} & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix},
\end{aligned}
$$

with positive definite coefficient matrices. Lastly, we have that

$$
\begin{aligned}
P^a_{\mathbf{Q},3} &= \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}, \\
P^t_{\mathbf{Q},3} &= \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{3}{2} & -\frac{1}{3} & \frac{3}{4} \\ -\frac{3}{2} & \frac{14}{3} & -\frac{1}{4} & -3 \\ -\frac{1}{3} & -\frac{1}{4} & 1 & 0 \\ \frac{3}{4} & -3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix},
\end{aligned}
$$

with positive definite coefficient matrices. In the three cases, only a positive semi-definite coefficient matrix for the auxiliary polynomials is required. Hence, the proof. ∎

Lemma 17 confirms that the recursion in Eq. (5.7) is a well-defined forward recursion and we convert it into the autonomous LTI system

$$\begin{bmatrix} \rho'^*_{s+2} \\ \rho'^*_{s+1} \\ \rho'^*_{s} \\ \rho'^*_{s-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}''_1 & \mathbf{Q}''_0 & \mathbf{Q}''_{-1} & \mathbf{Q}''_{-2} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \rho'^*_{s+1} \\ \rho'^*_{s} \\ \rho'^*_{s-1} \\ \rho'^*_{s-2} \end{bmatrix}, \tag{5.9}$$

where $\mathbf{Q}''_\sigma = (\mathbf{Q}'_2)^{-1}\mathbf{Q}'_\sigma$, for $\sigma = 1, 0, -1, -2$. We write system equation (5.9) in short as $\bar{\rho}^*_{s+2} = \mathbf{A}_\rho \bar{\rho}^*_{s+1}$. The subscript s of $\bar{\rho}^*_s$ indicates the largest separation s at which some linear combination of the steady state probability distribution is known, given $\bar{\rho}^*_{s'}$ for $s' \leq s$.

We now use the LTI system to compute

$$\Sigma^*_{\bar{s}} = \sum_{s=\bar{s}}^{\infty} \left(p^*(\delta^s_{00}) + p^*(\delta^s_{01}) + p^*(\delta^s_{11})\right), \tag{5.10}$$

where $p^*(\delta)$ is the steady state version of $p^k(\delta)$. It can be verified that $\mathbf{A}_\rho$ is diagonalizable. In particular, we have that $\mathbf{V}_\rho^{-1}\mathbf{A}_\rho\mathbf{V}_\rho = \Lambda_\rho$, where ith column of $\mathbf{V}_\rho$ is the ith eigenvector of $\mathbf{A}_\rho$ and where

$$\Lambda_\rho = \text{diagonal}(\lambda_1, \ldots, \lambda_{12})$$

with λ_i the ith eigenvalue of $\mathbf{A}_\rho$ (for $i = 1, \ldots, 12$). The system has a set of stable modes and a set of unstable modes. Unlike for matrix $\mathbf{A}$ in Section 4.2.6 (see Lemma 14), it proved hard to establish that the number of unstable modes is constant with p. However, for all values of p tested, there are nine unstable system modes. We divide $\mathbf{V}_\rho$, $\mathbf{V}_\rho^{-1}$ and Λ_ρ accordingly into submatrices. Specifically,

$$\mathbf{V}_\rho = \begin{bmatrix} \mathbf{V}_{\rho 1} & \mathbf{V}_{\rho 2} \end{bmatrix},$$

where the columns of $\mathbf{V}_{\rho 1}$ and $\mathbf{V}_{\rho 2}$ are the stable and unstable mode eigenvectors, respectively; similarly,

$$\Lambda_\rho = \text{diagonal}(\Lambda_{\rho 1}, \Lambda_{\rho 2})$$

and

$$\mathbf{V}_\rho^{-1} = \left[\begin{array}{cc} \bar{\mathbf{V}}_{\rho 1}^T & \bar{\mathbf{V}}_{\rho 2}^T \end{array} \right]^T .$$

Let

$$\bar{\rho}_s'^* = \mathbf{V}_\rho^{-1} \bar{\rho}_s^* . \tag{5.11}$$

Then, the diagonalized system equation becomes $\bar{\rho}_{s+2}'^* = \Lambda_\rho \bar{\rho}_{s+1}'^*$. We have that

$$\Sigma_{\bar{s}}^* = \left\| \sum_{k=0}^{\infty} \mathbf{p}_{\bar{s}+k}^* \right\| ,$$

where $\|(\cdot)\|$ denotes the sum of the rows of $(\cdot)$. Let $\mathbf{T}_3 \in \mathbb{R}^{3\times 12}$ be such that $\rho_s^* = \mathbf{T}_3 \bar{\rho}_s^*$. Then, with Eqs (5.5) and (5.11), we have that

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathbf{p}_{\bar{s}+k}^* &= \mathbf{R}_\rho^{-1} \sum_{k=0}^{\infty} \rho_{\bar{s}+k}^* , \\
&= \mathbf{R}_\rho^{-1} \sum_{k=2}^{\infty} \mathbf{T}_3 \bar{\rho}_{\bar{s}+k}^* , \\
&= \mathbf{R}_\rho^{-1} \mathbf{T}_3 \sum_{k=2}^{\infty} \mathbf{V}_{\rho 1} \bar{\rho}_{\bar{s}+k,1}'^* , \\
&= \mathbf{R}_\rho^{-1} \mathbf{T}_3 \mathbf{V}_{\rho 1} (\mathbf{I} - \Lambda_{\rho 1})^{-1} \bar{\mathbf{V}}_{\rho 1} \bar{\rho}_{\bar{s}+2}^* ,
\end{aligned}$$

where $\bar{\rho}_{\bar{s}+k,1}'^*$ denotes the part of $\bar{\rho}_{\bar{s}+k}'^*$ associated with the stable system modes, and where we use the fact that the unstable system modes are not excited, since otherwise $\mathbf{p}_s^*$ diverges which is impossible. For

$$\mathbf{W} = \|\mathbf{R}_\rho^{-1} \mathbf{T}_3 \mathbf{V}_{\rho 1} (\mathbf{I} - \Lambda_{\rho 1})^{-1} \bar{\mathbf{V}}_{\rho 1}\| ,$$

we have that

$$\Sigma_{\bar{s}}^* = \mathbf{W} \bar{\rho}_{\bar{s}+2}^* . \tag{5.12}$$

In words, Eq. (5.12) provides the sum of the steady state probabilities for $s \geq \bar{s}$ as a linear equation in $\bar{\rho}_{\bar{s}+2}^*$.

We now formulate a set of linear equalities with $\mathbf{p}_s^*$ for $s \leq \bar{s}+1$ as solution. In particular, let

$$\mathbf{p}^* = \left[\begin{array}{cccccc} (\mathbf{p}_0^*)^T & \dots & (\mathbf{p}_{\bar{s}+1}^*)^T & \rho_{\bar{s}+2}^{1,*} & \rho_{\bar{s}+2}^{2,*} & \rho_{\bar{s}+3}^{1,*} \end{array} \right]^T.$$

The first $2 + 3(\bar{s} - 1)$ equalities in Eq. (5.1) are then

$$\begin{bmatrix} \mathbf{Q}_{0,\mathbf{I}}^0 & \mathbf{Q}_1^0 & \mathbf{Q}_2^0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{-1}^1 & \mathbf{Q}_{0,\mathbf{I}}^1 & \mathbf{Q}_1^1 & \mathbf{Q}_2^1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{-2}^2 & \mathbf{Q}_{-1}^2 & \mathbf{Q}_{0,\mathbf{I}}^2 & \mathbf{Q}_1^2 & \mathbf{Q}_2^2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ & & \vdots & & & & \ddots & & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Q}_{-2}^{\bar{s}-1} & \mathbf{Q}_{-1}^{\bar{s}-1} & \mathbf{Q}_{0,\mathbf{I}}^{\bar{s}-1} & \mathbf{Q}_1^{\bar{s}-1} & \mathbf{Q}_2^{\bar{s}-1} & \mathbf{0} \end{bmatrix} \mathbf{p}^* = \mathbf{0}, \tag{5.13}$$

where $\mathbf{Q}_{0,\mathbf{I}}^s = \mathbf{Q}_0^s - \mathbf{I}$. We adapt the equalities of Eq. (5.1) associated with $\mathbf{p}_{\bar{s}}^*$ and $\mathbf{p}_{\bar{s}+1}^*$ using Eqs (5.3) and (5.4) and the definition of ρ_s^* [Eq. (5.5)], which yields

$$\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}_{-2} & \mathbf{Q}_{-1} & \mathbf{Q}_0 - \mathbf{I} & \mathbf{Q}_1 & \mathbf{q}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{-2} & \mathbf{Q}_{-1} & \mathbf{Q}_0 - \mathbf{I} & \mathbf{q}_{11} & \mathbf{q}_{12} & \mathbf{q}_2 \end{bmatrix} \mathbf{p}^* = \mathbf{0}. \tag{5.14}$$

Further, we add a set of equalities that ensures that the unstable system modes are not excited. In particular, with $\mathbf{T}_\rho$ such that $\bar{\rho}_{\bar{s}+2} = \mathbf{T}_\rho \mathbf{p}$, we have as necessary condition

$$\bar{\mathbf{V}}_{\rho 2} \mathbf{T}_\rho \mathbf{p}^* = \mathbf{0}. \tag{5.15}$$

In the set of equations (5.13-5.15), we have $n_e = 6 + 3\bar{s}$ equations and an equal number of unknowns. However, less than n_e equations are linearly independent, since otherwise only the zero vector is a solution, which is impossible. We believe that for most relevant p and α, there are exactly $n_e - 1$ linearly independent equations in the set, which is the subject of further investigation. The n_eth linearly independent equation originates from the fact that the probabilities sum up to one and, with Eq. (5.10), can be written as

$$\sum_{s=0}^{\bar{s}-1} \mathbf{p}_s^* + W \mathbf{T}_\rho \mathbf{p} = 1.$$

We now have a set of n_e linearly independent equations in n_e unknowns, with as unique non trivial solution $\mathbf{p}_s^*$ for $s = 0, 1, \ldots, \bar{s} + 1$. For separations $s \geq \bar{s} + 2$, we simulate the autonomous LTI system in Eq. (5.9) with $\mathbf{T}_\rho \mathbf{p}^*$ as initial condition. From LTI system (5.9), it is clear that for $s \to \infty$, $\mathbf{p}_s^*$ decays exponentially to zero.

5.2 Remarks on Other Problem Instances

In this section, we present some aspects regarding the computation of the state probability distribution under an optimal policy in steady state for the several other problem instances considered in this thesis. Hereby, we focus on those specifics that differ from the solution presented in this chapter regarding the two-agent cone-shaped local observation zone problem instance.

Two Agents with Short Tunnel-Shaped Local Observation Zones

For the two-agent problem with short tunnel-shaped local observation zones, we establish in Section 4.3 an upper bound of three to the agent separation in steady state. Hence, the Markov chain representing the system state evolution under an optimal policy has a finite number of states, one set of states (with one state associated to each possible local observation zone set) per separation $0 \leq s \leq 3$. Therefore, standard methods to compute the steady state probability distribution over the states of a Markov chain suffice (we refer to [39] for further details).

Two Agents with Extended Tunnel-Shaped Local Observation Zones

The algorithm presented in Section 5.1, for the two-agent problem with cone-shaped local observation zones applies to the extended tunnel-shaped local observation zone problem version. We remark that the conversion from a mixed forward-backward recursion to a forward recursion is not as straightforward as in the cone-shaped local observation zone case, but can be executed along the lines of the method employed for a similar conversion in Section 4.4. Note that we only consider values for $p \geq 0.42$, the range of p for which the recursion at large separations holds. We further mention that extracting the unstable system modes presents some additional difficulties in the shape of numerical instabilities. In fact, the equivalent of the matrix $\mathbf{V}_\rho$ has a high condition number and is not reliably invertible for the values of p for which computed. We use a linear transformation of the

Schur type, which produces a block upper diagonal transformation, separating the stable from the unstable modes. We consider it outside the scope of the thesis to address the latter numerical instabilities, a problem that could be alleviated by selecting more involved numerical algorithms.

For the purpose of the problem analysis, we compute an approximation to the state probability distribution in steady state under an optimal policy by imposing an upper bound of 15 to the graph width. The thus obtained results indicate a probability of the order of 10^{-10}, for the agents to be located at separation ten at the values of p for which computed. Properties of the recursion describing the evolution of the steady state probability distribution at large separations, indicate the latter probability decreases exponentially with s, allowing us to be fairly confident in the accuracy of the results obtained.

Three Agents with Short Tunnel-Shaped Local Observation Zones

The algorithm presented in Section 5.1 applies as well for the three-agent problem with short tunnel-shaped local observation zones, under the conditions presented in Section 4.5. Again, the conversion from a mixed forward-backward recursion to a forward recursion proves more involved but can be executed along the lines of the method employed for a similar conversion in Section 4.4. Note that we we only consider values for $p > 0.384$, the range of p for which the recursion at large separations holds. Further, we mention that contrary to the two-agent extended tunnel-shaped local observation zone problem instance, the matrix $\mathbf{V}_\rho$ is invertible for the problem at hand. However, numerical instabilities arise in the formulation of the equivalent of the set of equations (5.13-5.15) for the problem at hand. In fact, the coefficient matrix of the latter set of equations has a high condition number and the matrix is singular. More sophisticated methods are required.

For the purpose of problem analysis, we compute an approximation to the state probability similarly as in the two-agent extended tunnel-shaped local observation zone problem instance, with similar accuracy.

5.3 Chapter Summary

In this chapter, we use the optimal policy associated with various problem instances to formulate the corresponding Markov chain and compute the state probability distribution under an optimal policy in steady state. We develop the algorithm in detail for the two-agent cone-shaped local observation zone problem instance, and present algorithm extensions and differences for the other problem instances considered in this thesis. The algorithm exhibits similarities with the algorithm to compute the optimal value function, presented in the previous chapter. In particular, we compute the steady state probabilities for the system states associated with the separations $s < \bar{s}$ as the solution of a set of equations. Note that at this point, $\bar{s}$ is known. As in the previous chapter, a subset of the latter set of equations provides sufficient information on the optimal agent policy at the separations $s \geq \bar{s}$ for the solution to represent the adequate state probability distribution at the states associated with $s < \bar{s}$. Further, the state probability distribution for $s \geq \bar{s}$ is described implicitly by a discrete linear time invariant system, whose initial condition is part of the solution of the equation set. The connection constraints in the equation set ensure that the initial condition, through simulation of the LTI system, leads to the state probability distribution under an optimal policy in steady state, for $s \geq \bar{s}$.

Chapter 6

The Efficiency of Cooperation

In Chapter 4, we presented an algorithm which efficiently solves the simple, though non-trivial, multi-agent navigation problems at hand. Then, Chapter 5 describes the Markov decision processes associated with the latter, and, given the optimal policy, presents a method to compute the state probability distribution in steady state under an optimal policy. For multi-agent problems, where the state space is infinitely large, both the optimal value function and the state probability distribution in steady state are described implicitly by means of a recursive equation for large separations, and given explicitly for small separations.

In this chapter, we use obtain optimal policy properties for a set of multi-agent problems, adequately chosen to address the main objectives of this thesis. In Section 6.1 we make some preliminary remarks and introduce the notation. In Section 6.2, we study how the group performance evolves as a function of the group size, for different local observation zones. Then, in Section 6.3, we address the effectiveness of particular local observation zones, for different group sizes. Further, in Section 6.4, we describe agent spatial distribution properties for different group sizes and local observation zones. Finally, in Section 6.5, we present the implications for the agents communication structure.

6.1 Preliminaries and Notation

In this section, we introduce the variables adequate for performance comparison. For any navigation problem of the type studied in this thesis, the state space $\mathcal{S}$ contains a possibly infinite, though countable number of states. Hence, we can meaningfully associate an index

$i \in \mathbb{N}$ to each $x \in \mathcal{S}$. Then, we have that $\mathcal{S} = \{x_i\}$, for $i = 0, 1, \ldots, n_s$, where $n_s = |\mathcal{S}|$ for a finite size state space, and where $n_s = \infty$ for the problems with an infinite size state space. Let $\mathbf{X}^*$ denote a vector with as many entries as there are elements in $\mathcal{S}$, such that its ith entry contains $J^*(x_i)$. Further, let $\mu^*(x)$ denote an optimal policy, with associated state transition matrix $\mathbf{P}_{\mu^*} = [p_{ij}(\mu^*(x_i))]$, for $i, j = 1, 2, \ldots, n_s$, and with $p_{ij}(\mu^*(x_i))$ denoting the transition probability from x_i to x_j under optimal decision $\mu^*(x_i)$. Let $\mathbf{g}_{\mu^*}$ be a vector with $g(x_i, \mu^*(x_i))$ as its ith entry, for $i = 1, 2, \ldots, n_s$, where $g(x_i, \mu^*(x_i))$ denotes the local cost incurred at state x_i under optimal decision $\mu^*(x_i)$. Then, we have that

$$\mathbf{X}^* = \sum_{k=0}^{\infty} \alpha^k \mathbf{P}_{\mu^*}^k \mathbf{g}_{\mu^*}, \tag{6.1}$$

where $0 < \alpha < 1$ is the problem discount factor. Let $\mathbf{p}_{\mu^*}$ denote the vector of steady state occupancy probabilities under an optimal policy μ^*. Note that $\mathbf{p}_{\mu^*}$ is the eigenvector of $\mathbf{P}_{\mu^*}$ corresponding to eigenvalue $\lambda = 1$, and is computed by the procedure outlined in chapter 5. Left-multiplying Eq. (6.1) by $\mathbf{p}_{\mu^*}^T$ yields

$$\begin{aligned} \mathbf{p}_{\mu^*}^T \mathbf{X}^* &= \sum_{k=0}^{\infty} \alpha^k \mathbf{p}_{\mu^*}^T \mathbf{P}_{\mu^*}^k \mathbf{g}_{\mu^*}, \\ &= \frac{1}{1-\alpha} \gamma N, \end{aligned} \tag{6.2}$$

where $\gamma = \frac{1}{N} \mathbf{p}_{\mu^*}^T \mathbf{g}_{\mu^*}$ represents the average cost per stage and per agent under optimal policy μ^*, and where N denotes the agent group size in the problem at hand. Indeed, $\mathbf{p}_{\mu^*}$ is the vector of steady-state occupancy probabilities under optimal policy μ^*. Its product with $\mathbf{g}_{\mu^*}$, the vector of locally incurred costs at each state under optimal policy μ^*, yields the average cost per stage under optimal policy μ^*. Rearranging terms in Eq. (6.2) yields

$$\gamma = \frac{(1-\alpha)}{N} \mathbf{p}_{\mu^*}^T \mathbf{X}^*.$$

In the sections 6.2.1-6.2.3, we compute γ for a set of different navigation problems and to avoid confusion, we add possible subscripts $1, 2$, or 3 to γ, to indicate its association with a one-, two- or three-agent problem, respectively. Similarly, we use an additional superscript d, t_2, or t_3 to indicate a cone-shaped, short tunnel-shaped, or extended tunnel-shaped local observation zone, respectively.

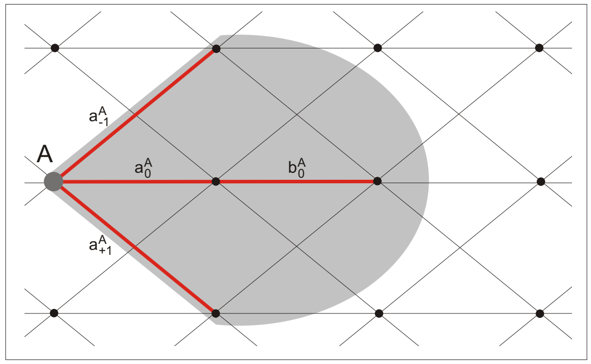

Figure 6-1: The cone-shaped local observation zone associated with agent A (grey dot), consisting of the red edges. The symbols associated with the latter represent their cost.

Finally, we note that we by no means claim that the policy μ^* minimizing the sum of discounted local costs is also the policy minimizing the average cost per stage. We only use the average local cost per agent, γ, as a means of comparing the performance of different navigation problems, of the type studied in this thesis.

6.2 Performance as Function of the Group Size

In this section, we present how the group performance evolves as a function of the number of group members for three distinct local observation zones. First, in Section 6.2.1, we consider the *cone-shaped* local observation zone, and compare the single-agent performance to the performance per agent for a two-agent problem. Then, in Section 6.2.2, we present results regarding the *short tunnel-shaped* local observation zone, comparing the per agent performance of a group with one, two and three agents. In Section 6.2.3, we study the *extended tunnel-shaped* local observation zone and compare the per agent performance of a group with one and two agents. Finally, in Section 6.2.4, we touch some aspects regarding suboptimal agent policies for larger agent groups. The results regarding the cone- and short tunnel-shaped local observation zones are previously presented in [25].

6.2.1 Cone-Shaped Local Observation Zone

We first consider the cone-shaped local observation zone case (see Fig. 6-1), and compare the single-agent performance to the performance of two agents. In particular, Fig. 6-2 shows γ_1^c (blue) and γ_2^c (red), the average cost per agent and per stage in steady state,

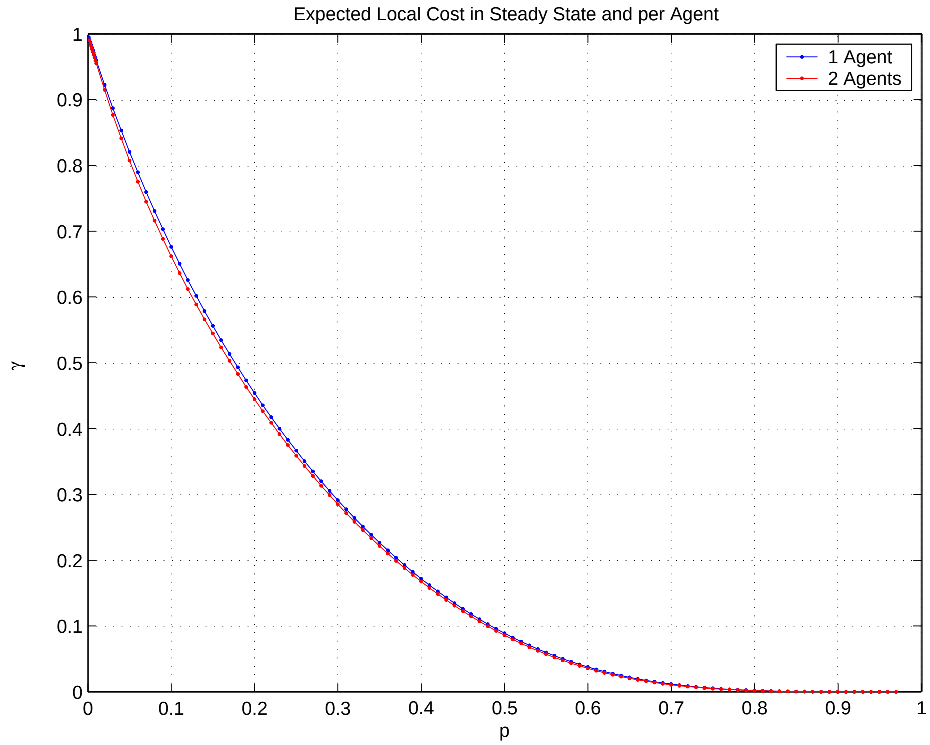

Figure 6-2: The average cost per stage and per agent under an optimal policy in steady state, as a function of p, for a single (blue) and a two- agent problem (red) with cone-shaped local observation zones.

for a set of values of p. To better present the asymptotic behavior of γ_1^c and γ_2^c as p approaches zero, we use a higher resolution for $0 \leq p \leq 0.01$; specifically, we display values for $p = 0.001, 0.002, \ldots, 0.01$. For $0.01 \leq p \leq 1$, we use a lower resolution, and display values for $p = 0.01, 0.02, \ldots, 0.97$. Numerical problems occur for $p > 0.97$, outlined more in detail later in this section.

We observe that, as expected, two agents perform better than a single agent. In particular, the average cost per stage and per agent is smaller for two agents than for a single agent. The two extreme cases where p approaches zero and one, respectively, exhibit to-be-expected behavior. As p approaches zero, the navigation graph consists primarily of edges with cost one, and thus, the average local cost per stage approaches one. A similar reasoning holds for the case where p approaches one.

The performance per agent is compared more effectively in a relative sense rather than in absolute sense. Specifically, we compute

$$r_{1,2}^c = \frac{\gamma_1^c - \gamma_2^c}{\gamma_1^c},$$

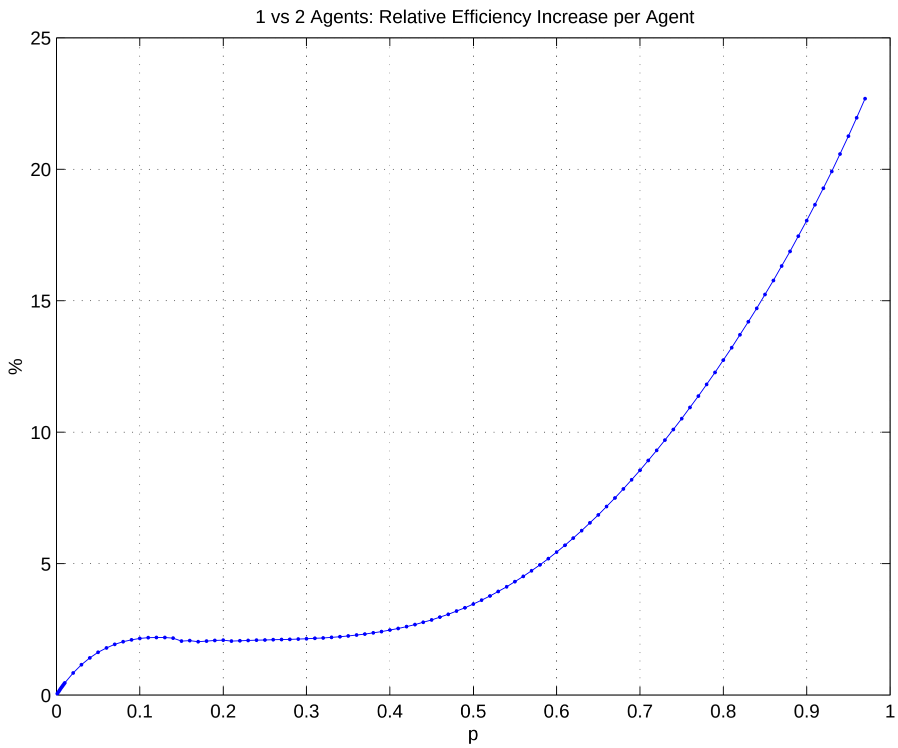

Figure 6-3: Decrease in the average cost per stage and per agent as a function of p (in percent) for a two-agent group compared to a single agent, for the case of cone-shaped local observation zones.

and show $100r^c_{1,2}$ (in percent) in Fig. 6-3, for the same set of p's as before. In words, Fig. 6-3 shows by what percentage the average cost per stage and per agent decreases when adding a second agent to a single agent. For example, for $p = 0.6$ the average cost per stage and per agent decreases by approximately 5.5%, when adding a second agent to a single-agent system.

Since $\lim_{p\to 0} \gamma^c_1 = \lim_{p\to 0} \gamma^c_2 = 1$, it is clear that $\lim_{p\to 0} r^c_{1,2} = 0$. In words, as the probability for encountering a zero edge cost decreases to zero, the benefit of two-agent cooperation disappears, and two agents perform as efficiently as a single agent. Further, for $p \leq 0.5$, the average cost per stage and per agent decreases only by 3% or less. Hence, in graphs where the majority of the edges are of cost one, a two-agent group is only slightly better ($< 3\%$) than a single agent. On the other hand, as p approaches one, and the graph primarily contains edges with zero cost, a rather surprising and interesting observation is made: an additional agent provides a cost decrease approaching 25%. In other words, when the probability of encountering an edge of cost one is small (p close to one), a two-agent group is up to 25% more efficient than a single agent. The surprise arises from the intuition that as the graph contains few rewards (edges with zero cost, p close to zero), more agents seem to be required to increase the number of locally observed edges and thus increase the probability of encountering an edge with zero cost. The results point in the opposite direction, namely, they indicate that two agents with cone-shaped local observation zones, are much more efficient detecting rare but high costs. Indeed, we note that the case where p is close to one corresponds to the practical situation where the environment is mostly safe, with very scarce but very damaging costs (for example, a mine field). In particular, notice that the maximum local cost incurred is one, while we have that for the average cost per stage in both the single-agent case as in the two-agent case $\gamma_1^c, \gamma_2^c << 1$.

Limit as p Approaches One

As previously mentioned, Fig. 6-3 only shows the values of $r_{1,2}^c$ for $p \leq 0.97$. For $p > 0.97$, numerical problems arise, and the value of γ_1^c is not sufficiently accurate to compute $r_{1,2}^c$. In particular, $\gamma_1^c = O(10^{-8})$ for $p > 0.97$, which is of order of the accuracy of the LP software CPLEX. However, with the single-agent and two-agent optimal agent policy for p approaching one, it is possible to compute

$$\bar{r}_{1,2}^c = \lim_{p \to 1} r_{1,2}^c = 0.25.$$

In particular, with $\varepsilon = 1 - p$, we first show rigorously that $\gamma_1^c = \varepsilon^4 + O(\varepsilon^5)$, so that $\gamma_1^c \approx \varepsilon^4$ for $\varepsilon << 1$. Then, we provide an intuitive, insightful explanation for this result. Finally, by a similar intuitive reasoning, we obtain that $\gamma_2^c \approx \frac{3}{4}\varepsilon^4$, for $\varepsilon << 1$, without a rigorous proof, which is more involved and left for future efforts. Specifically, in Section 7.2.1, we indicate a possible method for the latter.

We first review some notation. Recall the definition of $\mathcal{C}^A$ in Eq. (4.1):

$$\mathcal{C}^A = \begin{bmatrix} a_0^A & b_0^A & a_{-1}^A & a_1^A \end{bmatrix}^T,$$

which gathers the costs of the edges in the cone-shaped local observation zone associated with agent A at its current position. Fig. 6-1 indicates the edges are associated with the elements in $\mathcal{C}^A$; for further details, we refer to Section 3.1.1. The vector $\mathcal{C}^B$ is defined similarly.

We have the following lemma.

Lemma 18 *We have that* $\gamma_1^c = \varepsilon^4 + O(\varepsilon^5)$. $\square$

Proof: The proof goes as follows. We use the result of Lemma 8, which establishes optimality of the policy μ_1 (defined in Section 4.2.2) for all $0 < p, \alpha < 1$. Hence, we have that in the limit for ε approaching zero, policy $\mu_1^{*,c} = \mu_1$ is optimal. We compute the corresponding state probability distribution in steady state, which provides enough information to find the expression for γ_1^c.

Let the state $x = (a_1, b_1, a_{-1}, a_{+1}) \in \mathcal{S}_1^c$, *i.e.*, x contains the costs of the edges in the local observation zone at the current agent position. We associate an index to each state, which yields

$$
\begin{aligned}
x_1 &= (0,0,0,0),\\
x_2 &= (0,0,0,1),\\
&\dots\\
x_{16} &= (1,1,1,1),
\end{aligned}
$$

following an ordering as if the edge cost sets were binary numbers. With policy $\mu_1^{*,c}$, we can then write the state transition matrix $\mathbf{P}_{\mu_1^{*,c}} \in \mathbb{R}^{16\times 16}$, where entry $q_{ij}^{1,c}$ on row i and column j $(1 \le i, j \le 16)$ represents the probability of reaching state x_i from state x_j under policy $\mu_1^{*,c}$ and in one time step. In particular, we have that

$$
\mathbf{P}_{\mu_1^{*,c}} = \begin{bmatrix} \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & 0 & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c & \mathbf{q}_1^c \\ 0 & 0 & 0 & 0 & \mathbf{q}_2^c & \mathbf{q}_2^c & \mathbf{q}_2^c & \mathbf{q}_1^c & \mathbf{q}_2^c & \mathbf{q}_2^c & \mathbf{q}_2^c & 0 & \mathbf{q}_2^c & \mathbf{q}_2^c & \mathbf{q}_2^c & \mathbf{q}_2^c \end{bmatrix},
$$

with

$$
\begin{aligned}
\mathbf{q}_1^c &= \mathbf{t}_1 + O(\varepsilon)\mathbf{t}_1,\\
\mathbf{q}_2^c &= \varepsilon \mathbf{q}_1^c,
\end{aligned}
$$

and with

$$
\mathbf{t}_1 = \begin{bmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 & \varepsilon & \varepsilon^2 & \varepsilon^2 & \varepsilon^3 \end{bmatrix}^T.
$$

We now compute the eigenvector $\mathbf{p}_{\mu_1^{*,c}}$ of $\mathbf{P}_{\mu_1^{*,c}}$ corresponding to the unique eigenvalue equal to one. We have that

$$\mathbf{p}_{\mu_1^{*,c}} = \mathbf{t}_2 + O(\varepsilon)\mathbf{t}_2,$$

with

$$\mathbf{t}_2 = \begin{bmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 & \varepsilon & \varepsilon^2 & \varepsilon^2 & \varepsilon^3 & \varepsilon^2 & \varepsilon^3 & \varepsilon^3 & \varepsilon^4 & \varepsilon^3 & \varepsilon^4 & \varepsilon^4 & \varepsilon^5 \end{bmatrix}^T.$$

The eigenvector is scaled such that the sum of its entries converges to one, in the limit for ε approaching zero. Entry i of $\mathbf{p}_{\mu_1^{*,c}}$ (for $1 \leq i \leq 16$) represents the probability that the single-agent system state is x_i in steady state and under an optimal policy. Finally, we need the vector $\mathbf{g}_{\mu_1^{*,c}}$ of local costs under the policy $\mu_1^{*,c}$:

$$\mathbf{g}_{\mu_1^{*,c}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}^T,$$

which yields

$$\begin{aligned} \gamma_1^c &= \mathbf{g}_{\mu_1^{*,c}}^T \mathbf{p}_{\mu_1^{*,c}} \\ &= \varepsilon^4 + O(\varepsilon^5). \end{aligned}$$

Hence, for $\varepsilon << 1$, we have that $\gamma_1^c \approx \varepsilon^4$, which concludes the proof. ■

We now provide an intuitive explanation for the previous result, providing guidelines for the two-agent case. In particular, we study the most probable periodical sequence of states and compute the average cost incurred per stage. Note that under policy $\mu_1^{*,c}$ only two states lead to a local cost of one, namely the states $x_{12} = (1,0,1,1)$ and $x_{16} = (1,1,1,1)$. Therefore, we study the most likely sequence of states computing the probability of encountering x_{12} or x_{16}. Most commonly, the agent encounters a local observation zone where $a_0 = b_0 = 0$. In this case, the optimal decision is to traverse edge a_0 and we can compute the average number of stages $\bar{n}_1$ it takes for the agent to encounter an edge b_0 of cost one. In particular, we have that

$$\bar{n}_1 = \sum_{i=1}^{\infty} i\,\varepsilon(1-\varepsilon)^{i-1},$$

$$= \frac{1}{\varepsilon}.$$

Upon encountering an edge b_0 of cost one, the agent traverses a diagonal edge of cost zero, reaching a vertex on which previously no edge in the local information zone is observed. Therefore, every $\bar{n}_1$ edges, there is a probability of ε^3 that the agent incurs a local cost of one, which yields

$$\begin{aligned} \tilde{\gamma}_1^c &= \frac{\varepsilon^3}{\bar{n}_1}, \\ &= \varepsilon^4, \end{aligned}$$

where $\tilde{\gamma}_1^c$ denotes the average cost incurred in the particular state sequence presented above. Any other scenario leads to terms in γ_1^c which are at least one order higher in ε. For example, in the case where $a_{-1} = a_{+1} = b_0 = 1$ and $a_0 = 0$, then the optimal agent decision is to traverse edge a_0, reaching a node where with probability ε^2 the agent must incur a local cost of one. The latter scenario happens every $\bar{n}_1$ stages with a probability of ε^4, and hence leads to a term in the expression for γ_1^c that is of order five (specifically, a term $\frac{\varepsilon^4}{\bar{n}_1} = \varepsilon^5$). Hence, we conclude that $\gamma_1^c = \varepsilon^4 + O(\varepsilon^5)$ and for $\varepsilon << 1$, we have that $\gamma_1^c \approx \varepsilon^4$.

We now follow a similar intuitive reasoning for the two-agent case, studying the most likely scenario, without providing a rigorous proof. For the latter, we refer to Section 7.2.1 for suggestions regarding future efforts.We assume that the policy $\mu_2^{*,c}$ optimal for $\varepsilon = 0.03$ is also optimal in the limit for ε approaching one. Under optimal policy $\mu_2^{*,c}$, an agent incurs a local cost of one in case the three edges leading to the next stage are of cost one. Hence, we compute the probability of reaching such state. Policy $\mu_2^{*,c}$ is such that the agents A and B traverse the edges $a_0^A = 0$ and $a_0^B = 0$, respectively, until an edge $b_0^{(\cdot)}$ of cost one is observed. Depending on the agent separation, we have the following three possibilities:

1. Separation $s = 0$. From the single-agent case, we have that the average number of stages it takes for the agents to encounter an edge b_0 of cost one is $\bar{n}_2^0 = \frac{1}{\varepsilon}$. The subscript of $\bar{n}$ indicates the number of agents, while the superscript is the separation of the case considered. At this occurrence, the agents split along the two diagonal edges to reach two vertices where previously no edges are observed. The expected local cost $\bar{g}_2^0$ the two agents incur in this situation, is $\bar{g}_2^0 = 2((\varepsilon^3 + O(\varepsilon^4)) - \varepsilon^6) + 2\varepsilon^6$, where the first and the second term are associated with the cases one agent and two

agents face three edges of cost one, respectively. The term $O(\varepsilon^4)$ accounts for those states where one agent observes the costs of the edges leading to the next stage to be one, while the other agent observes the latter costs to be anything but three ones or three zeros. This term also includes those scenarios where the separation at the next stage equals two. Hence, we have that $\bar{g}_2^0 = 2\varepsilon^3 + O(\varepsilon^4)$.

2. Separation $s = 1$. In a similar computation as for the single-agent case, we find that the average number of stages it takes for one of the two agents to encounter an edge $b_0^{(\cdot)}$ of cost one is $\bar{n}_2^1 = \frac{1}{2\varepsilon}$. In the most likely case where only one edge is of cost one, we conclude that no local cost is incurred at $s = 1$. Indeed, suppose that all edges have zero cost, except for b_0^A which has cost one. Then agent A chooses the diagonal edge that leads to B's lane, encountering an edge cost $(a_0^A)' = (a_0^B)' = b_0^B = 0$ at separation zero, where $(\cdot)'$ indicates the value of $(\cdot)$ at the next stage. Hence, we write that the expected local cost the two agents incur in this situation, is $\bar{g}_2^1 = 0 + O(\varepsilon^5)$, where the second term arises from much less probable scenarios. One of those is when an agent A encounters two diagonal edge costs of one, while edge cost $b_0^A = 1$ and the two diagonal edge costs at the next stage on the same lane equal one as well. The second term further includes scenarios where the separation at the next stage equals two or three.

3. Separation $s = 2$. From the case where $s = 1$, we have that the average number of stages it takes for one of the agents to encounter an edge $b_0^{(\cdot)} = 1$ is $\bar{n}_2^2 = \frac{1}{2\varepsilon}$. Suppose that b_0^A is the only edge of cost one. Then, under policy $\mu_2^{*,c}$, agent A traverses the diagonal edge that leads towards B and to separation one, reaching a vertex where previously no edges are observed. Hence, the expected local cost the two agents incur is $\bar{g}_2^2 = \varepsilon^3 + O(\varepsilon^4)$, where the second term arises from less likely scenarios, including the ones where the separation at the next stage is zero, three or four.

We are now ready to compute γ_2^c by combining these results. In particular, we have that

$$
\begin{aligned}
\gamma_2^c &= \frac{1}{2}\frac{\bar{g}_2^0 + \bar{g}_2^1 + \bar{g}_2^2}{\bar{n}_2^0 + \bar{n}_2^1 + \bar{n}_2^2}, \\
&= \frac{3}{4}\varepsilon^4 + O(\varepsilon^5).
\end{aligned}
$$

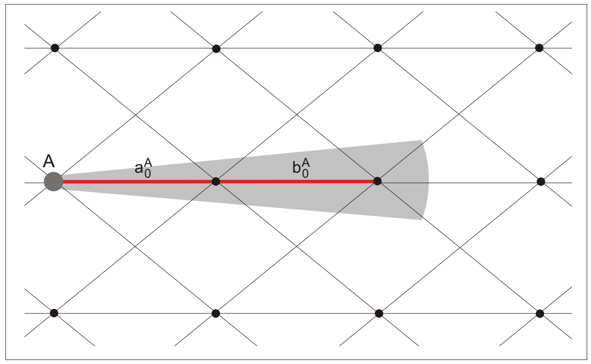

Figure 6-4: The short tunnel-shaped local observation zone associated with agent A (grey dot), consisting of the red edges. The symbols associated with the latter represent their cost.

Hence, we have that

$$\begin{aligned}
\bar{r}^c_{1,2} &= \lim_{\varepsilon\to 0}\frac{\gamma_1^c-\gamma_2^c}{\gamma_1^c}, \\
&= \lim_{\varepsilon\to 0}\frac{(\varepsilon^4+O(\varepsilon^5))-(\frac{3}{4}\varepsilon^4+O(\varepsilon^5))}{\varepsilon^4+O(\varepsilon^5)}, \\
&= \frac{1}{4}.
\end{aligned}$$

We conclude that based on the presented intuitive reasoning, a group of two agents is 25% more efficient than a single agent, in the limit for p approaching one.

6.2.2 Short Tunnel-Shaped Local Observation Zone

We now consider the short tunnel-shaped local observation zone case (see Fig. 6-4) and compare the performance of a single agent, two agents and three agents. In particular, Fig. 6-5 shows $\gamma_1^{t_2}$ (blue), $\gamma_2^{t_2}$ (red) and $\gamma_3^{t_2}$ (green), the average cost per agent and per stage for a single agent, a two-agent group, and a three-agent group, respectively, for a set of values of p. Again, to provide greater accuracy as p approaches zero, we use a small step size of 0.001 for $p \leq 0.01$. We use the same step size for $0.36 \leq p \leq 0.37$, where an apparent discontinuity appears in the curve for γ_1^c. Else, we use a normal step size of 0.01.

We observe that, as expected, three agents perform better than two agents, which in turn perform better than a single agent. In particular, the average cost per stage and per agent decreases the more agents the group contains. The two extreme cases where p approaches zero and one, respectively, exhibit to-be-expected behavior. As p approaches zero, the

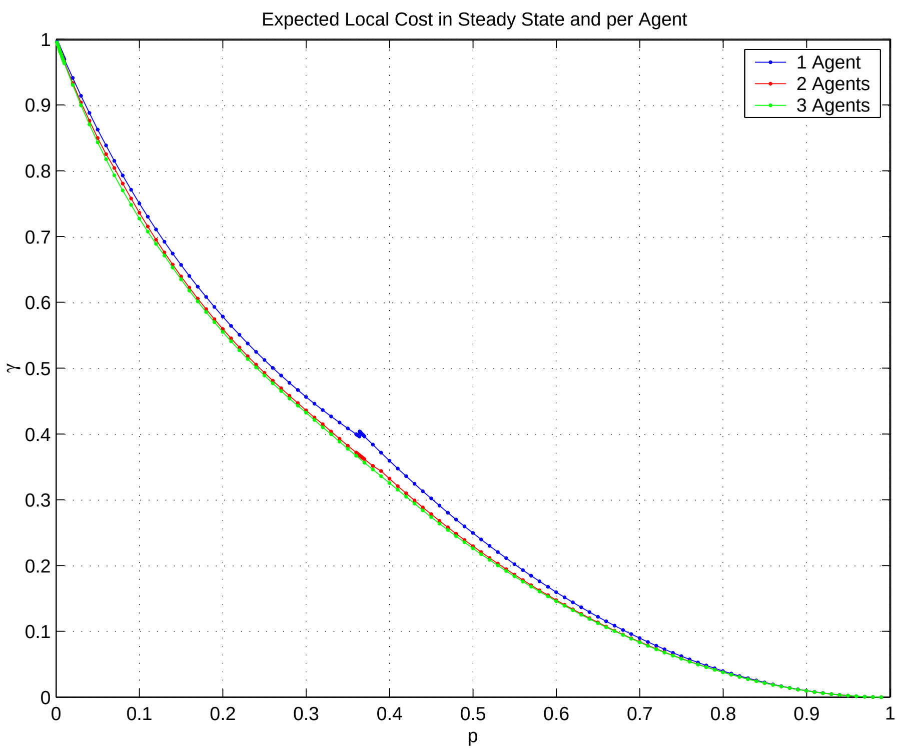

Figure 6-5: The average cost per stage and per agent under an optimal policy in steady state, as a function of p, for a single (blue), a two- (red), and a three- agent problem (green) with short tunnel-shaped local observation zones.

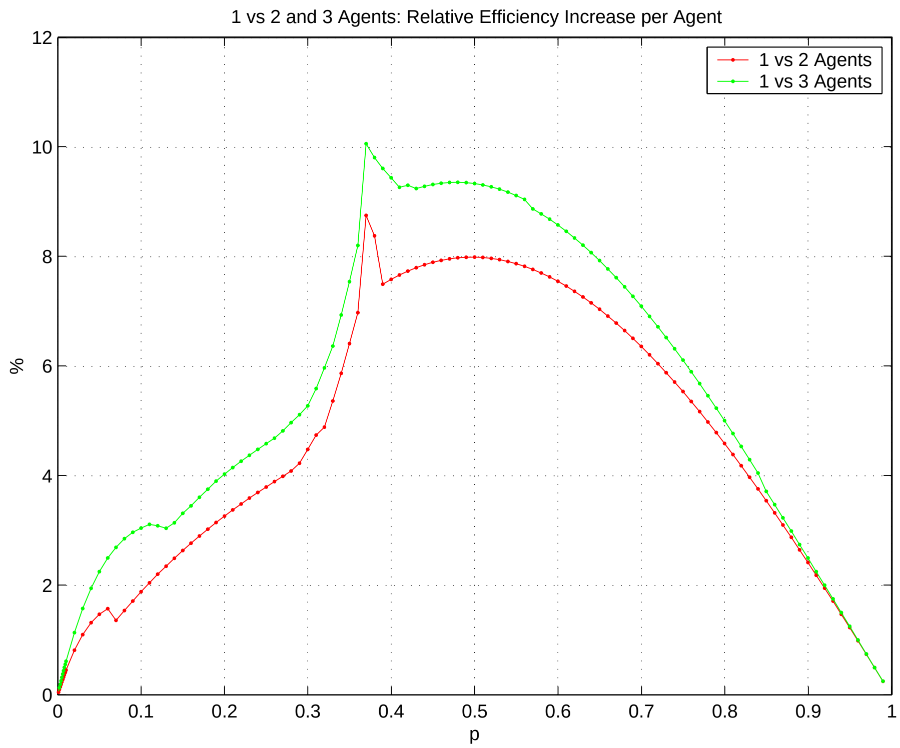

Figure 6-6: Decrease in the average cost per stage and per agent as a function of p (in percent) for a two-(red) and a three-agent group (green) compared to a single agent, for the case of short tunnel-shaped observation zones.

navigation graph consists primarily of edges with cost one, and thus, the average local cost per stage approaches one. A similar reasoning holds for the case where p approaches one.

The performance per agent is compared more effectively in a relative sense rather than in absolute sense. Specifically, we compute

$$r_{1,2}^{t_2} = \frac{\gamma_1^{t_2} - \gamma_2^{t_2}}{\gamma_1^{t_2}},$$

and

$$r_{1,3}^{t_2} = \frac{\gamma_1^{t_2} - \gamma_3^{t_2}}{\gamma_1^{t_2}},$$

for $p = 0.001, 0.002, \ldots, 0.009$, and for $p = 0.01, 0.02, \ldots, 0.99$, and show $100r_{1,2}^{t_2}$ (green) and $100r_{1,3}^{t_2}$ (red) (in percent) in Fig. 6-6. In words, Fig. 6-6 shows by what percentage the average cost per stage and per agent decreases when adding a second and a third agent to a single agent, respectively. For example, for $p = 0.5$, two and three agents, under an optimal policy, incur an average cost per stage and per agent that is approximately 8% and 9.3% smaller than the single-agent average cost per stage and per agent, respectively.

Since $\lim_{p\to 0}\gamma_1^{t_2} = \lim_{p\to 0}\gamma_2^{t_2} = \lim_{p\to 0}\gamma_3^{t_2} = 1$, it is clear that $\lim_{p\to 0} r_{1,2}^{t_2} = \lim_{p\to 0} r_{1,3}^{t_2} = 0$. In words, as the fraction of edges of cost one increases to one, two or three agents are as efficient as a single agent. Then, for $0.35 \leq p \leq 0.72$ two agents incur a cost per stage and per agent that is between 6% and 9% smaller than the cost per stage a single agent incurs. More interestingly, we observe that for approximately $0.05 \leq p \leq 0.7$, three agents incur a cost per stage and per agent that is about 1% smaller than the cost per stage and per agent two agents incur. For, $p \geq 0.7$, that difference in performance converges to zero, leading to the conclusion that for $p \geq 0.7$, three agents provide only a small to no benefit over two agents. The presented quantitative results, together with the fixed cost of an agent, provide the means to decide on the best group size for the type of task these problems model.

We now outline in further detail why $\lim_{p\to 1} r_{1,2}^{t_2} = \lim_{p\to 1} r_{1,3}^{t_2} = 0$, following an intuitive reasoning. First, we note that the optimal three-agent policy, for $p = 0.99$ is such that two out of three agents operate at separation zero. Hence, under the assumption that this policy is optimal as p approaches one, three agents essentially behave as two, from which we have that $\lim_{p\to 1} r_{1,2}^{t_2} = \lim_{p\to 1} r_{1,3}^{t_2}$. Next, we consider the single-agent case, and compute $\gamma_1^{t_2}$, *i.e.* the average cost per stage. In an analysis similar to the one in Section 6.2.1, we find that the average number of stages it takes for an agent to encounter an edge a_0 with cost one is $\frac{1}{\varepsilon}$, upon which the agent chooses a diagonal edge, incurring an average cost of ε. Hence, on average, the agent incurs a cost of ε, every $\frac{1}{\varepsilon}$ stages, and we have that $\gamma_1^{t_2} = \varepsilon^2$. A similar reasoning leads to an expression for $\gamma_2^{t_2}$. In particular, at $s = 0$, it takes an average of $\frac{1}{\varepsilon}$ stages to encounter an edge $a_0^A = a_0^B$ of cost one, upon which the agents diverge, choosing different diagonal edges and incurring a total average cost of 2ε. At $s = 2$, it takes an average of $\frac{1}{2\varepsilon}$ stages for one of the agents to encounter an edge $a_0^{(\cdot)}$ of cost one. On this occurrence and following optimal policy $\mu_2^{t_2}$, the latter agent traverses the diagonal edge that leads to $s' = 1$, incurring a average cost of $\varepsilon + O(\varepsilon^2)$, where the second term accounts for less likely scenarios. Finally and similarly, at $s = 1$, it takes an average of $\frac{1}{2\varepsilon}$ stages for one of the agents to encounter an edge $a_0^{(\cdot)}$ of cost one. Here, the latter agent chooses a diagonal edge converging or diverging from the other agent, incurring a total average cost of $\varepsilon + O(\varepsilon^2)$, where as before, the second term accounts for less likely scenarios. Combining the previous, yields

$$\gamma_2^{t_2} = \gamma_3^{t_2} = \varepsilon^2 + O(\varepsilon^3),$$

and hence,

$$\lim_{\varepsilon \to 0} r_{1,2}^{t_2} = \lim_{\varepsilon \to 0} \frac{\gamma_1^{t_2} - \gamma_2^{t_2}}{\gamma_1^{t_2}} = 0.$$

We conclude that both $r_{1,2}^{t_2}$ and $r_{1,3}^{t_2}$ approach zero, as p approaches one.

We now discuss changes in the optimal policy as a function of p. In the multi-agent case, we additionally focus on the optimal policy as a function of p for infinite agent separation. For the single-agent problem we have that for the values of $p \leq 0.3640$ for which we computed the optimal value function, the same policy is optimal. A similar statement holds for the (different) optimal policy for $0.3641 \leq p$. Hence, the optimal policy changes at least once, for $p \in [0.3640, 0.3641]$. Although the optimal value function is continuous as a function of p, the steady-state separation probability distribution changes in non-continuous fashion when the optimal policy changes. Therefore, a single jump appears at $p = 0.3641$ in the graph of $\gamma_1^{t_2}$. Further, a jump also appears in the graph of $r_{1,2}^{t_2}$ and $r_{1,3}^{t_2}$. We avoid the term 'discontinuity', since we represent the evolution of these variables at a discrete and not a continuous set of values for p.

Similarly, the optimal policy for the two-agent case changes at various values of p. In particular, by means of the following matrix, we indicate in which intervals for p the optimal policy changes at least once, at separations $0 \leq s \leq 3$. We only consider the latter separation range, since in steady state, we established that under an optimal policy the agent separation is upper bounded by three (see Section 4.3). Each column of the following matrix corresponds to an interval, bounded by the entries in the first and second row:

$$\mathbf{G}^2 = \begin{bmatrix} 0.060 & 0.068 & 0.120 & 0.286 & 0.312 & 0.314 & 0.318 & 0.324 & 0.332 & 0.382 \\ 0.062 & 0.070 & 0.122 & 0.288 & 0.314 & 0.316 & 0.320 & 0.326 & 0.334 & 0.384 \end{bmatrix}.$$

In each of these intervals, the optimal two-agent policy changes, leading to jumps in the graphs of both $\gamma_2^{t_2}$ and $r_{1,2}^{t_2}$, not always visible due the resolution employed. Note that at large separations, the two-agent optimal policy changes for at least one $p \in [0.3640, 0.3641]$,

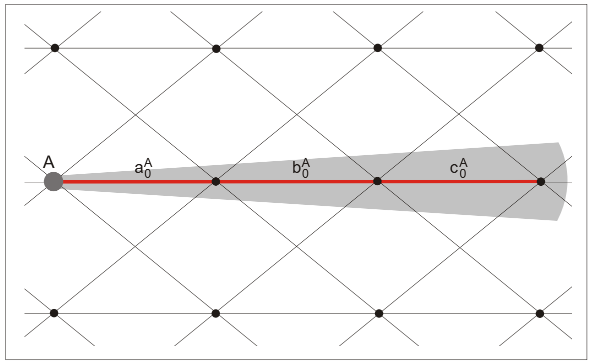

Figure 6-7: The extended tunnel-shaped local observation zone associated with agent A (grey dot), consisting of the red edges. The symbols associated with the latter represent their cost.

since in that interval the single-agent optimal policy changes at least once. Finally, we note that jumps in the graphs of $\gamma_3^{t_2}$ and $r_{1,3}^{t_2}$ arise for more values of p, since the optimal three-agent policy also changes at small separation as a function of p.

6.2.3 Extended Tunnel-Shaped Local Observation

Finally, we consider the extended tunnel-shaped local observation zone case (see Fig. 6-7) and compare the performance of a single agent and two agents. In particular, Fig. 6-8 shows $\gamma_1^{t_3}$ (blue) and $\gamma_2^{t_3}$ (red), the average cost per agent and per stage for a single- and a two-agent group, respectively, for a set of values of p. We notice that two agents perform better than a single agent in the sense that the average cost per stage and per agent is smaller for two agents than for a single agent. The extreme case where p approaches one, exhibits to-be-expected behavior, similar as in the two-agent case with short tunnel-shaped local observation zones.

Note that Fig. 6-8 and subsequent figures associated with the extended tunnel-shaped local observation zone problem instance only show results for values of $p \geq 0.42$, for reasons related to the computation of the optimal value function. We refer to Section 4.4 for further details.

The agent performance is compared more effectively in relative sense rather than in absolute sense. In particular, we compute

$$r_{1,2}^{t_3} = \frac{\gamma_1^{t_3} - \gamma_2^{t_3}}{\gamma_1^{t_3}},$$

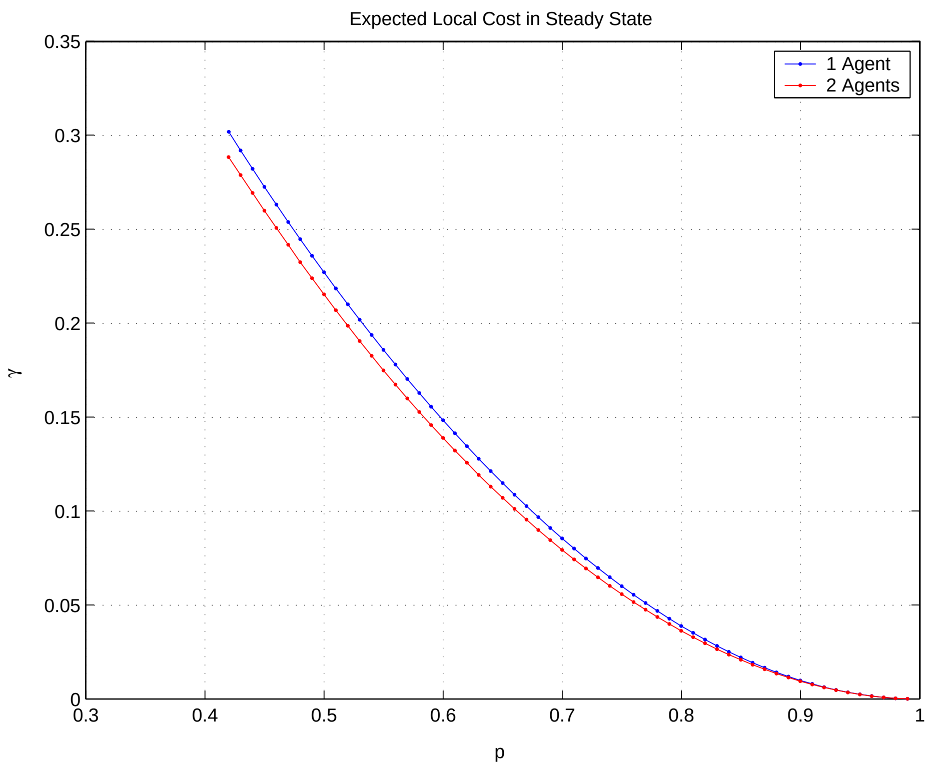

Figure 6-8: The average cost per stage and per agent under an optimal policy in steady state, as a function of p, for a single (blue) and a two- agent problem (red) with extended tunnel-shaped local observation zones.

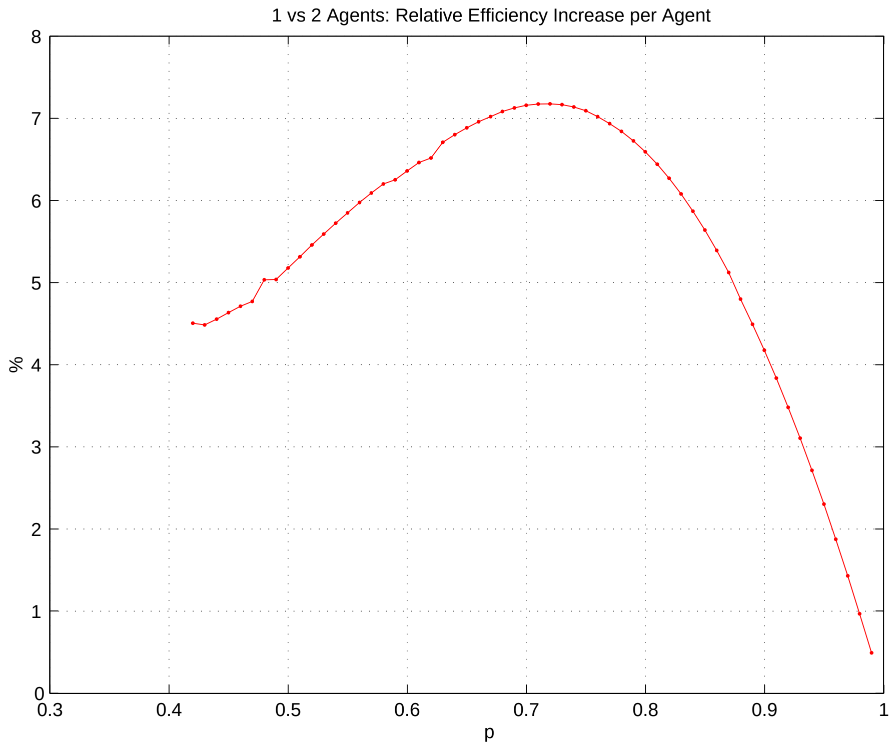

Figure 6-9: Decrease in the average cost per stage and per agent as a function of p (in percent) for two agents compared to a single agent, for the case of extended tunnel-shaped observation zones.

for $p \geq 0.42$ and show $100r_{1,2}^{t_3}$ in Fig. 6-9. In words, Fig. 6-9 shows by what percentage the average cost per stage and per agent decreases when adding a second agent to a single agent. We observe that for $p \sim 0.72$, the relative performance improvement reaches a maximum: two agents perform about 7% better than a single agent. We further note that as p approaches one, $r_{1,2}^{t_3}$ approaches zero, the explanation for which is similar to the one involving two agents with short tunnel-shaped local observation zones (see Section 6.2.2). In particular, one can see that by the same reasoning, we have that $\gamma_1^{t_3} = \gamma_2^{t_3} = \varepsilon^2 + O(\varepsilon^3)$, from which we have that

$$\lim_{\varepsilon \to 0} r_{1,2}^{t_3} = \lim_{\varepsilon \to 0} \frac{\gamma_1^{t_3} - \gamma_2^{t_3}}{\gamma_1^{t_3}} = 0.$$

Further, note that the apparent discontinuities in the Figs 6-8,6-9 arise from changes in the optimal policy as a function of p for some $s < \bar{s}$.

Finally, we compare the performance increases for the short and extended tunnel-shaped local observation zones. In particular, we point to a fairly peculiar difference. The maximum performance increase occurs at $p \sim 0.5$ for the short tunnel-shaped local observation zones and at $p \sim 0.72$ for the extended tunnel-shaped local observation zones, a difference which is cumbersome to clarify intuitively.

6.2.4 Large Agent Groups

In this section, we make some remarks regarding larger agent groups. For large groups, one traditionally uses approximate methods such as neuro-dynamic programming [7] to design multi-agent trajectories that are hopefully close to optimal. However, from the previous results, we can design suboptimal policies for moderately large agent groups, for which we can determine an upper bound on the optimality gap under reasonable assumptions. For example, for tunnel-shaped local observation zones, two agents increase the single-agent performance by approximately 5.5%, a third agents increases the performance further by approximately 0.5%, for $p = 0.75$ (see Fig. 6-6). We can conservatively assume that a fourth, fifth and sixth agent, each increase the performance by 0.5%. Hence, a suboptimal policy for a six-agent group that consists of a combination of two optimal three-agent policies, possibly misses out on a 1.5% performance increase, providing an indication of the size of the optimality gap. Fig. 6-10 shows an example six-agent trajectory set for $p = 0.75$ under a policy that consists of two three-agent optimal policies. Note how at from horizontal position 5 to 6 and from 24 to 27, the two groups of three intertwine, but do not interact. We can obtain a less conservative estimate of the optimality gap by extrapolating the evolution of the performance increase as the group size increases. By the latter reasoning, and for the example considered, the optimality gap is approximately 0.6% (assuming that the performance increase per additional agent decreases by a factor ten approximately).

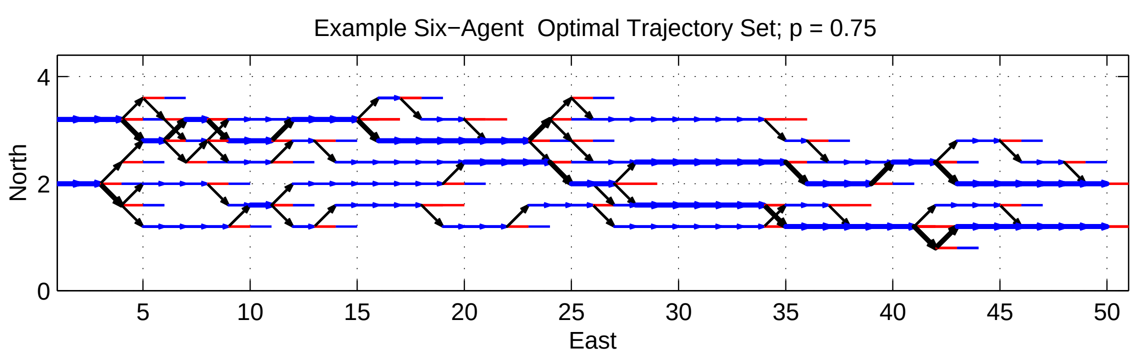

Figure 6-10: Six-agent example trajectory set for $p = 0.75$ under a policy that consists of two three-agent optimal policies. Blue (red) lines and arrows indicate edge costs of zero and one, respectively; black arrows indicates an agent traversing a previously unobserved edge.

We note that the success of the previous method, that is, splitting large groups into small subgroups, depends highly on the size of the performance increase for small agent groups. Indeed, for the cone-shaped local observation zone, and for p close to one, two agents provide a performance increase of approximately 25%, rendering estimating the performance evolution for larger groups cumbersome.

6.3 Performance as Function of the local Observation Zone Structure

In this section, we present how the agent performance evolves as a function of the local observation zone shape for a single- and a two-agent group. Specifically, in the Sections 6.3.1 and 6.3.2, we consider a single- and a two-agent group, respectively, and compare the performance for a short tunnel-shaped, extended tunnel-shaped and cone-shaped local observation zones. The notation adopted in the previous section, serves us in this section as well.

6.3.1 Single Agent

We first consider the single-agent problem and compare the performance for different local observation zones. In particular, Fig. 6-11 shows the average cost per stage for a set of values of p, for a short tunnel-shaped ($\gamma_1^{t_2}$, blue), an extended tunnel-shaped ($\gamma_1^{t_3}$, red), and a cone-shaped local observation zone (γ_1^{d}, green). These three local observation zones are shown in the Figs 6-4, 6-7, and 6-1, respectively. We observe that, as expected, a single agent with an extended tunnel-shaped or a cone-shaped local observation zone performs better than a with a short tunnel-shaped local observation zone. Somewhat less obvious, but not surprising, is the fact that two diagonal edges added to the short tunnel-shaped local observation zone lead to a more efficient agent compared to adding an extra edge straight ahead.

The performance for the different local observations is better compared in a relative sense, rather than in absolute sense. Specifically, we compute

$$\begin{aligned} r_1^{t_2,t_3} &= \frac{\gamma_1^{t_2} - \gamma_1^{t_3}}{\gamma_1^{t_2}}, \\ r_1^{t_2,c} &= \frac{\gamma_1^{t_2} - \gamma_1^{c}}{\gamma_1^{t_2}}, \end{aligned}$$

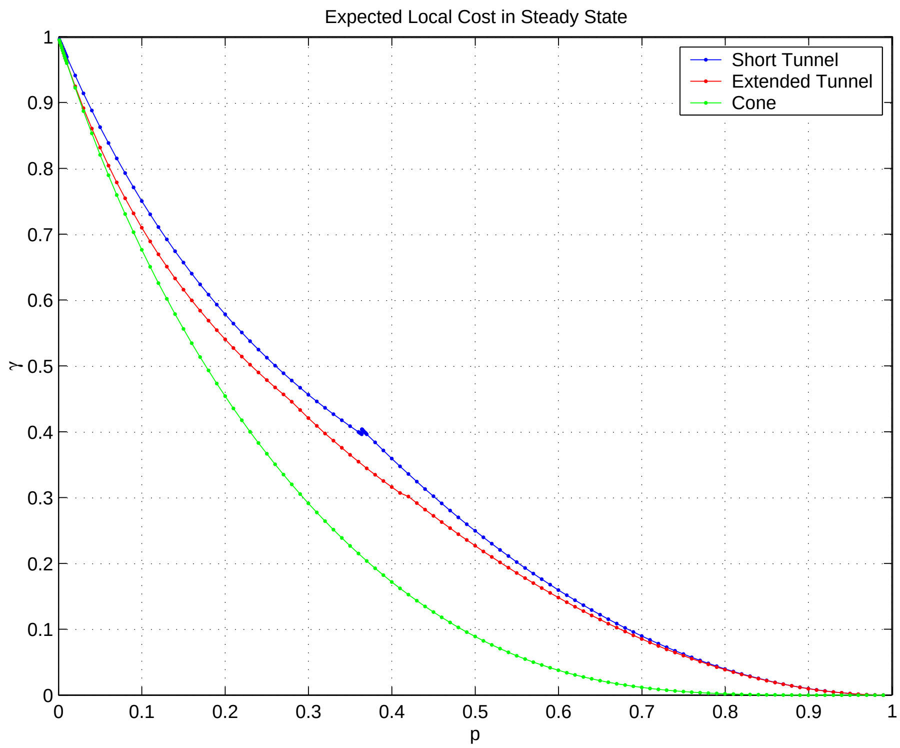

Figure 6-11: The average cost per stage under an optimal policy in steady state, as a function of p, for a single agent with short (blue), extended (red) tunnel-shaped and cone-shaped (green) local observation zones, respectively.

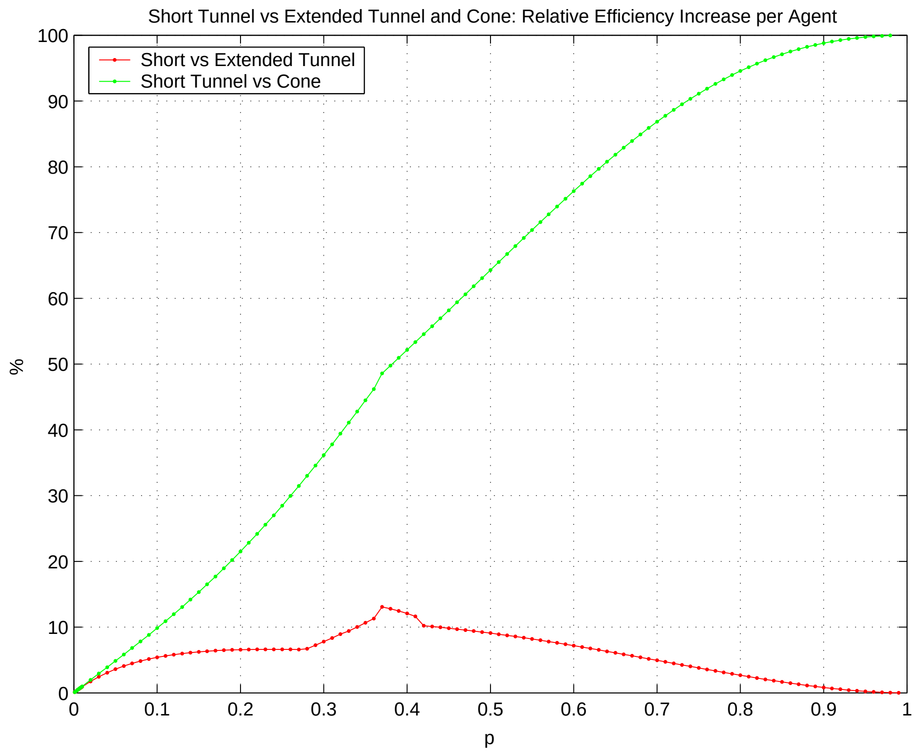

Figure 6-12: Decrease in the average cost per stage as a function of p (in percent) for a single agent with a short tunnel-shaped observation zone compared to a extended (red) tunnel-shaped and cone-shaped (green) local observation zone, respectively.

and show $100r_1^{t_2,t_3}$ (green) and $100r_1^{t_2,c}$ (red) (in percent) in Fig. 6-12. In words, Fig. 6-12 shows by what percentage the average cost per stage and per agent decreases when adding diagonal edges or a third edge straight ahead, respectively.

We observe that a single agent with an extended tunnel-shaped local observation zone performs between 5% and 13% more efficiently, for $0.07 \leq p \leq 0.7$ with a maximum at $p \simeq 0.37$. In the limit for p approaching zero, *i.e.* the probability of encountering a zero edge cost decreases to zero, an extra edge cost straight ahead provides no advantage. In an environment consisting mainly of edges of cost one, this is to be expected. Further, we note that in the limit as p approaches one, *i.e.* an environment where the probability of a zero edge cost approaches one, an extra edge straight ahead does not provide a significant advantage either.

Regarding the cone-shaped local observation zone, the results are different. In particular, a single agent with the latter local observation zone incurs an average cost per stage from 0% till 100% less as p goes from zero to one. Clearly, diagonal edges lead to a decrease in average cost per stage that is significantly larger than in the case of a local observation zone

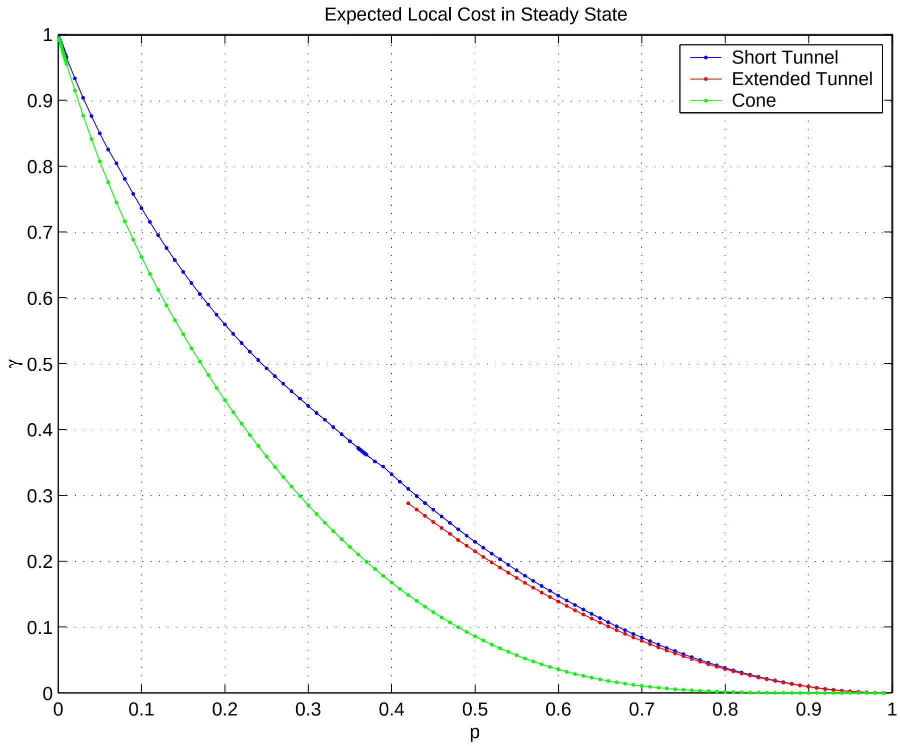

Figure 6-13: The average cost per stage under an optimal policy in steady state, as a function of p, for two agents with short (blue), extended (red) tunnel-shaped and cone-shaped (green) local observation zones, respectively.

with an extra edge straight ahead. The limit of $r_1^{t_2,c}$ as p approaches one, using the results for $\tilde{\gamma}_1^{t_2}$ and $\tilde{\gamma}_1^c$ obtained in the Sections 6.2.2 and 6.2.1, respectively, yields

$$\begin{aligned}\bar{r}_1^{t_2,c} &= \lim_{p\to 1} r_1^{t_2,c}\\ &= \lim_{\varepsilon\to 0}\frac{\tilde{\gamma}_1^{t_2}-\tilde{\gamma}_1^{c}}{\tilde{\gamma}_1^{t_2}},\\ &= \frac{\varepsilon^2-\varepsilon^4}{\varepsilon^2},\\ &= 1.\end{aligned}$$

Hence, as p approaches one, we observe a 100% decrease in average cost per stage incurred.

6.3.2 Two Agents

We now consider the two-agent problem and compare different sets of local observation zones. In particular, Fig. 6-13 shows the average cost per stage for a set of values of p, for short tunnel-shaped ($\gamma_2^{t_2}$, blue), extended tunnel-shaped ($\gamma_2^{t_3}$, red), and cone-shaped local

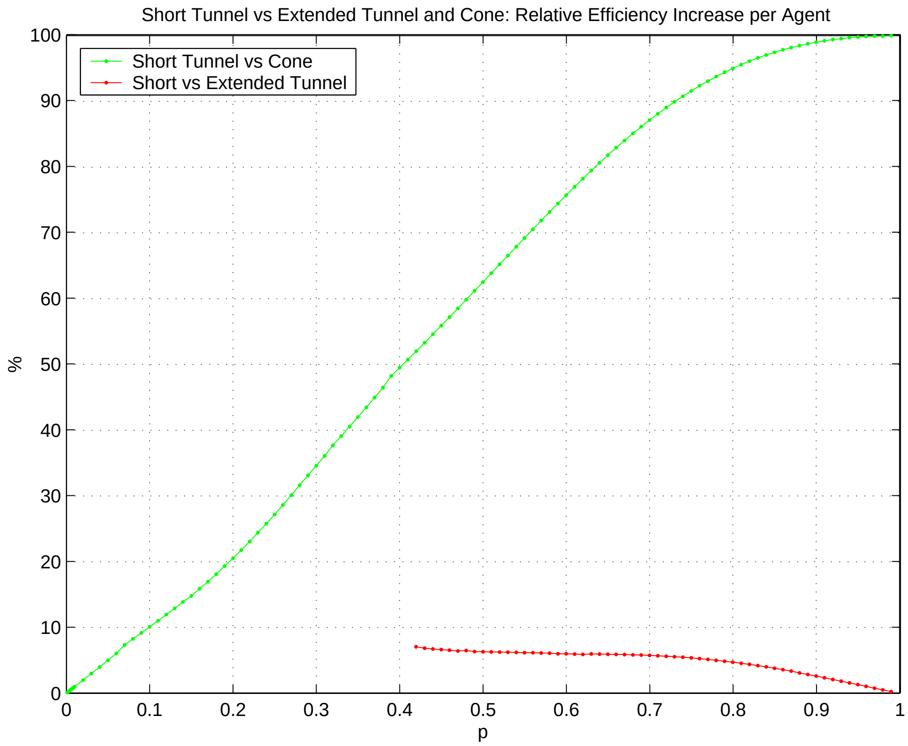

Figure 6-14: Decrease in the average cost per stage as a function of p (in percent) for two agents with a short tunnel-shaped observation zone compared to a extended (red) tunnel-shaped and cone-shaped (green) local observation zone, respectively.

observation zones (γ_2^c, green). Only values of $\gamma_2^{t_3}$ for $p \geq 0.42$ are presented for reasons described in Section 6.2.3. We observe that as expected, both the extended tunnel-shaped and the cone-shaped local observation zones yield enhanced performance in comparison to the short tunnel-shaped local observation zones. Further, similarly as in the single-agent case, two diagonal edges added to the short tunnel-shaped local observation zones lead to a more efficient agent pair than in the case only an extra edge straight ahead is added.

The performance is more effectively compared in relative sense. In fact, we compute

$$\begin{aligned} r_2^{t_2,t_3} &= \frac{\gamma_2^{t_2} - \gamma_2^{t_3}}{\gamma_2^{t_2}}, \\ r_2^{t_2,c} &= \frac{\gamma_2^{t_2} - \gamma_2^{c}}{\gamma_2^{c}}, \end{aligned}$$

and show $100 r_2^{t_2,t_3}$ (red) and $100 r_2^{t_2,c}$ (green) (in percent) in Fig. 6-14. The conclusions are similar to the ones in the single-agent case. In particular, we observe that two agents with extended tunnel-shaped local observation zones perform around 7% more efficiently at $p \sim 0.42$, decreasing towards 0% as p approaches one. In words, in the limit as p

approaches one, local information on an extra edge straight ahead is of little to no use, as in the single-agent case. Further, we observe that two agents with cone-shaped local observation zones incur a local cost per stage and per agent that is 0% to 100% smaller than in the short tunnel-shaped local observation zone case, as p evolves from zero to one, respectively. Clearly, as in the single-agent case, two agents with diagonal local information are far more efficient than two agents with enhanced straight ahead local information. In the limit as p approaches one, $r_2^{t_2,c}$ approaches one. Specifically, since $\tilde{\gamma}_2^{t_2} = \varepsilon^2$ and $\tilde{\gamma}_2^c = \frac{3}{4}\varepsilon^4$ (see Section 6.2), we have that

$$\lim_{\varepsilon \to 0} r_2^{t_2,c} = \lim_{\varepsilon \to 0} \frac{\tilde{\gamma}_2^{t_2} - \tilde{\gamma}_2^c}{\tilde{\gamma}_2^{t_2}} = 0.$$

In the next section, we focus on the quantifying the performance benefits different observed edges provide.

6.4 Spatial Distribution Characteristics

In this section, we present optimal policy properties related to the agent spatial distribution for different problems and as a function of parameter p. In particular, in Section 6.4.1, we consider the two-agent problem with cone-shaped local observation zones. Further, in Section 6.4.2, we present results regarding the two- and three-agent problems with short tunnel-shaped local observation zones. Finally, in Section 6.4.3, we handle the two-agent case with extended tunnel-shaped local observation zones.

6.4.1 Two Agents with Cone-Shaped Local Observation Zones

Recall that Theorem 2 allows us to solve for the two-agent optimal value function in two simple steps. First, solve $\mathcal{LP}_{in}$ to obtain the optimal value function at separations $s \leq \bar{s}-1$. Then, use its solution to compute the initial condition for the LTI system in Eq. (4.34) which is simulated to obtain the optimal value function for any $s \geq \bar{s}$. For $p = 0.7$ and $\alpha = 0.9$, Fig. 6-15 shows $E\mathbf{x}_{00}^{s,*}$, $E\mathbf{x}_{01}^{s,*}$ and $E\mathbf{x}_{11}^{s,*}$ as a function of s. As expected, as $s \to \infty$, the two-agent optimal value function converges to the dashed lines which represent the two-agent value function in case both agents adopt a single-agent optimal policy at all separations $s \geq 0$. For this particular case, $\bar{s} = 3$ is the smallest $\bar{s}$ for which we obtain the optimal value function. Fig. 6-16 (bottom) shows a sample optimal two-agent trajectory set for $p = 0.5$.

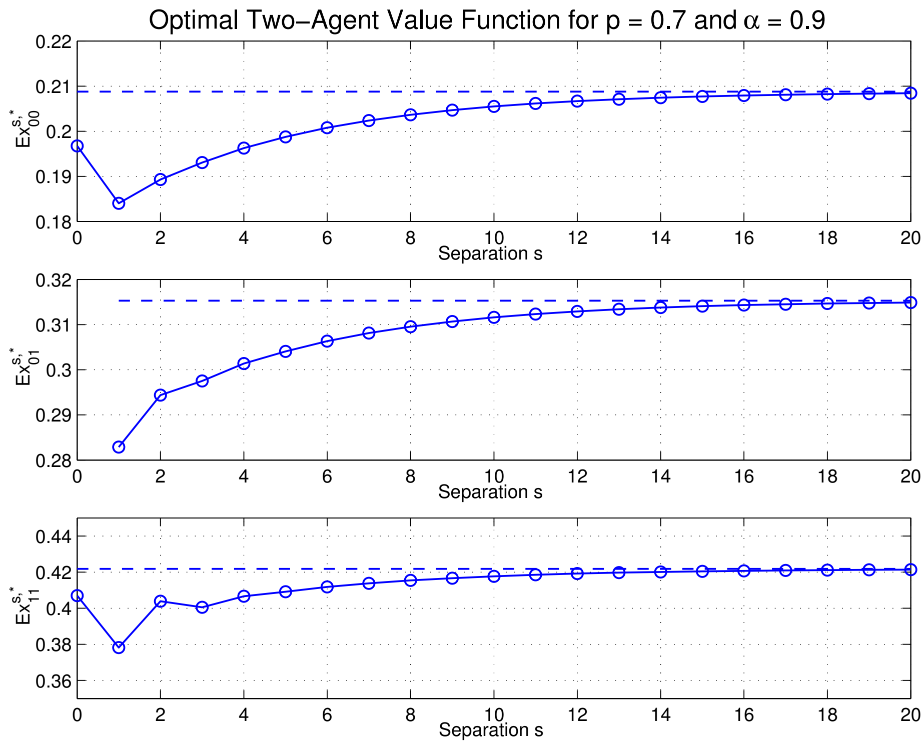

Figure 6-15: Two-agent problem with cone-shaped local observation zones. The optimal value function ($E\mathbf{x}_{00}^{s,*}$, $E\mathbf{x}_{01}^{s,*}$ and $E\mathbf{x}_{11}^{s,*}$) as a function of the agent separation s ($p = 0.7$ and $\alpha = 0.9$). The dashed line in each figure denotes the optimal two-agent value function for non-cooperating agents.

After the initial transition phase where, in expected sense, the agent separation decreases, the steady state is reached, where with high probability the agent separation remains small. Fig. 6-16 (top) shows the corresponding steady state probability distribution of the agent spatial distribution.

Fig. 6-17 shows the probability distribution of the two-agent separation in steady state and under an optimal policy for $p = 0.1, 0.2, \ldots, 0.9$. Later in this section, we focus on the agent behavior in the two extreme cases where p is close to zero and close to one. For all p, it is clear that the agents remain close to each other with high probability. In fact, as $s \to \infty$, the probability of encountering the agents at separation s decreases exponentially. Fig. 6-18 shows the expected value of agent separation in steady state. For p close to one, the expected agent separation is the smallest and equals approximately 0.75. On the other hand, for p close to zero, we have approximately the largest expected agent separation, equal to 2. This is further support for the intuition that as the environment is more hostile, the agents spread out to increase the size of the region where exact edge cost is observed, and thus increasing

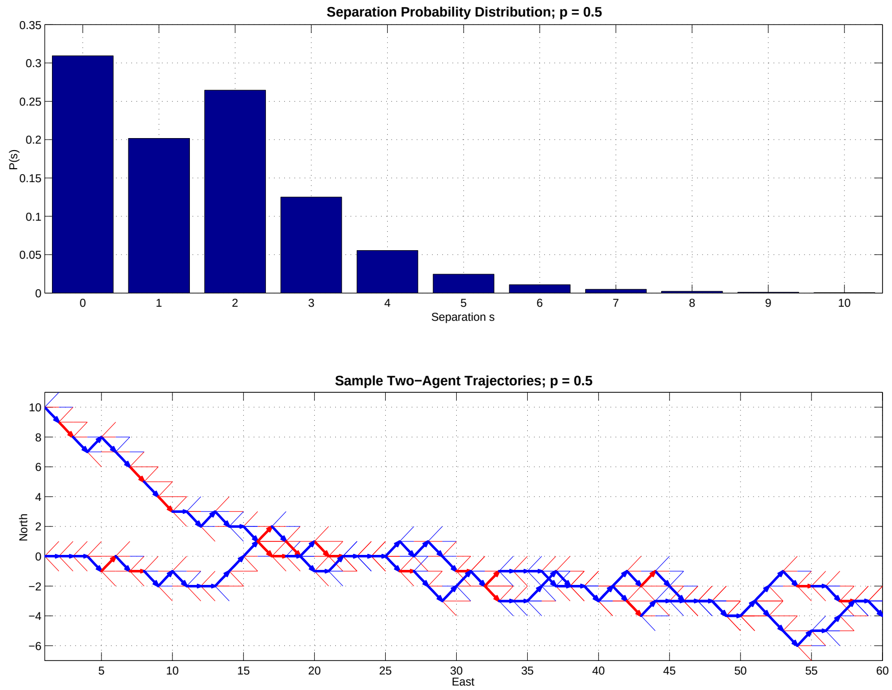

Figure 6-16: Two-agent problem with cone-shaped local observation zones. An example set of trajectories. Blue and red edges denote an edge cost of zero and one, respectively ($p = 0.7$ and $\alpha = 0.9$).

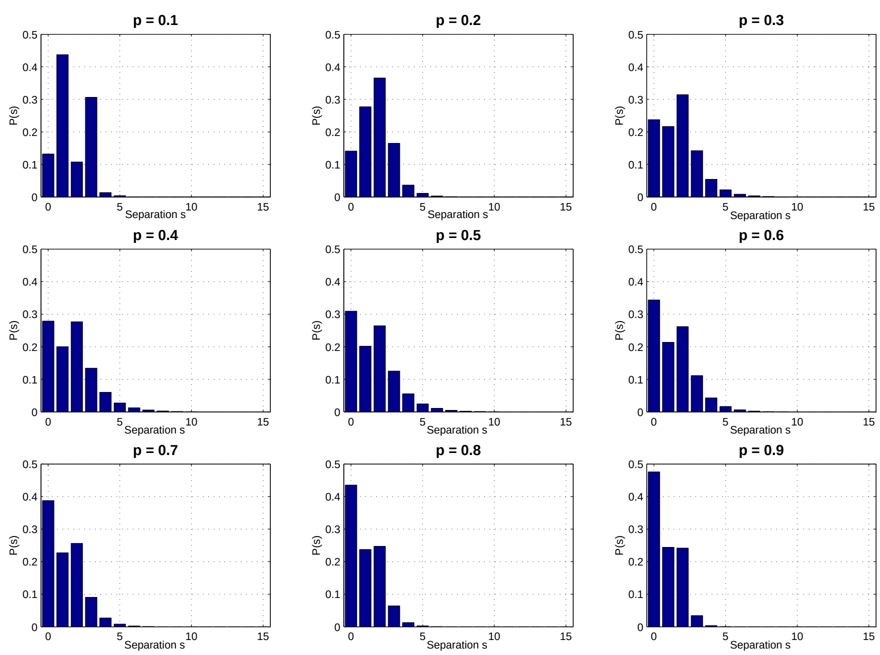

Figure 6-17: Two-agent problem with cone-shaped local observation zones. The steady state agent separation probability distributions for $p = 0.1, 0.2, \ldots, 0.9$, under an optimal policy.

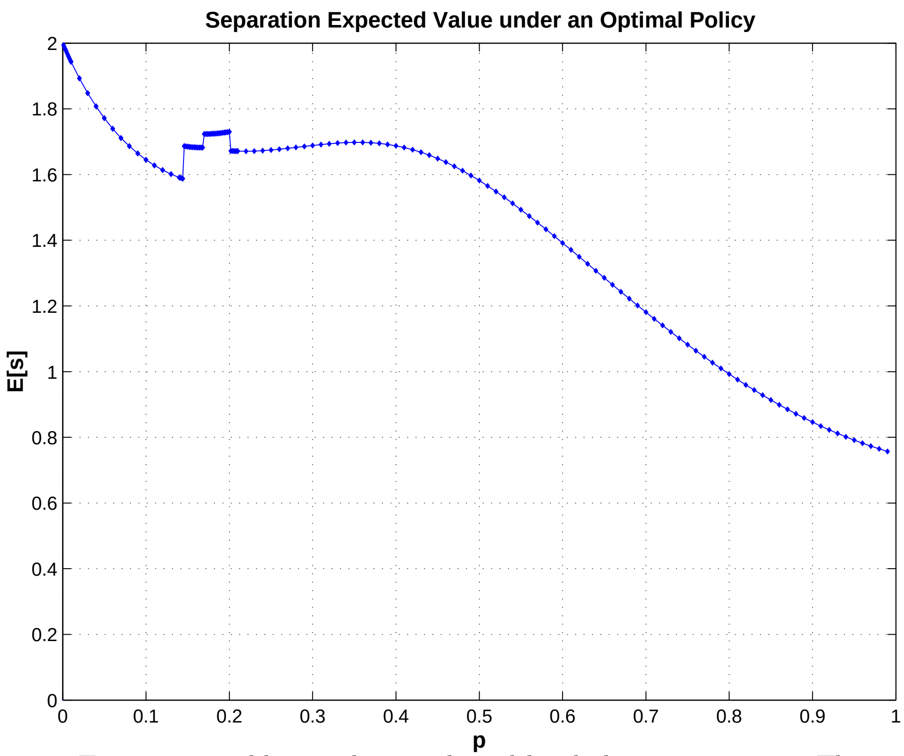

Figure 6-18: Two-agent problem with cone-shaped local observation zones. The expected value of the agent separation under an optimal policy for several values of p.

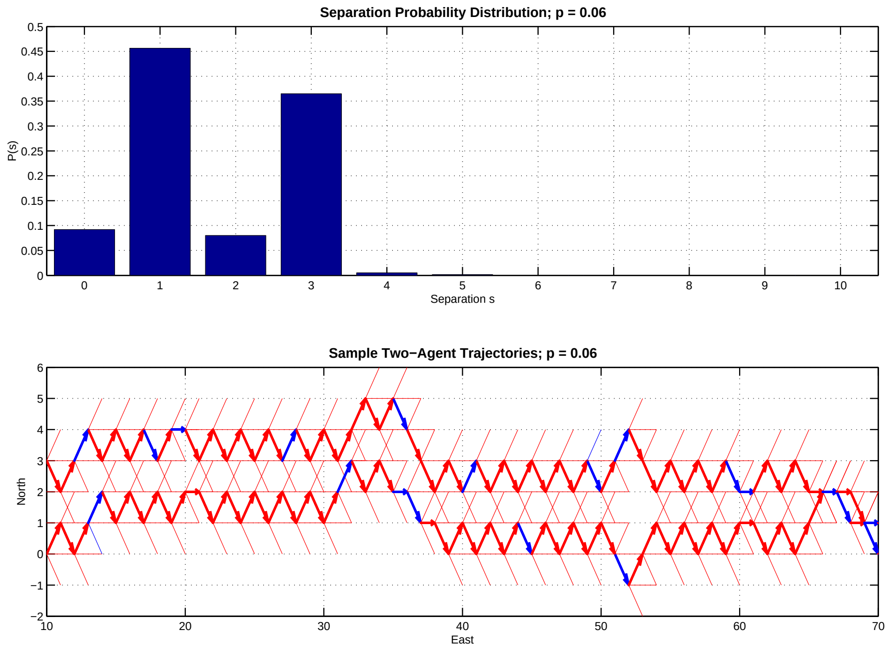

Figure 6-19: Two-agent problem with cone-shaped local observation zones. Top: Steady state agent separation probability distribution for a p close to zero ($p = 0.06$), under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.06$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively.

the probability of encountering a zero edge cost. The apparent discontinuous changes in the curve reflect changes in the optimal two-agent policy at the separations $s = 0, 1, 2$.

Fig. 6-19 and Fig. 6-20 show two extreme but instructive cases, for $p = 0.06$ and $p = 0.94$, respectively; in each figure, the top depicts the probability distribution of the separation in steady state, while the bottom depicts a set of sample trajectories. For $p = 0.06$, *i.e.* the case where ones are abundant, the separations $s = 1$ and $s = 3$ are most probable. The sample trajectories indicate the mechanics of cooperation. In particular, the agents tend to the most favorable separation $s = 1$, where one agent can leverage opportunities the other agent observes. However, with high probability, only ones are observed, driving the agents apart to the separation $s = 3$, where a set of eight previously unobserved edges is observed, thus maximizing the probability of encountering a zero. With high probability, only ones are in sight, and the agents converge again to the favorable separation $s = 1$, where eight previously unobserved edges enter the observation zone.

For $p = 0.94$, the case where zeros are abundant, the separation $s = 0$ is most likely, followed by $s = 1$ and $s = 2$, both equally likely. Again, the sample trajectories indicate the

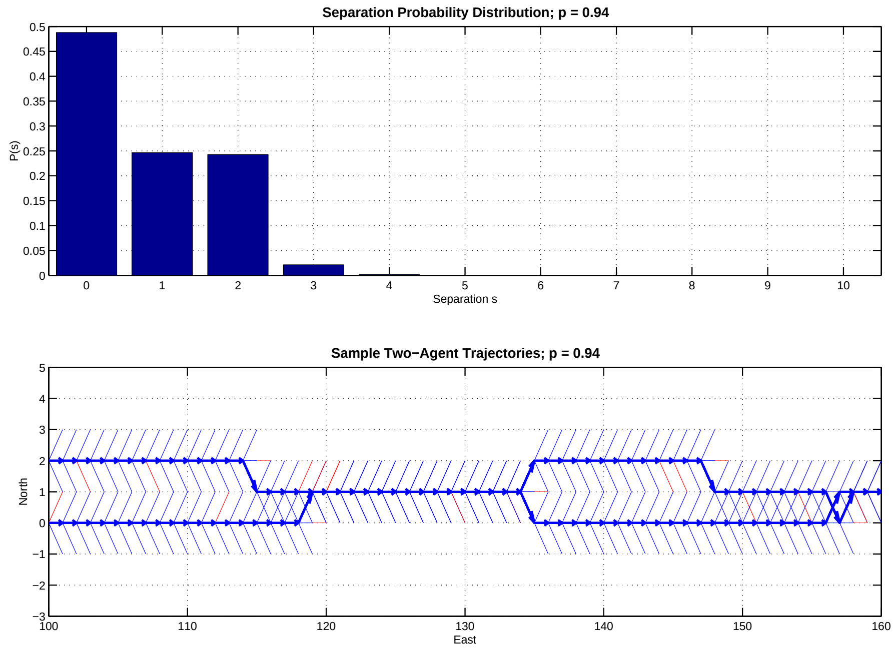

Figure 6-20: Two-agent problem with cone-shaped local observation zones. Top: Steady state agent separation probability distribution for a p close to one ($p = 0.94$). under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.94$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively.

mechanics of cooperation. In particular, let the agents start at $s = 1$, the most favorable separation. Most likely, only zeros are observed, and agents continue straight ahead. With probability $p_1 = 2p(1-p)$ ($= 0.11.$ for $p = 0.94$), an edge of cost one enters an observation zone one stage ahead (see for example stage 118 in Fig. 6-20). As a consequence, the agents converge to $s = 0$, which is maintained till again a one appears one stage further, which happens with with probability $p_2 = 1 - p$ ($= 0.06$ for $p = 0.94$). The agents split to $s = 2$, maximizing the number newly observed edges. Again, one edge cost of one enters an observation zone one stage ahead with probability p_1, causing the agents to converge back to $s = 1$. The difference in magnitude of p_1 and p_2 clarifies the difference of the probabilities with which agent are at separation $s = 0$ and at the separations $s = 1$ and $s = 2$.

6.4.2 Two and Three Agents with Short Tunnel-Shaped Local Observation Zones

We first present results regarding the two-agent problem and proceed with the three-agent problem.

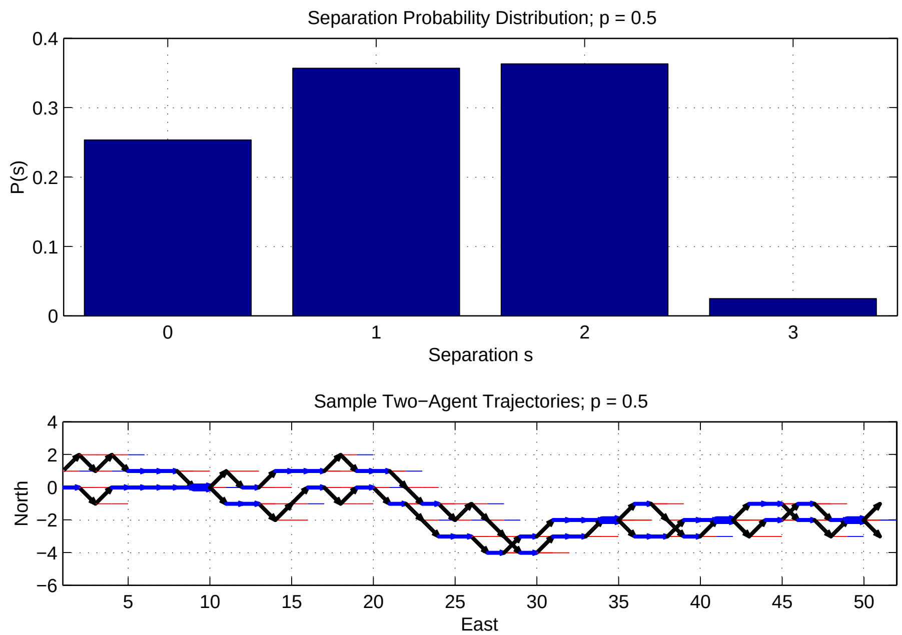

Figure 6-21: Two-agent problem with short tunnel-shaped local observation zones. Top: Steady state agent separation probability distribution for a $p = 0.50$, under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.50$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates a traversed edge with no previously known information, a bold arrow indicates an edge traversed by two agents.

Two Agents

From the exposition in Section 4.3, we have that for $0 \leq p \leq 1$ and for $0 \leq \alpha < 1$ the agent separation is upper bounded by three, in steady state. Hence, we can compute the optimal value function at the separations $s = 0, 1, 2, 3$ by considering a problem with finite state space, imposing the latter upper bound in Bellman's equation. Fig. 6-21 shows the steady state agent separation probability distribution for $p = 0.5$ (top), and a corresponding set of agent trajectories under an optimal policy (bottom). We observe indeed that in this particular sample trajectory set, the two agents remain within a separation of three. Correspondingly, only the steady state probability at separations $s = 0, 1, 2, 3$ is shown; at larger separations, this probability is zero. Note that one agent uses information the other agent observes at several occasions. Specifically, at the horizontal positions $8, 22, 28, 30, 33, 38, 40, 45, 47$ and 48, an agent (A) traverses the diagonal edge leading to the lane the other agent (B) is positioned at, ready to incur a cost agent B previously observed.

Fig. 6-22 shows the expected value of the agent separation under an optimal policies for a set of values of $0 < p < 1$ (dots) and for $\alpha = 0.9$. We notice that the smallest average

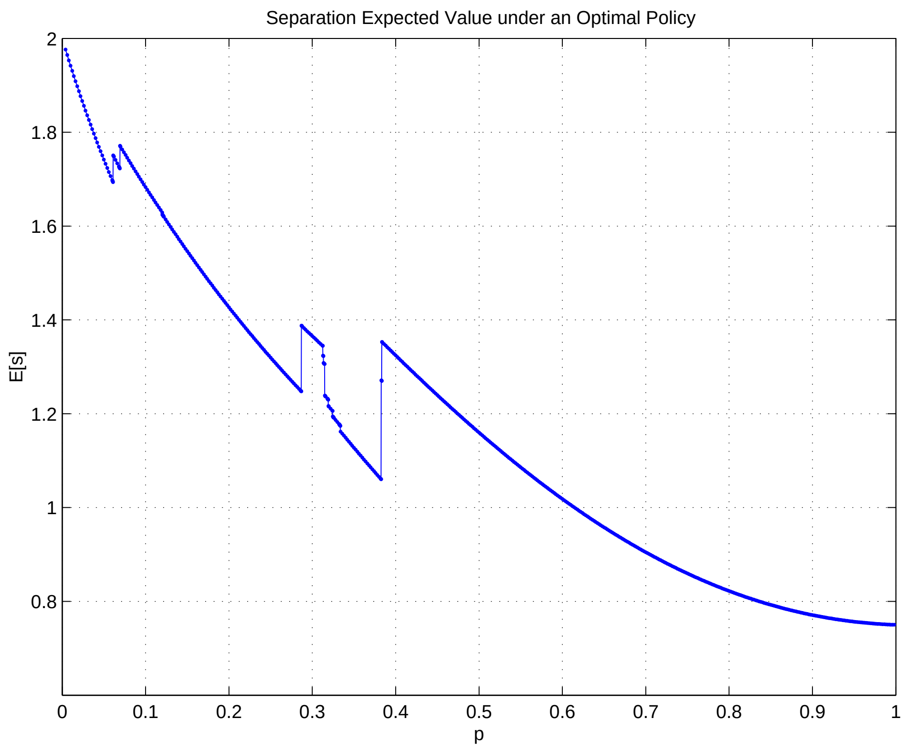

Figure 6-22: Two-agent problem with short tunnel-shaped local observation zones. The expected value of the agent separation under an optimal policy for several values of $0 < p < 1$ (dots on the curve)

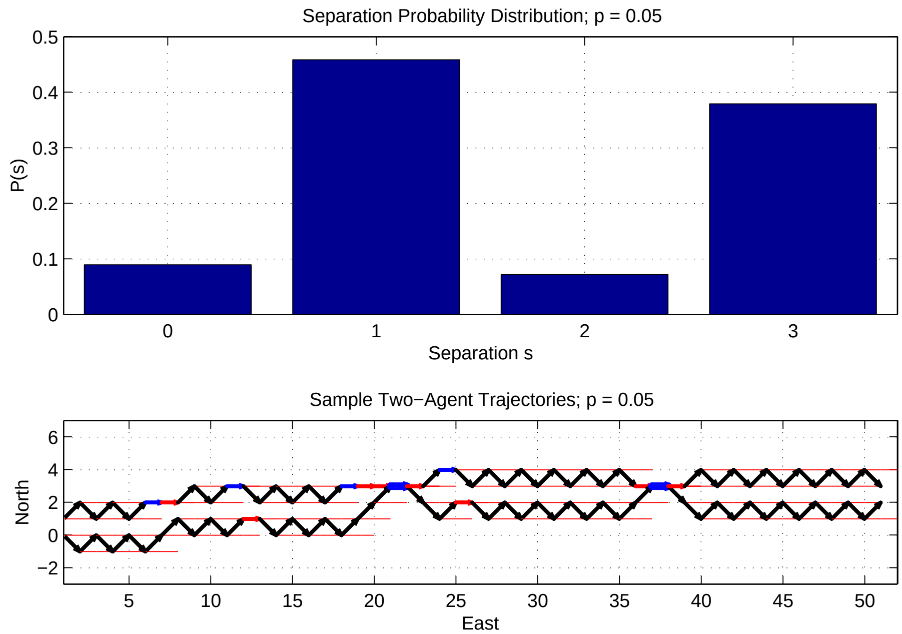

Figure 6-23: Two-agent problem with short tunnel-shaped local observation zones. Top: Steady state agent separation probability distribution for a $p = 0.05$, under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.05$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates a traversed edge with no previously known information, a bold arrow indicates an edge traversed by two agents.

separation is reached in the limit for $p \to 1$, and equals 0.75, the largest average separation equals two, and is reached in the limit for $p \to 0$. As in the two-agent case with cone-shaped local observation zones, the decreasing graph in Fig. 6-22 confirms the intuition that as less zeros are available in the environment (p close to zero), the agents tend to spread out, enlarging the union of the local observation zones and thus increasing the probability of encountering a zero. Similarly, as the environment contains few ones, there is no need to spread out and gather extra information. Specifically, the preferred separation is $s = 1$ so that when one agent encounters a rare edge of cost one, the other agent's observation of edge b_0 is of use to the first. Note that the apparent discontinuities in Fig. 6-22 reflect changes in the optimal policy.

In the rest of this section, we study the extreme cases for $p \to 0$ and $p \to 1$ more in detail. In particular, Fig. 6-23 (top) shows the steady state separation distribution under an optimal policy for p close to zero ($p = 0.05$). An example of an optimal trajectory set is depicted in Fig. 6-23 (bottom). A clear pattern emerges. In the event that no edges of zero cost are observed, the agents oscillate between $s = 1$ and $s = 3$. Indeed, if at $s = 1$, all observed edges are of cost one, the agents diverge, reaching two vertices at $s = 3$ where four

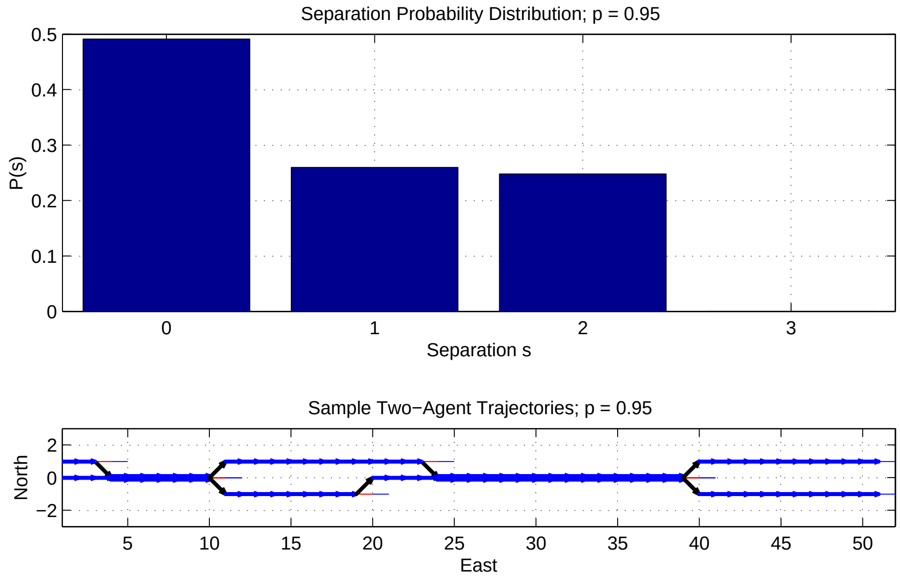

Figure 6-24: Two-agent problem with short tunnel-shaped local observation zones. Top: Steady state agent separation probability distribution for a $p = 0.95$, under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.95$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates a traversed edge with no previously known information, a bold arrow indicates an edge traversed by two agents.

previously unobserved edges enter the local observation zones. By this decision, the agents opt to gather information, rather than to stay close. Similarly, at $s = 3$, the agents decide to converge to $s = 1$, again observing four new edges, and with the additional advantage that if $b_0^A = 0$, agent B can take advantage as well by converging to $s = 0$. The latter situation occurs at horizontal positions 20 and 36 in Fig. 6-23 (bottom). This pattern is reflected in the steady state separation distribution, where we note that, indeed, the agents are at predominantly at $s = 1$ and $s = 3$, with equal probability.

Finally, we discuss the agent behavior as $p \to 1$. In particular, Fig. 6-24 (top) shows the steady state separation distribution under an optimal policy for $p = 0.95$. An example of an optimal trajectory set is depicted in Fig. 6-24 (bottom). Similarly as for $p = 0.05$, a clear pattern emerges. In fact, there is a clear similarity of the trajectory structure and the associated steady state separation probability distribution between this problem and the two-agent problem with cone-shaped local observation zones when p is close to one. The latter is described in Section 6.4.1 and applies here in identical fashion.

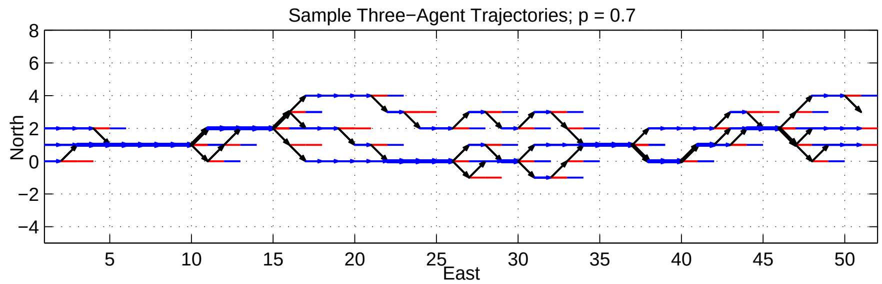

Figure 6-25: Three-agent problem with short tunnel-shaped local observation zones. A set of sample trajectories for $p = 0.70$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates an edge traversed without previous information on its cost, and a bold arrow indicates an edge traversed by more than one agent.

Three Agents

For the three-agent problem, recall that we assume an upper bound of three on the smallest agent-pair separation s_1. Although we provide no procedure to formally prove correctness of this assumption, numerical computations indicate that most likely the upper bound holds. We refer to Section 4.5 for details, and discuss results related to the three-agent spatial distribution here. Fig. 6-25 shows an example of a three-agent trajectory set under an optimal policy for $p = 0.7$. Note how in many occasions two out of three or all three agents converge to separation zero and traverse an edge with zero cost. Fig. 6-26 shows the corresponding steady state agent separation probability distribution for different separations s_1 and s_2. Since, without loss of generality, by using problem symmetry, we assumed that $s_1 \leq s_2$, probabilities appear for those pairs of s_1 and s_2 for which the previous relation holds. From the exposition in Section 5.2, we have that the probability for the agents to be at s_1 and s_2 decreases exponentially with increasing s_2. Hence, from the figure, we conclude that with a probability of approximately 0.99, the agents remain within a total separation of four.

Fig. 6-27 shows the expected separation s_1 (blue) and s_1+s_2 (red) for a set of values for p (dots). In other words, for any particular $0 < p < 1$, the average agent configuration is such that one agent sits on the p-axis, the second agent sits on the blue curve, while the third agent is located on the red curve. As for the two-agent problem with cone-shaped or short tunnel-shaped observation zones, we have that the tendency to spread is more pronounced in an environment predominantly containing edges with cost one ($p < 0.5$). Similarly, the tendency to converge dominates for $p > 0.5$, the case where the environment mainly consists

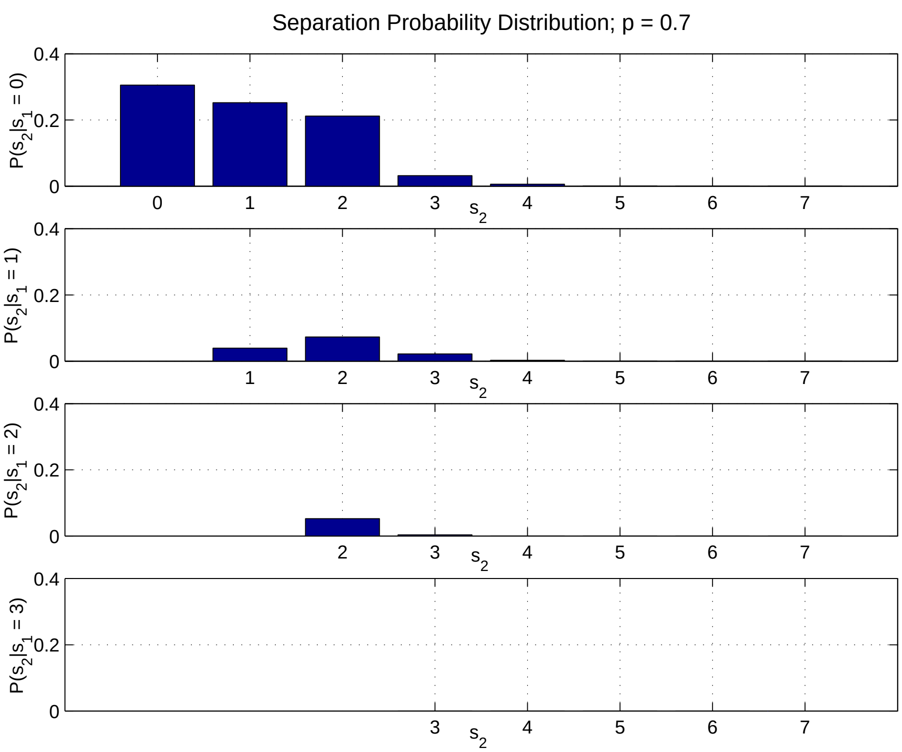

Figure 6-26: Three-agent problem with short tunnel-shaped local observation zones. The steady state agent separation probability distribution for $p = 0.7$. From top to bottom, the figures represent the distribution for $s_1 = 0, 1, 2, 3$, respectively; the horizontal axis represents s_2 ($\geq s_1$).

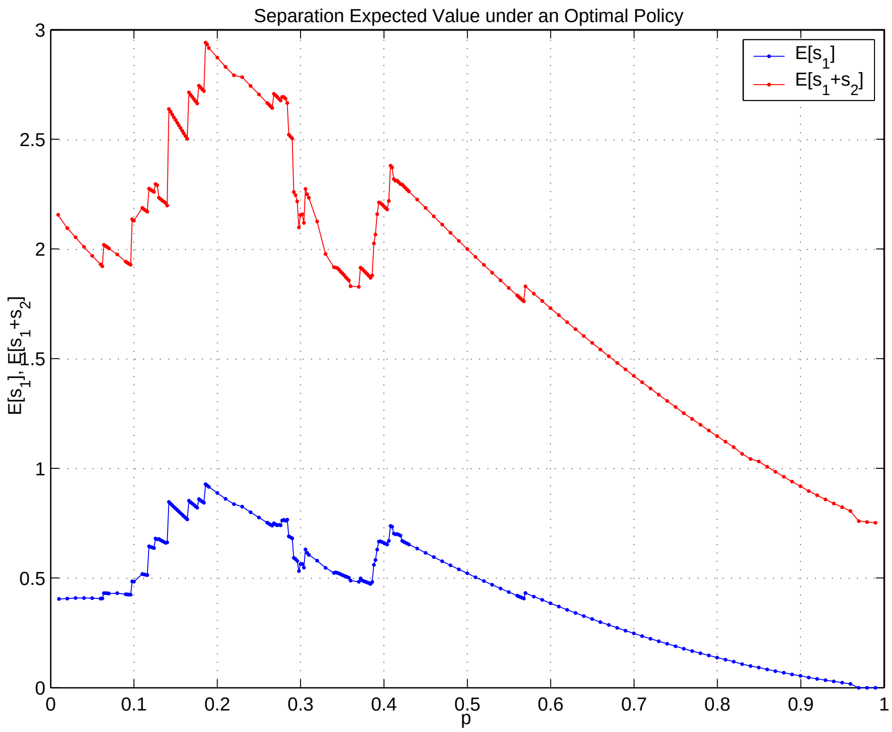

Figure 6-27: Three-agent problem with short tunnel-shaped local observation zones. The expected value of the agent separation s_1 (blue) and $s_1 + s_2$ (red) under an optimal policy for several values of $0 < p < 1$ (dots on the curve).

of edges with zero cost. Again, this confirms and quantifies the intuition that spreading and information gathering is essential in environments with few rewards (zero cost edges, p close to one). Further note that as p approaches one, separation s_1 approaches zero, indicating that for this type of environment, two out of three agents stay at separation zero. This confirms the result in Section 6.2.2, where we conclude that the benefit of a third agent disappears for p approaching one. Indeed, for such p, we have that two out of three agents essentially behave as a single agent, by sticking together under an optimal policy. Finally, note that the apparent discontinuities in Fig. 6-27 reflect policy changes as p varies.

We now focus on the agent behavior in two extreme cases, for small and large values of p. Specifically, Fig. 6-28 shows a set of three-agent trajectories under an optimal policy, for $p = 0.01$. We notice two different emerging patterns, equally present in Fig. 6-29, which shows the steady state separation probability distribution, under an optimal policy and for $p = 0.01$. The first pattern appears from horizontal position 1 to 13, and from 19 to 31. Here, separation $s_1 = 1$ and $s_2 = 3$, where the two outer agents switch role at each stage. This situation occurs when exclusively edges of cost one are observed and the agents tend

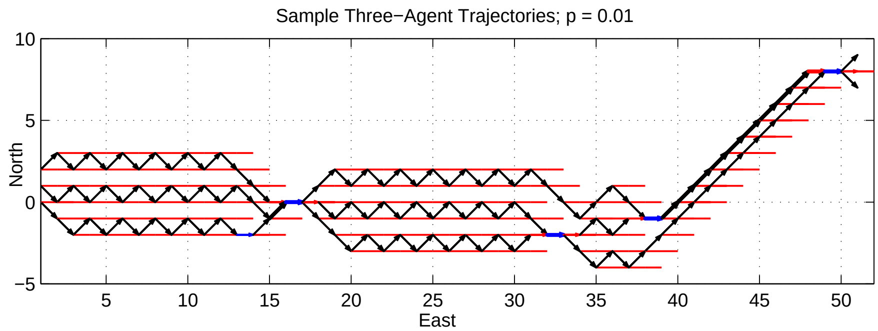

Figure 6-28: Three-agent problem with short tunnel-shaped local observation zones. A set of sample trajectories for $p = 0.01$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates an edge traversed without previous information on its cost, and a bold arrow indicates an edge traversed by more than one agent.

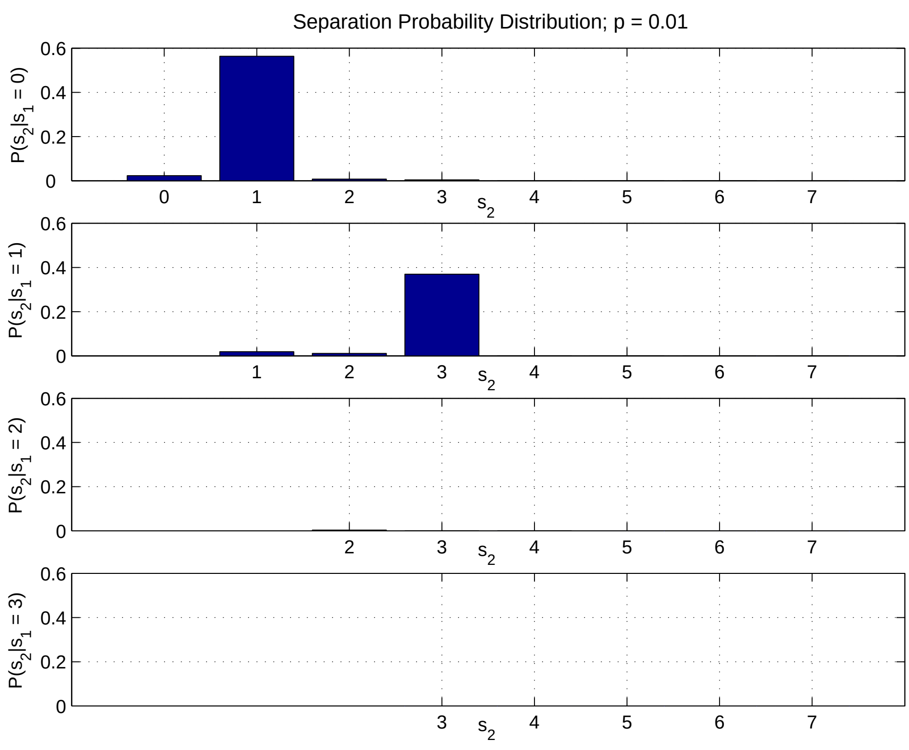

Figure 6-29: Three-agent problem with short tunnel-shaped local observation zones. The steady state agent separation probability distribution for $p = 0.01$. From top to bottom, the figures represent the distribution for $s_1 = 0, 1, 2, 3$, respectively; the horizontal axis represents s_2 ($\geq s_1$).

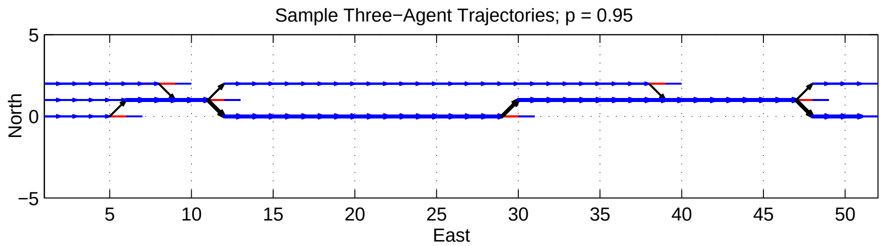

Figure 6-30: Three-agent problem with short tunnel-shaped local observation zones. A set of sample trajectories for $p = 0.95$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates an edge traversed without previous information on its cost, and a bold arrow indicates an edge traversed by more than one agent.

to spread, maximizing the number of new edges observed, thus increasing the probability of encountering an edge of zero cost. Note that at each stage two agents are at separation one, so that two out of three agents have information on three reachable edges rather than two, which is the case for the third agent, at separation three. The second pattern appears from horizontal position 39 to 48, where $s_1 = 0$ and $s_2 = 1$. In this case, all three agents have information on three rather than two reachable edges. This is better than in the first pattern, but comes at a price, namely at each stage only three new edges appear in the local observation zones, rather than six in the case of the first pattern. Different initial conditions lead to the different patterns. On average, the first and the second pattern occur at about 40% and 60% of the stages, respectively, as shown in Fig. 6-29. Finally note that this strategy pays off at four occasions in Fig. 6-28. In particular, at horizontal locations 16, 32, 38, and 49, at least two agents traverse an edge of cost zero, previously observed by only one agent. At these occasions, cooperation pays off.

Finally, we study the case where p is close to one. Fig. 6-30 shows an example set of three-agent trajectories under an optimal policy, for $p = 0.95$. After effects due to initial conditions have died out, the same pattern emerges as in the two-agent case with both cone-shaped and short tunnel-shaped local observation zones. Indeed, two out of three agents are at separation zero at all stages (in steady state), and we have a separation sequence as in the previously cited two-agent cases. The fact that $s_1 = 0$ at all stages reappears in Fig. 6-31 depicting the steady state separation probability distribution, for $p = 0.95$. It also appears in Fig. 6-27 where for p close to zero, we observe that the average separation $s_1 = 0$. For the description of the mechanics of cooperation in this case, we refer to the last case discussed in Section 6.4.1, the two-agent problem with cone-shaped local information zones, for p close to one.

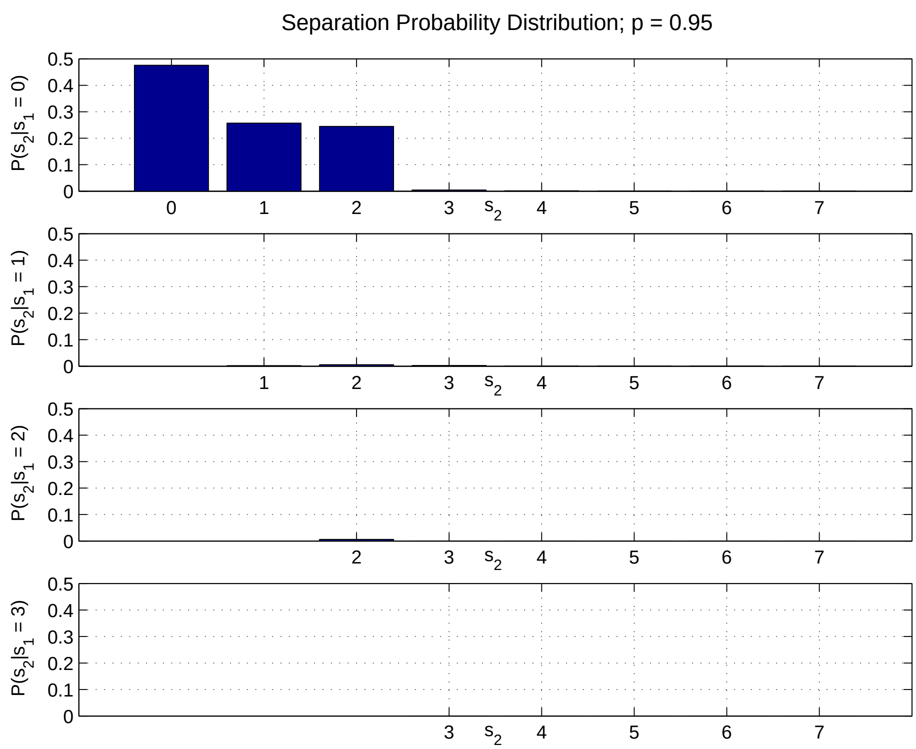

Figure 6-31: Three-agent problem with short tunnel-shaped local observation zones. The steady state agent separation probability distribution for $p = 0.95$. From top to bottom, the figures represent the distribution for $s_1 = 0, 1, 2, 3$, respectively; the horizontal axis represents s_2 ($\geq s_1$).

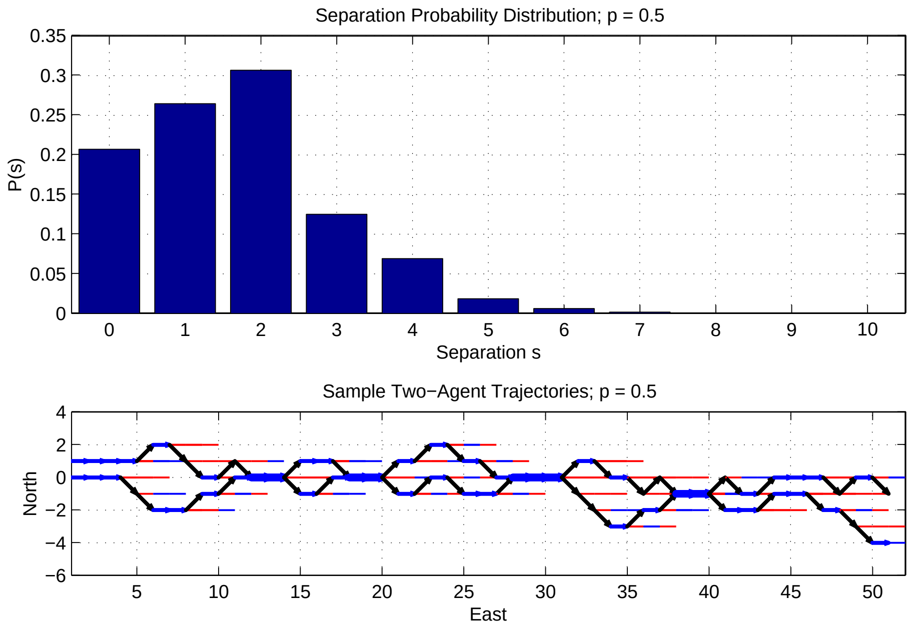

Figure 6-32: Two-agent problem with extended tunnel-shaped local observation zones. Top: Steady state agent separation probability distribution for a $p = 0.5$, under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.5$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates a traversed edge with no previously known information, a bold arrow indicates an edge traversed by two agents.

6.4.3 Two Agents with Extended Tunnel-Shaped local Observation Zones

In this section, we study the behavior of two agents with extended tunnel-shaped local observation zones. In particular, Fig. 6-32 (bottom) shows an example of a two-agent trajectory under an optimal policy for $p = 0.5$, and Fig. 6-32 (top) shows the corresponding steady state separation probability distribution. From the exposition in Section 5.2, we have that the steady state separation probability decreases exponentially with the separation s. Hence, for $p = 0.5$, with a probability of about 99%, the agents remain within a separation of five.

Fig. 6-33 shows the expected agent separation in steady state and under an optimal policy for a set of values for p. As in the cone-shaped and short tunnel-shaped local observation zone cases, the tendency to spread is more dominant as p decreases, and the edges of cost one are more probable than edges of cost zero. Similarly, the tendency to converge dominates for $p > 0.5$, the case where the environment mainly consists of edges of zero cost. As also demonstrated in the problems with cone-shaped and short tunnel-shaped local observation zones, we confirm and quantify the intuition that as rewards (edges with

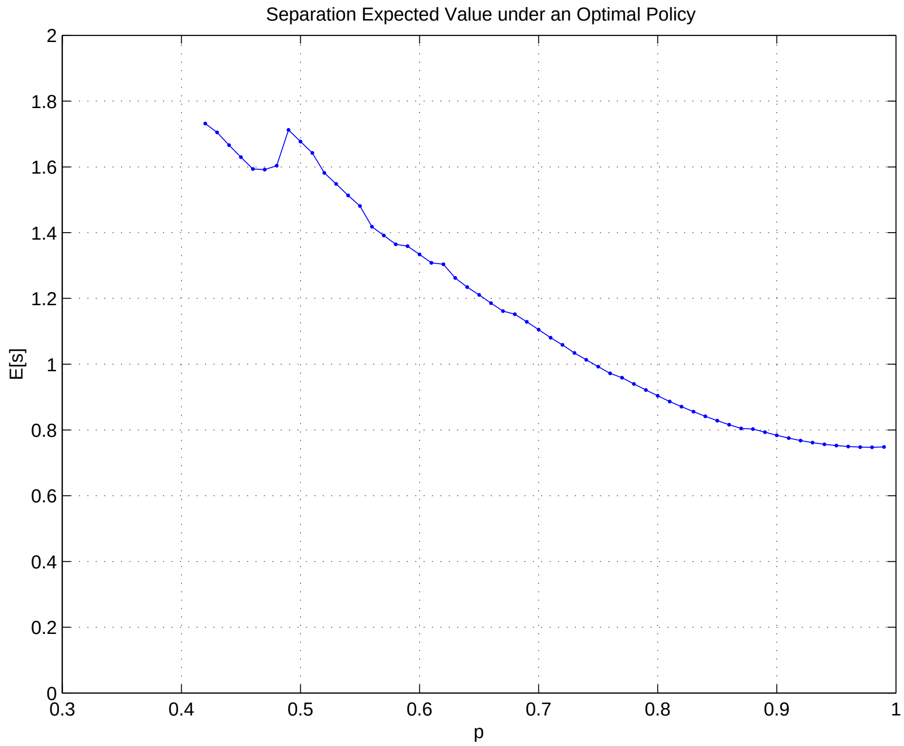

Figure 6-33: Two-agent problem with extended tunnel-shaped local observation zones. The expected value of the agent separation under an optimal policy for several values of p.

zero cost) are abundant (p approaching one), it is not necessary for the agents to spread and increase the size of the observed area. We note that in the limit, as p approaches one, the average separation equals 0.75, illustrated more in detail later in this section. Finally, the apparent discontinuities in Fig. 6-33 reflect policy changes as p varies.

We now study the extreme case where p is close to one. In particular, Fig. 6-34 (bottom) shows an example of a set of two agent trajectories under an optimal policy for $p = 0.95$. Fig. 6-34 (top) shows the associated steady state separation probability distribution. We observe identical behavior as in the two-agent problems with cone-shaped and short tunnel-shaped local observation zones. For a description of the mechanics of cooperation for this case, we refer to the last case discussed in Section 6.4.1, where the two-agent problem with cone-shaped local observation zones is discussed.

Finally, we discuss the extreme case where p is close to zero. In particular, Fig. 6-35 (bottom) shows an example of a set of two trajectories under an optimal policy for $p = 0.05$. Fig. 6-35 (top) shows the associated steady state separation probability distribution. The results presented for $p = 0.05$ are approximate. That is, we compute an upper bound and

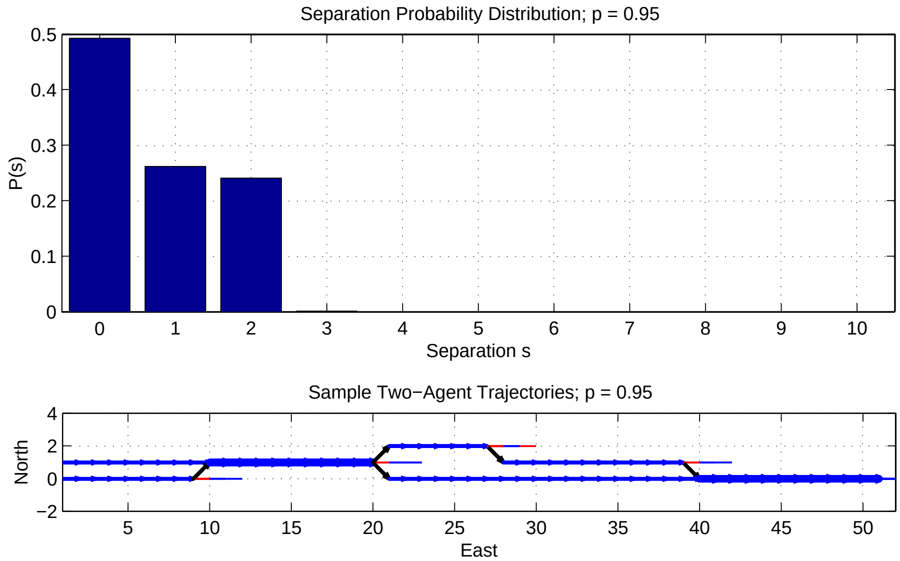

Figure 6-34: Two-agent problem with extended tunnel-shaped local observation zones. Top: Steady state agent separation probability distribution for a $p = 0.95$, under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.95$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates a traversed edge with no previously known information, a bold arrow indicates an edge traversed by two agents.

a lower bound to the optimal value function by considering a bounded and a cylindrically shaped navigation graph, respectively. If the upper and lower bound are sufficiently tight, here of the order of 10^{-8}, we can obtain the optimal policy at small separations. For $p = 0.05$, we find that the optimal two-agent policy is such that the agent separation is upper bounded by four, *i.e.* under an optimal policy the agents remain at all times in steady state within a separation of four. Note that for $p = 0.5$ (see Fig. 6-35) the agent separation is potentially unbounded under an optimal policy. Hence, there exists a $0.05 \leq \bar{p} \leq 0.5$ at which the optimal policy changes. Specifically, for $p < \bar{p}$, the optimal policy is such that the agent separation is upper bounded, and for $p > \bar{p}$, there is no separation upper bound. Numerical experiments indicate that computing $\bar{p}$ by finding an upper and lower bound to the optimal value function, is extremely sensitive to the choice of the width of the graph for the upper bound, and the diameter of the cylindrical graph for the lower bound. The algorithm to compute the optimal value function presented in this thesis is necessary and, hence, we present no approximate results for $p < 0.42$.

We conclude with some remarks on the mechanics of cooperation for the case where p is close to zero. In Fig. 6-35 (bottom), we notice two different modes. First, from horizontal coordinates 8 to 17, from 24 to 28, and from 38 to 41, the agents move diagonally at

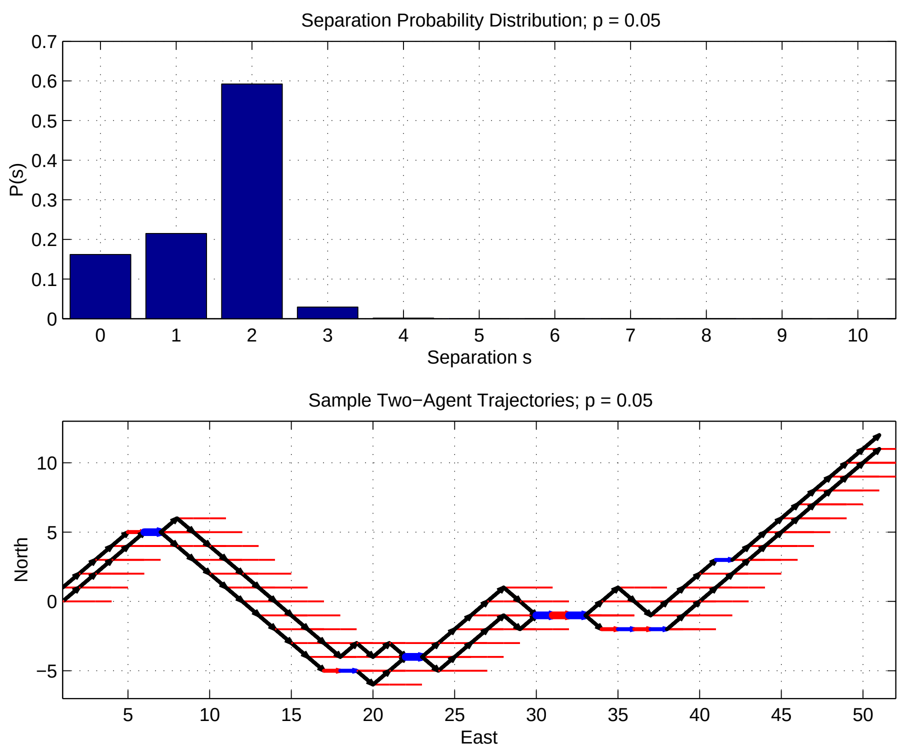

Figure 6-35: Two-agent problem with extended tunnel-shaped local observation zones. Top: Steady state agent separation probability distribution for a $p = 0.05$, under an optimal policy. Bottom: a set of sample trajectories, for $p = 0.05$. Red (blue) arrows and lines indicate traversed and observed edges of cost one (zero), respectively. A black arrow indicates a traversed edge with no previously known information, a bold arrow indicates an edge traversed by two agents.

separation two. In this mode, at each step, five previously unobserved edges enter the local observation zones. Further, at separation two, the third edge straight ahead agent A observes is reachable, and thus relevant to agent B, and vice versa. In the second mode, from horizontal coordinates 1 to 5 and 42 to 51, the agents move diagonally at separation one. In this mode, at each step, four previously unobserved edges enter the local observation zones. Further, at separation one, the second and third edge straight ahead agent A observes are reachable and hence relevant to agent B, and vice versa. The first and second mode occur with a probability of about 0.6 and 0.2, respectively. Both modes indicate a balance between the number of newly observed edges at each stage, and the number of edges one agent observes that are relevant to the other agent. In the first mode, five new edges are observed at each stage, and one edge observed by one agent is reachable by the other agent; in the second mode, only four new edges are observed at each stage, while two edges observed by one agent are relevant to the other agent. Finally, with probability 0.2, the agents are at separation zero, where both agents traverse one or more edges of costs zero, in general discovered while in one of the two modes. This occurs at horizontal coordinates 6, 22 and 30.

6.5 Implications for Communication

In this section, we discuss the requirements for agent communication under an optimal policy. We assume a distributed system whereby each agent has access to a computation unit with the optimal value function in its memory. Then, for each agent to determine its optimal decision, it generally requires access to the complete system state, including the current agent separation and the costs of the edges the other agent(s) observes. In case multiple decisions are optimal, communication is required to reach a consensus. This scheme is equivalent to centralized decision making.

However, consider a two-agent system. It is intuitively clear that when the agent separation is large, there is no need to communicate the costs of the observed edges. Indeed, edge costs agent A observes is of no use to agent B since the latter cannot reach and traverse the edges the former observes. Further, from the algorithm presented in this thesis, we have that for $s \geq \bar{s}$, each agent uses a single-agent optimal policy, reducing the agent separation in case multiple decisions are optimal (policy μ_∞). Hence, for $s \geq \bar{s}$, only information on

p	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	···	0.99
$\bar{s}$	11	9	7	6	5	5	4	3	···	3

Figure 6-36: The value of $\bar{s}$ for $p = 0.01, 0.02, \ldots, 0.99$, for the two-agent problem with cone-shaped local observation zones.

the relative position of the other agent is sufficient for each agent to determine its optimal decision. In other words, agent A only needs information on whether agent B is on its left or right. We denominate this as a *quasi-decentralized* optimal policy, since each agent behaves independently as a single agent under an optimal single-agent policy, breaking ties by choosing the decision that minimizes the resulting agent separation. On the other hand, for $s < \bar{s}$, the optimal agent policy differs from the combination of two optimal single-agent policies. Communication is required to convert the distributed computing into a *de facto* centralized decision making scheme.

In what follows, we discuss the size of $\bar{s}$ for the different multi-agent systems considered in this thesis, including the three-agent system, where s_2 plays the role of the separation s between the agents of a two-agent system.

Two Agents: Cone-Shaped Local Observation

We start with the two-agent problem with cone-shaped local observation zones. The table in Fig. 6-36 shows $\bar{s}$ for $p = 0.01, 0.02, \ldots, 0.99$. The separation $\bar{s}$ equals three for $p \geq 0.08$ at the values of p for which $\bar{s}$ is determined. For $p < 0.08$, we have that $\bar{s}$ increases as p decreases. Hence, we conclude that for $p \geq 0.08$, communication and thus centralized decision making, is only required when the agents are at separation $s \leq 2$; for smaller p centralized decision is required until somewhat larger separations.

We now give an intuitive explanation for the increase in $\bar{s}$ for small p by comparing the policy μ_∞ to the optimal two-agent policy at $3 \leq s < \bar{s}$ in the cases where $\bar{s} > 3$. An example of a state ($s > 3$) where there is a difference between μ_∞ and μ^* is shown in Fig. 6-37, where blue and red arrows or lines indicate an edge cost of zero and one, respectively. Further, the gray arrows indicate the most probable future agent decisions, given the high probability for encountering ones (only for $p < 0.08$ does it occur that $\bar{s} > 3$). The optimal decision for the bottom agent under an optimal two-agent policy, given the current separation and observation zones, is the same as under policy μ_∞. However, for the top agent, the solid arrows represent the optimal two-agent decision sequence, as opposed

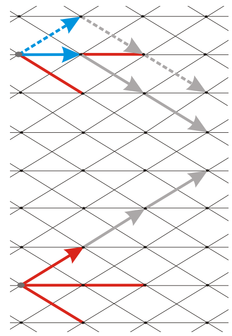

Figure 6-37: For any p such that $\bar{s} > 3$, the optimal policy at $s < \bar{s}$ (solid arrows) differs from a single-agent optimal policy (dashed arrows). A red arrow or line indicates an edge of cost one, while a blue arrow or line represents a zero edge cost.

to the dashed arrows, which represent the decision sequence the top agent were to make if it adopted policy μ_∞. Under μ^*, the agents reach a separation of one, three stages from the current stage, while under μ_∞, the agents reach separation two. Separation one is more favorable compared to separation two, particularly when the probability of encountering zeros is low, since at separation one, each agent potentially can leverage opportunities the other agent observes. Hence, odd separations are preferred over even ones, increasingly as the environment contains less edges of cost zero. Therefore, the smaller p, the larger the separations at which the agents anticipate, choosing those decisions that lead as fast as possible to a separation of one.

Two Agents: Short Tunnel-Shaped Local Observation

Next, we discuss the two-agent case with short tunnel-shaped local observation zones. From Section 4.3, we have that in steady state and under an optimal policy, the agent separation is bounded by a separation of three. Therefore, in steady state, the agents communicate at all times, thus employing a centralized decision making scheme. In transient state, with an initial agent separation greater than three, again a quasi-decentralized organization suffices.

Two Agents: Extended Tunnel-Shaped Local Observation

The table in Fig. 6-38 shows $\bar{s}$ for the two-agent problem with extended tunnel-shaped local observation, for $p = 0.42, 0.43, \ldots, 0.99$. Note that, as outlined in Section 6.2.3, we only compute exact information for the latter values of p. A similar conclusion as in the two-agent problem with short tunnel-shaped local observation zones holds: a quasi-decentralized

p	0.42	0.43	···	0.99
$\bar{s}$	3	3	···	3

Figure 6-38: The value of $\bar{s}$ for $p = 0.42, 0.43, \ldots, 0.99$, for the two-agent problem with extended tunnel-shaped local observation zones.

p	0.39	0.40	···	0.49	0.50	0.51	···	0.99
$\bar{s}$	7	5	···	5	4	5	···	5

Figure 6-39: The value of $\bar{s}$ for $p = 0.39, 0.40, \ldots, 0.99$, for the three-agent problem with short tunnel-shaped local observation zones.

decision scheme suffices when the agent separation $s \leq 3$. At the latter separations, the locally observed edge costs and the direction of decreasing separations suffice for optimal decision making.

Three Agents: Short Tunnel-Shaped Local Observation

Lastly, we consider the three-agent problem with short tunnel-shaped local observation zones. In Section 6.2.2, we assume that the smaller agent separation $s_1 \leq 3$, while computations show that the larger agent separation s_2 can grow unbounded. Hence, the two closer agents communicate at each decision step in steady state. The communication studied here, is between the group of two close agents and the third agent. The traditional $\bar{s}$ represents the separation that is such that for $s_2 \geq \bar{s}$, the three-agent group behaves as an agent pair employing a two-agent optimal policy, and a single agent, using a single-agent optimal policy. In particular, the table in Fig. 6-39 shows $\bar{s}$ for $p = 0.39, 0.40, \ldots, 0.99$. Again, from Section 6.2.2, we describe that we obtain exact results only at the latter values of p. Similarly as before, we notice that a quasi-decentralized strategy suffices for separations s_2 larger than a fairly small lower bound ($\bar{s}$). From Section 6.2.2, we have that the optimal policy at large separations s_2 changes at least once for $p \in [0.38, 0.39]$. We observe that $\bar{s}$ is larger for $p = 0.39$ than at larger values of p. We attribute this phenomenon, without further study, to the influence of the presence of some $p \in [0.38, 0.39]$ where two policies are optimal. Lastly, we notice that at $p = 0.50$, the separation $\bar{s} = 4$, different from its values for smaller and greater values of p. Without going in further detail, we mention that this occurrence is perfectly possible due to two or more policy changes at small separations for $p \in [0.49, 0.51]$.

6.6 Chapter Summary

In this chapter, we analyze and compare the performance of several multi-agent problems. In particular, we compute the optimal value function and corresponding optimal policy, and use the latter to determine the system state probability distribution in steady state.

First, we compare the multi-agent performance for groups with one, two and three agents and for the cone-shaped, the short and extended tunnel-shaped local observation zones. We confirm and quantify the intuition that more agents perform better. Two conclusions stand out: two agents with cone-shaped local information reduce the average locally incurred cost per agent by 25% in a mine field type environment, *i.e.* in the limit for p approaching one, when the environment contains scarce, though high costs. Second, in the short tunnel-shaped local observation zone case, three agents provide a performance increase of maximum 10% as compared to two agents. The latter performance benefit decreases rapidly to zero for $p > 0.8$, indicating that two agents suffice, particularly in a mine field type environment.

Then, we compare the multi-agent performance for various local observation zones, in the single- and two-agent cases. Both offer similar results. An extra observed edge cost straight ahead (extended vs. short tunnel-shaped local observation zones) provides a limited performance increase of maximum 10%, decreasing to zero for p approaching both zero and one. However, adding diagonal edges provides a much larger performance increase. Indeed, the performance increase ranges from 0% to 100% as p goes from zero to one.

Then, we study particular properties of the agent separation in steady state for the different problem instances. In general, we notice that the expected value of the agent separation increases as p decreases, *i.e.* as the environment contains more edges of cost one. This confirms and quantifies the intuition that as zeros become scarcer, the agents spread out to gather information on a larger local environment area. Further, for the two-agent problem with short tunnel-shaped local information, the agent separation is upper bounded by three in steady state and under an optimal policy, for any relevant value of p. Specific patterns arise in the optimal trajectories for the extreme cases where p approaches zero and one.

Finally, the study of optimal policies reveals that for the agent separation $s \geq \bar{s}$, the agents adopt single-agent optimal policies in the two-agent cases, rendering the communication of locally observed environment information unnecessary. Since for $s < \bar{s}$, communication of locally observed information is necessary to compute optimal decisions, we denote the policy as *quasi*-decentralized.

Chapter 7

Conclusions

We present a thesis summary with concluding remarks in Section 7.1, and make suggestions for continuing research efforts in Section 7.2.

7.1 Summary

Multi-agent systems are in general believed to be more efficient, robust and versatile than single agents. We distinguish two types of multi-agent problems: collaborative search for globally and individually rewarding resources. In this thesis, we focus on the version with individually rewarding resources where more than one agent can incur the same reward. We find an example of this scheme in the multi-agent navigation problem through a threat cluttered environment where it is potentially beneficial for more than one agent to choose the same safe path segment. Map building of a previously unknown environment, is an example of the search for globally rewarding resources: indeed, it is generally of no benefit for a second agent to map the same environment area.

We identify four key issues in multi-agent system design in the context of the search for individually rewarding resources, addressed to some extent in this thesis. First and foremost, it is of central importance to determine the optimal agent group composition (possibly heterogeneous, with agents of different capabilities) for the problem at hand. This involves a trade-off between cost and performance increase per added agent. Then, truly autonomous systems solely rely on on-board environment measurements. Here, there is trade-off between the cost and benefits of observations, calling for a study that quantifies the performance changes for different locally observed areas. Third, agents share observations, key to efficient cooperation. However, not all information is of equal importance, possibly reducing communication requirements. Finally, the system robustness is highly enhanced if a decentralized decision scheme is employed. It is of interest to study under which circumstances a decentralized scheme allows for optimal multi-agent strategies (Chapter 1).

The vehicle of our analysis is a multi-agent target acquisition problem on a structured unbounded, partially unknown graph. We solve several instances of this problem for various group sizes, and local observation zones, in order to rigorously address the issues presented in the previous paragraph. In particular, we consider three types of local observation zones, focusing on the impact of different local areas around the agent positions on the agent performance: short and extended tunnel-shaped and cone-shaped local observation zones. With the former two, we study the effect of observations in the general navigation direction, while the latter is employed to study the benefits of lateral information. Further, we restrict the agent motion to be spatially synchronous, leaving the lateral agent separation as the only controlled degree of freedom (Chapter 2).

We divide the rest of the summary in two parts. In the first part, we summarize the development of the algorithm that computes optimal multi-agent policies, the main technical thesis contribution. In the second part, we present analysis results of the graph traversal problems posed.

Development of Policy Design and Analysis Tools

Considering an infinitely long and wide graph ensures problem spatial invariance in both the lateral and longitudinal direction, resulting in a computationally simpler problem. An additional advantage is that the unattractive effects of boundary conditions in the form of a graph of finite size are avoided. Insight regarding the agent behavior at large agent separations allows for the development of an efficient solution method for this problem with an infinite state space. In particular, at infinite, or very large separations, each resulting agent group behaves independently of each other. For example, the two agents of a two-agent group adopt optimal single-agent strategies whenever their separation is large (Chapter 2).

We formulate the problem as a Dynamic Programming, but use the Linear Programming equivalent for our analysis. In particular, we decompose the LP with infinitely many constraints into an infinite set of LPs with finitely many constraints and prove that the unique bounded function that simultaneously solves the coupled LPs in the latter set represents the problem optimal value function. In this manner, we convert solving one LP with infinitely many constraints into solving infinitely many LPs with finitely many constraints each, formulated as finding the unique bounded fixed point of a properly defined function F (Chapter 3).

For the two-agent problem with cone-shaped local observation zones, we present a procedure to compute the fixed point of the function F explicitly for $s < \bar{s}$ and implicitly in the form of a recursion for $s \geq \bar{s}$. Here, $\bar{s}$ is defined as the smallest separation where the optimal two-agent policy consists of a combination of two single-agent policies. For $s \geq \bar{s}$, we use the insight regarding the agent behavior at large separations to express the evolution of the optimal value function as an autonomous Linear Time-Invariant system. For $s < \bar{s}$, we formulate an LP, denoted $\mathcal{LP}_{in}$, with finitely many constraints and whose solution determines the optimal value function for $s < \bar{s}$. However, the two parts of the state space need proper connection. Indeed, we establish that $\mathcal{LP}_{in}$ provides the optimal value function for $s < \bar{s}$ if supplemented with a set of *connection* constraints that guarantees that the initial state of the LTI system excites only the stable system modes. Finally, we define the output of the LTI system such that its positivity indicates the correct choice of $\bar{s}$, and satisfaction of the fixed point requirement for $s \geq \bar{s}$. We employ this method as well for the two-agent problem with extended tunnel-shaped local observation zones and for the three-agent problem with short tunnel-shaped local observation zones. For the two-agent version of the latter problem, a coupling argument establishes an upper bound to the agent separation under an optimal policy and in steady state, rendering the previous procedure unnecessary in this particular case (Chapter 4).

Finally, for the analysis of optimal policy properties, it is essential to compute the state probability distribution in steady state and under an optimal policy. For the two-agent problem with cone-shaped local observation zones, we present a procedure to this end, in fact computing the eigenvector corresponding to the unit eigenvalue of an infinitely large state transition matrix. In particular, we exploit the fact that for $s \geq \bar{s}$, the optimal two-agent policy is stationary (with s) and formulate an autonomous LTI system with the steady state probability at separation s as LTI state. The requirements that only stable system modes are excited, and that the sum of all probabilities equals one, allow for the formulation of a set of equations which provides as solution the steady state probabilities of the states with $s < \bar{s}$. The two-agent problem with short tunnel-shaped local observation zones is straightforward due to the separation upper bound in steady state. Further, we present peculiarities of this method for the two- and three-agent problems with extended and short tunnel-shaped local observation zones, respectively (Chapter 5).

Results of Analysis

First, we discuss the performance as a function of the group size for different problem instances. We confirm and quantify the intuition that the larger the agent group, the better the performance. Two observations stand out. First, for cone-shaped local information and in a mine field type environment where costs are scarce though high, two agents reduce the average locally incurred cost per agent by 25% as compared to a single agent. Second, in the short tunnel-shaped observation zone case, three agents provide a performance increase of less than 2%, decreasing fast to zero for $p \geq 0.8$, where p denotes the probability of encountering a zero edge cost; this indicates that a third agent is unnecessary, especially as the environment becomes more mine field like.

Second, we address the performance change for the different local observation zones, in the single- and two-agent cases. Both offer similar results. An extra observed edge cost straight ahead (short vs. extended tunnel-shaped local observation zones) provides a decrease of the average locally incurred cost per agent of maximum 10%, decreasing to zero as p approaches both zero and one. However, adding diagonal edges to the local observation zones provides a much larger benefit. Indeed, the performance increase ranges from 0% to 100%, roughly linearly, as p goes from zero to one. Intuition is confirmed and, more importantly, we quantify the performance increases exactly.

Then, we study particular properties of the agent separation in steady state for the different problem instances. In general, we notice that the expected value of the agent separation increases as p decreases, *i.e.* as the environment contains more edges of cost one. This confirms and quantifies the intuition that as zeros become scarcer, the agents spread out to gather information on a larger local environment area. Further, for the two-agent problem with short tunnel-shaped local information, the agent separation is upper bounded by three in steady state and under an optimal policy, for any relevant value of p. For the problems considered, the agent separation is smaller than five with a probability of approximately 90%, its exact value depending on the problem specifics, and with an exponentially decreasing probability (with varying speeds) as the agent separation increases.

Further, specific patterns arise in the optimal trajectories for the extreme cases where p approaches zero and one.

Finally, the study of optimal policies reveals that for the agent separation $s \geq \bar{s}$, the agents adopt single-agent optimal policies in the two-agent cases, rendering the communication of locally observed environment information unnecessary. For $s \geq \bar{s}$, it is sufficient for each agent to have information on the direction of the location of the other agent. Again, this is intuitively fairly straightforward, but we quantify exactly beyond which separation communication is of no use. Since for $s < \bar{s}$, communication of locally observed information is necessary to compute optimal decisions, we denote the policy as *quasi*-decentralized (Chapter 6).

7.2 Future Work

In this section, we suggest directions for future research. Some suggestions are rather technical, others are extensions that involve relaxing some of the constraints introduced for simplification.

7.2.1 Two Agents with Cone-Shaped Local Information: Limit as p Approaches One

In Section 6.2.1, we introduce the agent performance as a function of parameter p for the single- and two-agent cases with cone-shaped local observation zones. In the limit as $\varepsilon = 1 - p$ approaches zero, the performance increase of the two-agent group as compared to a single agent becomes numerically cumbersome to compute since the optimal value function at some states is of the order of magnitude of the LP-solver accuracy. However, we believe it is possible to compute the limit of $r^c_{1,2}$ explicitly in analytic form to show that $r^c_{1,2}$ converges to 0.25, or a 25% performance increase, as ε approaches zero, *i.e.* in a mine-field type environment.

One way to establish the result is to follow the development in the Chapters 4-5, neglecting higher order terms in ε in summations. In particular, one needs to compute the optimal value function for $\varepsilon << 1$, and establish the optimal two-agentpolicy in the limit. Then, it is necessary to determine the state probability distribution in steady state and under an optimal policy. Two methods seem possible: one can argue that the probability decreases exponentially with s and hence establish that artificially upper bounding the agent separation for $\varepsilon << 1$ yields the proper result; or, one can determine the recursion for the state probability as s is large and formulate the set of equations which yields the optimal state probability distribution at small separations. This information suffices to compute $r^c_{1,2}$ as ε approaches zero.

7.2.2 Three Agents with Short Tunnel-Shaped Local Observation Zones: Separation Upper Bound

For the three-agent problem with short tunnel-shaped local observation zones, we assume that under an optimal policy, the two closest agent are within separation three at all times. This assumption is partly based on the agent behavior when two agents are separated by a large distance from the third agent. In this situation, the two agents adopt a two-agent optimal policy, while the third agent navigates under a single-agent optimal policy. Two agents under an optimal policy do not reach a separation larger than three, motivating in part our assumption, confirmed by numerical experiments, though not established rigorously.

A procedure that proves the assumption correctness could involve the following. As in Section 4.5, we compute the candidate optimal value function at the separations $0 \leq s_1 \leq 3$ and $s_2 = s_1, s_1 + 1, \ldots$, where s_1 and s_2 denote the two separations between three agents. Here, we assume that $s_1 \leq s_2$, without loss of generality, and that $s_1 \leq 3$ under an optimal policy. Then, it is necessary to formulate a set of autonomous LTI systems, which sequentially compute the candidate optimal value function for $3 < s_1 \leq s_2$, one LTI system for each $s_2 = 3, 4, \ldots$. The output of each LTI system can be defined such that its positivity indicates that the computed value function is indeed a fixed point of F and hence is optimal. How to precisely formulate the set of LTI systems in order to provide a token for optimality, remains open.

7.2.3 Extension to Larger Agent Groups

On a more general note, we provide some comments regarding the natural extension to larger agent groups. Remember that we solve the two-agent problem by relying on the solution of the single-agent problem for $s \geq \bar{s}$. Similarly, we believe it is possible to solve the N-agent case, by solving a relatively small $\mathcal{LP}_{in}$-type LP with as solution the optimal N-agent value function where the maximum separation between two neighboring agents equals

$\bar{s}$. To $\mathcal{LP}_{in}$ are added a set of constraints so that for each way the agent cluster can split, only the stable modes are excited in the corresponding LTI-system. These LTI-systems are determined from the optimal agent behavior of cluster sizes $N-1$ and smaller, providing a hierarchical solution method whereby the N-agent solution relies on the solution of the problems involving agent groups of smaller size. Although the idea seems attractive, its implementation might prove cumbersome. We should note here, that even for three and four agent problems, the size of $\mathcal{LP}_{in}$ can become large, calling for approximate methods [7]. The results presented in this thesis translate into much needed heuristics, improving the quality of the approximations. An initial step towards this is taken in Section 6.2.4, where a large agent group with short tunnel-shaped local information is split into a set of smaller groups, each operating independently under an optimal policy. We estimate the performance loss. We cannot apply this to all types of local observation zones and environment characteristics. Therefore, we envision to solve $\mathcal{LP}_{in}$ approximately, for example using the constraint sampling method in [61], providing initial conditions to the respective LTI systems and a set of approximate recursions.

7.2.4 Spatial Asynchrony and Time-Varying Environments

A second natural extension involves removing the spatial synchrony constraint, allowing a longitudinal agent separation. For the case of time independent edge costs, the best multi-agent strategy seems for each agent to traverse the graph sequentially. That is, the first agent traverses the graph, observing edge costs on its way, until reaching the target. Then, the second agent traverses the graph, observing extra edges but exploiting the observations of the first agent, etc. However, in a time dependent graph, matters change drastically. For example, we can model the variation in time of each edge cost by a simple Markov chain that is such that if an edge cost is observed at time zero, the reliability of that information decreases with time. We envision a trade-off between two competing tendencies that leads to an equilibrium longitudinal agent separation: for a larger longitudinal separation, the trailing agent can enjoy information on more edges, but that information is less reliable. Simplified versions of the relaxation to spatial asynchrony, could involve fixing the longitudinal agent separation at different values and analyzing the group performance.

7.2.5 Non-Homogeneous Local Observation Zones and Irregular Graphs

An interesting avenue for future exploration concerns considering non-homogeneous local observation zone problems. Here, each agent observes different local areas of the environment. For example, one could consider a two-agentproblem with cone-shaped and extended tunnel-shaped local observation zones, respectively. This problem extension leans closely to the work presented in this thesis and does not require major changes to the approach. Considering agents with differing local observation zones seems to model groups where agents have different capabilities. It is of interest to study whether the differing capabilities naturally lead to an optimal policy where each agent takes on a particular role in the group.

Finally, in order to model real environments more accurately, it is necessary to consider irregular graphs rather then the structured ones considered here, which provide advantages regarding the computational complexity. Rather than graph invariance in one or more directions, we could envision graphs where certain statistical properties are invariant in some directions. For example, the probability distribution of the number of edges leaving a particular vertex is invariant. This extension, in contrary to the previous one, seems rather more involved.

Appendix A

Proofs

Lemma 19 *The inequality in Eq. (4.45), namely*

$$-P_6'(p,\alpha) > 0, \tag{A.1}$$

holds for $0 < p, \alpha < 1$. $\square$

Proof: Let $C_0(p,\alpha) = -P_6'(p,\alpha)$. Then, we have that

$$C_0(p,\alpha) = 1 + \sum_{i=1}^{3} C_{0,i}(p)\alpha^i,$$

with $-P_6'(p,\alpha) = C_0(p,\alpha)$, and with

$$\begin{aligned}
C_{0,1}(p) &= -1 + 5p - 14p^2 + 28p^3 - 47p^4 + 62p^5 - 54p^6 + 24p^7 - 4p^8 \\
C_{0,2}(p) &= p(-1 + 10p - 53p^2 + 186p^3 - 483p^4 + 993p^5 - 1676p^6 + 2316p^7 - \ldots \\
&\quad 2522p^8 + 2042p^9 - 1152p^{10} + 420p^{11} - 88p^{12} + 8p^{13}) \\
C_{0,3}(p) &= -p^3(-1 + 5p - 14p^2 + 18p^3 - 10p^4 + 2p^5)^3.
\end{aligned}$$

For Eq. (A.1) to hold, it is sufficient to prove that

$$D_\alpha C_0(p,\alpha) < 0, \text{ for } 0 < p, \alpha < 1, \tag{A.2}$$

where D_α is an operator which represents the first derivative of its argument to parameter

α, and that

$$C_0(p,1) > 0 \text{ for } 0 < p < 1. \tag{A.3}$$

For Eq. (A.3), we have that

$$C_0(p,1) = -4pC_{0,f}(p),$$

and negativity of $C_{0,f}(p)$ for $0 < p < 1$ is assured by noticing that $C_{0,f}(1) = 0$ and by proving that $D_pC_{0,f}(p) > 0$ for $0 < p < 1$. We have that,

$$\begin{aligned} D_pC_{0,f}(p) &= 1 + 12p - 93p^2 + 304p^3 - 425p^4 - 834p^5 + 6643p^6 - 20808p^7 + \ldots \\ &\quad 42750p^8 - 62730p^9 + 67122p^{10} - 52404p^{11} + 29497p^{12} - \ldots \\ &\quad 11648p^{13} + 3060p^{14} - 480p^{15} + 34p^{16}. \end{aligned}$$

To show negativity of $D_pC_{0,f}(p)$ for $0 < p < 1$, we find a *Sum of Squares* (SOS) auxiliary polynomial $C^{p,a}_{0,f}(p)$ such that

$$C^{p,t}_{0,f}(p) = -D_pC_{0,f}(p) + p(p-1)C^{p,a}_{0,f}(p) > 0.$$

In particular, we require $C^{p,t}_{0,f}(p)$ to be SOS with a positive definite coefficient matrix to prove the strict inequality. Using the SOSTOOLS Sum of Squares Optimization Toolbox for Matlab [63], we have that for

$$C^{p,a}_{0,f}(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 19 & -33 \\ -33 & 60 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix},$$

where the coefficient matrix is positive definite, the polynomial $C^{p,t}_{0,f}(p)$, is SOS, with a positive definite coefficient matrix. In particular,

$$C^{p,t}_{0,f}(p) = (Z^{p,t}_{0,f})^T \begin{bmatrix} \mathbf{C}^{p,t,1}_{0,f} & \mathbf{C}^{p,t,2}_{0,f} \end{bmatrix} Z^{p,t}_{0,f},$$

where

$$Z_{0,f}^{p,t} = \begin{bmatrix} 1 & p & p^2 & p^3 & p^4 & p^5 & p^6 & p^7 & p^8 \end{bmatrix}^T,$$

with

$$\mathbf{C}_{0,f}^{p,t,1} = \begin{bmatrix} 1 & -\frac{7}{2} & -\frac{2575}{192} & \frac{499}{8} & -\frac{1649}{16} & \frac{2023559}{20160} & -\frac{13821}{224} & \frac{1341}{64} \\ -\frac{7}{2} & \frac{1807}{96} & \frac{213}{8} & -\frac{463}{2} & \frac{29297}{64} & -\frac{211847}{448} & \frac{18315}{64} & -\frac{54323}{576} \\ -\frac{2575}{192} & \frac{213}{8} & \frac{2433}{8} & -\frac{62409}{64} & \frac{569829}{448} & -\frac{34739}{32} & \frac{390691}{576} & -\frac{15791}{64} \\ \frac{499}{8} & -\frac{463}{2} & -\frac{62409}{64} & \frac{289423}{56} & -\frac{308017}{32} & \frac{1428721}{144} & -\frac{192139}{32} & \frac{124399}{64} \\ -\frac{1649}{16} & \frac{29297}{64} & \frac{569829}{448} & -\frac{308017}{32} & \frac{438368839}{20160} & -\frac{402027}{16} & \frac{1009357}{64} & -\frac{322703}{64} \\ \frac{2023559}{20160} & -\frac{211847}{448} & -\frac{34739}{32} & \frac{1428721}{144} & -\frac{402027}{16} & \frac{505927}{16} & -\frac{1337931}{64} & \frac{1097823}{160} \\ -\frac{13821}{224} & \frac{18315}{64} & \frac{390691}{576} & -\frac{192139}{32} & \frac{1009357}{64} & -\frac{1337931}{64} & \frac{1159759}{80} & -\frac{316585}{64} \\ \frac{1341}{64} & -\frac{54323}{576} & -\frac{15791}{64} & \frac{124399}{64} & -\frac{322703}{64} & \frac{1097823}{160} & -\frac{316585}{64} & \frac{21067}{12} \\ -\frac{26}{9} & \frac{817}{64} & \frac{1147}{32} & -\frac{8147}{32} & \frac{51089}{80} & -\frac{56151}{64} & \frac{15653}{24} & -240 \end{bmatrix},$$

$$\mathbf{C}_{0,f}^{p,t,2} = \begin{bmatrix} -\frac{26}{9} & \frac{817}{64} & \frac{1147}{32} & -\frac{8147}{32} & \frac{51089}{80} & -\frac{56151}{64} & \frac{15653}{24} & -240 & 34 \end{bmatrix}^T,$$

where the coefficient matrix of $C_{0,f}^{p,t}(p)$ is positive definite, as required. Therefore, $C_{0,f}^{p,t}(p) > 0$ and the inequality in Eq. (A.3) holds.

We have left to show that the inequality in Eq. (A.2) holds. For this, it is sufficient to show that

$$D_\alpha^2 C_0(p,\alpha) < 0, \text{ for } 0 < p, \alpha < 1, \tag{A.4}$$

and that

$$D_\alpha C_0(p,0) < 0, \text{ for } 0 < p < 1. \tag{A.5}$$

For the inequality in Eq. (A.5), we find an auxiliary SOS polynomial $C_0^{\alpha,a}(p)$ such that

$$C_0^{\alpha,t}(p) = -D_\alpha C_0(p,0) + p(p-1)\, C_0^{\alpha,a}(p) > 0.$$

As earlier, we require $C_0^{\alpha,t}(p)$ to be SOS, with a positive definite coefficient matrix. In particular, we have that the previous requirements are satisfied for

$$C_0^{\alpha,a}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix}^T \begin{bmatrix} 2 & -3 & -\frac{1}{3} & 0 \\ -3 & \frac{26}{3} & -1 & -\frac{13}{3} \\ -\frac{1}{3} & -1 & \frac{14}{3} & -1 \\ 0 & -\frac{13}{3} & -1 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix},$$

where the coefficient matrix is positive definite. Then, $C_0^{\alpha,t}(p)$ is a SOS polynomial. Specifically,

$$C_0^{\alpha,t}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{7}{2} & \frac{23}{6} & -\frac{3}{2} & \frac{1}{2} \\ -\frac{7}{2} & \frac{43}{3} & -\frac{39}{2} & \frac{19}{2} & -\frac{3}{2} \\ \frac{23}{6} & -\frac{39}{2} & 37 & -\frac{57}{2} & 7 \\ -\frac{3}{2} & \frac{19}{2} & -\frac{57}{2} & 38 & -18 \\ \frac{1}{2} & -\frac{3}{2} & 7 & -18 & 14 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \end{bmatrix},$$

with indeed a positive definite coefficient matrix. Therefore, $C_0^{\alpha,t}(p) > 0$ and hence the equality in Eq. (A.5) holds.

We now focus on establishing the inequality in Eq. (A.4). We have that

$$D_\alpha^2 C_0(p,\alpha) = 2p\, C_{0,f_1}^{\alpha\alpha}(p)\, C_{0,f_2}^{\alpha\alpha}(p,\alpha),$$

with

$$\begin{aligned} C_{0,f_1}^{\alpha\alpha}(p) &= 1 - 5p + 14p^2 - 18p^3 + 10p^4 - 2p^5 \\ C_{0,f_2}^{\alpha\alpha}(p,\alpha) &= -1 + 5p + p^2(-14+3\alpha) + p^3(28-30\alpha) + p^4(-47+159\alpha) + \ldots \\ &\quad p^5(62-528\alpha) + 54p^6(-1+22\alpha) - 24p^7(-1+76\alpha) + \ldots \\ &\quad 4p^8(-1+468\alpha) - 1248p^9\alpha + 516p^{10}\alpha - 120p^{11}\alpha + 12p^{12}\alpha. \end{aligned}$$

It suffices to show that

$$C_{0,f_1}^{\alpha\alpha}(p) > 0 \text{ for } 0 < p < 1, \tag{A.6}$$

and that

$$C^{\alpha\alpha}_{0,f_2}(p,0) \quad < \quad 0, \text{ for } 0<p<1, \tag{A.7}$$

$$C^{\alpha\alpha}_{0,f_2}(p,1) \quad < \quad 0, \text{ for } 0<p<1, \tag{A.8}$$

since $C^{\alpha\alpha}_{0,f_2}(p,\alpha)$ is linear in α. We follow a similar procedure as before to establish the last three inequalities. Specifically, for the inequality in Eq. (A.6), we find a SOS auxiliary polynomial $C^{\alpha\alpha,a}_{0,f_1}(p)$ such that

$$C^{\alpha\alpha,t}_{0,f_1}(p) = C^{\alpha\alpha}_{0,f_1}(p) + p(p-1)\,C^{\alpha\alpha,a}_{0,f_1}(p) > 0.$$

For

$$C^{\alpha\alpha,a}_{0,f_1}(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 4 & -4 \\ -4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix},$$

with a positive definite coefficient matrix, we have that

$$C^{\alpha\alpha,t}_{0,f_1}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{5}{2} & \frac{5}{2} & -1 \\ -\frac{5}{2} & 9 & -10 & \frac{7}{2} \\ \frac{5}{2} & -10 & 15 & -\frac{15}{2} \\ -1 & \frac{7}{2} & -\frac{15}{2} & 5 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \end{bmatrix},$$

with a positive definite coefficient matrix, and thus $C^{\alpha\alpha,t}_{0,f_1}(p) > 0$, and hence the equality in Eq. (A.6) holds.

Next, for the inequality in Eq. (A.7), we find a SOS auxiliary polynomial $C^{\alpha\alpha,a}_{0,f_2,0}(p)$ such that

$$C^{\alpha\alpha,t}_{0,f_2,0}(p) = -C^{\alpha\alpha}_{0,f_2}(p,0) + p(p-1)\,C^{\alpha\alpha,a}_{0,f_2,0}(p) > 0.$$

This inequality is satisfied requiring $C^{\alpha\alpha,t}_{0,f_2,0}(p)$ to be SOS with a positive definite coefficient matrix. Specifically, we have that

$$C^{\alpha\alpha,a}_{0,f_2,0}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}^T \begin{bmatrix} 2 & -3 & -1 \\ -3 & 9 & -5 \\ -1 & -5 & 11 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix},$$

with a positive definite coefficient matrix, and

$$C^{\alpha\alpha,t}_{0,f_2,0}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{7}{2} & 3 & -\frac{3}{4} & \frac{1}{5} \\ -\frac{7}{2} & 16 & -\frac{79}{4} & \frac{46}{5} & -\frac{7}{4} \\ 3 & -\frac{79}{4} & \frac{226}{5} & -\frac{159}{4} & \frac{25}{3} \\ -\frac{3}{4} & \frac{46}{5} & -\frac{159}{4} & \frac{145}{3} & -12 \\ \frac{1}{5} & -\frac{7}{4} & \frac{25}{3} & -12 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \end{bmatrix},$$

with a positive definite coefficient matrix, guaranteeing that $C^{\alpha\alpha,t}_{0,f_2,0}(p) > 0$, and hence the inequality in Eq. (A.7).

Finally, for the inequality in Eq. (A.8), we find a SOS auxiliary polynomial $C^{\alpha\alpha,a}_{0,f_2,1}(p)$ such that

$$C^{\alpha\alpha,t}_{0,f_2,1}(p) = -C^{\alpha\alpha}_{0,f_2}(p,1) + p(p-1)\, C^{\alpha\alpha,a}_{0,f_2,1}(p) > 0.$$

This inequality is satisfied requiring $C^{\alpha\alpha,t}_{0,f_2,1}(p)$ to be SOS with a positive definite coefficient matrix. Specifically, we have that

$$C^{\alpha\alpha,a}_{0,f_2,1}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \\ p^5 \end{bmatrix}^T \begin{bmatrix} 8 & -23 & -5 & 29 & 21 & -\frac{85}{3} \\ -23 & 87 & -39 & -71 & -\frac{109}{3} & 86 \\ -5 & -39 & 176 & -\frac{520}{3} & 63 & -54 \\ 29 & -71 & -\frac{520}{3} & 596 & -547 & 213 \\ 21 & -\frac{109}{3} & 63 & -547 & 1209 & -781 \\ -\frac{85}{3} & 86 & -54 & 213 & -781 & 637 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \\ p^5 \end{bmatrix},$$

with a positive definite coefficient matrix, and

$$C^{\alpha\alpha,t}_{0,f_2,1}(p) = \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \\ p^5 \\ p^6 \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{13}{2} & 7 & \frac{53}{4} & -\frac{227}{10} & 4 & \frac{111}{28} \\ -\frac{13}{2} & 51 & -\frac{295}{4} & -\frac{371}{5} & \frac{769}{4} & -\frac{2311}{28} & -\frac{13}{2} \\ 7 & -\frac{295}{4} & \frac{894}{5} & -\frac{45}{4} & -\frac{10389}{28} & 374 & -104 \\ \frac{53}{4} & -\frac{371}{5} & -\frac{45}{4} & \frac{4441}{14} & -\frac{305}{2} & -\frac{1215}{4} & 212 \\ -\frac{227}{10} & \frac{769}{4} & -\frac{10389}{28} & -\frac{305}{2} & \frac{2087}{2} & -\frac{2013}{2} & \frac{1295}{4} \\ 4 & -\frac{2311}{28} & 374 & -\frac{1215}{4} & -\frac{2013}{2} & \frac{4067}{2} & -\frac{2079}{2} \\ \frac{111}{28} & -\frac{13}{2} & -104 & 212 & \frac{1295}{4} & -\frac{2079}{2} & 625 \end{bmatrix} \begin{bmatrix} 1 \\ p \\ p^2 \\ p^3 \\ p^4 \\ p^5 \\ p^6 \end{bmatrix},$$

with a positive definite coefficient matrix, guaranteeing that $C^{\alpha\alpha,t}_{0,f_2,1}(p) > 0$ and hence the equality in Eq. (A.8). This concludes the proof. ∎

Lemma 20 *The inequality in Eq. (4.46), namely*

$$P_4'(p,\alpha) > 0, \tag{A.9}$$

holds for $0 < p, \alpha < 1$. □

Proof: Let $C_1(p,\alpha) = P_4'(p,\alpha)$. Then,

$$C_1(p,\alpha) = 3 + \sum_{i=1}^{3} C_{1,i}(p)\alpha^i,$$

with

$$\begin{aligned} C_{1,1}(p) &= 5 - 37p + 142p^2 - 312p^3 + 403p^4 - 308p^5 + 124p^6 - 24p^7 + 4p^8, \\ C_{1,2}(p) &= p(5 - 46p + 213p^2 - 550p^3 + 557p^4 + 1473p^5 - 7722p^6 + 17446p^7 - \ldots \\ &\quad 24744p^8 + 23416p^9 - 14680p^{10} + 5816p^{11} - 1312p^{12} + 128p^{13}), \\ C_{1,3}(p) &= 3(1-p)^3p^3(1 - 12p + 54p^2 - 64p^3 - 528p^4 + 3328p^5 - 10264p^6 + \ldots \\ &\quad 20400p^7 - 27952p^8 + 26832p^9 - 17936p^{10} + 8160p^{11} - 2408p^{12} + \ldots \\ &\quad 416p^{13} - 32p^{14}). \end{aligned}$$

Proving that the inequality in Eq. (A.9) holds, for $0 < p, \alpha < 1$ is equivalent to proving that

$$D_pC_1(p,\alpha) < 0 \text{ for } 0 < p, \alpha < 1, \tag{A.10}$$

and that $C_1(1,\alpha) > 0$. We have that $C_1(1,\alpha) = 3 - 3\alpha$, and therefore $C_1(1,\alpha) > 0$ for $0 < \alpha < 1$. We have left to establish the inequality in Eq. (A.10). Specifically, taking the first derivative to p of $C_1(p,\alpha)$ yields $D_pC_1(p,\alpha) = \alpha C^p_{1,f}(p,\alpha)$, where

$$C^p_{1,f} = \sum_{i=0}^{2} C^p_{1,f,i}\alpha^i,$$

with

$$\begin{aligned}
C^p_{1,f,0} &= -37 + 284p - 936p^2 + 1612p^3 - 1540p^4 + 744p^5 - 168p^6 + 32p^7,\\
C^p_{1,f,1} &= 5 - 92p + 639p^2 - 2200p^3 + 2785p^4 + 8838p^5 - 54054p^6 + 139568p^7 - \ldots\\
&\quad 222696p^8 + 234160p^9 - 161480p^{10} + 69792p^{11} - 17056p^{12} + 1792p^{13},\\
C^p_{1,f,2} &= 9p^2 - 180p^3 + 1395p^4 - 4734p^5 - 3402p^6 + 111984p^7 - 587736p^8 + \ldots\\
&\quad 1851120p^9 - 4067976p^{10} + 6557472p^{11} - 7904832p^{12} + 7157472p^{13} - \ldots\\
&\quad 4838760p^{14} + 2402688p^{15} - 849864p^{16} + 202608p^{17} - 29184p^{18} + 1920p^{19}.
\end{aligned}$$

It suffices to show that $C^p_{1,f}(p,\alpha) < 0$ for $0 < p, \alpha < 1$. To this end, we find two SOS auxiliary polynomials, $C^{p,a,1}_{1,f}(p,\alpha)$ and $C^{p,a,2}_{1,f}(p)$ such that

$$C^{p,t}_{1,f}(p,\alpha) = -C^p_{1,f} + p(p-1)\,C^{p,a,1}_{1,f}(p,\alpha) + \alpha(\alpha-1)\,C^{p,a,2}_{1,f}(p) > 0.$$

For the condition $C^{p,t}_{1,f}(p,\alpha) > 0$ to hold, we require $C^{p,t}_{1,f}(p,\alpha)$ to be SOS with a positive definite coefficient matrix. The auxiliary polynomial $C^{p,a,1}_{1,f}(p,\alpha)$ is as follows:

$$C^{p,a,1}_{1,f}(p,\alpha) = (Z^{p,a,1}_{1,f})^T \begin{bmatrix} \mathbf{C}^{p,a,1,1}_{1,f} & \mathbf{C}^{p,a,1,2}_{1,f} \end{bmatrix} Z^{p,a,1}_{1,f},$$

with

$$Z^{p,a,1}_{1,f} = \begin{bmatrix} 1 & p & p^2 & p^3 & \alpha p & \alpha p^2 & \alpha p^3 & \alpha p^4 & \alpha p^5 & \alpha p^6 & \alpha p^7 & \alpha p^8 & \alpha p^9 \end{bmatrix}^T,$$

$$
\mathbf{C}_{1,f}^{p,a,1,1} = \begin{bmatrix}
52 & 0 & -541 & \frac{60296}{105} & -377 & 1315 & -\frac{148913}{140} \\
0 & 36 & -\frac{87}{4} & -\frac{160}{3} & 10 & -\frac{479}{3} & \frac{729}{4} \\
-541 & -\frac{87}{4} & \frac{17855}{3} & -\frac{12739}{2} & \frac{10222}{3} & -\frac{47755}{4} & 9631 \\
\frac{60296}{105} & -\frac{160}{3} & -\frac{12739}{2} & 7051 & -\frac{13655}{4} & 12194 & -\frac{1043411}{105} \\
-377 & 10 & \frac{10222}{3} & -\frac{13655}{4} & 6778 & -25549 & 22254 \\
1315 & -\frac{479}{3} & -\frac{47755}{4} & 12194 & -25549 & 100098 & -92632 \\
-\frac{148913}{140} & \frac{729}{4} & 9631 & -\frac{1043411}{105} & 22254 & -92632 & \frac{496676}{5} \\
-\frac{43}{4} & 34 & -\frac{129}{4} & \frac{57}{4} & 2235 & -\frac{33994}{5} & -6040 \\
5 & \frac{71}{4} & -\frac{279}{4} & 122 & -\frac{3854}{5} & 6353 & -11924 \\
\frac{23}{4} & \frac{69}{4} & -90 & 81 & -313 & 2675 & -\frac{14737}{4} \\
\frac{153}{4} & -1 & -16 & -183 & -3443 & \frac{50919}{4} & -9185 \\
-31 & -22 & 232 & -200 & \frac{2731}{4} & -4546 & \frac{41607}{4} \\
24 & -49 & 68 & -163 & -2286 & \frac{37875}{4} & -\frac{68458}{7}
\end{bmatrix},
$$

$$
\mathbf{C}_{1,f}^{p,a,1,2} = \begin{bmatrix}
-\frac{43}{4} & 5 & \frac{23}{4} & \frac{153}{4} & -31 & 24 \\
34 & \frac{71}{4} & \frac{69}{4} & -1 & -22 & -49 \\
-\frac{129}{4} & -\frac{279}{4} & -90 & -16 & 232 & 68 \\
\frac{57}{4} & 122 & 81 & -183 & -200 & -163 \\
2235 & -\frac{3854}{5} & -313 & -3443 & \frac{2731}{4} & -2286 \\
-\frac{33994}{5} & 6353 & 2675 & \frac{50919}{4} & -4546 & \frac{37875}{4} \\
-6040 & -11924 & -\frac{14737}{4} & -9185 & \frac{41607}{4} & -\frac{68458}{7} \\
25384 & -\frac{38913}{4} & -4603 & \frac{9659}{4} & -\frac{36650}{7} & 1470 \\
-\frac{38913}{4} & 31264 & -\frac{6649}{4} & -\frac{109191}{7} & 1372 & \frac{249}{5} \\
-4603 & -\frac{6649}{4} & \frac{144216}{7} & -2842 & -\frac{59096}{5} & 1053 \\
\frac{9659}{4} & -\frac{109191}{7} & -2842 & \frac{117694}{5} & -1053 & -\frac{16711}{3} \\
-\frac{36650}{7} & 1372 & -\frac{59096}{5} & -1053 & \frac{33422}{3} & 0 \\
1470 & \frac{249}{5} & 1053 & -\frac{16711}{3} & 0 & 6688
\end{bmatrix},
$$

where the coefficient matrix of $C_{1,f}^{p,a,1}(p,\alpha)$ is positive definite. Further, we obtain for the second auxiliary polynomial $C_{1,f}^{p,a,2}(p)$:

$$
C_{1,f}^{p,a,2}(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 24 & -37 \\ -37 & 74 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix},
$$

with a positive definite coefficient matrix. This yields

$$C^{p,t}_{1,f}(p,\alpha) = (Z^{p,t}_{1,f})^T \left[\begin{array}{cc} \mathbf{C}^{p,t,1}_{1,f} & \mathbf{C}^{p,t,2}_{1,f} \end{array} \right] Z^{p,t}_{1,f},$$

with

$$Z^{p,t}_{1,f} = \left[\begin{array}{cccccccccccccccc} 1 & p & p^2 & p^3 & p^4 & \alpha & \alpha p & \alpha p^2 & \alpha p^3 & \alpha p^4 & \alpha p^5 & \alpha p^6 & \alpha p^7 & \alpha p^8 & \alpha p^9 & \alpha p^{10} \end{array} \right]^T,$$

$$\mathbf{C}^{p,t,1}_{1,f} = \left[\begin{array}{cccccccc}
37 & -168 & 30 & \frac{96809}{210} & -\frac{1836}{5} & -\frac{29}{2} & 11 & \frac{1777}{6} \\
-168 & 928 & -744 & -\frac{7066}{5} & 1428 & 72 & -\frac{1565}{6} & -772 \\
30 & -744 & \frac{14749}{5} & -4170 & \frac{5924}{3} & -\frac{89}{6} & 935 & -\frac{415243}{105} \\
\frac{96809}{210} & -\frac{7066}{5} & -4170 & \frac{44408}{3} & -9911 & -199 & -\frac{11727}{10} & \frac{57374}{5} \\
-\frac{1836}{5} & 1428 & \frac{5924}{3} & -9911 & 7051 & \frac{1583}{10} & \frac{2469}{5} & -7146 \\
-\frac{29}{2} & 72 & -\frac{89}{6} & -199 & \frac{1583}{10} & 24 & -37 & -\frac{1808}{3} \\
11 & -\frac{1565}{6} & 935 & -\frac{11727}{10} & \frac{2469}{5} & -37 & \frac{3811}{3} & -5140 \\
\frac{1777}{6} & -772 & -\frac{415243}{105} & \frac{57374}{5} & -7146 & -\frac{1808}{3} & -5140 & \frac{10623689}{210} \\
-566 & \frac{25893}{10} & \frac{21114}{5} & -18908 & 12830 & 1841 & \frac{1023749}{210} & -104807 \\
\frac{1283}{10} & -\frac{15691}{5} & 7335 & -5287 & 1068 & -1929 & 7056 & \frac{11375519}{210} \\
-\frac{536}{5} & 8139 & -30930 & 43051 & -\frac{103479}{5} & 2266 & -\frac{161874}{7} & 15517 \\
\frac{118019}{210} & -14159 & 46407 & -\frac{290614}{5} & \frac{130201}{5} & -\frac{27012}{7} & 31869 & 8143 \\
-428 & 10674 & -\frac{169024}{5} & \frac{202991}{5} & -17527 & 3419 & -24984 & -27887 \\
202 & -\frac{21759}{5} & \frac{62521}{5} & -13434 & 5237 & -1597 & 9844 & 22206 \\
-\frac{524}{5} & \frac{7496}{5} & -3539 & 2638 & -518 & 451 & -2526 & -\frac{37859}{5} \\
\frac{156}{5} & -329 & 785 & -347 & -163 & 24 & \frac{4446}{5} & -4669
\end{array} \right],$$

$$\mathbf{C}_{1,f}^{p,t,2}=\begin{bmatrix}
-566 & \frac{1283}{10} & -\frac{536}{5} & \frac{118019}{210} & -428 & 202 & -\frac{524}{5} & \frac{156}{5} \\
\frac{25893}{10} & -\frac{15691}{5} & 8139 & -14159 & 10674 & -\frac{21759}{5} & \frac{7496}{5} & -329 \\
\frac{21114}{5} & 7335 & -30930 & 46407 & -\frac{169024}{5} & \frac{62521}{5} & -3539 & 785 \\
-18908 & -5287 & 43051 & -\frac{290614}{5} & \frac{202991}{5} & -13434 & 2638 & -347 \\
12830 & 1068 & -\frac{103479}{5} & \frac{130201}{5} & -17527 & 5237 & -518 & -163 \\
1841 & -1929 & 2266 & -\frac{27012}{7} & 3419 & -1597 & 451 & 24 \\
\frac{1023749}{210} & 7056 & -\frac{161874}{7} & 31869 & -24984 & 9844 & -2526 & \frac{4446}{5} \\
-104807 & \frac{11375519}{210} & 15517 & 8143 & -27887 & 22206 & -\frac{37859}{5} & -4669 \\
\frac{1921018}{7} & -239292 & 159148 & -296099 & 307332 & -\frac{787024}{5} & 45167 & 5211 \\
-239292 & 390512 & -611869 & 922130 & -\frac{3893309}{5} & 352493 & -95278 & -\frac{2826}{7} \\
159148 & -611869 & 1569644 & -\frac{489509747}{210} & 1759037 & -739400 & \frac{1388375}{7} & 3677 \\
-296099 & 922130 & -\frac{489509747}{210} & 3549554 & -2708333 & \frac{7894812}{7} & -299053 & -4405 \\
307332 & -\frac{3893309}{5} & 1759037 & -2708333 & \frac{15026216}{7} & -905968 & \frac{48555149}{210} & 710 \\
-\frac{787024}{5} & 352493 & -739400 & \frac{7894812}{7} & -905968 & 396244 & -102014 & -3330 \\
45167 & -95278 & \frac{1388375}{7} & -299053 & \frac{48555149}{210} & -102014 & 35844 & -\frac{903839}{210} \\
5211 & -\frac{2826}{7} & 3677 & -4405 & 710 & -3330 & -\frac{903839}{210} & 6688
\end{bmatrix},$$

The coefficient matrix of $C_{1,f}^{p,t}(p,\alpha)$ is positive definite and therefore $C_{1,f}^{p,t}(p,\alpha) > 0$. Hence, the inequality in Eq. (A.10) holds and the proof is complete. ■

Lemma 21 *The inequality in Eq. (4.47), namely*

$$P_4'(p,\alpha) > -P_6'(p,\alpha), \tag{A.11}$$

holds for $0 < p, \alpha < 1$. □

Proof: Let $C_2(p,\alpha) = P_4'(p,\alpha) + P_6'(p,\alpha)$. Then,

$$C_2(p,\alpha) = 2 + \sum_{i=1}^{3} C_{2,i}(p)\alpha^i,$$

with

$$\begin{aligned}
C_{2,1}(p) &= 2(3 - 21p + 78p^2 - 170p^3 + 225p^4 - 185p^5 + 89p^6 - 24p^7 + 4p^8), \\
C_{2,2}(p) &= 2(1-p)^2 p(3 - 22p + 86p^2 - 174p^3 + 86p^4 + 586p^5 - 1937p^6 + \ldots
\end{aligned}$$

$$
\begin{aligned}
& 3105p^7 - 2964p^8 + 1654p^9 - 492p^{10} + 60p^{11}), \\
C_{2,3}(p) \quad = \quad & 2(1-p)^3 p^3 (1 - 12p + 42p^2 + 68p^3 - 1281p^4 + 6048p^5 - 17060p^6 + \ldots \\
& 32472p^7 - 43386p^8 + 41008p^9 - 27156p^{10} + 12288p^{11} - 3616p^{12} + \ldots \\
& 624p^{13} - 48p^{14}).
\end{aligned}
$$

Proving that the inequality in Eq. (A.11) holds, is equivalent to showing that

$$
D_p C_2(p, \alpha) < 0 \text{ for } 0 < p, \alpha < 1, \tag{A.12}
$$

and that $C_2(1, \alpha) > 0$ for $0 < \alpha < 1$. We have that $C_2(1, \alpha) = 2(1 - \alpha)$, and therefore $C_2(1, \alpha) > 0$ for $0 < \alpha < 1$. We have left to establish the inequality in Eq. (A.12). In particular, we have that $D_p C_2(p, \alpha) = 2\alpha C^p_{2,f}(p, \alpha)$ where

$$
C^p_{2,f}(p, \alpha) = \sum_{i=0}^{2} C^p_{2,f,i}(p) \alpha^i,
$$

with

$$
\begin{aligned}
C^p_{2,f,1}(p) \quad = \quad & -21 + 156p - 510p^2 + 900p^3 - 925p^4 + 534p^5 - 168p^6 + 32p^7, \\
C^p_{2,f,2}(p) \quad = \quad & (1-p)(3 - 53p + 346p^2 - 1126p^3 + 1474p^4 + 2914p^5 - 18247p^6 + \ldots \\
& 42273p^7 - 57726p^8 + 49144p^9 - 25260p^{10} + 7116p^{11} - 840p^{12}), \\
C^p_{2,f,3}(p) \quad = \quad & 3(1-p)^2 p^2 (1 - 18p + 98p^2 + 24p^3 - 3193p^4 + 20398p^5 - 73356p^6 + \ldots \\
& 176480p^7 - 299794p^8 + 366500p^9 - 322716p^{10} + 202176p^{11} - 87712p^{12} + \ldots \\
& 25024p^{13} - 4224p^{14} + 320p^{15}).
\end{aligned}
$$

To establish the inequality in Eq. (A.12), we find two SOS auxiliary polynomials, $C^{p,a,1}_{2,f}(p, \alpha)$ and $C^{p,a,2}_{2,f}(p)$ such that

$$
C^{p,t}_{2,f}(p, \alpha) = -D_p C_2(p, \alpha) + p^7 (p-1)^5 \, C^{p,a,1}_{2,f}(p, \alpha) + \alpha(\alpha - 1) \, C^{p,a,2}_{2,f}(p) > 0.
$$

For the condition $C^{p,t}_{2,f} > 0$ to hold, we require $C^{p,t}_{2,f}$ to be SOS with a positive definite coefficient matrix. The auxiliary polynomial $C_{2,f}^{p,a,1}(p,\alpha)$ is as follows:

$$C_{2,f}^{p,a,1}(p,\alpha) = (Z_{2,f}^{p,a,1})^T \mathbf{C}_{2,f}^{p,a,1}\, Z_{2,f}^{p,a,1},$$

with

$$Z_{2,f}^{p,a,1} = \begin{bmatrix} 1 & p & p^2 & p^3 & \alpha & \alpha p & \alpha p^2 & \alpha p^3 & \alpha p^4 \end{bmatrix}^T,$$

$$\mathbf{C}_{2,f}^{p,a,1} = \begin{bmatrix} 6039 & 0 & -11774 & -12599 & -8666 & 4990 & 4132 & \frac{1959}{8} & 19830 \\ 0 & 23548 & -10716 & -\frac{113434}{3} & -4545 & 10398 & -\frac{83617}{8} & -2037 & -14751 \\ -11774 & -10716 & \frac{226868}{3} & 0 & 1 & -\frac{17409}{8} & 20724 & -23028 & -39715 \\ -12599 & -\frac{113434}{3} & 0 & 154306 & \frac{324927}{8} & -43489 & -4383 & 18452 & -3851 \\ -8666 & -4545 & 1 & \frac{324927}{8} & 37161 & -31512 & -42733 & \frac{128689}{4} & -7738 \\ 4990 & 10398 & -\frac{17409}{8} & -43489 & -31512 & 38541 & \frac{140265}{4} & -37276 & -22254 \\ 4132 & -\frac{83617}{8} & 20724 & -4383 & -42733 & \frac{140265}{4} & 82073 & -53152 & -16586 \\ \frac{1959}{8} & -2037 & -23028 & 18452 & \frac{128689}{4} & -37276 & -53152 & 55475 & 36991 \\ 19830 & -14751 & -39715 & -3851 & -7738 & -22254 & -16586 & 36991 & 135459 \end{bmatrix}$$

where the coefficient matrix $\mathbf{C}_{2,f}^{p,a,1}$ is positive definite. Further, we obtain for the second auxiliary polynomial $C_{2,f}^{p,a,2}(p)$:

$$C_{2,f}^{p,a,2}(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 11 & -\frac{23}{2} \\ -\frac{23}{2} & 15 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix},$$

with a positive definite coefficient matrix. This yields

$$C_{2,f}^{p,t}(p,\alpha) = (Z_{2,f}^{p,t})^T \begin{bmatrix} \mathbf{C}_{2,f}^{p,t,1} & \mathbf{C}_{2,f}^{p,t,2} & \mathbf{C}_{2,f}^{p,t,3} \end{bmatrix} Z_{2,f}^{p,t},$$

with

$$Z_{2,f}^{p,t} = \begin{bmatrix} 1 & p & p^2 & p^3 & p^4 & p^5 & p^6 & p^7 & p^8 & p^9 & \alpha & \alpha p & \alpha p^2 & \alpha p^3 & \ldots \\ & & & & & & & & \alpha p^4 & \alpha p^5 & \alpha p^6 & \alpha p^7 & \alpha p^8 & \alpha p^9 & \alpha p^{10} \end{bmatrix}^T,$$

$$\mathbf{C}_{2,f}^{p,t,1} = \begin{bmatrix}
21 & -78 & -\frac{1331}{12} & \frac{5181}{8} & -\frac{3771}{8} & -\frac{4409}{12} & 101 \\
-78 & \frac{4391}{6} & -\frac{8781}{8} & -\frac{22979}{8} & \frac{183533}{24} & -3206 & -\frac{10999}{8} \\
-\frac{1331}{12} & -\frac{8781}{8} & \frac{15225}{2} & -\frac{181123}{24} & -\frac{36437}{2} & \frac{126397}{4} & -\frac{147413}{18} \\
\frac{5181}{8} & -\frac{22979}{8} & -\frac{181123}{24} & 42815 & -\frac{135059}{4} & -\frac{3205547}{72} & \frac{270242}{5} \\
-\frac{3771}{8} & \frac{183533}{24} & -\frac{36437}{2} & -\frac{135059}{4} & \frac{2595211}{18} & -\frac{3994579}{40} & -\frac{194333}{4} \\
-\frac{4409}{12} & -3206 & \frac{126397}{4} & -\frac{3205547}{72} & -\frac{3994579}{40} & \frac{1021083}{4} & -\frac{4067781}{32} \\
101 & -\frac{10999}{8} & -\frac{147413}{18} & \frac{270242}{5} & -\frac{194333}{4} & -\frac{4067781}{32} & \frac{1131585}{4} \\
\frac{4039}{8} & -\frac{303514493}{73920} & \frac{44787}{5} & 3289 & -\frac{261533}{32} & -\frac{124037}{8} & -\frac{1791287}{8} \\
-\frac{3143}{18} & \frac{67359}{10} & -\frac{52905}{2} & -\frac{314641}{32} & \frac{899993}{8} & -\frac{62181}{2} & \frac{1868997}{20} \\
-\frac{2889}{40} & -\frac{19029}{8} & \frac{430435}{32} & -\frac{18185}{8} & -\frac{442413}{8} & \frac{348711}{10} & -\frac{84921}{4} \\
-7 & \frac{133}{8} & 60 & -\frac{1427}{8} & \frac{1007}{20} & \frac{6295}{48} & -\frac{29}{4} \\
\frac{183}{8} & -\frac{66361679}{221760} & \frac{4177}{8} & \frac{40708483}{44352} & -\frac{125459}{48} & 1214 & \frac{8903}{64} \\
\frac{129}{4} & \frac{110436479}{221760} & -\frac{127191}{40} & \frac{132775}{48} & \frac{58749}{8} & -\frac{697345}{64} & \frac{246329}{72} \\
-\frac{423}{4} & \frac{36839}{40} & \frac{67435}{48} & -\frac{81045}{8} & \frac{367631}{64} & \frac{401377}{36} & -\frac{932689}{80} \\
-\frac{47}{5} & -\frac{116111}{48} & \frac{41539}{4} & -\frac{167321}{64} & -\frac{1121423}{36} & \frac{2446181}{80} & -\frac{217861}{80} \\
\frac{505}{48} & 1651 & -\frac{836625}{64} & \frac{1359683}{72} & \frac{1514811}{80} & -\frac{3751481}{80} & \frac{258569}{8} \\
\frac{27719999}{221760} & -\frac{59825}{64} & \frac{47027}{18} & -\frac{353929}{80} & \frac{361919}{80} & \frac{44655}{8} & -\frac{912441}{64} \\
\frac{5135}{64} & \frac{5023}{9} & -\frac{178889}{80} & -\frac{63741}{80} & \frac{28403}{4} & -\frac{441313}{64} & -\frac{163131}{8} \\
-\frac{3425}{36} & -\frac{10149}{80} & \frac{270809}{80} & -\frac{30575}{8} & -\frac{529673}{64} & 7815 & \frac{141901}{8} \\
-\frac{15249}{80} & \frac{48499}{80} & \frac{10497}{4} & -\frac{497937}{64} & -\frac{50023}{8} & \frac{228863}{8} & -\frac{71639}{8} \\
\frac{10969}{80} & -\frac{3773}{8} & -\frac{160857}{64} & \frac{57519}{8} & 4636 & -\frac{81751}{4} & \frac{17343}{4}
\end{bmatrix},$$

$$
\mathbf{C}_{2,f}^{p,t,2} = \begin{bmatrix}
\frac{4039}{8} & -\frac{3143}{18} & -\frac{2889}{40} & -7 & \frac{183}{8} & \frac{129}{4} & -\frac{423}{4} \\
-\frac{303514493}{73920} & \frac{67359}{10} & -\frac{19029}{8} & \frac{133}{8} & -\frac{66361679}{221760} & \frac{110436479}{221760} & \frac{36839}{40} \\
\frac{44787}{5} & -\frac{52905}{2} & \frac{430435}{32} & 60 & \frac{4177}{8} & -\frac{127191}{40} & \frac{67435}{48} \\
3289 & -\frac{314641}{32} & -\frac{18185}{8} & -\frac{1427}{8} & \frac{40708483}{44352} & \frac{132775}{48} & -\frac{81045}{8} \\
-\frac{261533}{32} & \frac{899993}{8} & -\frac{442413}{8} & \frac{1007}{20} & -\frac{125459}{48} & \frac{58749}{8} & \frac{367631}{64} \\
-\frac{124037}{8} & -\frac{62181}{2} & \frac{348711}{10} & \frac{6295}{48} & 1214 & -\frac{697345}{64} & \frac{401377}{36} \\
-\frac{1791287}{8} & \frac{1868997}{20} & -\frac{84921}{4} & -\frac{29}{4} & \frac{8903}{64} & \frac{246329}{72} & -\frac{932689}{80} \\
\frac{7480381}{10} & -\frac{3094459}{4} & \frac{58632771119}{221760} & -\frac{2569}{64} & \frac{33709}{36} & -\frac{714739}{80} & \frac{1305839}{80} \\
-\frac{3094459}{4} & \frac{6085591}{6} & -385765 & -\frac{827}{9} & -\frac{241577027}{221760} & \frac{1256299}{80} & -\frac{103151}{4} \\
\frac{58632771119}{221760} & -385765 & 154306 & \frac{5321}{80} & \frac{19909}{80} & -\frac{54069}{8} & \frac{778031}{64} \\
-\frac{2569}{64} & -\frac{827}{9} & \frac{5321}{80} & 11 & -\frac{23}{2} & -\frac{1243}{8} & \frac{713}{2} \\
\frac{33709}{36} & -\frac{241577027}{221760} & \frac{19909}{80} & -\frac{23}{2} & \frac{1291}{4} & -\frac{653}{2} & -2674 \\
-\frac{714739}{80} & \frac{1256299}{80} & -\frac{54069}{8} & -\frac{1243}{8} & -\frac{653}{2} & 5324 & -\frac{12027}{2} \\
\frac{1305839}{80} & -\frac{103151}{4} & \frac{778031}{64} & \frac{713}{2} & -2674 & -\frac{12027}{2} & \frac{1486157}{28} \\
\frac{173027}{8} & -\frac{2489993}{64} & \frac{61055}{4} & -\frac{381}{2} & \frac{12391}{2} & -\frac{952341}{56} & -\frac{380877}{4} \\
-\frac{3849881}{64} & \frac{158725}{2} & -\frac{124653}{4} & 103 & -\frac{62653}{14} & \frac{270431}{8} & \frac{5199931}{72} \\
-\frac{104951}{8} & \frac{155641}{4} & -\frac{151465}{8} & -\frac{2399}{7} & \frac{18957}{8} & -\frac{106408}{9} & -\frac{1156543}{20} \\
84076 & -\frac{21915237959}{221760} & \frac{8275972319}{221760} & \frac{1013}{4} & -\frac{241091}{72} & -\frac{14762}{5} & \frac{3656269}{44} \\
-\frac{224021}{8} & \frac{116175}{8} & -\frac{73993}{24} & -\frac{9989}{72} & \frac{100729}{40} & -\frac{547325}{88} & -\frac{2790019}{40} \\
-\frac{195673}{8} & \frac{407867}{24} & -\frac{10427}{8} & \frac{11186683}{44352} & -\frac{61081}{88} & \frac{238631}{40} & \frac{1019209}{36} \\
\frac{140779}{12} & -\frac{5637}{8} & -3851 & -\frac{3019}{22} & \frac{27337463}{221760} & -\frac{20885}{36} & -5592
\end{bmatrix},
$$

$$\mathbf{C}_{2,f}^{p,t,2} = \begin{bmatrix}
-\frac{47}{5} & \frac{505}{48} & \frac{27719999}{221760} & \frac{5135}{64} & -\frac{3425}{36} & -\frac{15249}{80} & \frac{10969}{80} \\
-\frac{116111}{48} & 1651 & -\frac{59825}{64} & \frac{5023}{9} & -\frac{10149}{80} & \frac{48499}{80} & -\frac{3773}{8} \\
\frac{41539}{4} & -\frac{836625}{64} & \frac{47027}{18} & -\frac{178889}{80} & \frac{270809}{80} & \frac{10497}{4} & -\frac{160857}{64} \\
-\frac{167321}{64} & \frac{1359683}{72} & -\frac{353929}{80} & -\frac{63741}{80} & -\frac{30575}{8} & -\frac{497937}{64} & \frac{57519}{8} \\
-\frac{1121423}{36} & \frac{1514811}{80} & \frac{361919}{80} & \frac{28403}{4} & -\frac{529673}{64} & -\frac{50023}{8} & 4636 \\
\frac{2446181}{80} & -\frac{3751481}{80} & \frac{44655}{8} & -\frac{441313}{64} & 7815 & \frac{228863}{8} & -\frac{81751}{4} \\
-\frac{217861}{80} & \frac{258569}{8} & -\frac{912441}{64} & -\frac{163131}{8} & \frac{141901}{8} & -\frac{71639}{8} & \frac{17343}{4} \\
\frac{173027}{8} & -\frac{3849881}{64} & -\frac{104951}{8} & 84076 & -\frac{224021}{8} & -\frac{195673}{8} & \frac{140779}{12} \\
-\frac{2489993}{64} & \frac{158725}{2} & \frac{155641}{4} & -\frac{21915237959}{221760} & \frac{116175}{8} & \frac{407867}{24} & -\frac{5637}{8} \\
\frac{61055}{4} & -\frac{124653}{4} & -\frac{151465}{8} & \frac{8275972319}{221760} & -\frac{73993}{24} & -\frac{10427}{8} & -3851 \\
-\frac{381}{2} & 103 & -\frac{2399}{7} & \frac{1013}{4} & -\frac{9989}{72} & \frac{11186683}{44352} & -\frac{3019}{22} \\
\frac{12391}{2} & -\frac{62653}{14} & \frac{18957}{8} & -\frac{241091}{72} & \frac{100729}{40} & -\frac{61081}{88} & \frac{27337463}{221760} \\
-\frac{952341}{56} & \frac{270431}{8} & -\frac{106408}{9} & -\frac{14762}{5} & -\frac{547325}{88} & \frac{238631}{40} & -\frac{20885}{36} \\
-\frac{380877}{4} & \frac{5199931}{72} & -\frac{1156543}{20} & \frac{3656269}{44} & -\frac{2790019}{40} & \frac{1019209}{36} & -5592 \\
\frac{17533517}{36} & -\frac{3886392}{5} & \frac{46045155}{88} & -\frac{9946569}{40} & \frac{24636311}{72} & -\frac{641471}{2} & \frac{802247}{8} \\
-\frac{3886392}{5} & \frac{70950221}{44} & -\frac{24695077}{20} & \frac{7090805}{18} & -\frac{4502743}{8} & \frac{3000101}{4} & -\frac{1131483}{4} \\
\frac{46045155}{88} & -\frac{24695077}{20} & \frac{42773629}{36} & -\frac{4177701}{8} & \frac{1800989}{4} & -569326 & \frac{928055}{4} \\
-\frac{9946569}{40} & \frac{7090805}{18} & -\frac{4177701}{8} & \frac{2360105}{4} & -\frac{1651851}{4} & \frac{1096855}{8} & -\frac{113613}{8} \\
\frac{24636311}{72} & -\frac{4502743}{8} & \frac{1800989}{4} & -\frac{1651851}{4} & \frac{2377283}{4} & -\frac{3800279}{8} & \frac{1099995}{8} \\
-\frac{641471}{2} & \frac{3000101}{4} & -569326 & \frac{1096855}{8} & -\frac{3800279}{8} & \frac{2986305}{4} & -\frac{604273}{2} \\
\frac{802247}{8} & -\frac{1131483}{4} & \frac{928055}{4} & -\frac{113613}{8} & \frac{1099995}{8} & -\frac{604273}{2} & \frac{30039387839}{221760}
\end{bmatrix},$$

where the coefficient matrix of $C_{2,f}^{p,t}(p,\alpha)$ is positive definite. Therefore, $C_{2,f}^{p,t}(p,\alpha) > 0$ and the inequality in Eq. (A.12) holds, which completes the proof. ■

Lemma 22 *The inequality in Eq. (4.48), namely*

$$-P_2'(p,\alpha) > P_0'(p,\alpha), \tag{A.13}$$

holds for $0 < p, \alpha < 1$. □

Proof: Let $C_3(p,\alpha) = -P_2'(p,\alpha) - P_0'(p,\alpha)$. Then,

$$C_3(p,\alpha) = 2 + \sum_{i=1}^{3} C_{3,i}(p)\alpha^i,$$

with

$$
\begin{aligned}
C_{3,1}(p) &= 6-42p+156p^2-340p^3+450p^4-370p^5+178p^6-48p^7+8p^8, \\
C_{3,2}(p) &= 6p-88p^2+554p^3-2176p^4+5872p^5-11808p^6+18882p^7-24742p^8+\ldots \\
&\quad 26162p^9-21186p^{10}+12264p^{11}-4684p^{12}+1048p^{13}-104p^{14}, \\
C_{3,3}(p) &= 2p^3-30p^4+162p^5-190p^6-2310p^7+15498p^8-51510p^9+111018p^{10}-\ldots \\
&\quad 167484p^{11}+182036p^{12}-143508p^{13}+81324p^{14}-32280p^{15}+8520p^{16}-\ldots \\
&\quad 1344p^{17}+96p^{18}.
\end{aligned}
$$

Proving that the inequality in Eq. (A.13) holds, is equivalent to showing that

$$
D_pC_3(p,\alpha) < 0, \text{ for } 0 < p, \alpha < 1, \tag{A.14}
$$

and that $C_3(1,\alpha) > 0$, for $0 < \alpha < 1$. We have that $C_3(1,\alpha) = 2 - 2\alpha$, and therefore $C_3(1,\alpha) > 0$ for $0 < \alpha < 1$. We have left to establish the inequality in Eq. (A.14). In particular, we have that $D_pC_3(p,\alpha) = 2\alpha C^p_{3,f}(p,\alpha)$, where

$$
C^p_{3,f}(p,\alpha) = \sum_{i=1}^{3} C^p_{3,f,i}(p)\alpha^i,
$$

with

$$
\begin{aligned}
C_{3,f,1} &= -21+156p-510p^2+900p^3-925p^4+534p^5-168p^6+32p^7, \\
C_{3,f,2} &= -3+88p-831p^2+4352p^3-14680p^4+35424p^5-66087p^6+98968p^7-\ldots \\
&\quad 117729p^8+105930p^9-67452p^{10}+28104p^{11}-6812p^{12}+728p^{13}, \\
C_{3,f,3} &= 3(1-p)^2p^2(1-18p+98p^2+24p^3-2745p^4+15150p^5-44220p^6+\ldots \\
&\quad 81440p^7-99954p^8+82724p^9-45532p^{10}+15968p^{11}-3232p^{12}+288p^{13}).
\end{aligned}
$$

To establish the inequality in Eq. (A.14), we find two SOS auxiliary polynomials $C^{p,a,1}_{3,f}(p,\alpha)$ and $C^{p,a,2}_{3,f}(p)$ such that

$$
C^{p,t}_{3,f}(p,\alpha) = -D_pC_3(p,\alpha) + p^7(p-1)^5\, C^{p,a,1}_{3,f}(p,\alpha) + \alpha(\alpha-1)\, C^{p,a,2}_{3,f}(p) > 0.
$$

For the condition $C^{p,t}_{3,f}(p,\alpha) > 0$ to hold, we require $C^{p,t}_{3,f}(p,\alpha)$ to be SOS with a positive definite coefficient matrix. The auxiliary polynomial $C^{p,a,1}_{3,f}(p,\alpha)$ is as follows:

$$C^{p,a,1}_{3,f}(p,\alpha) = (Z^{p,a,1}_{3,f})^T\, \mathbf{C}^{p,a,1}_{3,f}\, Z^{p,a,1}_{3,f},$$

with

$$Z^{p,a,1}_{3,f} = \begin{bmatrix} 1 & p & p^2 & p^3 & \alpha & \alpha p & \alpha p^2 & \alpha p^3 \end{bmatrix}^T,$$

$$\mathbf{C}^{p,a,1}_{3,f} = \begin{bmatrix} 15616 & 0 & -\frac{74272}{3} & -30221 & -101206 & 247509 & -43956 & -128710 \\ 0 & \frac{148544}{3} & -37279 & -72835 & -18254 & -7080 & 53072 & 5557 \\ -\frac{74272}{3} & -37279 & 145670 & 0 & \frac{130034}{3} & -55970 & 7051 & -48275 \\ -30221 & -72835 & 0 & 323695 & 379903 & -870959 & 43849 & 524791 \\ -101206 & -18254 & \frac{130034}{3} & 379903 & 1097557 & -2801676 & \frac{1471891}{3} & 1652080 \\ 247509 & -7080 & -55970 & -870959 & -2801676 & \frac{22100092}{3} & -1618848 & -\frac{12483214}{3} \\ -43956 & 53072 & 7051 & 43849 & \frac{1471891}{3} & -1618848 & \frac{3334055}{3} & 244593 \\ -128710 & 5557 & -48275 & 524791 & 1652080 & -\frac{12483214}{3} & 244593 & 3104765 \end{bmatrix},$$

where the coefficient matrix $\mathbf{C}^{p,a,1}_{3,f}$ is positive definite. Further, we obtain for the second auxiliary polynomial $C^{p,a,2}_{3,f}(p)$:

$$C^{p,a,2}_{3,f}(p) = \begin{bmatrix} 1 \\ p \end{bmatrix}^T \begin{bmatrix} 19 & -23 \\ -23 & 30 \end{bmatrix} \begin{bmatrix} 1 \\ p \end{bmatrix},$$

with a positive definite coefficient matrix. This yields

$$C^{p,t}_{3,f}(p,\alpha) = (Z^{p,t}_{3,f})^T \begin{bmatrix} \mathbf{C}^{p,t,1}_{3,f} & \mathbf{C}^{p,t,2}_{3,f} & \mathbf{C}^{p,t,3}_{3,f} \end{bmatrix} Z^{p,t}_{3,f},$$

with

$$Z^{p,t}_{3,f} = \begin{bmatrix} 1 & p & p^2 & p^3 & p^4 & p^5 & p^6 & p^7 & p^8 & p^9 & \alpha & \alpha p & \alpha p^2 & \alpha p^3 & \ldots \\ & & & & & & & & \alpha p^4 & \alpha p^5 & \alpha p^6 & \alpha p^7 & \alpha p^8 & \alpha p^9 \end{bmatrix}^T,$$

$$
\mathbf{C}_{3,f}^{p,t,1} = \begin{bmatrix}
21 & -78 & -\frac{2371}{12} & \frac{65675}{64} & -\frac{31257}{64} & -\frac{72417}{32} & \frac{333359}{112} \\
-78 & \frac{5431}{6} & -\frac{94475}{64} & -\frac{181259}{32} & \frac{1046743}{64} & -\frac{855103}{224} & -\frac{5845141}{256} \\
-\frac{2371}{12} & -\frac{94475}{64} & \frac{423375}{32} & -\frac{918997}{64} & -\frac{23341893}{448} & \frac{28888991}{256} & -\frac{1881827}{64} \\
\frac{65675}{64} & -\frac{181259}{32} & -\frac{918997}{64} & \frac{23756295}{224} & -\frac{24963365}{256} & -\frac{7051705}{32} & \frac{6750027}{16} \\
-\frac{31257}{64} & \frac{1046743}{64} & -\frac{23341893}{448} & -\frac{24963365}{256} & \frac{8608809}{16} & -\frac{26958299}{64} & -\frac{313947431}{576} \\
-\frac{72417}{32} & -\frac{855103}{224} & \frac{28888991}{256} & -\frac{7051705}{32} & -\frac{26958299}{64} & \frac{441420935}{288} & -\frac{69278521}{64} \\
\frac{333359}{112} & -\frac{5845141}{256} & -\frac{1881827}{64} & \frac{6750027}{16} & -\frac{313947431}{576} & -\frac{69278521}{64} & \frac{232281783359}{80640} \\
-\frac{13140067}{40320} & \frac{167335}{8} & -\frac{2466487}{32} & -\frac{38279821}{288} & \frac{6778091}{8} & -\frac{130467709}{224} & -\frac{113503817}{64} \\
-\frac{72501}{64} & -\frac{67753}{32} & \frac{34203385}{576} & -\frac{4176121}{32} & -\frac{3462757}{14} & \frac{62381977}{64} & -\frac{12731985}{64} \\
\frac{29551}{64} & -\frac{1263167}{576} & -\frac{692525}{64} & \frac{16207225}{224} & -\frac{152275}{4} & -\frac{9842815}{32} & \frac{22206005}{64} \\
-11 & \frac{1751}{64} & \frac{10143}{64} & -\frac{16191}{32} & -\frac{2667}{64} & \frac{5105}{4} & -\frac{59877}{56} \\
\frac{2537}{64} & -\frac{36981}{64} & \frac{48949}{64} & \frac{71061}{16} & -\frac{931196699}{80640} & \frac{2521331}{448} & \frac{111047}{16} \\
-\frac{357}{32} & \frac{55595}{32} & -\frac{146231}{16} & -\frac{9031}{32} & \frac{6766391}{112} & -\frac{1294651}{16} & -\frac{13595}{36} \\
\frac{14498819}{80640} & -\frac{104811}{64} & \frac{340771}{16} & -\frac{12388463}{224} & \frac{5235}{64} & \frac{36776737}{576} & \frac{528561}{8} \\
-\frac{30801}{32} & \frac{358139}{64} & -\frac{8975231}{448} & \frac{5343967}{32} & -\frac{19822415}{36} & \frac{5459775}{8} & -\frac{20946701}{32} \\
\frac{21943}{16} & -\frac{2402377}{112} & \frac{1680011}{64} & -\frac{139676003}{576} & \frac{81584903}{64} & -\frac{119301937}{64} & \frac{117552577}{64} \\
-\frac{274851}{224} & \frac{938753}{32} & -\frac{24389729}{576} & \frac{10479417}{64} & -\frac{8560503}{8} & \frac{61104175}{32} & -\frac{517430631}{224} \\
\frac{66181}{32} & -\frac{334700893}{26880} & \frac{916464779}{80640} & \frac{569301}{16} & \frac{1039257}{8} & -\frac{196041269}{224} & \frac{105615724829}{80640} \\
-\frac{1345139}{576} & -\frac{201327}{64} & \frac{877037}{32} & -\frac{3985755}{32} & \frac{124643035}{448} & \frac{72616931}{384} & -\frac{3398343}{16} \\
\frac{28407}{32} & \frac{82303}{32} & -\frac{503645}{32} & \frac{23443447}{448} & -\frac{42620299}{384} & -\frac{1010153}{32} & -\frac{10580693}{256}
\end{bmatrix},
$$

$$
\mathbf{C}^{p,t,2}_{3,f} = \begin{bmatrix}
-\frac{13140067}{40320} & -\frac{72501}{64} & \frac{29551}{64} & -11 & \frac{2537}{64} & -\frac{357}{32} & \frac{14498819}{80640} \\
\frac{167335}{8} & -\frac{67753}{32} & -\frac{1263167}{576} & \frac{1751}{64} & -\frac{36981}{64} & \frac{55595}{32} & -\frac{104811}{64} \\
-\frac{2466487}{32} & \frac{34203385}{576} & -\frac{692525}{64} & \frac{10143}{64} & \frac{48949}{64} & -\frac{146231}{16} & \frac{340771}{16} \\
-\frac{38279821}{288} & -\frac{4176121}{32} & \frac{16207225}{224} & -\frac{16191}{32} & \frac{71061}{16} & -\frac{9031}{32} & -\frac{12388463}{224} \\
\frac{6778091}{8} & -\frac{3462757}{14} & -\frac{152275}{4} & -\frac{2667}{64} & -\frac{931196699}{80640} & \frac{6766391}{112} & \frac{5235}{64} \\
-\frac{130467709}{224} & \frac{62381977}{64} & -\frac{9842815}{32} & \frac{5105}{4} & \frac{2521331}{448} & -\frac{1294651}{16} & \frac{36776737}{576} \\
-\frac{113503817}{64} & -\frac{12731985}{64} & \frac{22206005}{64} & -\frac{59877}{56} & \frac{111047}{16} & -\frac{13595}{36} & \frac{528561}{8} \\
\frac{105808815}{32} & -\frac{130108405}{64} & \frac{10187599}{24} & -\frac{559}{8} & -\frac{2201809}{288} & \frac{671163}{32} & -\frac{4267471}{32} \\
-\frac{130108405}{64} & \frac{28655801}{12} & -\frac{1618475}{2} & \frac{77981}{288} & \frac{121461}{64} & \frac{502307}{16} & \frac{164599}{16} \\
\frac{10187599}{24} & -\frac{1618475}{2} & 323695 & -\frac{2211}{64} & \frac{3689}{64} & -\frac{762491}{32} & \frac{400699}{14} \\
-\frac{559}{8} & \frac{77981}{288} & -\frac{2211}{64} & 19 & -23 & -\frac{41603}{96} & \frac{70103}{64} \\
-\frac{2201809}{288} & \frac{121461}{64} & \frac{3689}{64} & -23 & \frac{42899}{48} & -\frac{68183}{64} & -\frac{2261833}{160} \\
\frac{671163}{32} & \frac{502307}{16} & -\frac{762491}{32} & -\frac{41603}{96} & -\frac{68183}{64} & \frac{4216707}{160} & -\frac{1330131}{32} \\
-\frac{4267471}{32} & \frac{164599}{16} & \frac{400699}{14} & \frac{70103}{64} & -\frac{2261833}{160} & -\frac{1330131}{32} & \frac{118699367}{224} \\
\frac{81089955}{64} & -\frac{642581479}{448} & \frac{206838707}{384} & \frac{242159}{320} & \frac{2987307}{64} & -\frac{23658721}{112} & -\frac{42947705}{32} \\
-\frac{1719321619}{448} & \frac{1780253123}{384} & -\frac{115855103}{64} & -\frac{308805}{64} & -\frac{12174621}{224} & \frac{23390945}{32} & \frac{37444128913}{26880} \\
\frac{1622422907}{384} & -\frac{299433963}{64} & \frac{454484831}{256} & \frac{2095799}{448} & \frac{2012555}{64} & -\frac{418801069}{576} & -\frac{54867945}{64} \\
-\frac{3452845}{8} & -\frac{107436017}{256} & \frac{24045695}{96} & -\frac{4603}{64} & -\frac{3559903}{144} & \frac{1172409}{64} & \frac{266958085}{288} \\
-\frac{529010841}{256} & \frac{153481585}{48} & -\frac{41087207}{32} & -\frac{1098109}{576} & \frac{777715}{32} & \frac{50989979}{144} & -\frac{237047749}{256} \\
\frac{92513279}{96} & -\frac{43020985}{32} & 524791 & \frac{57241799}{80640} & -\frac{5198815}{576} & -\frac{37965725}{256} & \frac{147864499}{448}
\end{bmatrix},
$$

$$
\mathbf{C}_{3,f}^{p,t,2} = \begin{bmatrix}
-\frac{30801}{32} & \frac{21943}{16} & -\frac{274851}{224} & \frac{66181}{32} & -\frac{1345139}{576} & \frac{28407}{32} \\
\frac{358139}{64} & -\frac{2402377}{112} & \frac{938753}{32} & -\frac{334700893}{26880} & -\frac{201327}{64} & \frac{82303}{32} \\
-\frac{8975231}{448} & \frac{1680011}{64} & -\frac{24389729}{576} & \frac{916464779}{80640} & \frac{877037}{32} & -\frac{503645}{32} \\
\frac{5343967}{32} & -\frac{139676003}{576} & \frac{10479417}{64} & \frac{569301}{16} & -\frac{3985755}{32} & \frac{23443447}{448} \\
-\frac{19822415}{36} & \frac{81584903}{64} & -\frac{8560503}{8} & \frac{1039257}{8} & \frac{124643035}{448} & -\frac{42620299}{384} \\
\frac{5459775}{8} & -\frac{119301937}{64} & \frac{61104175}{32} & -\frac{196041269}{224} & \frac{72616931}{384} & -\frac{1010153}{32} \\
-\frac{20946701}{32} & \frac{117552577}{64} & -\frac{517430631}{224} & \frac{105615724829}{80640} & -\frac{3398343}{16} & -\frac{10580693}{256} \\
\frac{81089955}{64} & -\frac{1719321619}{448} & \frac{1622422907}{384} & -\frac{3452845}{8} & -\frac{529010841}{256} & \frac{92513279}{96} \\
-\frac{642581479}{448} & \frac{1780253123}{384} & -\frac{299433963}{64} & -\frac{107436017}{256} & \frac{153481585}{48} & -\frac{43020985}{32} \\
\frac{206838707}{384} & -\frac{115855103}{64} & \frac{454484831}{256} & \frac{24045695}{96} & -\frac{41087207}{32} & 524791 \\
\frac{242159}{320} & -\frac{308805}{64} & \frac{2095799}{448} & -\frac{4603}{64} & -\frac{1098109}{576} & \frac{57241799}{80640} \\
\frac{2987307}{64} & -\frac{12174621}{224} & \frac{2012555}{64} & -\frac{3559903}{144} & \frac{777715}{32} & -\frac{5198815}{576} \\
-\frac{23658721}{112} & \frac{23390945}{32} & -\frac{418801069}{576} & \frac{1172409}{64} & \frac{50989979}{144} & -\frac{37965725}{256} \\
-\frac{42947705}{32} & \frac{37444128913}{26880} & -\frac{54867945}{64} & \frac{266958085}{288} & -\frac{237047749}{256} & \frac{147864499}{448} \\
\frac{2892769015}{288} & -\frac{370139355}{16} & \frac{12704059997}{576} & -\frac{1139532537}{256} & -\frac{2594381497}{448} & \frac{224319967199}{80640} \\
-\frac{370139355}{16} & \frac{4532273299}{72} & -\frac{16319852709}{256} & \frac{4412607485}{448} & \frac{1432051361}{64} & -\frac{834970144679}{80640} \\
\frac{12704059997}{576} & -\frac{16319852709}{256} & \frac{15129347145}{224} & -\frac{835641321}{64} & -\frac{174295335}{8} & \frac{84345575}{8} \\
-\frac{1139532537}{256} & \frac{4412607485}{448} & -\frac{835641321}{64} & \frac{177027387}{16} & -\frac{44759699}{8} & \frac{121755499}{96} \\
-\frac{2594381497}{448} & \frac{1432051361}{64} & -\frac{174295335}{8} & -\frac{44759699}{8} & \frac{905557445}{48} & -\frac{15035503}{2} \\
\frac{224319967199}{80640} & -\frac{834970144679}{80640} & \frac{84345575}{8} & \frac{121755499}{96} & -\frac{15035503}{2} & 3104765
\end{bmatrix},
$$

where the coefficient matrix of $C_{3,f}^{p,t}(p,\alpha)$ is positive definite, and therefore $C_{3,f}^{p,t}(p,\alpha) > 0$. Hence, the inequality in Eq. (A.14) holds, and the proof is complete. ■

Bibliography

[1] T. Arai, E. Pagello, and L. E. Parker. Editorial: Advances in multi-robot systems. *IEEE Transactions on Robotics and Automation*, 18(5):655–661, October 2002.

[2] M. Athans and P. Falb. *Optimal Control*. McGraw-Hill, 1966.

[3] H. Axelsson, A. Muhammad, and M. Egerstedt. Autonomous formation switching for multiple, mobile robots. In *Proceedings of the IFAC Conference on Analysis and Design of Hybrid Systems*, 2003.

[4] R. Beard and T. McLain. Multiple UAV cooperative search under collision avoidance and limited range communication constraints. In *Proceedings of the IEEE Conference on Decision and Control*, 2003.

[5] R. Beard, T. McLain, M. Goodrich, and E. Anderson. Coordinated target assignment and intercept for unmanned air vehicles. *IEEE Transactions on Robotics and Automation*, 18(6), December 2002.

[6] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA, 2nd edition, 2001.

[7] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro–Dynamic Programming*. Athena Scientific, 1996.

[8] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Belmont, MA, 1997.

[9] D. Bertsimas and G. Van Ryzin. A stochastic and dynamic vehicle routing problem in the euclidean plane. *Operations Research*, 39(4), July-August 1991.

[10] D. Bertsimas and G. Van Ryzin. Stochastic and dynamic vehicle routing in the euclidean plane with multiple capacitated vehicles. *Operations Research*, 41(1), January-February 1993.

[11] V. Blondel, J. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *IEEE Conference on Decision and Control (submitted)*, 2005.

[12] M. Brand and D. Nikovski. Optimal parking in group elevator control. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 1, pages 1002–1008, 2004.

[13] A. Bryson and Y. Ho. *Applied Optimal Control*. Hemisphere Publishing, New York, 1975.

[14] F. Bullo and J. Cortés. Adaptive and distributed coordination algorithms for mobile sensing networks. In V. Kumar, N. E. Leonard, and A. S. Morse, editors, *Cooperative Control. (Proceedings of the 2003 Block Island Workshop on Cooperative Control)*, volume 309 of *Lecture Notes in Control and Information Sciences*, pages 43–62. 2004.

[15] W. Burgard, D. Fox, M. Moors, R. Simmons, and S. Thrun. Collaborative multi-robot exploration. In *Proceedings of the IEEE International Conference on Robotics and Automation, San Francisco, CA*, 2000.

[16] Y. Cao, A. Fukunaga, and A. Kahng. Cooperative mobile robotics: Antecedents and directions. *Autonomous Robots*, 4:1–23, 1997.

[17] R. Cassinis, G. Bianco, A. Cavagnini, and P. Ransenigo. Strategies for navigation of robot swarms to be used in land mines detection. In *Proceedings of Eurobot*, 1999.

[18] H. Choset. *Sensor based motion planning: The hierarchical generalized voronoi graph*. PhD thesis, California Institute of Technology, 1996.

[19] J. Cortés, S. Martínez, and F. Bullo. Coordinated deployment of mobile sensing networks with limited-range interactions. In *Proceedings of the IEEE Conference on Decision and Control*, 2004.

[20] J. Cortés, S. Martínez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, April 2004.

[21] B. DasGupta, J. Hespanha, and E. Sontag. Computational complexity of honey-pot searching with local sensory information. In *Proceedings of the American Control Conference*, 2004.

[22] B. Davey and H. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990.

[23] J. De Mot and E. Feron. Optimal agent cooperation with local cost information. Technical Report LIDS-P-2633, MIT, November 2004.

[24] J. De Mot and E. Feron. Optimal two-agent navigation with local environment information. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, October 2004.

[25] J. De Mot and E. Feron. Optimal agent group sizes for cooperative navigation with cone- and tunnel-shaped local observation. In *IEEE Conference on Decision and Control (submitted)*, December 2005.

[26] J. De Mot and E. Feron. Spatial distribution statistics for two-agent optimal navigation with cone-shaped local observation. In *Proceedings of the American Control Conference*, June 2005.

[27] W. Dunbar and R. M. Murray. Model predictive control of coordinated multi-vehicle formations. In *Proceedings of the IEEE Conference on Decision and Control*, Las Vegas, NV, 2002.

[28] M. Egerstedt and X. Hu. Formation constrained multi-agent control. In *Proceedings of the IEEE Conference on Robotics and Automation*, Seoul, Korea, May 2001.

[29] A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. In *Proceedings of the 15th IFAC Conference*, pages 283–288, 2002.

[30] M. Flint, E. Fernández-Gaucherand, and M. Polycarpou. Cooperative control for uav's searching risky environments for targets. In *Proceedings of the IEEE Conference on Decision and Control*, Maui, Hawaii, December 2003.

[31] E. Frazzoli and F. Bullo. Decentralized algorithms for vehicle routing in a stochastic time-varying environment. In *Proceedings of the IEEE Conference on Decision and Control*, 2004.

[32] E. Frazzoli, M. A. Dahleh, and E. Feron. Real-time motion planning for agile autonomous vehicles. *AIAA Journal of Guidance, Control, and Dynamics*, 25(1):116–129, 2002.

[33] J. Fredslund and M. Mataric. A general, local algorithm for robot formations. *IEEE Transactions on Robotics and Automation*, 18(5):837–846, October 2002.

[34] Y. Fukada. Speed to fly with management of the risk of landing out. *Technical Soaring*, XXV(3):88–94, 2000.

[35] B. Gerkey and M. Matarić. Multi-robot task allocation: Analyzing the complexity and optimality of key architectures. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, September 2003.

[36] J. P. Hespanha, H. Kim, and S. Sastry. Multiple-agent probabilistic pursuit-evasion games. In *Proceedings of the IEEE Conference on Decision and Control*, volume 3, pages 2432–2437, December 1999.

[37] J. P. Hespanha and M. Prandini. Optimal pursuit under partial information. In *Proceedings of the Mediterranean Conference on Control and Automation*, July 2002.

[38] J. P. Hespanha, M. Prandini, and S. Sastry. Probabilistic pursuit-evasion games: A one-step nash approach. In *Proceedings of the IEEE Conference on Decision and Control*, volume 3, pages 2272–2277, December 2000.

[39] F. Hillier and G. Lieberman. *Introduction to Operations Research*. McGraw-Hill, sixth edition edition, 1995.

[40] J. Ho and E. Loute. A comparative study of two methods for staircase linear programs. *ACM Transactions on Mathematical Software*, 6:17–30, 1980.

[41] J. Ho and R. Sundarraj. Distributed nested decompostion of staircase linear programs. *ACM Transactions on Mathematical Software*, 23(2):148–173, June 1997.

[42] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[43] R. Horn and C. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[44] F. Kluegl, J. Wahle, A. Bazzan, and M. Schreckenberg. Towards anticipatory traffic forecast-modelling of route choice behaviour. In *Proceedings of the Workshop on Agents in Traffic and Transportation (Autonomous Agents 2000)*, Barcelona, Spain, June 2000.

[45] J.-C. Latombe. *Robot Motion Planning.* Kluwer Academic Publishers, 1991.

[46] S. M. LaValle. *Planning Algorithms.* [Online], 2004. Available at http://msl.cs.uiuc.edu/planning/.

[47] W. Li and C. Cassandras. Distributed cooperative coverage control of sensor networks. In *IEEE Conference on Decision and Control (submitted)*, 2005.

[48] L. Liusternik and V. Sobolev. *Elements of Functional Analysis.* Ungar, NY, 1961.

[49] D. Luenberger. *Optimization by Vector Space Methods.* Wiley, NY, 1969.

[50] F. Mazur, R. Chrobok, S. Hafstein, A. Pottmeier, and M. Schreckenberg. Furture of traffic information - online-simulation of a large scale freeway network. In *IADIS International Conference*, pages 665–672, Madrid, Spain, October 2004.

[51] P. McCready. An optimal airspeed selector. *Soaring*, 1954.

[52] N. Moshtagh, A. Jadbabaie, and K. Daniilidis. Vision-based distributed coordination of multiagent systems. In *IEEE Conference on Robotics and Automation (submitted)*, Barcelona, Spain, 2005.

[53] D. Nikovski and M. Brand. Decision-theoretic group elevator scheduling. In *Proceedings of the International Conference on Automated Planning and Scheduling*, 2003.

[54] U. Ogras, O. Dagci, and U. Ozguner. Cooperative control of mobile robots for target search. In *Proceedings of the American Control Conference*, 2003.

[55] P. Ogren, E. Fiorelli, and N. Leonard. Formations with a mission: Stable coordination of vehicle group maneuvers. In *Proceedings of the Symposium on Mathematical Theory of Networks and Systems*, August 2002.

[56] P. Ogren, E. Fiorelli, and N. Leonard. Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment. *IEEE Transactions on Automatic Control*, 49(8):1292–1302, August 2004.

[57] R. Olfati-Saber and R. M. Murray. Flocking for multi-agent dynamic systems: Algorithms and theory. Technical Report CIT-CDS 2004-005, California Institute of Technology, June 2004.

[58] G. Olsder and G. Papavassilopoulos. About when to use a searchlight. *Journal of Mathematical Analysis and Applications*, 136:466–478, 1988.

[59] G. Olsder and G. Papavassilopoulos. A markov chain game with dynamic information. *J. Optim. Theory Appl.*, 59:467–486, 1988.

[60] G. Oriolo, G. Ulivi, and M. Vendittelli. Real-time map building and navigation for autonomous robots in unknown environments. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 28(3):316–333, June 1998.

[61] D. P. de Farias and B. Van Roy. On constraint sampling for the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.

[62] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Ser. B*, 96(2):293–320, 2003.

[63] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, 2004.

[64] D. Pynadath and M. Tambe. The communicative multiagent team decision problem: Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, 16:389–423, 2002.

[65] S. Rajko and S.M. Lavalle. A pursuit-evasion bug algorithm. In *Proceedings of the IEEE International Conference on robotics and Automation*, pages 1954–1960, 2001.

[66] N. S. V. Rao, N. Manickam, and V. Protopopescu. Cooperative terrain model acquisition by a team of two or three point-robots. In *Proceedings of the International Conference on Robotics and Automation*, volume 2, pages 1427–1433, 1996.

[67] H. Reichmann. *Cross-Country Soaring*. Thomson Publications, 1978.

[68] G. Ribichini and E. Frazzoli. Energy-efficient coordination of multiple-aircraft systems. In *Proceedings of the IEEE Conference on Decision and Control*, 2003.

[69] P. Rybski, S. Stoeter, M. Erickson, M. Gini, D. Hougen, and N. Papanikolopoulos. A team of robotic agents for surveillance. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 9–16, Barcelona, Catalonia, Spain, 2000. ACM Press.

[70] M. Saptharisthi, C. Oliver, C. Diehl, K. Bhat, J. Dolan, A. Trebi-Ollennu, and P. Khosla. Distributed surveillance and reconnaissance using multiple autonomous atvs: Cyberscout. *IEEE Transactions on Robotics and Automation: Special Issue on Advances in Multirobot Systems*, 18(5), 2002.

[71] S. Sastry. *Nonlinear Systems: Analysis, Stability, and control*. Springer-Verlag, New York, 1999.

[72] T. Schouwenaars, J. How, and E. Feron. Receding horizon path planning with implicit safety guarantees. In *Proceedings of the American Control Conference*, Boston, MA, July 2004.

[73] R.G. Simmons, D. Apfelbaum, W. Burgard, D. Fox, M. Moors, S. Thrun, and H. Younes. Coordination for multi-robot exploration and mapping. In *AAAI/IAAI*, pages 852–858, 2000.

[74] M. Sipser. *Introduction to the Theory of Computation*. PWS Publishing Company, Boston, MA, 1997.

[75] V. Sujan and S. Dubowsky. Visually built task models for robot teams in unstructured environments. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, May 2002.

[76] H. Tanner, A. Jadbabaie, and G. Pappas. Stable flocking of mobile agents part I: Fixed topology. In *Proceedings of the IEEE Conference on Decision and Control*, Maui, Hawaii, December 2003.

[77] H. Tanner, A. Jadbabaie, and G. Pappas. Stable flocking of mobile agents part II: Dynamic topology. In *Proceedings of the IEEE Conference on Decision and Control*, Maui, Hawaii, December 2003.

[78] S. Thrun. A probabilistic online mapping algorithm for teams of mobile robots. *International Journal of Robotics Research*, 20(5):335–363, 2001.

[79] J. N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Department of EECS, MIT, November 1984.

[80] Y. Yang, A. Minai, and M. Polycarpou. Decentralized cooperative search by networked UAVs in an uncertain environment. In *Proceedings of the American Control Conference*, 2004.