A Theory of Satisficing Control

Michael A. Goodrich

Department of Electrical and Computer Engineering

Ph.D. Degree, August 1996

ABSTRACT

The existence of an optimal control policy and the techniques for finding it are grounded fundamentally in a superlative perspective. These techniques can be of limited value when the global behavior of the system is difficult to characterize, as it may be when the system is nonlinear, when the input is constrained, or when only partial information is available regarding system dynamics or the environment. Satisficing control theory is an alternative approach that is compatible with such systems. This theory is extended by the introduction of the notion of strongly satisficing to provide a rigorous, systematic procedure for the design of satisficing controllers which are consistent with optimal control theory.

Because they are often difficult to solve optimally, one application of satisficing control theory is to nonlinear control problems. Of particular interest are the nonlinear quadratic regulator and nonlinear minimum time problems. A controller synthesis procedure and resulting solutions are presented for single agent nonlinear control problems. The application of the satisficing principle to single agent control problems establishes the validity of the principle in a control setting, and motivates its use in multiple agent problems.

Multiple agent control problems have proven difficult to solve using classical control methods. Agents may disagree about what constitutes optimality and, even if they agree, an optimal solution may be unknown or computationally impractical. These difficulties are to a significant degree resolved by applying satisficing control theory. A controller synthesis procedure is developed for a general three-team problem with pursuit-evasion and nonlinear regulator characteristics. Simulations demonstrate that the resulting controllers not only perform well in both cooperative and non-cooperative environments, but also that performance degrades gracefully in the presence of partial information.

COMMITTEE APPROVAL:

Wynn C. Stirling
Committee Chairman

Richard L. Frost
Committee Member

Todd K. Moon
Committee Member

A. Lee Swindlehurst
Committee Member

David J. Comer
Committee Member

Wynn C. Stirling
Graduate Coordinator

A Theory of Satisficing Control

A Dissertation

Presented to the

Department of Electrical and Computer Engineering

Brigham Young University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Michael A. Goodrich

August 1996

This dissertation by Michael A. Goodrich is accepted in its present form by the Department of Electrical and Computer Engineering of Brigham Young University as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Wynn C. Stirling
Committee Chairman

Richard L. Frost
Committee Member

Todd K. Moon
Committee Member

A. Lee Swindlehurst
Committee Member

David J. Comer
Committee Member

Date

Wynn C. Stirling
Graduate Coordinator

DEDICATION

To my sweetheart
Patricia
and our sons
Daniel and David

ACKNOWLEDGEMENTS

I appreciate Dr. Wynn C. Stirling for his thoughtful mentoring during the completion of this research. The academic exercises and personal interactions with him, and Dr. Richard L. Frost, made my doctoral experience not only intellectual challenging and rewarding, but also pleasant. Thanks also go to Dr. Todd K. Moon, Dr. A. Lee Swindlehurst, and Dr. David J. Comer for their feedback and questions on my research and dissertation. I also appreciate the financial support of the Electrical and Computer Engineering Department at Brigham Young University.

All of my successes are associated with my wife Patricia. I feel a measure of success in finishing this research, but I feel more success in our happy marriage and caring relationship. While completing school, we have added two little boys to our family. If this work had interfered with their development I would not have proceeded. With the love from my wife and our boys, this is a family accomplishment.

Contents

Dedication **iii**

Acknowledgements **iv**

Glossary **xi**

Glossary of Acronyms **xv**

1 Introduction and Significance **1**
- 1.1 Problem Statement . . . 1
- 1.2 Historical Background and Related Literature . . . 2
- 1.3 Problem Definition . . . 5
- 1.4 Contributions . . . 6

2 Single Agent Satisficing Control: Theory **8**
- 2.1 Preliminaries: Epistemic Utility Theory . . . 8
- 2.2 Satisficing Control . . . 10
- 2.3 Strongly Satisficing Control . . . 16
- 2.4 An Important Special Case . . . 19

3 Single Agent Satisficing Control: Applications **24**
- 3.1 Receding Horizon Controller Synthesis . . . 24
- 3.2 Nonlinear Quadratic Regulator . . . 28
  - 3.2.1 Linear Quadratic Regulator . . . 30
  - 3.2.2 The Inverted Pendulum . . . 34
  - 3.2.3 The RTAC . . . 42
- 3.3 Nonlinear Minimum Time . . . 47
  - 3.3.1 Zermelo's Problem . . . 48
  - 3.3.2 Modified Zermelo's Problem . . . 51
- 3.4 Some Implementation Details . . . 52

4 Multiple Agent Satisficing Control: Theory **54**
- 4.1 Preliminaries . . . 54
- 4.2 Satisficing Control . . . 56
  - 4.2.1 Satisficing Decisions for Agent Coalitions . . . 56
  - 4.2.2 Consequences of Coalition Involvement . . . 60
  - 4.2.3 Satisficing Strategies: Coordination over Time . . . 62
- 4.3 Strongly Satisficing Control . . . 62
- 4.4 An Important Special Case . . . 63

5 Multiple Agent Satisficing Control: Applications 68
5.1 Multiple Agent Controller Synthesis 68
5.2 Pursuit-Evasion: Capture the Flag 72
5.2.1 Problem Formulation 72
5.2.2 CTF Epistemic System 74
5.2.3 Reactive Modeling 76
5.2.4 Strategic Decisions 76
5.2.5 Anticipatory Modeling 79
5.2.6 Cooperating Guards 81
5.2.7 Discussion of CTF 83
5.3 Shared Resource Quadratic Regulator 84
5.3.1 Problem Formulation 84
5.3.2 Six-Agent Constrained Coordination 87
5.3.3 128-Agent Constrained Coordination 90
5.4 Some Implementation Details 93

6 Conclusions 94
6.1 Results 94
6.2 Discussion 95

A Suggestions for Further Work 100

B Taxonomy of Multiple Agent Problems 103
B.1 Introduction 103
B.2 Multiple Agent Taxonomic Characteristics 104
B.2.1 Preference Knowledge 104
B.2.2 Coordination 105
B.2.3 Strategy 105
B.3 Game Theory 106
B.4 Capture the Flag 107
B.4.1 Reactive Modeling 108
B.4.2 Strategic Decisions 108
B.4.3 Anticipatory Modeling 109
B.4.4 Cooperating Guards 109
B.5 Robot Rugby 110
B.6 Discussion 111

C Accuracy and Rejectability Sets 113
C.1 Convex Utilities 113
C.2 Definitions of Admissibility 114
C.3 Discussion 115

Vitae 117

Bibliography **117**

List of Tables

3.1 Zermelo's problem results. 50

5.1 Simulation results for varying indexes of boldness for attacker and guards using reactive modeling: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$. 78

5.2 Simulation results for varying indexes of boldness for attacker using reactive modeling coupled with strategic decisions, and guards using reactive modeling: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$. . . . 79

5.3 Simulation results for varying indexes of boldness for attacker using reactive modeling coupled with strategic decisions, and guards using anticipatory modeling: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$. . 80

5.4 Simulation results for varying indexes of boldness for strategizing attacker using reactive modeling , and guards coordinating their decisions based on an anticipatory model of the attacker: (a) $b_a =$ 1.0, (b) $b_a = 0.5$, and (c) $b_a = 0.0$. 83

List of Figures

2.1 Satisficing regions for a concave $\bar{f}_A$ and convex $\bar{f}_R$ plotted as densities. 22
2.2 Satisficing regions for a concave $\bar{f}_A$ and convex $\bar{f}_R$ cross plotted. . . 23
3.1 Performance results for the single-agent linear regulator problem: (a) control history, (b) phase plane. 33
3.2 Inverted pendulum on a cart. 34
3.3 Phase planes for the inverted pendulum on a cart (m = constant): (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second). 37
3.4 Control inputs for the inverted pendulum on a cart (time measured in 0.01 sec increments). 38
3.5 Polarity switch phenomenon in terms of accuracy and rejectability densities: (a) before switch and (b) after switch. 39
3.6 Phase planes for the inverted pendulum on a cart (m = random): (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second). 39
3.7 Phase planes for the inverted pendulum on a cart with constrained input: (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second). 40
3.8 Control inputs for the inverted pendulum on a cart with constrained controls (time measured in 0.01 sec increments). 41
3.9 Phase planes for the inverted pendulum on a cart with control randomly selected from strongly satisficing set: (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second). 41
3.10 Control inputs for the inverted pendulum on a cart with control randomly selected from strongly satisficing set (time measured in 0.01 sec increments). 42
3.11 Translational oscillator with rotational actuator. 43
3.12 Phase planes for the satisficing RTAC controller: (a) rotational proof mass phase plane (in radians and radians per second), (b) translational cart phase plane (in meters and meters per second). . 45
3.13 Phase planes for Bupp's RTAC controller: (a) rotational proof mass phase plane (in radians and radians per second), (b) translational cart phase plane (in meters and meters per second). 46
3.14 Control history for the RTAC satisficing and Bupp's controllers. . . 46
3.15 Geometry for Zermelo's problem. 49
3.16 Performance results for the single-agent Zermelo's problem problem: (a) comparison of optimal, one-step satisficing, and two-step satisficing, and (b) comparison of optimal with two-step satisficing for varying index of boldness. 51

3.17 1-step satisficing solutions for modified Zermelo's problem. 52
4.1 The relationships between S^1, $G^1_{\mathcal{K}(k)}$, and $H^1_{\mathcal{K}(k)}$, assuming f^1_A is concave and f^1_R is convex. 61
4.2 Satisficing regions for a concave $\bar{f}_A$ and convex $\bar{f}_R$. 66
5.1 Typical results for $(b_a, b_g) = (1.0, 0.5)$ (a) reactive guards and reactive attacker, (b) reactive guards and strategizing attacker, (c) anticipating guards and strategizing attacker, and (d) coordinating anticipating guards and strategizing attacker. 77
5.2 Maximum number of wins and losses versus simulation set. 84
5.3 Phase plane performance results for the multi-agent shared-resource linear regulator problem: (a) constrained optimal, (b) satisficing (b=1.2). 88
5.4 Energy consumption for optimal and satisficing controllers. 89
5.5 Control history results for the multiple-agent shared-resource linear regulator problem: (a) optimal controller: (b) satisficing controller. 89
5.6 Binary tree of coordinating agents. 90
5.7 Phase plane performance results for the multi-agent shared-resource linear regulator problem: (a) constrained optimal, (b) satisficing for $b = 1.6$. 92
5.8 Control history for 128 agents sharing a bounded resource: constrained optimal versus constrained satisficing. 93
B.1 Taxonomy of coordination types. 103
B.2 Robot rugby playing field. 110

GLOSSARY

$\|\cdot\|$	Vector norm.
$\ll$	Absolutely continuous.
$\perp$	Mutually singular.
∇	Gradient.
$\propto$	Proportional to.
$^O, ^P, ^E$	When used as superscripts, team nominal, team pursuer, and team evader.
^{T}	When used as a superscript, transpose.
a	Attacker.
A, A	Linear system matrices.
$\mathcal{A}$	Convex set of accuracies.
b	Rejectivity index.
B, B	Linear system matrices.
$\mathcal{B}$	σ-field associates with decision space (single and multiple agent).
$\mathcal{C}$	Constraint set (single or multiple agent).
δ	Delta function.
D_t	Explicit influence horizon (single or multiple agent).
$\mathfrak{E}$	Equilibrium set (single or multiple agent).
$\varepsilon(G, \vartheta)$	Epistemic utility function.
$\mathbf{f}$	System dynamics vector.
$f_A(u; \vartheta)$	Accuracy transition probability density (single agent).
$f_{\mathbf{A}}^j(\mathbf{u}; \vartheta)$	Joint accuracy transition probability density from agent j's perspective (multiple agent).
$\bar{f}_A(u)$	Expected accuracy probability density (single agent).
$\bar{f}_{\mathbf{A}}^j(\mathbf{u}_{\mathcal{K}})$	Expected joint accuracy probability density from agent j's perspective (multiple agent).
$f_{\mathbf{BC}}^j(\mathbf{u}, \vartheta)$	Subjective belief density of other agents and state of nature from agent j's perspective (multiple agent).
$f_C(\vartheta)$	Subjective probability density of state of nature (single agent).
$f_{\mathbf{C}}^j(\vartheta)$	Subjective probability density of state of nature from agent j's perspective (multiple agent).
$f_R(u; \vartheta)$	Rejectability transition probability density (single agent).
$f_{\mathbf{R}}^j(\mathbf{u}; \vartheta)$	Joint rejectability transition probability density from agent j's perspective (multiple agent).
$\bar{f}_R(u)$	Expected accuracy probability density (single agent).
$\bar{f}_{\mathbf{R}}^j(\mathbf{u}_{\mathcal{K}})$	Joint expected accuracy probability density from agent j's perspective (multiple agent).

$\mathcal{F}$	σ-field associated with state of nature Θ (single and multiple agent).
F	Context dependent.
$\mathbf{g}$	System vector.
g_i	Guard i.
g_A	Shifted accuracy cost function.
G	Context dependent.
G_A	Normalizing term for accuracy.
G^j	From agent j's perspective, the set of possible decisions that may occur when agents in a coalition act independently, but in good faith.
g_R	Shifted rejectability cost function.
G_R	Normalizing term for rejectability.
$h(u)$	The Radon-Nikodym derivative of $\bar{P}_{R_1}$ and $\bar{P}_A$ on $\mathcal{B}$. Also called the likelihood ratio (single or multiple agent).
H^j	From agent j's perspective, the set of possible decisions that may occur when agents in a coalition choose controls that guarantee the coalition's decision is satisficing when made in good faith.
$I_W(v)$	Indicator function.
$J, \mathcal{J}$	Quadratic cost functions.
J_A	Accuracy cost function.
J_R	Rejectability cost function.
L	Path cost function.
Λ	Rejectability local performance index (single or multiple agent).
$\mathcal{M}$	Multiple agent epistemic system.
$m(G)$	Utility of rejecting G (single agent).
$\mu_b(\cdot)$	The discrimination measure or density (single or multiple agent).
$\nu(G; u)$	Levi's epistemic utility function (single agent).
Ω_r	The set of all decisions u that achieve rejectability level $r \in [0, 1]$ (single or multiple agent).
$\mathbf{p}$	Position vector.
P	Terminal state cost weighting matrix.
$P_A(G; \vartheta)$	Accuracy transition probability measure (single agent).
$P^j_{\mathbf{A}}(G; \vartheta)$	Joint accuracy transition probability measure from agent j's perspective (multiple agent).
$\bar{P}_A(G)$	Expected accuracy probability measure (single agent).
$\bar{P}^j_{\mathbf{A}}(G)$	Expected joint accuracy probability measure from agent j's perspective (multiple agent).

$P^j_{\mathbf{BC}}(G, F)$	Subjective belief probability measure of other agents and state of nature from agent j's perspective (multiple agent).
$P_C(F)$	Subjective probability measure of state of nature (single agent).
$P^j_{\mathbf{C}}(F)$	Subjective probability measure of state of nature from agent j's perspective (multiple agent).
$P_R(G; \vartheta)$	Rejectability transition probability measure (single agent).
$P^j_{\mathbf{R}}(G; \vartheta)$	Joint rejectability transition probability measure from agent j's perspective (multiple agent).
$\bar{P}_R(G)$	Expected accuracy probability measure (single agent).
$\bar{P}^j_{\mathbf{R}}(G)$	Joint expected accuracy probability measure from agent j's perspective (multiple agent).
Φ	Accuracy local performance index (single or multiple agent).
$q(u)$	Subjective probability density regarding u (single agent).
Q, Q	State cost to go weighting matrices.
$R_u, R_{\boldsymbol{u}}, \mathsf{R}_{\boldsymbol{u}}$	Control cost weighting matrices.
R, etc.	$R_{k_1} \cdots R_{k_K}$, where k_i set is defined from context.
$\Re$	Real numbers.
$\mathcal{R}$	Convex set of rejectabilities.
S_b	Maximal satisficing set (single agent).
$\mathfrak{S}_b$	Strongly satisficing set (single agent).
$S^{\parallel}$	Subset of the satisficing set on which accuracy is concentrated (single agent).
$S^{\perp}$	Subset of the satisficing set which is negligible with respect to accuracy (single agent).
$\mathcal{S}$	S-admissible set of controls.
$\tau(G; u)$	Utility of accepting G when u is true (single agent).
T	Sample time.
$\mathcal{T}$	T-admissible set of controls.
Θ	State of nature (single or multiple agents).
ϑ	Element of Θ (single or multiple agents).
U	Set of possible decisions (single agent).
u	Element of U (single agent)
$U_{1:M}$	Set of possible decisions for M-agent community (single or multiple agent).
$U_{\mathcal{K}}$	Set of possible decisions for community $\mathcal{K}$ (multiple agent).
$\mathbf{u}_{\mathcal{K}}$	$[u_{k_1}, \ldots, u_{k_K}]^T$, $\mathcal{K} = \{k_1, k_2, \ldots, k_K\}$ (multiple agent).
$\mathbf{u}_{\bar{\mathcal{K}}}$	$[u_{j_1}, \ldots, u_{j_{M-K}}]^T$, for each $j_i \in \{1, \ldots, M\} \setminus \mathcal{K}$ (multiple agent).
$\mathbf{v}$	Velocity vector.
$\mathcal{W}$	W-admissible set of controls.

W	Context dependent.
$\mathbf{x}$	State vector.
$\boldsymbol{\xi}$	Influence vector.
$\boldsymbol{\zeta}$	The constraint function.

GLOSSARY OF ACRONYMS

CTF	Capture The Flag
FLC	Fuzzy Logic Controller
JSLRT	Joint Satisficing Likelihood Ratio Test
LQR	Linear Quadratic Regulator
LRT	Likelihood Ratio Test
NQR	Nonlinear Quadratic Regulator
RTAC	Rotational/Translational ACtuator
SLRT	Satisficing Likelihood Ratio Test

Chapter 1

INTRODUCTION AND SIGNIFICANCE

1.1 Problem Statement

Control problems are typically characterized by two desiderata that may be in tension: (a) the commitment to achieving the fundamental goal of the controller, such as tracking, regulation, or terminal control, and (b) the commitment to a performance criterion, such as minimum control effort or minimum time. Usually, these two commitments are combined into a single performance index to be minimized by applying techniques based on Bellman's principle of optimality or Pontryagin's minimum principle. This view of control is designed to accomplish a very specific goal: obtain the best and only the best solution. Such solutions obey the *optimality* principle.

Optimal control should be preferred when possible, reasonable, and practical. For many problems, however, an optimal solution is either intractable, requires the imposition of possibly unwarranted assumptions, or is prohibitively expensive [69]. Controller designs for nonlinear systems and multiple agent problems are often difficult to obtain via the optimality paradigm. Usually in such cases, the engineering approach is either to find a modification of the problem such that the optimal solution to the modified problem is a satisfactory solution to the original problem (such as feedback linearization), or to adopt an *ad hoc* solution based on imprecise or vague notions of desirability. These approaches are, of course, problematic, and if they fail, the designer must seek an alternative paradigm.

The optimality principle and ad hoc design principles are not the only possible paradigms of rational choice. Economics has motivated the need for alternative paradigms that are commensurate with the available knowledge and capabilities: "broadly stated, the task is to replace ... global rationality ... with a kind of rational behavior that is compatible with the access to information and the computational capabilities that are actually possessed ..." [58]. Simon advocates the application of a minimum standard, or aspiration level: "If, instead of requiring that the pay-off be maximized, we require only that the pay-off exceed some given amount, then ...it is not necessary to determine whether there exists an alternative plan with a still higher pay-off" [58]. Solutions that meet a minimum standard, possibly obtained under constraints of partial information or restricted computation, are said to obey the *satisficing*[1] principle and, hence, are called satisficing solutions. The concept of satisficing is more general than the concept of optimality. An optimal control is clearly satisficing, but the notion of a minimum standard may persist even if a best solution either does not exist or is not attainable. Furthermore, a satisficing solution is distinct from a "suboptimal" solution, since the latter concept presupposes the existence of an optimal solution and usually represents a simplification of it. Since a satisficing solution must meet a minimum standard of performance, it is also different from *ad hoc* solutions, which are typically based on vague notions of desirability.

This dissertation presents a theory of satisficing control that:

1. Incorporates a performance index or other design principles.

2. Is amenable to a systematic design procedure.

3. Is capable of meaningful operation in the presence of limited, or local, system models and information.

4. Does not require restrictive modeling assumptions such as linearity, time-invariance, or perfect communication.

5. Meets a minimum standard of performance, and provides adequate performance when applied to problems for which optimal solutions are available.

1.2 Historical Background and Related Literature

Isaac Levi developed an epistemological theory based on the concept of *avoiding error* rather than *seeking truth*. The formal tool in this theory was the

[1] *Satisfice*=*satis*fy+ suf*fice*

set-valued epistemic utility function that was a combination of seeking truth and rejecting decisions with little information [33, 34].

The application of epistemic utility theory to engineering began at Brigham Young University in the 1980's. The first application to engineering was in Morrell's dissertation [46] which developed a theory of set-valued estimation. Morrell exploited the set-valued decision philosophy of Levi and the set theoretic estimation principles of others [11, 57], but did not explicitly apply epistemic utility theory [51]. The introduction and advocacy of epistemic utility theory as a decision rule applicable to estimation in engineering problems occured first in [64] and then subsequently in [26, 48, 62]. The application of epistemic utility theory to multiple-target tracking was presented in [42, 43].

Stirling and, later, Frost then began the development of a theory of satisficing control based on epistemic utility theory [14, 60, 61, 63]. Kenney's work on stochastic satisficing control is an important link in the development of the theory [27] and is a more complete treatment of the generalized decision theory presented in [47]. Since satisficing control is a natural fit for multiple agent problems, preliminary work in this context appeared in the early publications. The application of epistemic utility theory to multiple agents was also found in [44] where the theory was applied to the prisoner's dilemma.

The synthesis procedure developed in this dissertation is based on a receding horizon concept similar to Mayne and Michalska's work in [39, 41]. However, whereas their work is grounded in the optimality principle and *converges to the optimal* for sufficiently simple problems, this research is based on the satisficing principle and is shown to be *consistent with the optimal* when information is sufficient. Satisficing control theory has been developed for a receding horizon controller, but many other control techniques, such as ordinal optimization [22–24] and fuzzy logic [18], could be used in conjunction with the satisficing principle.

The satisficing control theory presented herein is a type of set theoretic controller. Although the concept of set theoretic controller is not new [15], there is no other epistemic utility based set-theoretic controller known to the author. The advantage of the epistemic utility based approach is the ability to specify controller admissibility using both the purpose of the controller and the cost associated with each control. This is in contrast to the work in [15] which considers only the purpose of the controller and largely ignores the cost of control.

Single agent satisficing control theory has not been designed to compete with existing solution techniques, but rather as a supplement to them. For single agent controllers, representative problems from linear and nonlinear control are presented. Of particular interest are the inverted pendulum and RTAC problems. Although solutions exist for the inverted pendulum problem, these solutions impose restrictions on the starting position of the pendulum, or do not completely regulate the cart about the origin [10, 28, 59, 67]. The RTAC problem serves as a benchmark for nonlinear controllers. The problem statement and one solution are given in [7].

The multiple agent controllers developed herein are relatively new and, consequently, literature is scarce. Game theory forms the basis for one multi-agent decision methodology [37]. It is, perhaps, the most complete treatment of multiple agent decision problems. However, game theory requires restrictive modeling assumptions which make it difficult to apply to real-time multiple agent control problems. Additionally, different agents may disagree about what constitutes optimality in a game (a non-zero sum N-person game may not have a unique equilibrium) which complicates the application of optimal control.

Other work on multiple agent controllers include developments in a linear quadratic context [56]. One idea of coordination is presented in [9], but this only applies to cooperation and not competition. Work on coalition formation can be found in [5], and a discussion of multiple agent resource allocation is available in [45].

There is a growing consensus that progress on many problems may be more rapid if the requirement for optimal behavior is relaxed in favor of *intelligent* behavior. Intelligent behavior is somewhat loosely defined, but is often understood as behavior which imitates in some meaningful way that of a competent biological organism in similar circumstances. Familiar intelligent control methodologies include fuzzy logic, artificial neural networks, genetic algorithms, and artificial intelligence [54]. These methods focus on various aspects of intelligent behavior; for example, fuzzy logic emphasizes making good decisions in the presence of imprecision, while artificial neural networks exploit a metaphor from biology to learn general patterns from examples. While satisficing control theory appears compatible with and complementary to other intelligent control techniques [17, 18], this dissertation considers solutions only for problems describable by state-space formulations. Multiple agent coordinated controllers using principles of artificial intelligence include [2, 3].

1.3 Problem Definition

The purpose of this dissertation is to present a theory of satisficing control and develop a synthesis procedure for constructing satisficing controllers. An agent is defined as a decision-maker capable of anticipating consequences of potential decisions, capable of forming a numerical utility representing preferences for these consequences, and required to choose a single decision. Single agent problems have only one such decision-maker whereas multiple agent problems have more than one decision-maker. In multiple agent problems, the potential consequences may be a result of the group decision. In developing a theory and a synthesis procedure, the following problems have been addressed:

- Theoretical Foundations:
  - Extend Levi's epistemic utility theory to single-agent control problems with uncountably infinite control spaces using measure theoretic principles. Define a satisficing set using this extension.
  - Refine the single-agent satisficing set by eliminating all actions for which an obviously better action exists. The refined satisficing set is called the strongly satisficing set.
  - Develop multiple agent satisficing decision theory. Treat non-cooperative coordinated decision-making and cooperative coordinated decision-making as important special cases of the general theory.
- Application to Control:
  - Present a synthesis method for single agent control based on a receding horizon control concept. Restrict attention to problems described

in state-space. Include nonlinear regulator and minimum time performance objectives.

– Present a synthesis method for multiple agent control based on a receding horizon control concept. Restrict attention to problems describable in state space. Include pursuit-evasion and regulator considerations.

– Demonstrate and validate the resulting controllers on representative control problems.

By restricting problems to state-space descriptions, the scope of the problem is limited. The receding horizon controller formulation is chosen for the same reason. These restrictions limit the way knowledge is formalized (state-space) and the computational aspects of finding a solution (receding horizon). Because the satisficing principle is the basis for a decision engine, the solutions presented using these restrictions should not be considered the only possible satisficing controller.

1.4 Contributions

Satisficing control theory is a new controller design paradigm based on the satisficing principle. This is in contrast to optimal control theory which is based on the superlative principle. Satisficing control theory is important because it can be applied to single agent nonlinear problems with success and, more importantly, because it can be applied to multiple agent control problems.

This dissertation presents a satisficing control theory. The theory and development includes the following contributions to theoretical control:

- Applies the satisficing principle to control problems.
- Introduces and develops the strongly satisficing and equilibrium sets.
- Establishes the satisficing likelihood ratio test using measure theoretic arguments.
- Develops a synthesis procedure for single agent satisficing control problems described in state space.

- Establishes consistency of single agent satisficing control with optimal control for both the nonlinear quadratic regulator and the nonlinear minimum time problem.
- Develops multiple agent satisficing control theory, including multi-agent strongly satisficing and equilibrium sets.
- Develops a synthesis procedure for multiple agent satisficing control problems described in state space and incorporating both pursuit-evasion and energy conservation characteristics.

These contributions coupled with the previous work done in satisficing estimation and decision theory establish the satisficing principle as applicable to many different problems and helps to justify future research pursuits.

Chapter 2

SINGLE AGENT SATISFICING CONTROL: THEORY

In this chapter, a review of Levi's epistemic utility theory is presented. Following this review, the application of epistemic utility theory to control is presented, including measure theoretic extensions to uncountable decision spaces[47]. This gives the foundations of satisficing control theory. Next, an important special case of this control theory is developed. All developments are restricted to single agent problems.

2.1 Preliminaries: Epistemic Utility Theory

The concept of optimal control admits a distinctive philosophical interpretation: seeking "the best and only the best" solution to a control problem is analogous to the epistemological stance of seeking "the truth and nothing but the truth" regarding an inquiry. This epistemological goal is a very ambitious one, and success cannot always be realized due to limitations of information and resources. In recognition of this fact, Isaac Levi proposed an alternative school of epistemological thought that professes a more modest goal than truth-seeking. The goal of this alternative school is *error avoidance*, and the methodology employed to achieve this goal is termed *epistemic utility theory* [34].

Epistemic utility theory employs two utilities, rather than one. The first utility is designed to characterize the truth content of the propositions being evaluated, and the second utility is designed to characterize the informational value of *rejecting* them. These two utilities are developed independently. For example, the truth content of rival scientific theories may be assessed through their conformance with observations, and their informational value may be assessed in terms of their simplicity, explanatory power, or predictive power. The fundamental content of an epistemic utility-based approach is that those propositions whose truth-content utility does not outweigh their informational-value-of-rejection utility should be rejected. All other propositions should be retained as being serious possibilities. In essence, this procedure acts to retain all propositions that are considered "good," without focusing exclusively on a search for one that is deemed "best." In this way, serious errors are avoided since no credible and valuable propositions will be eliminated from consideration. This approach results, in general, in a weaker decision than does a truth-seeking approach, since the set of unrejected propositions may not be a singleton set.

Formally, let V be a decision space and $\mathcal{B}_V$ the Boolean algebra defined from V. Levi considers elements of $W \in \mathcal{B}$ potential answers to an inquiry. Since the decision is based on *avoiding error* rather than *finding the optimal*, a set-valued answer to an inquiry is acceptable. The utility of avoiding error $\tau : \mathcal{B}_V \times V \mapsto \{0,1\}$ is defined as

$$\tau(W;v) = I_W(v) = \begin{cases} 1, & v \in W \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

which is unity when v, the true answer, is in W, and zero otherwise. This utility alone admits all decisions W that contain v.

Levi then considers the value of rejecting a decision W independent of its truth. This requires the definition of the utility of rejecting W, $m : \mathcal{B} \mapsto [0,1]$. The utility of accepting W, regardless of its truth value is denoted as

$$c(W) = 1 - m(W). \tag{2.2}$$

The epistemic utility function is the convex combination of these utilities

$$\nu(W;v) = \alpha\tau(W;v) + (1-\alpha)c(W). \tag{2.3}$$

Given a subjective probability measure q which relates the probability that W is true, the expected utility is, after a positive linear transformation of the epistemic utility,

$$E\nu(W;u) = q(W) - bm(W), \tag{2.4}$$

where $b = (1-\alpha)/\alpha$ is called the index of boldness and represents how bold an agent is in rejecting hypotheses. So that the value of rejecting a hypothesis is not preferred to truth, boldness is restricted to $[0,1]$. Levi then states that the set W_m which maximizes this expected utility is the decision that most likely avoids error. Consequently, any $v \in W_m$ is admissible according to the utility function.

2.2 Satisficing Control

For epistemic utility theory to have application to control, several adaptations must be made. In the cognitive context that motivated the origination of this theory, the goal is to make an assessment of truth (what to believe). For practical decision-making, the goal must be shifted to make an assessment of appropriateness (how to act). Also, in the cognitive context, the set of propositions to be decided upon may be viewed as the state of nature; i.e., the problem is an estimation problem. For control applications, however, decisions must be made over a set of actions that is usually distinct from the state of nature. Let V denote this set of actions, and let Π denote the state of nature. In the control context, Π is the set of all parameters that may influence the appropriateness of a control action. Thus, a separate action space must be introduced and the theory must be extended to account for the distinction between V and Π.

Epistemic utility theory is adapted to action by associating the subjective belief measure with the state of nature. Thus, the subjective belief measure $q(W)$ becomes a probability measure of the set W in the σ-field associated with Π. It is helpful to work in a derived probability space. Define $C : \Pi \mapsto \Theta$ as a one-to-one real random variable over $(\Pi, \mathcal{B}_\Pi, q)$ where $\mathcal{B}_\Pi$ represents the σ-field associated with Π. Define the derived probability measure $P_C(F) = q[C^{-1}(F)]$ for $F \subset \Theta$ and $F \in \mathcal{F}$ where $\mathcal{F}$ is the Borel field in Θ. The measure $P_C(F)$ represents belief that $F \in \mathcal{F}$ contains the actual state of nature.

The next step in adapting epistemic utility theory to action is to interpret the notions of truth and informational value in a practical setting. Again, it is useful to work in a derived probability space. To this end, the decision space V is associated with a subset of the real line U, and $\mathcal{B}$ is used to denote the Borel field in U.

The notion of "truth" is extended to the more general concept of "accuracy." The term *accuracy* is defined to mean *conformity to a given standard.* In the epistemological context, the standard is truth; in the practical arena of control engineering, the standard is the achievement of the fundamental goal of the controller. Consequently, control actions that have high accuracy will achieve the fundamental design goal of the controller (for example, set-point regulation). A utility function must be defined to characterize the accuracy of a set $G \in \mathcal{B}$ as parameterized by the state of nature. Define $A : V \mapsto U$ as a one-to-one real random variable over $(V, \mathcal{B}_V, \tau)$, where, for reasons that will soon become apparent, τ is now allowed to be an arbitrary transition probability measure instead of the indicator function used in epistemic utility theory. Define the derived distribution $P_A : \mathcal{B} \times \Theta \mapsto [0,1]$ as $P_A(G; \vartheta) = \tau[A^{-1}(G); \vartheta]$ for $G \subset U, G \in \mathcal{B}$ such that for fixed $\vartheta \in \Theta$, $P_A(\cdot; \vartheta)$ is a probability measure, and for fixed $G \in \mathcal{B}$, $P_A(G; \cdot)$ is a measurable function.

The notion of informational value of rejection must also be adapted to a practical, rather than cognitive, context. The term *rejectability* is introduced to characterize the conformance of elements of $\mathcal{B}$ to the design principles used to generate the control policy (for example, control effort). This utility is also required to be a transition probability defined via the one-to-one real random variable $R : V \mapsto U$, yielding $P_R : \mathcal{B} \times \Theta \mapsto [0,1]$ defined as $P_R(G; \vartheta) = m[R^{-1}(G); \vartheta]$. $P_R(G; \vartheta)$ is the utility of *rejecting* the set $G \in \mathcal{B}$ when ϑ is the state of nature. Once again, the utility of rejecting has been modified from a probability measure $m(W)$ to a transition probability measure $m(W; \vartheta)$ to account for the dependence on the state of nature.

The *epistemic utility function* [34, 35, 60, 64] is the convex combination of accuracy utility and rejectability utility:

$$\varepsilon(G, \vartheta) = \alpha P_A(G; \vartheta) + (1 - \alpha)(1 - P_R(G; \vartheta)), \tag{2.5}$$

where $\alpha \in [0, 1]$. After a positive linear transformation, a more convenient form for this utility function is

$$\varepsilon_a(G, \vartheta) = P_A(G; \vartheta) - bP_R(G; \vartheta), \tag{2.6}$$

where

$$b = (1 - \alpha)/\alpha. \tag{2.7}$$

The parameter, $b \in [0, \infty)$, is termed the *rejectivity index*, in the sense that large values of b imply an increased willingness to reject propositions. The condition $b \approx 0$ ($\alpha \approx 1$) corresponds to minimal rejectivity in that rejectability is ignored. The condition $b \approx 1$ ($\alpha \approx \frac{1}{2}$) corresponds to equal weighting of accuracy and rejectability. As $b \to \infty$ ($\alpha \to 0$), rejectability dominates accuracy, and eventually all propositions with nonzero rejectability value are rejected.

For fixed ϑ, $P_R(G;\vartheta)$ and $P_A(G;\vartheta)$ are probability measures, but neither represents the subjective *belief* concerning the event G. They are best understood as utility functions that happen to possess the structure of transition probabilities. By using utilities that are probability measures, it is possible to compare these utilities on the same scale so that one unit of accuracy corresponds to one unit of rejectability. The function $P_A(G;\vartheta)$ is a measure of how much accuracy is possessed by G when ϑ is the state of nature, and $P_R(G;\vartheta)$ is a measure of how much value accrues if G is rejected when ϑ is the state of nature. If the state of nature is known, then these utilities may be evaluated. Generally, however, the state of nature is viewed as a random variable represented by the subjective belief measure $P_C(F)$ for $F \in \mathcal{F}$. In [19, 27], Kenney defines the expected value of the epistemic utility function (2.6) as, for $G \in \mathcal{B}$,

$$\overline{\varepsilon}_a(G) = \int_\Theta [P_A(G;\vartheta) - bP_R(G;\vartheta)]dP_C = \bar{P}_A(G) - b\bar{P}_R(G), \tag{2.8}$$

where

$$\bar{P}_A(G) = \int_\Theta P_A(G;\vartheta)dP_C \tag{2.9}$$

$$\bar{P}_R(G) = \int_\Theta P_R(G;\vartheta)dP_C. \tag{2.10}$$

Let[1]

$$S_b = \arg\sup_{G\in\mathcal{B}} \overline{\varepsilon}_a(G) = \arg\sup_{G\in\mathcal{B}} \{\bar{P}_A(G) - b\bar{P}_R(G)\}. \tag{2.11}$$

If $\overline{\varepsilon}_a(S_b) > 0$, then S_b consists of the largest measurable set of propositions in U for which expected accuracy exceeds (after scaling by b) expected rejectability; that is, $\bar{P}_A(S_b) - b\bar{P}_R(S_b) > 0$. If $\bar{P}_A$ is absolutely continuous with respect to $\bar{P}_R$, then, as $b \to \infty$, $\overline{\varepsilon}_a(G) < 0$ for all $G \in \mathcal{B}$ such that $\bar{P}_R(G) > 0$. In this case, (2.11) implies that S_b must be a set of $\bar{P}_R$ measure zero (and also of $\bar{P}_A$ measure zero—see Theorem 1). S_b is termed the *maximal satisficing set* for rejectivity index b; if $G \subset S_b$, then G will be termed a *satisficing set*.

[1]The extension of this definition when controls are constrained is given in Chapter 4.

The epistemic utility function (2.6) is a generalization of the epistemic utility proposed by Levi. Setting $\Theta \equiv U$, $P_R(G; u) = m(G)$, and $P_A(G; u) \equiv I_G(u)$, where I_G is the indicator function for G, (2.6) is equivalent to the original epistemic utility function of(2.4). See [27] for details. Levi's original rule of epistemic utility [34, Page 53] generalizes to become: *Given a set of actions U, a rejectability utility P_R, an accuracy utility P_A, a probability P_C, and a rejectivity index, b, none of the elements of S_b given by (2.11) should be rejected* [19, 27].

The octuple $\{U, \mathcal{B}, \Theta, \mathcal{F}, P_A, P_R, P_C, b\}$ is termed an *epistemic system.* It is composed of three probability spaces: $\{\Theta, \mathcal{F}, P_C\}$, $\{U, \mathcal{B}, P_R(\cdot; \vartheta)\}$, and $\{U, \mathcal{B}, P_A(\cdot; \vartheta)\}$. The parameter b relates the latter two spaces and establishes the satisficing threshold.

The measures P_R and P_A, and, therefore, $\bar{P}_R$ and $\bar{P}_A$, are obtained independently. It is of interest, however, to establish any natural connections between these two measures. Performing the Lebesgue decomposition of $\bar{P}_R$ relative to $\bar{P}_A$ gives $\bar{P}_R = \bar{P}_{R_1} + \bar{P}_{R_2}$, where $\bar{P}_{R_1} \ll \bar{P}_A$ ($\bar{P}_{R_1}$ is absolutely continuous with respect to $\bar{P}_A$) and $\bar{P}_{R_2} \perp \bar{P}_A$ ($\bar{P}_{R_2}$ and $\bar{P}_A$ are mutually singular). By the definition of mutually singular measures, there exist disjoint sets B_1 and B_2 such that $\bar{P}_A(B_2) = \bar{P}_{R_2}(B_1) = 0$ and $U = B_1 \cup B_2$. Define $S^{\parallel} = S_b \cap B_1$ and $S^{\perp} = S_b \cap B_2$.

By the Radon-Nikodym theorem [53], there exists a nonnegative measurable function, h, termed the Radon-Nikodym derivative of $\bar{P}_{R_1}$ with respect to $\bar{P}_A$, such that for any $G \in \mathcal{B}$,

$$\bar{P}_{R_1}(G) = \int_G h(u) d\bar{P}_A.$$

The function h is also called the *likelihood ratio* between $\bar{P}_{R_1}$ and $\bar{P}_A$ on $\mathcal{B}$. It is convenient to define the *discrimination measure*,

$$\mu_b(G) = \bar{P}_A(G) - b\bar{P}_R(G), \tag{2.12}$$

for $G \in \mathcal{B}$, where the subscript b indicates that rejectivity is a parameter. The discrimination measure may be written

$$\begin{aligned} \mu_b(G) &= \bar{P}_A(G) - b\bar{P}_R(G) = \bar{P}_A(G) - b[\bar{P}_{R_1}(G) + \bar{P}_{R_2}(G)] \\ &= \bar{P}_A(G) - b\int_G h(u) d\bar{P}_A - b\bar{P}_{R_2}(G). \end{aligned} \tag{2.13}$$

Although (2.11) defines the maximal satisficing set, it is useful to characterize its properties in terms of the Radon-Nikodym derivative.

Theorem 1 *Let* $S^* = \{u \in U : h(u) \leq \frac{1}{b}\} \cap S_b$*. Then* $\bar{P}_A(S_b - S^*) = 0$*; that is, the subset of the maximal satisficing set over which the Radon-Nikodym derivative exceeds the bound* $\frac{1}{b}$ *has* $\bar{P}_A$ *measure zero.*

<u>*Proof*</u> First, observe that $S^* = (S^* \cap B_1) \cup (S^* \cap B_2)$. Since by definition, B_2 is $\bar{P}_A$-negligible, it may be assumed without loss of generality that $S^* \subset B_1$. Since S_b is the union of the disjoint sets $S^{\|}$ and $S^{\perp}$ and $\bar{P}_A(S^{\perp}) = 0$, only the set $S^{\|}$ need be considered. Let $S_* = S^{\|} - S^*$. Then $\mu_b(S_*) = \bar{P}_A(S_*) - b\int_{S_*} h(u)d\bar{P}_A$. On S_*, however, $h(u) > \frac{1}{b}$, so $b\int_{S_*} h(u)d\bar{P}_A \geq \int_{S_*} d\bar{P}_A = \bar{P}_A(S_*)$, so $\mu_b(S_*) \leq \bar{P}_A(S_*) - \bar{P}_A(S_*) = 0$. But $S_* \subset S_b$ implies $\mu_b(S_*) \geq 0$, hence $\mu_b(S_*) = 0$, which means that $\bar{P}_A(S_*) = b\int_{S_*} h(u)d\bar{P}_A$. Since $h(u) > \frac{1}{b}$ on S_*, the only way equality can hold is for $\bar{P}_A(S_*) = 0$. ■

For satisficing controller synthesis, it is useful to reconstruct the previous theoretical developments in terms of accuracy and rejectability densities. To this end, define the action, or control, space, U as the interval $U = [u_{min}, u_{max}] \subset \Re$. Assume that the accuracy and rejectability distribution functions assign all of the mass to this interval. Let $F_A(u;\vartheta) = P_A([u_{min}, u);\vartheta)$, and $F_R(u;\vartheta) = P_R([u_{min}, u);\vartheta)$ denote these distribution functions. The expected accuracy distribution, $\bar{F}_A(u) = \bar{P}_A([u_{min}, u))$, and the expected rejectability distribution, $\bar{F}_R(u) = \bar{P}_{R_1}([u_{min}, u))$ can also be constructed.

In this dissertation, $\Theta \subset \Re^n$ for $n \geq 1$ and $F_C(\vartheta)$ is defined as the distribution function corresponding to the measure P_C. The probability, P_C, corresponds to a (possibly subjective) distribution over the states of nature, Θ, that represents beliefs regarding the elements of Θ. Additionally, attention is restricted to continuous distributions so that F_A, F_R and F_C are all differentiable, with density functions $f_A(u;\vartheta) = \frac{d}{du}F_A(u;\vartheta)$, $f_R(u;\vartheta) = \frac{d}{du}F_R(u;\vartheta)$ and $f_C(\vartheta) = \frac{d}{d\vartheta}F_C(\vartheta)$, respectively. Rewriting (2.9) and (2.10) in terms of density functions, the expected accuracy and expected rejectability densities are

$$\bar{f}_A(u) = \int_\Theta f_A(u;\vartheta)f_C(\vartheta)d\vartheta \tag{2.14}$$

$$\bar{f}_R(u) = \int_\Theta f_R(u;\vartheta)f_C(\vartheta)d\vartheta. \tag{2.15}$$

In a deterministic environment, the state of nature is known, and the probability density function will concentrate all of the mass on the true value; that is, if ϑ_0 is the true state of nature, then $f_C(\vartheta) = \delta(\vartheta - \vartheta_0)$, so $\bar{f}_A(u) = f_A(u; \vartheta_0)$ and $\bar{f}_R(u) = f_R(u; \vartheta_0)$. In a stochastic environment, f_C will characterize the uncertainty regarding the state of nature. This uncertainty may enter, for example, through process noise, sensor noise, or other random phenomenon in the system. Throughout the remainder of this dissertation, only the deterministic problem will be treated. It is assumed that either the true state of nature, ϑ_0, is known or that f_C places all of its mass on one particular estimate $\hat{\vartheta}_0$. Kenney treats the stochastic case in [27].

It is straightforward to see that the Radon-Nikodym derivative of $\bar{P}_{R_1}$ with respect to $\bar{P}_A$ becomes $h(u) = \frac{\bar{f}_R(u)}{\bar{f}_A(u)}$ almost everywhere with respect to $\bar{P}_A$. If $\bar{P}_A(S_b) > 0$, (2.11) becomes

$$\begin{aligned} S_b &= \arg\sup_{G \in \mathcal{B}} \left\{ \int_G [1 - bh(u)] \bar{f}_A(u) du \right\} \\ &= \arg\sup_{G \in \mathcal{B}} \left\{ \int_G [\bar{f}_A(u) - b\bar{f}_R(u)] du \right\}. \end{aligned} \tag{2.16}$$

If $\bar{P}_A(S_b) = 0$, (2.11) becomes

$$S_b = \arg \sup_{\substack{G \in \mathcal{B} \\ \bar{P}_A(G)=0}} \sup_{u \in G} \left\{ \bar{f}_A(u) - b\bar{f}_R(u) \right\}. \tag{2.17}$$

The Radon-Nikodym derivative is the basis for forming a Likelihood Ratio Test (LRT). In the context of satisficing control theory, applying the Radon-Nikodym derivative to form (2.16) or (2.17) gives the Satisficing Likelihood Ratio Test (SLRT). By defining the discrimination function as $\mu_b(u) = \bar{f}_A(u) - b\bar{f}_R(u)$, (2.16) and (2.17) become, respectively,

$$\begin{aligned} S_b &= \arg\sup_{G \in \mathcal{B}} \int_G \mu_b(u) du \\ S_b &= \arg \sup_{\substack{G \in \mathcal{B} \\ \bar{P}_A(G)=0}} \sup_{u \in G} \mu_b(u). \end{aligned}$$

In consequence of these definitions, $S_b \neq \emptyset$.

In Levi's original context, boldness represented a proclivity to reject a decision and is restricted to the set $b \in [0, 1]$. In the control context, the term *boldness* can be misinterpreted. Consequently, in the control engineering context the term *boldness* is replaced by the term *rejectivity index*. The rejectivity index still represents a proclivity to reject a decision, but in the control context b can be any nonnegative value, $b \in [0, \infty)$. The larger the rejectivity index, the larger the set of rejected elements of U, since $b_1 < b_2$ implies $S_{b_1} \supset S_{b_2}$. S_b will generally not be a singleton set, and there may even be a continuum of satisficing values. In a satisficing context, the rejectivity index may be interpreted as an aspiration threshold [58]. In contrast to utility maximizing procedures, this approach relaxes the requirement for a unique best decision, and admits as satisficing all decisions for which the relative accuracy-to-rejectability value (the SLRT) is at least as great as the rejectivity index. Obviously, only one control can actually be implemented, but, strictly from a satisficing point of view, one may choose any of the unrejected control decisions with confidence that the action will yield good. Thus the designer has considerable latitude in the ultimate choice of the control to be implemented. In Section 2.3 a refinement to the satisficing set that further restricts the choice of controls is introduced.

Accuracy and rejectability utility functions are mechanisms to implement the goals and design ideals of the problem. If information is available over the full extent of the problem, then the utilities may be designed from a global perspective. If a global solution is not available or is not implementable, whatever information is available may still be incorporated to design the accuracy and rejectability utilities. By permitting the designer to tailor the structure of the utilities according to what is actually known or defensibly assumed, the problem may be cast in its natural setting. This capability frees the designer from the need to make arbitrary assumptions simply to invoke a global solution technique.

2.3 Strongly Satisficing Control

Satisficing, as defined herein, is a liberal notion of performance: broadly speaking, a control is satisficing if the good (characterized by accuracy) outweighs the bad (characterized by rejectability). Furthermore, the maximal satisficing set, S_b, will generally not be a singleton set, and there may be many satisficing possibilities. When determining a control law which is to govern a plant, a single input is required. Since each control u in the maximally satisficing set S_b is justifiable by the satisficing principle as an acceptable input to the plant, how can a single control be selected[2]?

Although all controls in S_b are satisficing, they are not necessarily all equal. In selecting a control, if a choice exists between two controls of equal rejectability but differing accuracy, it is reasonable to select the one with higher accuracy. Similarly, if a choice exists between two controls of equal accuracy, it is reasonable to select the one with lower rejectability. The set

$$\mathfrak{E} = \{u \in U : \nexists v \in U \text{ for which } \bar{f}_R(v) < \bar{f}_R(u) \text{ and } \bar{f}_A(v) > \bar{f}_A(u)\} \qquad (2.18)$$

is termed the *equilibrium* set. Whereas the satisficing set is determined by comparing a control's accuracy against its rejectability, the equilibrium set is determined by comparing controls against each other. Decisions $u \in \mathfrak{E}$ possesses an important multi-dimensional equilibrium property: within the set, perturbations in u cannot increase accuracy without also increasing rejectability, nor can rejectability be decreased without also decreasing accuracy. The *strongly satisficing* set is defined as the intersection of the satisficing set and the equilibrium set:

$$\mathfrak{S}_b = \mathfrak{E} \cap S_b. \qquad (2.19)$$

Observe that the definition of the equilibrium set and the strongly satisficing set apply for U finite, countable, or uncountable.

Lemma 2 *If U is closed, then S_b is closed.*

<u>*Proof:*</u> Let $\rho(u,v) = |\bar{f}_A(u) - \bar{f}_A(v)| + b|\bar{f}_R(u) - \bar{f}_R(v)|$. Then $\rho(u,v)$ defines a metric on the equivalence class of points with the same accuracy and the same rejectability. Let v be a point of closure of S_b. Then by definition, for all $\varepsilon > 0$ there is a $u_\varepsilon \in S_b$ such that $\rho(u_\varepsilon, v) < \varepsilon$. Now

$$\begin{aligned}
\varepsilon &> \rho(u_\varepsilon, v) \\
&= |\bar{f}_A(u_\varepsilon) - \bar{f}_A(v)| + b|\bar{f}_R(u_\varepsilon) - \bar{f}_R(v)| \\
&\geq [\bar{f}_A(u_\varepsilon) - b\bar{f}_R(u_\varepsilon)] - [\bar{f}_A(v) - b\bar{f}_R(v)]
\end{aligned}$$

[2]The necessity of answering this question is not restricted to satisficing controllers, but also arises for fuzzy logic controllers (FLC's) [12, 18, 40]. The solution for FLC's is to apply a defuzzification procedure that performs some kind of weighted average using the utility of each admissible control in the weighting.

implies that

$$\begin{aligned} \varepsilon + [\bar{f}_A(v) - b\bar{f}_R(v)] &> [\bar{f}_A(u_\varepsilon) - b\bar{f}_R(u_\varepsilon)] \\ &\geq 0. \end{aligned}$$

The last inequality holds since $u_\varepsilon \in S_b$ implies $\bar{f}_A(u_\varepsilon) - b\bar{f}_R(u_\varepsilon) \geq 0$. This is true for any $\varepsilon > 0$ which means that

$$\bar{f}_A(v) - b\bar{f}_R(v) \geq 0$$

implying that $v \in S_b$. Thus, S_b is closed. ■

Theorem 3 *For U closed, the strongly satisficing set $\mathfrak{S}_b \neq \emptyset$.*

Proof: This result is established by constructing a specific element of $\mathfrak{S}_b$. Define the *most discriminating* satisficing control,

$$u_D = \arg\sup_{z \in S_b} \{\bar{f}_A(z) - b\bar{f}_R(z)\}. \tag{2.20}$$

u_D exists by Lemma 2, but may be an equivalence class if (2.20) does not have a unique solution. Since $S_b \neq \emptyset$, u_D exists. By construction, $\bar{f}_A(u_D) - b\bar{f}_R(u_D) \geq \bar{f}_A(u) - b\bar{f}_R(u) \quad \forall u \in U$, or, equivalently,

$$\bar{f}_A(u_D) - \bar{f}_A(u) \geq b[\bar{f}_R(u_D) - \bar{f}_R(u)] \quad \forall u \in U. \tag{2.21}$$

For $\bar{f}_A(u) > \bar{f}_A(u_D)$ this implies $\bar{f}_R(u) > \bar{f}_R(u_D)$. For $\bar{f}_R(u) < \bar{f}_R(u_D)$, (2.21) implies $\bar{f}_A(u) < \bar{f}_A(u_D)$. Hence, there is no $u \in U$ such that both $f_R(u_D) < f_R(u)$ and $f_A(u_D) > f_A(u)$ which means that $u_D \in \mathfrak{L}$. ■

Thus $\mathfrak{S}_b$ always contains at least one element, namely a most discriminating satisficing control. If $b = 0$, the most discriminating control is the *most accurate* satisficing control, $u_A = \arg\sup_{z \in S_b}\{\bar{f}_A(z)\}$ obtains. This limiting case represents a very aggressive stance to achieve the goal at the risk of excessive cost. A most accurate satisficing control may be considered for cases with large variations in $\bar{f}_A$ and small variations in $\bar{f}_R$. Another limiting case occurs as $b \to \infty$, resulting in a *least rejectable* satisficing control, $u_R = \arg\inf_{z \in S_b}\{\bar{f}_R(z)\}$. This procedure is very conservative, and reflects a willingness to compromise the fundamental goal in the interest of reducing cost. It may be appropriate when there are large variations in $\bar{f}_R$ relative to small variations in $\bar{f}_A$. In Chapter 3 it will be demonstrated that any strongly satisficing decision is a justifiable decision.

2.4 An Important Special Case

An important class of accuracy density functions is the set of density functions that are concave over the support, U. That is, for $\lambda \in [0,1]$, $\lambda \bar{f}_A(u_1)+(1-\lambda)\bar{f}_A(u_2) \leq \bar{f}_A(\lambda u_1+(1-\lambda)u_2)$ for all $u_1 \in U$, $u_2 \in U$. Similarly, an important class of rejectability density functions is the set of density functions that are convex over the support, U. That is, for $\lambda \in [0,1]$, $\lambda \bar{f}_R(u_1)+(1-\lambda)\bar{f}_R(u_2) \geq \bar{f}_R(\lambda u_1+(1-\lambda)u_2)$ for all $u_1 \in U$, $u_2 \in U$. For these classes, the following theorems are important in developing a synthesis procedure.

Lemma 4 *For continuous accuracy and continuous rejectability,* $\mathfrak{L} \subset \{u \in U : \bar{f}_A^+(u)\bar{f}_R^+(u) \geq 0\} \cup \{u_A, u_R\}$ *where*

$$\bar{f}_A^+(u) = \lim_{\lambda \downarrow 0} \frac{\bar{f}_A(u+\lambda) - \bar{f}_A(u)}{\lambda} \tag{2.22}$$

denotes the right derivative of $\bar{f}_A$*, and similarly define* $\bar{f}_R^+$ *as the right derivative of* $\bar{f}_R$*.*

<u>*Proof:*</u> Since u_A and u_R represent the most discriminating controls for $b = 0$ and $b \to \infty$, respectively, by the proof of Theorem 3, they are in $\mathfrak{L}$. For any other $u \in \mathfrak{L}$, accuracy and rejectability are either both nonincreasing or both nondecreasing in the neighborhood of u. Hence, $\bar{f}_A^+(u)\bar{f}_R^+(u) \geq 0$. ■

If, in addition to being continuous, both $\bar{f}_A$ and $\bar{f}_R$ are differentiable with derivatives denoted $\bar{f}_A'$ and $\bar{f}_R'$, respectively, the equilibrium set satisfies

$$\mathfrak{L} \subset \{u \in U : \bar{f}_A'(u)\bar{f}_R'(u) \geq 0\} \cup \{u_A, u_R\}. \tag{2.23}$$

Moreover, for concave $\bar{f}_A$ and convex $\bar{f}_R$ the following lemma establishes a necessary and sufficient condition for determining the equilibrium set.

Lemma 5 *For differentiable concave* $\bar{f}_A$ *and differentiable convex* $\bar{f}_R$

$$\mathfrak{L} = \{u \in U : \bar{f}_A'(u)\bar{f}_R'(u) \geq 0\} \cup \{u_A, u_R\}.$$

<u>*Proof:*</u> Let $u \in \{v \in U : \bar{f}_A'(v)\bar{f}_R'(v) \geq 0\} \cup \{u_A, u_R\}$. If $u = u_A$ or $u = u_R$ then $u \in \mathfrak{L}$ since the most discriminating control is always in the equilibrium set.

Otherwise, u must satisfy $\bar{f}'_A(u)\bar{f}'_R(u) \geq 0$ which implies that $\exists\lambda \geq 0$ such that $\bar{f}'_A(u) - \lambda\bar{f}'_R(u) = 0$. However, since the sum of two concave functions ($\bar{f}_A(u)$ and $-\lambda\bar{f}_R(u)$) is also concave then the u which satisfies $\bar{f}'_A(u) - \lambda\bar{f}'_R(u) = 0$ is a most discriminating control for $b = \lambda$. Hence, $u \in \mathfrak{S}_b$ which implies that $u \in \mathfrak{E}$. Thus,

$$\mathfrak{E} \supset \{u \in U : \bar{f}'_A(u)\bar{f}'_R(u) \geq 0\} \cup \{u_A, u_R\}.$$

This result, coupled with Lemma 4 establishes the desired result. ■

Theorem 6 *Let $U \subset \Re$. For $\bar{f}_A$ a concave density function and $\bar{f}_R$ a convex density function over U the maximal satisficing set S_b is convex. Moreover, for concave differentiable accuracy and convex differentiable rejectability, the equilibrium set $\mathfrak{E}$ and the strongly satisficing set $\mathfrak{S}_b$ are convex.*

Proof: Convexity of the maximal satisficing set is shown by establishing that, for $u_1 \in S_b$ and $u_2 \in S_b$, the point $v_\lambda = \lambda u_1 + (1-\lambda)u_2 \in S_b$ for any $0 \leq \lambda \leq 1$. For v_λ to be in S_b,

$$\bar{f}_A(v_\lambda) - b\bar{f}_R(v_\lambda) > 0 \tag{2.24}$$

must hold. Since convexity and concavity imply

$$\begin{aligned} \bar{f}_A[\lambda u_1 + (1-\lambda)u_2] &\geq \lambda\bar{f}_A(u_1) + (1-\lambda)\bar{f}_A(u_2) \\ -\bar{f}_R[\lambda u_1 + (1-\lambda)u_2] &\geq -\lambda\bar{f}_R(u_1) - (1-\lambda)\bar{f}_R(u_2), \end{aligned}$$

substitution of these quantities into $\bar{f}_A(v_\lambda) - b\bar{f}_R(v_\lambda)$ gives

$$\bar{f}_A[\lambda u_1 + (1-\lambda)u_2] - b\bar{f}_R[\lambda u_1 + (1-\lambda)u_2] \geq \lambda[\bar{f}_A(u_1) - b\bar{f}_R(u_1)] + (1-\lambda)[\bar{f}_A(u_2) - b\bar{f}_R(u_2)].$$

Since $u_1, u_2 \in S_b$ implies that $\bar{f}_A(u_1) - b\bar{f}_R(u_1) > 0$ and $\bar{f}_A(u_2) - b\bar{f}_R(u_2) > 0$, it follows that

$$\bar{f}_A[\lambda u_1 + (1-\lambda)u_2] - b\bar{f}_R[\lambda u_1 + (1-\lambda)u_2] > 0,$$

which implies that for $0 \leq \lambda \leq 1$, the point $\lambda u_1 + (1-\lambda)u_2 \in S_b$. Consequently, the maximal satisficing set is convex.

To establish the convexity of $\mathfrak{E}$, first note that by Lemma 4

$$\mathfrak{E} \subset \{u \in U : \bar{f}'_A(u)\bar{f}'_R(u) \geq 0\} \cup \{u_A, u_R\}. \tag{2.25}$$

If $u_A = u_R$ then $\mathfrak{L} = \{u_A\}$. Otherwise, let $u_1 < u_2$, and suppose $\bar{f}'_A(u_1)$ and $\bar{f}'_A(u_2)$ are of opposite sign. Since $\bar{f}_A$ is concave, this requires that $\bar{f}'_A(u_1) > 0$ and $\bar{f}'_A(u_2) < 0$. For u_1 and u_2 to be elements of $\mathfrak{L}$, requires that $\bar{f}'_R(u_1) > 0$ and $\bar{f}'_R(u_2) < 0$, which is impossible since $\bar{f}_R$ is convex. Thus the directional derivatives of $\bar{f}_A$ and $\bar{f}_R$ cannot change sign in $\mathfrak{L}$. Now let $u = \alpha u_1 + (1-\alpha)u_2$, where $\alpha \in [0,1]$, and u_1 and u_2 are elements of $\mathfrak{L}$. Since $\bar{f}_A$ is concave, there can be at most one sign change in the derivative, so it can be concluded that $\bar{f}'_A(u)$ has the same sign as $\bar{f}'_A(u_1)$ and $\bar{f}'_A(u_2)$. A similar argument holds for $\bar{f}'_R(u)$, and consequently $\bar{f}'_A(u)\bar{f}'_R(u) \geq 0$ so, by Lemma 5, $u \in \mathfrak{L}$.

Finally, since the intersection of convex sets is convex, $\mathfrak{S}_b$ is convex. ■

This theorem means that, for concave accuracy and convex rejectability functions defined on an interval $U = (u_{\min}, u_{\max})$, the maximally satisficing set is also an interval. Moreover, $\mathfrak{L} = [\min\{u_A, u_R\}, \max\{u_A, u_R\}]$. In the next chapter, it will be shown that this is a useful characteristic.

The following theorem establishes the equilibrium set as the set of controls that are maximal for a given rejectivity.

Theorem 7 *Let $U \in \Re$, let $\bar{f}_A$ be concave density function, differentiable over the interior of U, and let $\bar{f}_R$ be a convex density function, differentiable over the interior of U. Then for every $u \in \mathfrak{L}$ there exists a rejectivity value $b \in [0,\infty)$ such that u is a most discriminating satisficing control. Furthermore, if $\bar{f}_R - b\bar{f}_A$ is strictly concave for every $b \in [0,\infty)$, then for every such b there corresponds a unique most discriminating satisficing control.*

<u>*Proof:*</u> Let $u \in \mathfrak{L}$. Define $u_* = \min\{u_A, u_R\}$ and $u^* = \max\{u_A, u_R\}$. For these limiting cases, it has previously been established that by assigning rejectivity values of $b = \infty$ and $b = 0$, respectively, u_R and u_A are most discriminating satisficing controls. Fix $u \in (u_*, u^*)$. The function $\bar{f}_A(v) - b\bar{f}_R(v)$ is extremized when $\bar{f}'_A(v) - b\bar{f}'_R(v) = 0$. Evaluating this function at $v = u$ implies $b_u = \frac{\bar{f}'_A(u)}{\bar{f}'_R(u)}$ is the rejectivity required to render u a most discriminating satisficing control. Furthermore, for $b \in [0,\infty)$, $u = \arg\max_{v \in (u_*, u^*)}\{\bar{f}_A(v) - b\bar{f}_R(v)\}$ is the unique most discriminating satisficing control. ■

Figures 2.1 and 2.2 illustrate the maximal satisficing and strongly satisficing sets for a concave differentiable $\bar{f}_R$ and convex differentiable $\bar{f}_A$. In Fig-

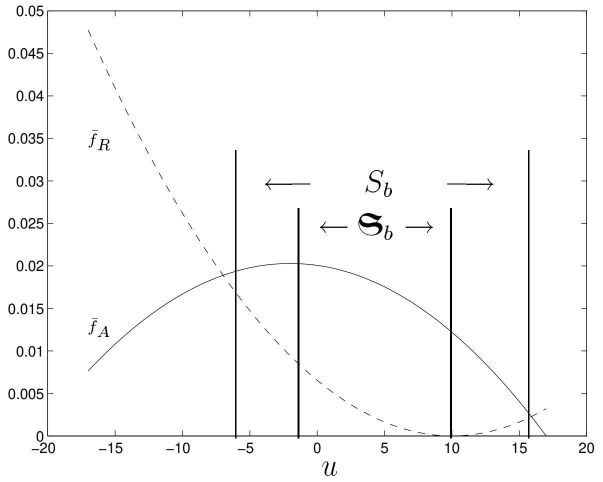


Figure 2.1: Satisficing regions for a concave $\bar{f}_A$ and convex $\bar{f}_R$ plotted as densities.

ure 2.1, the accuracy and rejectability utilities are plotted as functions of u. The satisficing and strongly satisficing sets are shown on the plot for $b = 1$. In the figure, S_b consists of those u for which $\bar{f}_A(u)$ exceeds $\bar{f}_R(u)$. The set $\mathbf{\mathfrak{S}}_b$ consists of those u for which no more accurate control exists for a given rejectability level and for which no less rejectable control exists for a given accuracy level. Observe that for $u \in \mathbf{\mathfrak{S}}_b$ the accuracy and rejectability utilities have slopes with the same sign whence controls in this region are in equilibrium.

In Figure 2.2, the accuracy and rejectability functions are cross-plotted as parametric functions of u. The "$\cdot$" symbol denotes the pair $(\bar{f}_R(u), \bar{f}_A(u))$ as u ranges over its admissible values. The "$\times$" symbol represents the point $(\bar{f}_R(u_A), \bar{f}_A(u_A))$, where the accuracy is a maximum. The "$\circ$" symbol represents the $(\bar{f}_R(u_R), \bar{f}_A(u_R))$, where the rejectability is minimum. The points to the left of the $\bar{f}_A = b\bar{f}_R$ line are the satisficing solutions. The graph may be divided into four regions. In Region I, $\bar{f}_R(u) > \bar{f}_R(u_R)$, but $\bar{f}_A(u) < \bar{f}_A(u_R)$, thus solutions in this region should be rejected since the minimally rejectable solution has both higher accuracy and lower rejectability. In Region II, $\bar{f}_A(u) < \bar{f}_A(u_A)$ but $\bar{f}_R(u) > \bar{f}_R(u_A)$, thus solutions in this region should be rejected since the maximally accurate solution has both higher accuracy and lower rejectability. Region III should also obviously be rejected since both the maximally accurate and minimally rejectable solutions are superior. Region IV is the region for which $\bar{f}_A(u) < \bar{f}_A(u_A)$ and $\bar{f}_R(u) > \bar{f}_R(u_R)$. Solutions in this region, the points on the arc between

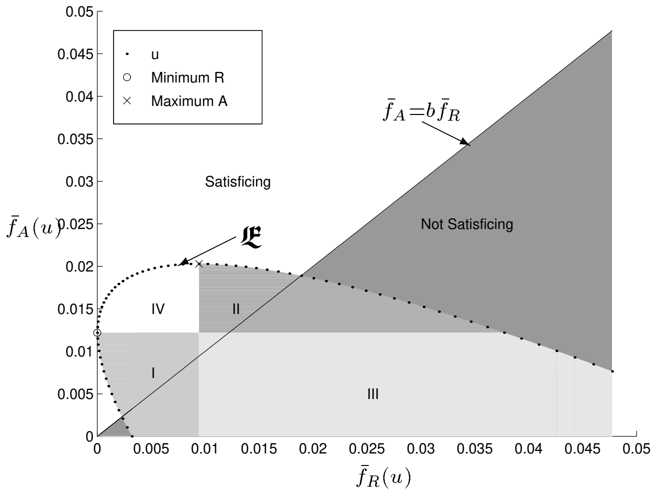


Figure 2.2: Satisficing regions for a concave $\bar{f}_A$ and convex $\bar{f}_R$ cross plotted.

$(\bar{f}_R(u_R), \bar{f}_A(u_R))$ (the "$\circ$" symbol) and $(\bar{f}_R(u_A), \bar{f}_A(u_A))$ (the "$\times$" symbol), correspond to the equilibrium set, $\mathfrak{L}$. If the line $\bar{f}_A = b\bar{f}_R$ intersects $\mathfrak{L}$, then only points above the line qualify as members of $\mathfrak{S}_b$ (in Figure 2.2, $\mathfrak{S}_b$ coincides with $\mathfrak{L}$ and the tangent to the curve with slope b in Region IV is identified with the most discriminating control for b).

Any element of the strongly satisficing set $\mathfrak{S}_b$ has merit as a control. If $\bar{f}_A$ is concave over its support and $\bar{f}_R$ is convex over its support, then u_R and u_A represent boundaries of a convex set of controls, and any convex combination $\lambda u_R + (1-\lambda)u_A$ of these controls is also strongly satisficing. The parameter λ is a design parameter to tune the controller to obtain satisfactory performance. Thus, in this case, there exists a viable synthesis procedure for controller design that parallels the synthesis procedures commonly used with optimal controller design.

Chapter 3

SINGLE AGENT SATISFICING CONTROL: APPLICATIONS

One approach to developing controllers for nonlinear plants is to linearize the system model about an operating point in state-space, and then to develop a linear controller for all time based on the linearized system model. This can be referred to as spatial localization. Depending upon the error introduced by the spatial localization, the resulting controllers may or may not be effective controllers of the nonlinear system.

An alternative to spatial localization is temporal localization. In temporal localization, controls are determined using only the immediate future. It is assumed that a designer can determine which controls are appropriate for the immediate future using the nonlinear system dynamics. Thus, the global nonlinear spatial dynamics are retained while the temporal design criterion is localized. The resulting controllers are referred to as *receding horizon* controllers.

In this chapter, a receding horizon methodology will be employed to develop a synthesis procedure for single agent satisficing controllers. Although the receding horizon methodology is not the only way a satisficing controller can be formed, it is an effective way. This methodology is applied to quadratic regulator problems and minimum time problems, and consistency is shown.

3.1 Receding Horizon Controller Synthesis

Consider the nonlinear, time-varying, single-input, n-dimensional system

$$\mathbf{x}(t+1) = \mathbf{f}[\mathbf{x}(t), u(t), t], \quad t = 0, 1, \ldots, t_f - 1 \tag{3.1}$$

together with the performance index

$$J = \phi[\mathbf{x}(t_f)] + \sum_{t=0}^{t_f-1} L[\mathbf{x}(t), u(t), t], \tag{3.2}$$

where ϕ and L are given performance indices (for example, quadratic forms), and where the terminal time, t_f, is unspecified. Suppose L is separable; that is, L is of the form

$$L[\mathbf{x}(t), u(t), t] = L_1[\mathbf{x}(t), t] + L_2[u(t), t]. \tag{3.3}$$

Since the initial state, $\mathbf{x}(0)$, cannot be affected by the input, the following performance index is equivalent to (3.2):

$$J = \phi'[\mathbf{x}(t_f)] + \sum_{t=0}^{t_f-1} \left\{ L_1[\mathbf{x}(t+1), t+1] + L_2[u(t), t] \right\}, \tag{3.4}$$

where $\phi'[\mathbf{x}(t_f)] = \phi[\mathbf{x}(t_f)] - L_1[\mathbf{x}(t_f), t_f)]$.

General global solutions for this nonlinear control problem are not easily obtainable, but many such problems may be addressed from a local perspective. One natural approach is to invoke a receding, or rolling, horizon control strategy [21, 39, 41]. This approach consists of implementing a feedback controller through a series of repeated open-loop calculations based on the instantaneous state. For a discrete-time receding horizon of length d, the next d values, $\{u(t), \ldots, u(t + d - 1)\}$, are computed as functions of the current state, $\mathbf{x}(t)$; i.e., the control $u(t)$ is determined by (possibly nonlinear) state feedback. The control $u(t)$ is implemented, producing a state $\mathbf{x}(t + 1)$, the horizon is shifted forward one time unit, and the process is repeated. For example, a one-step control horizon ($d = 1$) would require the design of only $u(t)$, the control for the current time increment. For $d = 2$, both $u(t)$ and $u(t + 1)$ are required. [39] provides a stability theorem for time-invariant nonlinear systems, but the theorem requires that the system satisfy a number of technical conditions that are difficult to verify. [21] develops error bounds for receding horizon controllers for nonlinear systems, but only for stationary, or time-invariant systems. If time-invariance does not hold, there seem to be few theoretical results regarding the performance of receding-horizon control techniques for nonlinear systems. Similar results apply for receding horizon H_∞ control [31, 32].

The use of a receding horizon makes it possible to develop a tractable satisficing controller using the principles of epistemic utility theory. Concentration is placed on the development of a satisficing controller for systems of the form (3.1) for which a global solution has proven difficult. A receding horizon is employed to specify f_A and f_R as functions of the control input, u and the state of nature, ϑ, such that f_A assigns high accuracy to those control values that tend to achieve the goal of the system, and f_R assigns low rejectability to those control values that tend to conform to the design principles. As the length of the control horizon, d, is increased, more of the future state values are taken into consideration for the calculation of the current control. It is therefore reasonable to expect that performance will improve with increasing d. When applied to problems for which an optimal solution exists, if both the optimal and the satisficing solutions use exactly the same information, the optimal solution must be attainable, at least in principle, via the satisficing procedure. This principle is called the *consistency principle* and is one of the strengths of the satisficing approach. Without consistency, it would be difficult to develop confidence in the satisficing solutions [63]. For the satisficing controllers developed herein, the subsequent sections will show that as the horizon increases, the strongly satisficing set includes the optimal solution and, hence, that the receding horizon satisficing controllers are consistent.

Formally, because the system model (3.1) may involve delays (it is assumed causal), it is possible that $u(t+\ell)$ will have an explicit influence on elements of the state vector beyond the $t+\ell+1$st step. Consequently, the *explicit influence horizon* of $u(t+\ell)$, denoted $D_{t+\ell}$, may be defined as the maximum number of time increments for which $u(t+\ell)$ has explicit influence on any component of the state.

Let $\xi_\ell(t+\tau)$ be the sub-state of $\mathbf{x}(t+\tau)$ that is an explicit function of $u(t+\ell)$, $\tau = 1, \ldots, D_{t+\ell}$. Continuing in this manner, d sequences of sub-states may be generated, denoted the *influence vectors*, of the form

$$\begin{aligned}
\boldsymbol{\xi}_0(u(t)) &= \{\xi_0(t+1), \ldots, \xi_0(t+D_t)\} \\
\boldsymbol{\xi}_1(u(t+1)) &= \{\xi_1(t+2), \ldots, \xi_1(t+1+D_{t+1})\} \\
&\vdots \\
\boldsymbol{\xi}_{d-1}(u(t+d-1)) &= \{\xi_{d-1}(t+d), \ldots, \xi_{d-1}(t+d-1+D_{t+d-1})\}.
\end{aligned} \tag{3.5}$$

The collection $\{\boldsymbol{\xi}_\ell(u(t+\ell)),\ 0 \leq \ell \leq d-1\}$ then represents the state elements that are explicit functions of $u(t), \ldots, u(t+d-1)$. Using these elements, *local performance indices*, $\Phi[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))]$ and $\Lambda[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))]$ may be formed and used to compute f_A and f_R. These local performance indices are formed according to the same principles that are used to define the global performance indices, ϕ and L. Specific examples of these functionals will be provided in subsequent sections.

From a local perspective, $t+d$ is viewed as the terminal time, t_f. Let the state of nature be the current state of the system, that is, $\vartheta = \mathbf{x}(t)$. The assumption that ϑ is known is satisfied if perfect state feedback is available. Because the accuracy utility f_A is associated with the goal, $f_A[\mathbf{u}; \mathbf{x}(t)]$ is designed to have large values for $\mathbf{u} = [u(t), \ldots, u(t+d-1)]^T$ such that $\Phi[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))]$ is small. The density f_A is obtained by negating and normalizing this function. Define

$$g_A[\mathbf{u}; \mathbf{x}(t)] = \sup_{\mathbf{z} \in U^d} \{\Phi[\boldsymbol{\xi}_0(z(t)), \ldots, \boldsymbol{\xi}_{d-1}(z(t+d-1))]\} - \Phi[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))] + \epsilon, \tag{3.6}$$

where $\mathbf{z} = [z(t), \ldots, z(t+d-1)]^T$, U^d is the d-dimensional control space $U \times \ldots \times U$, and ϵ is a small nonnegative number inserted to ensure that all of the values of g_A are nonzero. The probability density function f_A is obtained from g_A by normalizing:

$$f_A[\mathbf{u}; \mathbf{x}(t)] = \frac{g_A[\mathbf{u}; \mathbf{x}(t)]}{\int_{U^d} g_A[\mathbf{z}; \mathbf{x}(t)] d\mathbf{z}}. \tag{3.7}$$

f_A takes its maximum at the values $\mathbf{u}$ that drive Φ closest to zero, but also assigns significant accuracy to control values in the neighborhood of this control value.

The rejectability density function f_R is a measure of how well the control decision $\mathbf{u}$ complies with the incremental cost functional, $\Lambda[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))]$, from the local perspective. Consequently, define

$$g_R[\mathbf{u}; \mathbf{x}(t)] = \Lambda[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))] - \inf_{\mathbf{z} \in U} \{\Lambda[\boldsymbol{\xi}_0(z(t)), \ldots, \boldsymbol{\xi}_{d-1}(z(t+d-1))]\} + \epsilon. \tag{3.8}$$

Converting this into a probability density function by normalizing yields

$$f_R[\mathbf{u}; \mathbf{x}(t)] = \frac{g_R[\mathbf{u}; \mathbf{x}(t)]}{\int_{U^d} g_R[\mathbf{z}; \mathbf{x}(t)] d\mathbf{z}}. \tag{3.9}$$

This function represents the local cost of the decision to take action $\mathbf{u}$. Actions that result in large values of f_R will be more rejectable than actions that result in small values of f_R. These valuations are made independently of the likelihood that a given choice will achieve the desired goal—the only consideration is rejectability, which is equated with cost.

Concave $f_A(u;\vartheta)$ and convex $f_R(u;\vartheta)$ may naturally arise when the performance induces Λ and Φ are convex functions. When expectation is taken, this means that $\bar{f}_A(u)$ and $\bar{f}_R(u)$ will also be concave and convex, respectively. For the linear and nonlinear quadratic regulator, this property is exploited. This makes the design of the satisficing controllers easier by utilizing Theorem 6.

3.2 Nonlinear Quadratic Regulator

An important control problem is the nonlinear, nonstationary regulator problem of the form

$$\mathbf{x}(t+1) = \mathbf{f}[\mathbf{x}(t), u(t), t], \quad \mathbf{x}(0) = \mathbf{x}_0, \quad t = 0, 1, 2, \ldots \tag{3.10}$$

where $\mathbf{f}$ is a nonlinear vector function. It is desirable to regulate this system about the origin $\mathbf{x} = \mathbf{0}$ in such a way that the performance index (3.11)

$$J = \mathbf{x}^T(t_f)[P - Q(t_f)]\mathbf{x}(t_f) + \sum_{t=0}^{t_f-1} \left[\mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + R_u(t)u^2(t)\right] \tag{3.11}$$

is kept small. An optimal solution would minimize this performance index but, unlike the linear regulator system, no general systematic solution has been discovered for the general nonlinear problem. A conventional approach is to linearize this system about an equilibrium point and apply linear system techniques, resulting in at least a spatially local solution whose functionality may be problematic. Techniques, such as feedback linearization, adaptive control, and gain scheduling are based in large part on such assumptions.

The satisficing approach permits a straightforward synthesis procedure. For a horizon length of d, the local performance indices become

$$\Phi[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))] = \mathbf{x}^T(t+d)[P - Q(t_f)]\mathbf{x}(t+d) \tag{3.12}$$

$$\Lambda[\boldsymbol{\xi}_0(u(t)), \ldots, \boldsymbol{\xi}_{d-1}(u(t+d-1))] = \sum_{t=0}^{t+d-1} \left[\mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + R_u(t)u^2(t)\right]. \tag{3.13}$$

The form of the resulting controller will be investigated in more detail in subsequent sections. However, the following theorem establishes consistency for this formulation.

Theorem 8 (Consistency of scalar quadratic regulator) *For the deterministic quadratic regulator problem (3.1) and (3.11), if the control horizon spans the full extent of the problem, that is, for $d = t_f$, then there exists a rejectivity index b such that the most discriminating satisficing control is identical to the optimal control.*

Proof: This proof is for the scalar control case only. The vector-valued equivalent is given in Chapter 5. The most discriminating control, denoted $\mathbf{u}_D = [u(t), \ldots, u(t_f - 1)]^T$, is $\mathbf{u}_D = \arg\sup_{\mathbf{z}\in\Re^{t_f}} \{\bar{f}_A(\mathbf{z}) - b\bar{f}_R(\mathbf{z})\}$ and is always strongly satisficing by Theorem 3. From (3.7) and (3.9),

$$\arg\sup_{\mathbf{z}\in\Re^{t_f}} \{\bar{f}_A(\mathbf{z}) - b\bar{f}_R(\mathbf{z})\} = \arg\sup_{\mathbf{z}\in\Re^{t_f}} \{g_A(\mathbf{z}) - b' g_R(\mathbf{z})\} \tag{3.14}$$

where $b' = b\frac{\int_{U^{t_f}} g_A(\mathbf{z};\mathbf{x}(t))d\mathbf{z}}{\int_{U^{t_f}} g_A(\mathbf{z};\mathbf{x}(t))d\mathbf{z}}$. Also, from (3.6) and (3.8),

$$\begin{aligned} \arg\sup_{\mathbf{z}\in\Re^{t_f}} \{g_A(\mathbf{z}) - b' g_R(\mathbf{z})\} &= \arg\inf_{\mathbf{z}\in\Re^{t_f}} \{\Phi[\boldsymbol{\xi}_0(z(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(z_(t_f - 1))] \\ &\quad +b'\Lambda[\boldsymbol{\xi}_0(z(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(z(t_f - 1))]\} \end{aligned} \tag{3.15}$$

For the quadratic regulator problem with $d = t_f$, take $\Phi[\boldsymbol{\xi}_0(u(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(u(t_f - 1))] = \mathbf{x}^T(t_f)[P - Q(t_f)]\mathbf{x}(t_f)$ and $\Lambda[\boldsymbol{\xi}_0(u(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(u(t_f - 1))] = \sum_{t=0}^{t_f-1} \mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + R_u(t)z(t)^2$. Thus,

$$\mathbf{u}_D = \arg\inf_{\mathbf{z}\in\Re^{t_f}} \{\mathbf{x}^T(t_f)P\mathbf{x}(t_f) + b' \sum_{t=0}^{t_f-1} \mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + R_u(t)z^2(t)\}, \tag{3.16}$$

where $\mathbf{x}(t)$ is a function of $z(t)$ via the dynamics equation (3.1). Rejectivity b can be chosen so that $b' = 1$ yielding the optimal quadratic regulator solution. ■

The examples provided in Sections 3.2.1-3.2.3 illustrate some very desirable properties of satisficing control. First, the satisficing feedback controller is able to stabilize three unstable systems. Second, for the linear quadratic regulator with $d = 1$, a satisficing controller performs very nearly as well as the optimal. For the linear problem these results are perhaps not very surprising, since the satisficing receding horizon solution is similar to other bounded-memory approaches (for example, see [21, 39]) known to have similar properties. For the nonlinear problem, however, no other results of comparable generality are known to the author.

3.2.1 Linear Quadratic Regulator

Since the solution to the optimal linear quadratic regulator is well known, it provides a convenient benchmark against which to evaluate satisficing solutions. Consider the following time-varying single-input system:

$$\mathbf{x}(t+1) = A(t)\mathbf{x}(t) + B(t)u(t), \quad t = 0, 1, \ldots, t_f - 1, \tag{3.17}$$

where $\mathbf{x}(t)$ is an n-dimensional state vector, $A(t)$ is an $n \times n$ matrix, $u(t)$ is a scalar input, and $B(t)$ is an n-dimensional vector. It is desirable choose the control according to a quadratic performance index, so (3.4) becomes

$$J = \mathbf{x}^T(t_f)[P - Q(t_f)]\mathbf{x}(t_f) + \sum_{t=0}^{t_f-1}[\mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + R_u(t)u^2(t)], \tag{3.18}$$

where $P \geq 0$, $Q(t) \geq 0$ and $R_u(t) \geq 0$.

To demonstrate the application of satisficing control, a satisficing approach to a temporally local version of this problem is developed. Locality is invoked via a receding horizon control strategy. The most restrictive such version is a one-step control horizon, that is, $d = 1$. It is important to emphasize that, because the satisficing controller uses only temporally local information, it is insensitive to a time-invariance assumption. Thus, the matrices A, B, Q, and R may be time-varying without appreciable change to the procedure. Furthermore, these matrices need not even be known in advance of the instantaneous horizon, $t + d$. In this latter case, it would not be possible to obtain a globally optimal solution.

Let $U_m = (-u_m, u_m)$ be a finite open interval. Begin by specifying g_A and g_R according to (3.6) and (3.8). To calculate g_A, set $P = I$, $Q(t_f) = 0$. Assume that the explicit influence horizon is $D_t = 1$. Then the influence vector is $\boldsymbol{\xi}_0(u) = \mathbf{x}(t+1) = A(t)\mathbf{x}(t) + B(t)u$. Accuracy is based on the local performance index

$$\begin{aligned}\Phi[\boldsymbol{\xi}_0(u)] &= \mathbf{x}^T(t+1)[P - Q(t_f)]\mathbf{x}(t+1) \\ &= [A(t)\mathbf{x}(t) + B(t)u]^T[A(t)\mathbf{x}(t) + B(t)u].\end{aligned} \tag{3.19}$$

To calculate g_R, define the local performance index

$$\Lambda[\boldsymbol{\xi}_0(u)] = \mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + R_u(t)u^2$$

$$= \quad [A(t)\mathbf{x}(t) + B(t)u]^T Q(t+1)[A(t)\mathbf{x}(t) + B(t)u] + R_u(t)u^2. \tag{3.20}$$

Let

$$u_{\mathfrak{L}_A} = \arg\min_{u \in U_m} \Phi[\boldsymbol{\xi}_0(u)] \tag{3.21}$$
$$u_{\mathfrak{L}_R} = \arg\min_{u \in U_m} \Lambda[\boldsymbol{\xi}_0(u)]. \tag{3.22}$$

From an implementation perspective, it is helpful to restrict U_m to a smaller control space. Let

$$u_* = \min\{u_{\mathfrak{L}_A}, u_{\mathfrak{L}_R}\} \tag{3.23}$$
$$u^* = \max\{u_{\mathfrak{L}_A}, u_{\mathfrak{L}_R}\} \tag{3.24}$$
$$U = [u_*, u^*], \tag{3.25}$$

and assume $u_m \gg |u_*|$ and $u_m \gg |u^*|$ so that the boundaries of U_m do not affect these values. Observe that this restricted $U = \mathfrak{L}_b$.

Since $\Lambda[\boldsymbol{\xi}_0(u)]$ is quadratic, it assumes its unique minimum at $u_{\mathfrak{L}_R}$. Thus, (3.8) becomes

$$g_R[u; \mathbf{x}(t)] = \Lambda[\boldsymbol{\xi}_0(u)] - \Lambda[\boldsymbol{\xi}_0(u_{\mathfrak{L}_R})] + \epsilon \tag{3.26}$$

with normalizing term

$$G_R[\mathbf{x}(t)] = \int_{u_*}^{u^*} g_R(w; \mathbf{x}(t))dw. \tag{3.27}$$

From an implementation perspective, it is helpful to restrict attention to $\mathfrak{L}$ since only elements of this set will ever be considered as acceptable controls. Because $\Phi[\boldsymbol{\xi}_0(u)]$ is quadratic in u it achieves its maximum at $u_{\mathfrak{L}_R}$ when restricted to $\mathfrak{L}$. Hence, (3.6) becomes,

$$g_A[u; \mathbf{x}(t)] = \Phi[\boldsymbol{\xi}_0(u_{\mathfrak{L}_R})] - \Phi[\boldsymbol{\xi}_0(u)] + \epsilon, \tag{3.28}$$

with normalizing term

$$G_A[\mathbf{x}(t)] = \int_{u_*}^{u^*} g_A(w; \mathbf{x}(t))dw. \tag{3.29}$$

By construction, $\bar{f}_R$ and $\bar{f}_A$ are convex and concave, respectively, and the equilibrium set equals the interval (u_*, u^*). Even though $\bar{f}_A$ and $\bar{f}_R$ are functions of t, only a fixed value of t is used which means that the argument can be suppressed without risk of ambiguity. By restricting attention to $\mathfrak{L}$, $u_A = u_{\mathfrak{L}_A}$ and $u_R = u_{\mathfrak{L}_R}$ which simplify to

$$
\begin{aligned}
u_A &= -[B^T(t)B(t)]^{-1}B^T(t)A(t)\mathbf{x}(t) && (3.30)\\
u_R &= [R_u(t) + B^T(t)Q(t+1)B(t)]^{-1}B^T(t)Q(t+1)A(t)\mathbf{x}(t). && (3.31)
\end{aligned}
$$

Because Φ and Λ are convex, $\mathfrak{S}_b$ is convex, which implies that all possible strongly satisficing controls may be obtained via convex combinations of u_R and u_A. For $\lambda \in [0,1]$ define $u_\lambda = \lambda u_R + (1-\lambda)u_A$. For $\lambda \approx 0$, the control tends to reduce the accumulated cost at the expense of large terminal error, and for $\lambda \approx 1$, the control tends to reduce the terminal error at the expense of the accumulated cost. Thus, λ is a design parameter for a synthesizing procedure. There exists a λ_D such that the most discriminating control is given by $u_D = u_{\lambda_D}$. The most discriminating control u_D can be calculated directly directly in a manner similar to the calculations of u_A and u_R, yielding

$$u_D = -[B^T(t)(I+b'Q(t+1))B(t)+b'R_u(t)]^{-1}B^T(t)(I+b'Q(t+1))A(t)\mathbf{x}(t), \quad (3.32)$$

where $b' = b\frac{G_A[\mathbf{x}(t)]}{G_R[\mathbf{x}(t)]}$. Let $P_D(t+1) = I + b'Q(t+1)$ and define the satisficing gain

$$K_D(t) = [B^T(t)P_D(t+1)B(t) + b'R_u(t)]^{-1}B^T(t)P_D(t+1)A(t), \quad (3.33)$$

so that $u_D = -K_D(t)\mathbf{x}(t)$. The satisficing control for the linear quadratic regulator is a state-feedback control, and has a structure similar to the optimal feedback control. Because the gain K_D is a function of the state $\mathbf{x}(t)$ (via b'), however, the feedback is not linear.

When $A(t)$, $B(t)$, $Q(t)$ and $R_u(t)$ are defined for all time and known beforehand, the optimal solution can be obtained of the form $u(t) = -K(t)\mathbf{x}(t)$, where $K(t)$ is the Kalman gain (see [30])

$$K(t) = \{B^T(t)[Q(t+1) + P(t+1)]B(t) + R_u(t)\}^{-1}B^T(t)[Q(t+1) + P(t+1)]A(t), \quad (3.34)$$

and where $P(t) = A^T(t)[Q(t+1) + P(t+1)][A(t) - B(t)K(t)]$, with terminal condition $P(t_f) = P$. In the special case where $A(t)$, $B(t)$, $Q(t)$ and $R_u(t)$ are constant, the system is time-invariant and a steady-state solution exists of the form $u(t) = K_\infty \mathbf{x}(t)$, where the steady-state Kalman gain K_∞ is obtained via the algebraic Riccati equation.

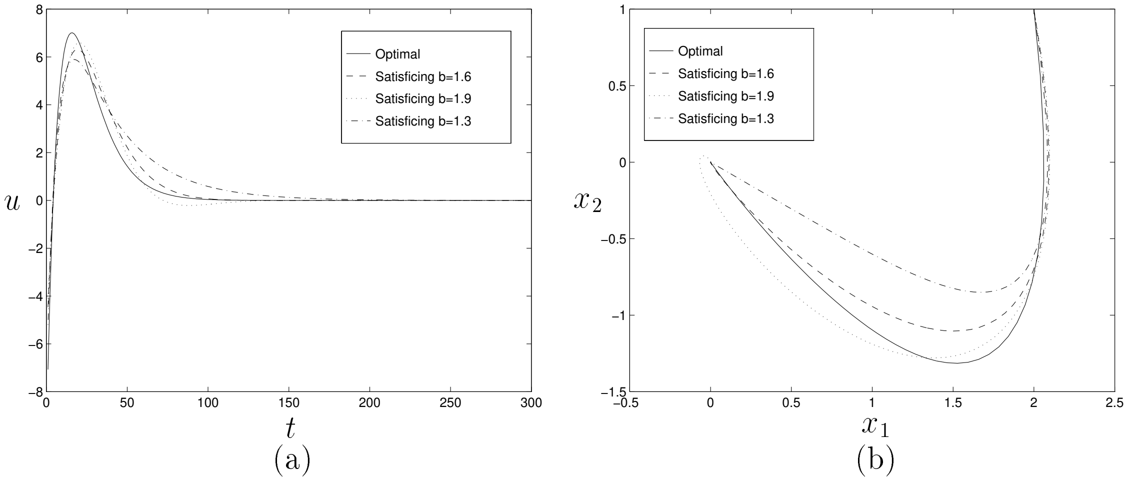


Figure 3.1: Performance results for the single-agent linear regulator problem: (a) control history, (b) phase plane.

To compare the optimal and satisficing control policies, an unstable second-order linear time-invariant system example taken from [29] is used. Let $R_u = [0.05]$, $P = I$,

$$A = \begin{bmatrix} 0.9974 & 0.0539 \\ -0.1078 & 1.1591 \end{bmatrix} \quad B = \begin{bmatrix} 0.0013 \\ 0.0539 \end{bmatrix} \quad Q = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.05 \end{bmatrix}$$

Since negligible performance is lost by using steady-state gains, comparisons were made with results obtained using $K_\infty = [-0.5522 \; -5.9690]$. The resulting control history is plotted in Figure 3.1(a), with the solid curve representing the optimal steady-state control, and the dashed curves representing the one-step satisficing control for three different values of boldness.

The steady-state optimal trajectory is shown in Figure 3.1(b) as the solid curve, and the solution of the one-step satisficing epistemic utility-based controller is displayed with the dashed curves for three different boldness values. The simulation begins at the region of state space designated by "$\circ$" and terminates

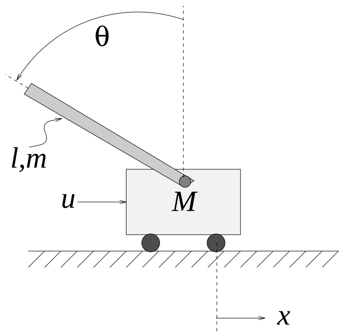


Figure 3.2: Inverted pendulum on a cart.

at the origin designated by "×". The most discriminating satisficing control, u_D, is employed. The steady-state optimal cost for this problem is $J_{ss} = 38.3$, and the cost for the one-step controllers J_b with rejectivity index b are $J_{1.3} = 42.5$, $J_{1.6} = 39.9$ and $J_{1.9} = 40.0$. Thus, performance degrades approximately 4% if a satisficing solution (b=1.6) is used.

3.2.2 The Inverted Pendulum

In this section, a satisficing receding horizon control is applied to a problem that has proven to be surprisingly difficult: The control of an inverted pendulum in a vertical plane with full circular freedom by applying a lateral force to the cart to which the pendulum is attached, while <u>simultaneously</u> regulating the position of the cart.

Consider the apparatus illustrated in Figure 3.2. The problem is to bring the pendulum from vertically downward to vertically upward by applying a force to the cart. This problem is prototypical of many nonlinear control problems, and thus makes a good test case for satisficing control. Traditional controllers linearize the dynamics model of the pendulum in a small region within, say 10 degrees of the vertical. More recently, a fuzzy controller trained by a genetic algorithm has been shown to balance the pendulum 90% of the time if the pendulum is given a random initial position within 80 deg of the vertical and a random initial velocity less that 80 deg/sec [28]. There also exists controllers which balance the pendulum from a vertical down starting position, but they only approximately regulate the cart about the origin and they fail to be generalizable to other problems [10, 67]. In this section, an epistemic utility-based controller with the control horizon $d = 1$ is designed which will control the pendulum given <u>any</u> set of initial conditions, while simultaneously positioning the cart at a desired point. The only restriction made is that the initial pendulum velocity be small enough so that the sample time is much less than the time it takes for the pendulum to rotate 2π radians ($\dot{\theta} \ll 2\pi/T$ where T is the sample time). To render the problem nonstationary, assume further that the mass of the pendulum, $m(t)$, may be a random walk whose future d values are known, but values $d+1, \ldots$ are possibly unknown.

Let $\mathbf{x} = [\theta, x, \dot{\theta}, \dot{x}]^T$ denote the state of the cart/pendulum system. The continuous-time dynamical equation for this problem is

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u, \tag{3.35}$$

where $\mathbf{f} = [f_1, f_2, f_3, f_4)]^T$ and $\mathbf{g} = [g_1, g_2, g_3, g_4)]^T$ are given by

$$\begin{aligned} f_1(\mathbf{x}) &= \dot{\theta} && (3.36)\\ f_2(\mathbf{x}) &= \dot{x} && (3.37)\\ f_3(\mathbf{x}) &= \frac{(M+m)g\sin\theta - ml\cos\theta\sin\theta\dot{\theta}^2}{l(M+m\sin^2\theta)} && (3.38)\\ f_4(\mathbf{x}) &= \frac{ml\sin\theta\dot{\theta}^2 - mg\cos\theta\sin\theta}{M+m\sin^2\theta}. && (3.39) \end{aligned}$$

$$\begin{aligned} g_1(\mathbf{x}) &= 0 && (3.40)\\ g_2(\mathbf{x}) &= 0 && (3.41)\\ g_3(\mathbf{x}) &= \frac{-\cos\theta}{l(M+m\sin^2\theta)} && (3.42)\\ g_4(\mathbf{x}) &= \frac{1}{M+m\sin^2\theta}. && (3.43) \end{aligned}$$

M is the mass of the cart, l is the length of the pendulum, m is the mass of the pendulum, θ is the angle from vertical (measured counterclockwise), x is the horizontal position of the cart, and u, the control input, is a lateral force applied to the cart.

Using Euler integration, the equivalent discrete-time dynamical expression is

$$\mathbf{x}(t+1) = \mathbf{x}(t) + T\{\mathbf{f}[\mathbf{x}(t)] + \mathbf{g}[\mathbf{x}(t)]u(t)\}, \tag{3.44}$$

for $t = 0, 1, \ldots$. With this discretization procedure, $\theta(t+1)$ and $x(t+1)$ are not explicit functions of $u(t)$, but $\theta(t{+}2)$ and $x(t{+}2)$ are explicit functions of $u(t)$. Thus $D_t = 2$, and the components of the influence vector $\boldsymbol{\xi}_0(u) = \{\xi_0(t+1),\ \xi_0(t+2)\}$ can be identified as

$$\xi_0(t+1) = \begin{bmatrix} \dot{\theta}(t+1) \\ \dot{x}(t+1) \end{bmatrix} = \begin{bmatrix} \dot{\theta}(t) + T\{f_3[\mathbf{x}(t)] + g_3[\mathbf{x}(t)]u(t)\} \\ \dot{x}(t) + T\{f_4[\mathbf{x}(t)] + g_4[\mathbf{x}(t)]u(t)\} \end{bmatrix}. \tag{3.45}$$

and

$$\xi_0(t+2) = \begin{bmatrix} \theta(t+2) \\ x(t+2) \end{bmatrix} = \begin{bmatrix} \theta(t) + 2T\dot{\theta}(t) + T^2\{f_3[\mathbf{x}(t)] + g_3[\mathbf{x}(t)]u(t)\} \\ x(t) + 2T\dot{x}(t) + T^2\{f_4[\mathbf{x}(t)] + g_4[\mathbf{x}(t)]u(t)\} \end{bmatrix}. \tag{3.46}$$

Adopt the same quadratic performance index (3.20) that was used for the previous linear quadratic regulator example. Since the inverted pendulum is a regulator problem, the goal of the system is the same as for that problem; namely, bring the system to rest at the desired point. The local performance indices then become

$$\Phi[\boldsymbol{\xi}_0(u)] = [\xi_0(t+2), \xi_0(t+1)]P \begin{bmatrix} \xi_0(t+2) \\ \xi_0(t+1) \end{bmatrix} \tag{3.47}$$

and

$$\Lambda[\boldsymbol{\xi}_0(u)] = [\xi_0(t+2), \xi_0(t+1)]Q \begin{bmatrix} \xi_0(t+2) \\ \xi_0(t+1) \end{bmatrix} + R_u(t)u^2. \tag{3.48}$$

From these equations, $\bar{f}_A$ and $\bar{f}_R$ may be calculated in accordance with (3.6)-(3.9). Thus, problems with linear dynamics and nonlinear dynamics are both treated exactly the same; neither linearity nor time-invariance are exploited. The resulting most discriminating controller is, after some calculations,

$$u_D = -\left[G^T[\mathbf{x}(t)](P + b'Q(t+1))G[\mathbf{x}(t)] + b'R_u(t)\right]^{-1} G^T[\mathbf{x}(t)](P + b'Q(t+1))F[\mathbf{x}(t)], \tag{3.49}$$

$$G[\mathbf{x}(t)] = \left[T^2 g_3[\mathbf{x}(t)] \quad T^2 g_4[\mathbf{x}(t)] \quad T g_3[\mathbf{x}(t)] \quad T g_4[\mathbf{x}(t)]\right]^T \tag{3.50}$$

$$F[\mathbf{x}(t)] = \begin{bmatrix} \theta(t) + 2T\dot{\theta}(t) + T^2 f_3[\mathbf{x}(t)] \\ x(t) + 2T\dot{x}(t) + T^2 f_4[\mathbf{x};(t)] \\ \dot{\theta}(t) + T f_3[\mathbf{x}(t)] \\ \dot{x}(t) + T f_4[\mathbf{x}(t)] \end{bmatrix}. \tag{3.51}$$

The following values were used in the simulations:

$$Q(t) = \begin{bmatrix} 30 & 1 & 0 & 0 \\ 1 & 0.3 & 0 & 0 \\ 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0.2 & 0.2 \end{bmatrix},$$

$P = I$, $R_u(t) = 10^{-4}$, $M = 0.455$ kg, $m = 0.21$ kg (nominally), $l = 0.61$ m, $T = 0.01$ s, and $U = [-1000, 1000]$. The off-diagonal terms in $Q(t)$ reflect coupling between θ and x necessitated by one control input but two degrees of freedom in the system, namely the rotational and translational components.

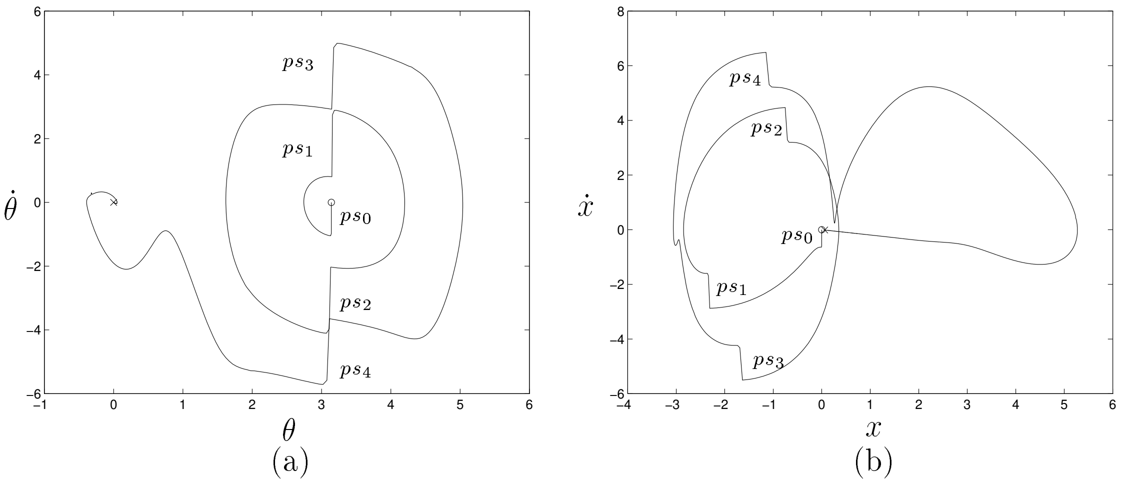


Figure 3.3: Phase planes for the inverted pendulum on a cart (m = constant): (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second).

As a baseline, the results for the time-invariant case, that is, $m(t) = m$, a constant are presented first; the results for $m(t)$ a random walk are then presented, followed by the results for $u(t)$ constrained by an unpredictable bound. Figures 3.3(a) and 3.3(b) illustrate the rotational (pendulum) and translational (cart) phase planes, respectively, for the constant mass case. The "$\circ$" symbol represents the initial conditions (the cart at the origin with the pendulum in the vertical down position) and the "$\times$" symbol represents the terminal conditions (the cart at the origin with the pendulum balanced in the vertical up position).

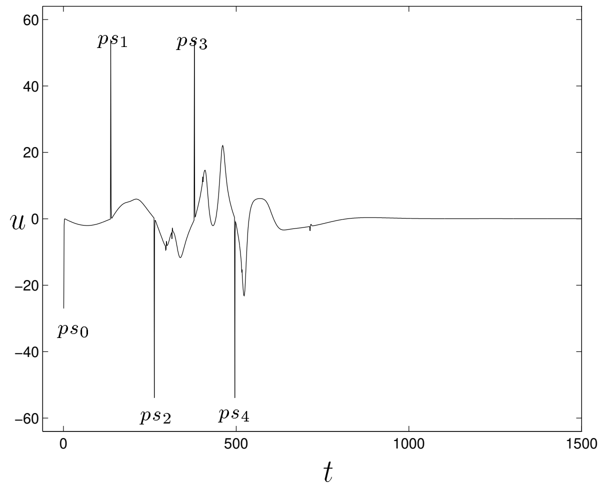
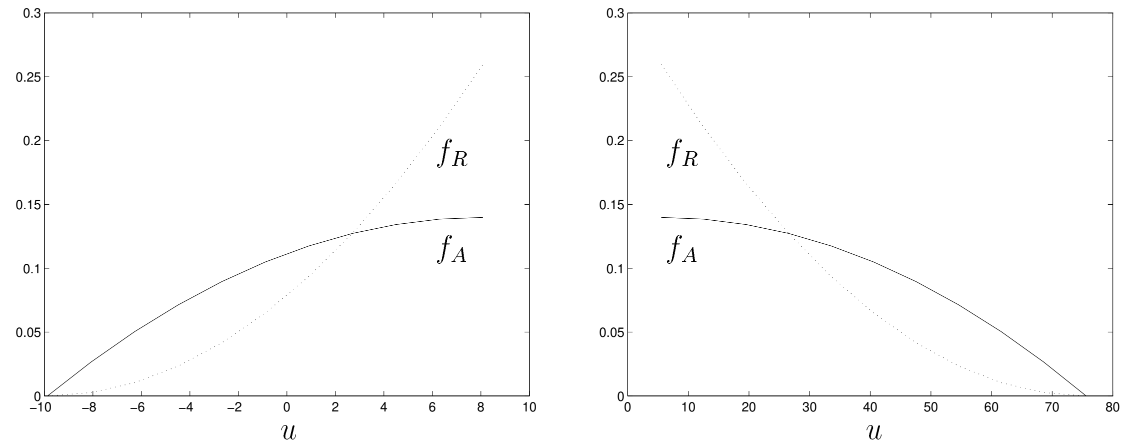


Figure 3.4: Control inputs for the inverted pendulum on a cart (time measured in 0.01 sec increments).

Figure 3.4 provides the control time history for this problem. The system achieves its desired objective of balancing the pendulum at the origin by swinging the pendulum back and forth while the cart oscillates around the origin. As the cart oscillates, the pendulum gathers momentum. In the translational and rotational phase planes, this motion is manifest as growing spirals. When the amplitude increases sufficiently, the oscillation ceases and the pendulum then converges to the vertical upright position. Finally, the cart returns to the origin.

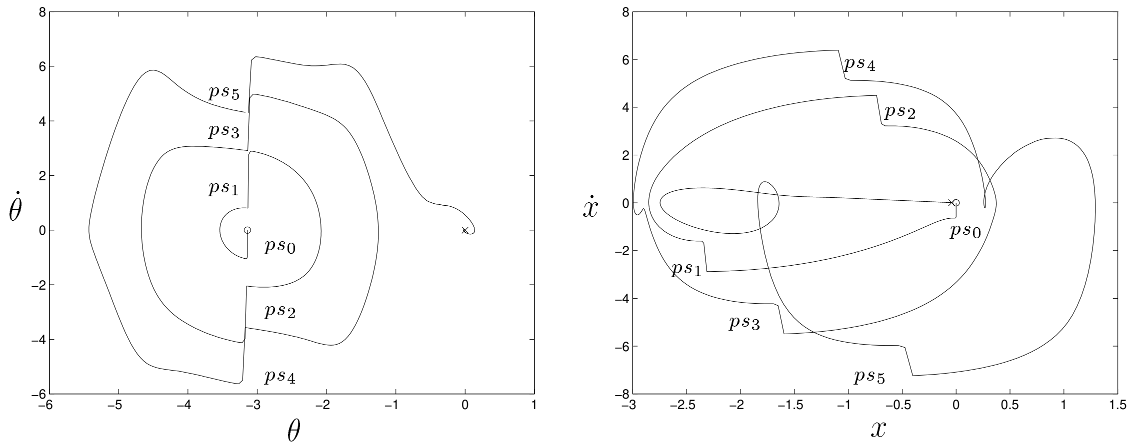
An interesting feature of this controller is that jumps occur in both translational and angular velocity when the pendulum swings through $\theta = -\pi$. This phenomenon is a consequence of the coupling between the translational and rotational positions in the rejectability density function, which causes a polarity switch, denoted by ps_i in the figures, to occur between u_R and u_A. On one side of the vertical, $u_R < u_A$, and on the opposite side, $u_R > u_A$. This phenomena is illustrated in Figure 3.5, where the accuracy and rejectability densities are shown for u in the interval between u_R and u_A. Initially, $\mathbf{x} = [3.14, -1.34, 2.54, -4.86]^T$ whereas after the satisficing control has been applied, $\mathbf{x} = [-3.12, -1.39, 2.68, -4.77]^T$. Since the angle θ has changed signs, the coupling between translational and rotational positions causes u_R to change signs. This polarity switch creates a large change in u_D, and the resulting change in rotational rate then acts to restore the control to near its value before the polarity switch. Thus, the phenomenon appears



(a) (b)

Figure 3.5: Polarity switch phenomenon in terms of accuracy and rejectability densities: (a) before switch and (b) after switch.

as an impulsive control input when the pendulum goes through the vertical down position, as illustrated in Figure 3.4.



(a) (b)

Figure 3.6: Phase planes for the inverted pendulum on a cart (m = random): (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second).

Figures 3.6(a) and 3.6(b) illustrate the rotational and translational phase planes for one realization of the random-mass case. Here, $m(t+1) = m(t) + \nu(t)$, with $\nu(t)$ an uncorrelated process with each time-sample drawn from a uniform distribution over the interval $(-0.01, 0.01)$, and with a reflecting boundary at $m_{\min} = 0.1$ kg. Note that, although the trajectories differ quantitatively from those of the constant-mass case, the qualitative behavior is very similar. The controller stabilizes the pendulum and positions the cart at the origin. The fact that the mass is random has no effect on either the controller structure or design procedure.

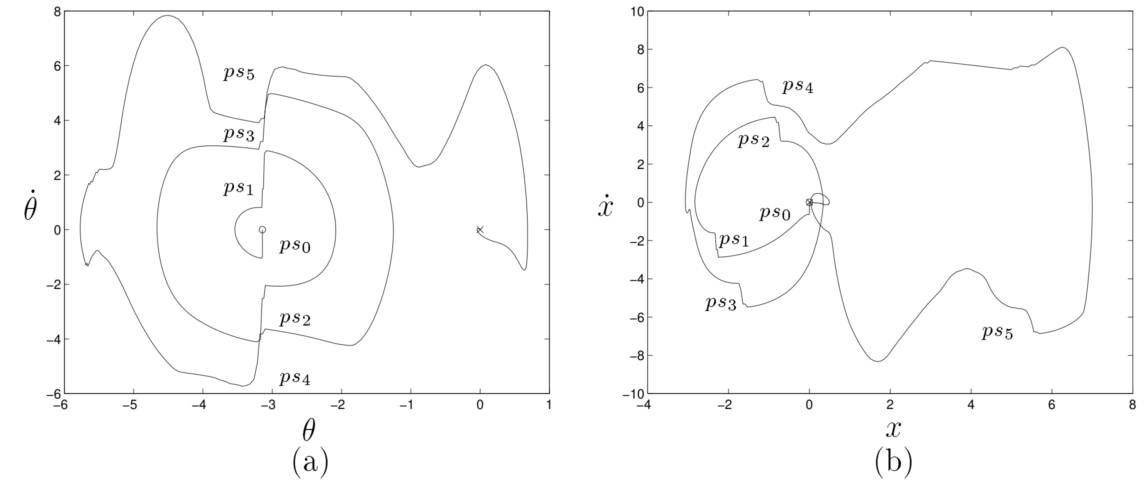


Figure 3.7: Phase planes for the inverted pendulum on a cart with constrained input: (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second).

As a further example of the power of the satisficing approach, take the case of an upper and lower bound on the control u at each time when the bound is not known *a priori*. Thus, restrict u to an interval $(-u_M(t), u_M(t))$ where $u_M(t)$ is a realization of a random walk initialized at $u_M(t_0)$ with reflecting boundaries at $\frac{u_M(t_0)}{2}$ and $\frac{3u_M(t_0)}{2}$. To demonstrate that the satisficing approach can work even when the constraint prevents the application of the unconstrained satisficing solution, the value $u_M(t_0) = 50$ was selected. A step size of ± 2 was chosen. The resulting phase plane plots and control history (including control bounds) for one realization are given in Figures 3.7 and 3.8, respectively. In Figure 3.8, the dotted lines represent the bound $(-u_M(t), u_M(t))$. The polarity switch indicators have been omitted from Figure 3.8 for clarity. Observe that the control always stays within the established bounds even when the bounds deny the use of control that was required in the previous simulations.

In Chapter 2 it was asserted that every decision in the strongly satisficing set was justifiable. Thus far only the most discriminating decision for a given

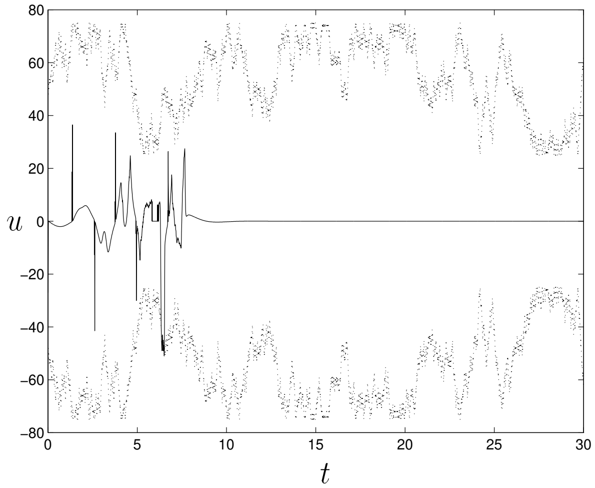


Figure 3.8: Control inputs for the inverted pendulum on a cart with constrained controls (time measured in 0.01 sec increments).

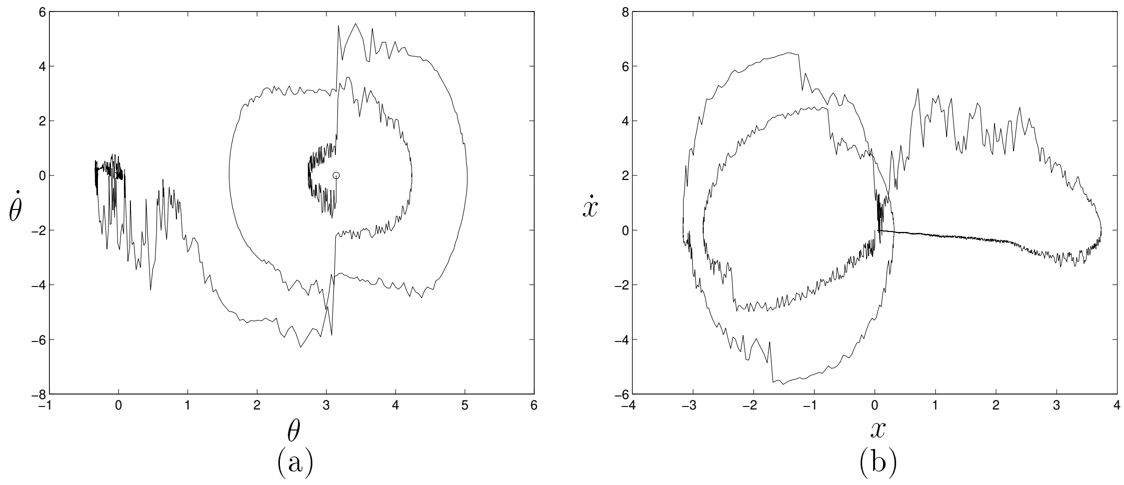


Figure 3.9: Phase planes for the inverted pendulum on a cart with control randomly selected from strongly satisficing set: (a) rotational phase plane (in radians and radians per second), (b) translational phase plane (in meters and meters per second).

rejectivity (selected to yield desired performance) has been applied as a control to a problem. To demonstrate that any $u \in \mathfrak{S}_b$ is justifiable as a control, suppose that controls are selected randomly from $\mathfrak{S}_b$ and applied to the inverted pendulum problem. Let $p(u)$ be a uniform density that assigns all of its mass to $\mathfrak{S}_b$. For each time, any control $u \in \mathfrak{S}_b$ was selected with probability $p(u)$ and applied to the inverted pendulum. The resulting phase plane plots and control history (including control bounds) for one realization are given in Figures 3.9 and 3.10, respectively. Observe that the pendulum behaves qualitatively like the unconstrained control,

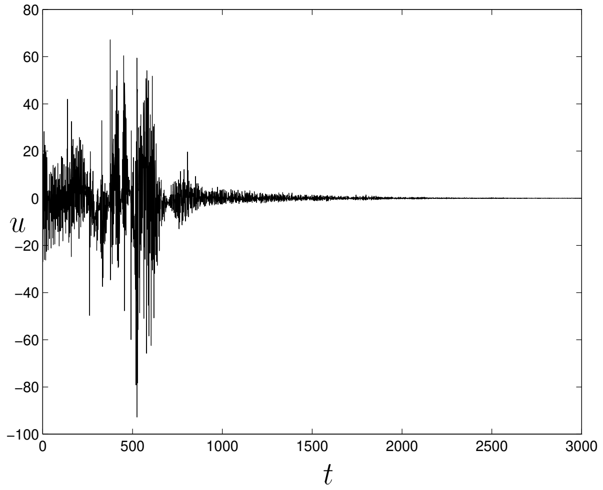


Figure 3.10: Control inputs for the inverted pendulum on a cart with control randomly selected from strongly satisficing set (time measured in 0.01 sec increments).

constant mass case depicted in Figures 3.3 and 3.4 indicating that the any control in $\mathfrak{S}_b$ is justifiable.

3.2.3 The RTAC

The Rotational Translational ACtuator (RTAC) problem serves as a benchmark for robust nonlinear controller design [7, 16]. The system shown in Figure 3.11, represents a translational oscillator with an eccentric rotational proof mass connected by a linear spring of stiffness k to a fixed wall. The cart is constrained to have one-dimensional travel. The proof mass actuator affixed to the cart has mass m and moment of inertia E about its center of mass. Its center of mass is located a distance l from the axis about which it rotates. It is assumed that motion occurs in a horizontal plane, so that there are no gravitational forces

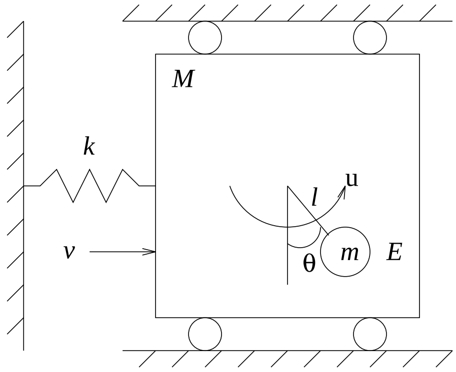


Figure 3.11: Translational oscillator with rotational actuator.

to be considered. u denotes control torque applied to the proof mass and d is the disturbance force on the cart.

Let x and $\dot{x}$ denote the translational position and velocity of the cart, and let θ and $\dot{\theta}$ denote the angular position and velocity of the rotational proof mass. Let $\mathbf{x} = [\theta, x, \dot{\theta}, \dot{x}]^T$ denote the state of the cart/rotational proof mass system. The continuous-time dynamical equation for this problem is

$$\dot{\mathbf{x}} = \boldsymbol{f}(\mathbf{x}) + \boldsymbol{g}(\mathbf{x})u + \mathbf{h}(\mathbf{x})d(t), \tag{3.52}$$

where $\boldsymbol{f} = [f_1, f_2, f_3, f_4]^T$, $\boldsymbol{g} = [g_1, g_2, g_3, g_4]^T$, and $\boldsymbol{h} = [h_1, h_2, h_3, h_4]^T$ are given by

$$\begin{aligned}
f_1(\mathbf{x}) &= \dot{\theta} && (3.53)\\
f_2(\mathbf{x}) &= \dot{\mathbf{x}} && (3.54)\\
f_3(\mathbf{x}) &= \frac{(-ml\cos\theta)(ml\dot{\theta}^2\sin\theta - kx)}{(M+m)(E+ml^2) - m^2l^2\cos^2\theta} && (3.55)\\
f_4(\mathbf{x}) &= \frac{(E+ml^2)(ml\dot{\theta}^2\sin\theta - kx)}{(M+m)(E+ml^2) - m^2l^2\cos^2\theta} && (3.56)\\
g_1(\mathbf{x}) &= 0 && (3.57)\\
g_2(\mathbf{x}) &= 0 && (3.58)\\
g_3(\mathbf{x}) &= \frac{1 - m^2l^2\cos^2\theta}{(E+ml^2)[(M+m)(E+ml^2) - m^2l^2\cos^2\theta]} && (3.59)\\
g_4(\mathbf{x}) &= \frac{-ml\cos\theta}{(M+m)(E+ml^2) - m^2l^2\cos^2\theta} && (3.60)\\
h_1(\mathbf{x}) &= 0 && (3.61)
\end{aligned}$$

$$h_2(\mathbf{x}) = 0 \quad (3.62)$$

$$h_3(\mathbf{x}) = \frac{-ml\cos\theta}{(M+m)(E+ml^2) - m^2l^2\cos^2\theta} \quad (3.63)$$

$$h_4(\mathbf{x}) = \frac{E+ml^2}{(M+m)(E+ml^2) - m^2l^2\cos^2\theta}. \quad (3.64)$$

Using Euler integration, the equivalent discrete-time dynamical expression is

$$\mathbf{x}(t+1) = \mathbf{x}(t) + T\{\boldsymbol{f}[\mathbf{x}(t)] + \boldsymbol{g}[\mathbf{x}(t)]u(t) + \boldsymbol{h}[\mathbf{x}(t)]d(t)\}, \quad (3.65)$$

for $t = 0, 1, \ldots$. With this discretization procedure, $\theta(t+1)$ and $x(t+1)$ are not explicit functions of $u(t)$, but $\theta(t+2)$ and $x(t+2)$ are explicit functions of $u(t)$. Thus, $D_t = 2$, and the influence vector $\boldsymbol{\xi}_0(u) = \{\xi_0(t+1),\ \xi_0(t+2)\}$ can be identified as

$$\xi_0(t+1) = \begin{bmatrix} \dot{\theta}(t+1) \\ \dot{x}(t+1) \end{bmatrix} = \begin{bmatrix} \dot{\theta}(t) + T\{f_3[\mathbf{x}(t)] + g_3[\mathbf{x}(t)]u(t) + h_3[\mathbf{x}(t)]\hat{d}_{t|t-1}\} \\ \dot{x}(t) + T\{f_4[\mathbf{x}(t)] + g_4[\mathbf{x}(t)]u(t) + h_4[xbf(t)]\hat{d}_{t|t-1}\} \end{bmatrix}, \quad (3.66)$$

and

$$\xi_0(t+2) = \begin{bmatrix} \theta(t+2) \\ x(t+2) \end{bmatrix} = \begin{bmatrix} \theta(t) + 2T\dot{\theta}(t) + T^2\{f_3[\mathbf{x}(t)] + g_3[\mathbf{x}(t)]u(t) + h_3[\mathbf{x}(t)]\hat{d}_{t|t-1}\} \\ x(t) + 2T\dot{x}(t) + T^2\{f_4[\mathbf{x}(t)] + g_4[\mathbf{x}(t)]u(t) + h_4[xbf(t)]\hat{d}_{t|t-1}\} \end{bmatrix}, \quad (3.67)$$

where $\hat{d}_{t|t-1}$ is an estimate of the disturbance at time t given observations of this disturbance given the past observations $t-1, t-2, \ldots 0$. Using the value $\hat{d}_{t|t-1}$ corresponds to a simplified version of the expectation used to obtain $\bar{f}_R$ and $\bar{f}_A$. The subjective belief density is $f_C(d) = \delta(d - \hat{d}_{t|t-1})$; i.e., all belief mass is placed on the estimated value. This presents a computational simplification that can be used. In the simulations, $d(t) = \hat{d}_{t|t-1} = 0$, but the disturbance is included for completeness.

The local performance indices become

$$\Phi[\boldsymbol{\xi}_0(u)] = [\xi_0(t+2), \xi_0(t+1)]P \begin{bmatrix} \xi_0(t+2) \\ \xi_0(t+1) \end{bmatrix} \quad (3.68)$$

and

$$\Lambda[\boldsymbol{\xi}_0(u)] = [\xi_0(t+2), \xi_0(t+1)]Q \begin{bmatrix} \xi_0(t+2) \\ \xi_0(t+1) \end{bmatrix} + R_u(t)u^2. \quad (3.69)$$

After finding f_A and f_R, the most discriminating controller is

$$\begin{aligned} u_D &= \left[G^T[\mathbf{x}(t)](P + b'Q(t+1))G[\mathbf{x}(t)] + b'R\right]^{-1} \times \\ & \quad G^T[\mathbf{x}(t)](P + b'Q(t+1))(F[\mathbf{x}(t)] + H[\mathbf{x}(t)]\hat{d}_{t|t-1}), \end{aligned} \tag{3.70}$$

where $G[\mathbf{x}(t)]$ and $F[\mathbf{x}(t)]$ are given in (3.50) and (3.51), respectively, and

$$H[\mathbf{x}(t)] = \left[T^2 h_3[\mathbf{x}(t)], T^2 h_4[\mathbf{x}(t)], T h_3[\mathbf{x}(t)], T h_4[\mathbf{x}(t)]\right]^T.$$

The following values were used in the simulations:

$$Q(t) = \begin{bmatrix} .05 & 0 & 0 & 0 \\ 0 & 5 & 0 & 30 \\ 0 & 0 & .05 & 0 \\ 0 & 30 & 0 & 5 \end{bmatrix}$$

$R(t) = 0.5$, $P = I$, $M = 1.856$ kg, $m = 0.065$ kg, $l = 0.062$ m, $E = 0.000014$ kg m^2, and $k = 203.7$ N/m. The off-diagonal units in $Q(t)$ represent a coupling between cart position and cart velocity.

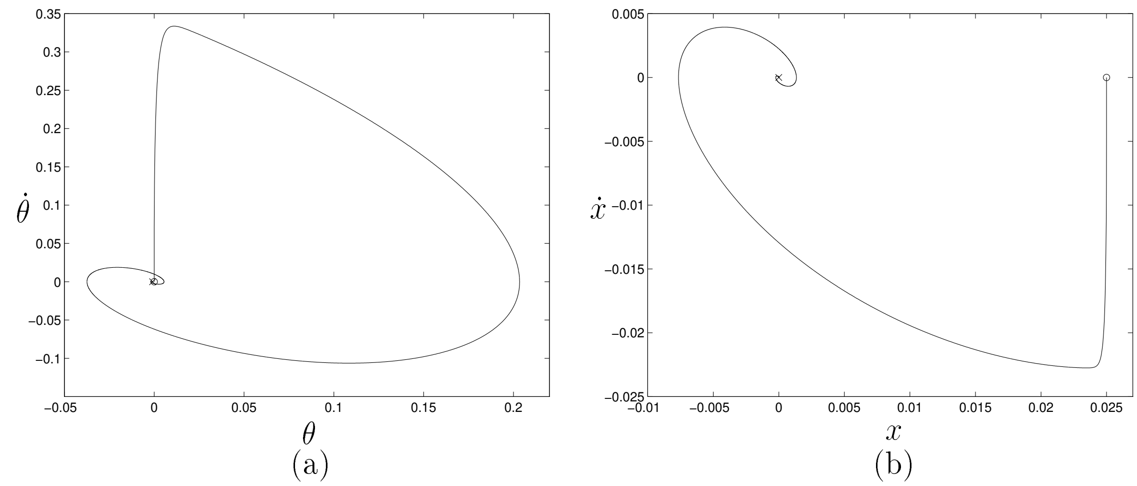


Figure 3.12: Phase planes for the satisficing RTAC controller: (a) rotational proof mass phase plane (in radians and radians per second), (b) translational cart phase plane (in meters and meters per second).

The phase plane plots of the rotational proof mass angular velocity versus angle and the cart velocity versus cart position are shown in Figure 3.12 (a) and (b), respectively. By comparison, simulation results using a discretized

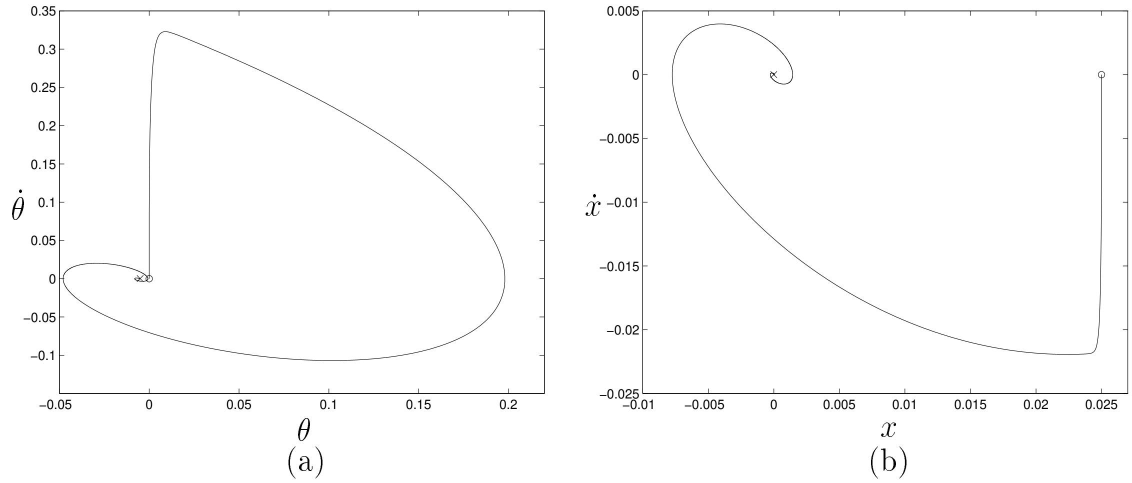


Figure 3.13: Phase planes for Bupp's RTAC controller: (a) rotational proof mass phase plane (in radians and radians per second), (b) translational cart phase plane (in meters and meters per second).

version of the stabilizing controller developed by Bupp, et al., in [7] are shown in Figure 3.13. The system begins in the region of state space designated by the "∘" and terminate at the origin designated by the "×".The two controllers give very similar results.

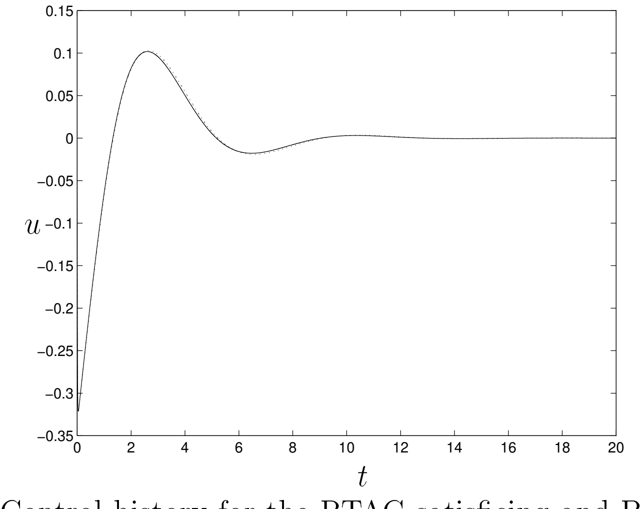


Figure 3.14: Control history for the RTAC satisficing and Bupp's controllers.

The control history for each controller is given in Figure 3.14. The solid line is the satisficing control and the dotted line is Bupp's control. The two are very similar, and yield costs of 11.81 and 12.18 for the satisficing and Bupp's controllers, respectively.

3.3 Nonlinear Minimum Time

For minimum-time problems with constrained final state, the performance index (3.2) is

$$J = \mathbf{x}^T(t_f)\mathbf{x}(t_f) + \sum_{t=0}^{t_f-1} 1 = \mathbf{x}^T(t_f)\mathbf{x}(t_f) + t_f. \tag{3.71}$$

The objective is to find a satisficing control that embodies the minimum-time performance criterion. From the local point of view, the essence of minimizing time is to maximize velocity. Consequently, a measure of velocity, or position change, must be included in the performance index. Adding and subtracting velocity related components to (3.71) yields

$$J = \mathbf{x}^T(t_f)\mathbf{x}(t_f) + t_f \pm \sum_{j=1}^{d}[\mathbf{x}(t_f-d+j)-\mathbf{x}(t_f-d+j-1)]^T[\mathbf{x}(t_f-d+j)-\mathbf{x}(t_f-d+j-1)], \tag{3.72}$$

which may be rearranged as

$$\begin{aligned} J \quad = \quad & \mathbf{x}^T(t_f)\mathbf{x}(t_f) + \sum_{j=1}^{d}[\mathbf{x}(t_f-d+j)-\mathbf{x}(t_f-d+j-1)]^T[\mathbf{x}(t_f-d+j)-\mathbf{x}(t_f-d+j-1)] \\ & +t_f - \sum_{j=1}^{d}[\mathbf{x}(t_f-d+j)-\mathbf{x}(t_f-d+j-1)]^T[\mathbf{x}(t_f-d+j)-\mathbf{x}(t_f-d+j-1)]. \end{aligned} \tag{3.73}$$

From a local perspective, $t+d$ is associated with the terminal time, t_f. Thus, a local terminal performance index may be formed from the first two terms on the right-hand side of (3.73) to obtain

$$\begin{aligned} \Phi[\boldsymbol{\xi}_0(\theta_0),\ldots,\boldsymbol{\xi}_{d-1}(\theta_{d-1})] \quad = \quad & \mathbf{x}^T(t+d)\mathbf{x}(t+d) \\ & + \sum_{j=1}^{d}[\mathbf{x}(t+j)-\mathbf{x}(t+j-1)]^T[\mathbf{x}(t+j)-\mathbf{x}(t+j-1)]. \end{aligned} \tag{3.74}$$

For $d=1$, (3.74) simplifies to

$$\Phi[\boldsymbol{\xi}_0(\theta_0) = 2\mathbf{x}_0^T(t+1)[\mathbf{x}(t)-\mathbf{x}(t+1)] + \mathbf{x}^T(t)\mathbf{x}(t). \tag{3.75}$$

The essence of (3.75) is the inner product $\mathbf{x}^T(t+1)[\mathbf{x}(t) - \mathbf{x}(t+1)]$. Its effect on accuracy is best understood by considering a few simple scenarios. Suppose that the present position, $\mathbf{x}(t)$, is positive (to the right of the origin). If a control, $\theta(t)$, produces a next state, $\mathbf{x}(t+1)$, that is more positive than the present position, then the inner product is negative and the corresponding accuracy, after shifting, is small. If, however, the next state is positive and to the left of the present state, then the inner product is positive and the corresponding accuracy, after shifting, is high. Similar results for other regions of state space indicate that this inner product yields a large accuracy for controls that move the system quickly toward the origin, but low accuracy for controls that move the system in the wrong direction or that move the system slowly. This inner product is the local basis for constructing the accuracy density function.

To construct the rejectability density function, the second two terms from (3.73) are used, and t_f is associated with $t + d$, to form

$$\Lambda[\boldsymbol{\xi}_0(\theta_0), \ldots, \boldsymbol{\xi}_{d-1}(\theta_{d-1})] = t+d-\sum_{j=1}^{d}[\mathbf{x}(t+j)-\mathbf{x}(t+j-1)]^T[\mathbf{x}(t+j)-\mathbf{x}(t+j-1)]. \tag{3.76}$$

The rejectability density function constructed from this places large rejectability on low speeds and small rejectability on large speeds, indicating that wherever the agent travels, it should go there speedily. This is similar to the optimal bang-bang controllers that originate from minimum time problems with bounded controls.

The following theorem establishes the result that in the limit as d approaches t_f, the satisficing control is equivalent to the optimal control. The proof is analogous to the proof of Theorem 8.

Theorem 9 (Consistency of minimum time controller) *For the deterministic minimum time problem defined by (3.10) and (3.71), if the control horizon spans the full extent of the problem, that is, for $d = t_f$, then the most discriminating satisficing control is identical to the optimal minimum-time control.*

3.3.1 Zermelo's Problem

Zermelo's problem is a minimum time problem with nonlinear dynamics for which an optimal solution is known [6, 36]. A satisficing solution to this problem

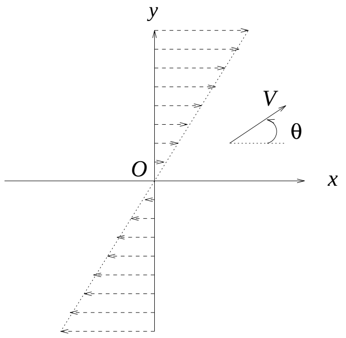


Figure 3.15: Geometry for Zermelo's problem.

is presented and compared to the performance of the optimal solution. Zermelo's problem involves a ship that must travel through a region of strong currents to reach an island, placed at the origin of a Cartesian coordinate system, in minimum time. The current vector, $[c_x, c_y]$, in Cartesian coordinates, is given as $c_x = 0$, $c_y = Vy/h$, as shown in Figure 3.15, where V is the magnitude of the ship's velocity relative to the water, and h is a fixed constant. The ship's heading angle, θ, is defined relative to the positive x axis, and is the control variable for this problem. Define $U = [0, 2\pi]$. The discretized equations of motion are

$$\begin{aligned} x(t+1) &= x(t) + TV[\cos\theta(t) + y(t)/h] \\ y(t+1) &= y(t) + TV\sin\theta(t). \end{aligned}$$

Observe that the explicit influence horizon for this problem is $D_t = 1$, since the control, $\theta(t)$, explicitly affects both $x(t+1)$ and $y(t+1)$. For a d-length control horizon, the influence vectors are $\boldsymbol{\xi}_\ell(u(t+\ell)) = \mathbf{x}(t+\ell+1)$, $0 \leq \ell \leq d-1$. Specifically, for $d = 1$,

$$\boldsymbol{\xi}_0(\theta) = \mathbf{x}(t+1) = \begin{bmatrix} x(t) + TV[\cos\theta) + y(t)/h] \\ y(t) + TV\sin\theta \end{bmatrix}.$$

For the case $d = 1$, substitution of (3.74) into (3.6) yields

$$\begin{aligned} g_A[\theta; \mathbf{x}(t)] &= \sup_{z\in U}\{2\mathbf{x}^T(t+1)[\mathbf{x}(t+1) - \mathbf{x}(t)]\} - 2\mathbf{x}^T(t+1)[\mathbf{x}(t+1) - \mathbf{x}(t)] + \epsilon \\ &= \sup_{z\in U}\left\{2\begin{bmatrix} x(t) + TV[\cos z + y(t)/h] \\ y(t) + TV\sin z \end{bmatrix}^T \begin{bmatrix} TV[\cos z + y(t)/h] \\ TV\sin z \end{bmatrix}\right\} \end{aligned}$$

$$- \begin{bmatrix} x(t) + TV[\cos\theta + y(t)/h] \\ y(t) + TV\sin\theta \end{bmatrix}^T \begin{bmatrix} TV[\cos\theta + y(t)/h] \\ TV\sin\theta \end{bmatrix} + \epsilon \quad (3.77)$$

Again, for the case $d = 1$, substituting (3.76) into (3.8) yields, after some manipulation,

$$\begin{aligned} g_R[\theta; \mathbf{x}(t)] &= \sup_{z\in U}\{[\mathbf{x}(t+1) - \mathbf{x}(t)]^T[\mathbf{x}(t+1) - \mathbf{x}(t)]\} - \\ &\quad [\mathbf{x}(t+1) - \mathbf{x}(t)]^T[\mathbf{x}(t+1) - \mathbf{x}(t)] + \epsilon \\ &= \sup_{z\in U}\left\{ \begin{bmatrix} TV[\cos z + y(t)/h] \\ TV\sin z \end{bmatrix}^T \begin{bmatrix} TV[\cos z + y(t)/h] \\ TV\sin z \end{bmatrix} \right\} \\ &\quad - \begin{bmatrix} TV[\cos\theta + y(t)/h] \\ TV\sin\theta \end{bmatrix}^T \begin{bmatrix} TV[\cos\theta + y(t)/h] \\ TV\sin\theta \end{bmatrix} + \epsilon. \end{aligned} \quad (3.78)$$

Control Method	Δt
Optimal	1.40
1-step satisficing ($b = 0.7$)	1.98
2-step satisficing ($b = 0.5$)	1.66
2-step satisficing ($b = 0.7$)	1.64
2-step satisficing ($b = 0.9$)	1.80

Table 3.1: Zermelo's problem results.

For these dynamics, the accuracy and rejectability are formed from (3.77) and (3.78) and the resulting boat trajectories are plotted for $V = 4.0$, $h = 1.0$, $T = 0.01$, $x(0) = 4.9h$, and $y(0) = 1.66h$ in Figure 3.16. Trajectories are given for the optimal solution, a one-step control horizon satisficing solution, and a two-step control horizon satisficing solution, where the two-step satisficing solution is given for three different values of boldness. The corresponding times are given in Table 3.1.

The one-step satisficing solution does not recognize that by crossing over the $y = 0$ axis the boat can gain speed, since the rejectability for the one-step solution is uniform at $y(t) = 0$. The two-step satisficing solution, by contrast, takes advantage of the $y = 0$ crossing and decreases the amount of time required. High

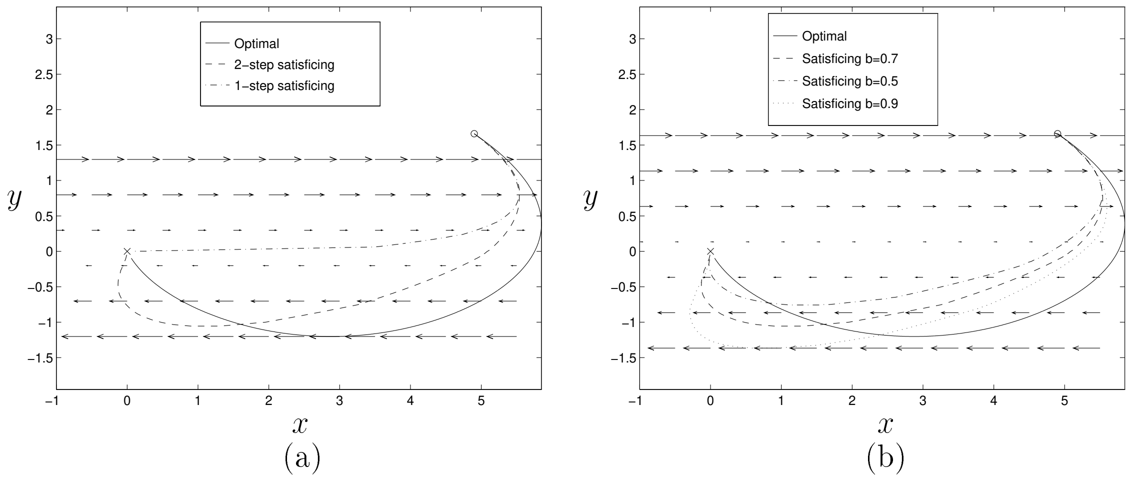


Figure 3.16: Performance results for the single-agent Zermelo's problem problem: (a) comparison of optimal, one-step satisficing, and two-step satisficing, and (b) comparison of optimal with two-step satisficing for varying index of boldness.

boldness ($b = 0.9$) causes the boat to overshoot the island and back-track. Lower boldness ($b = 0.7$) eliminates the overshoot and decreases time, but decreasing the boldness further ($b = 0.5$) increases time since currents are not fully exploited. Although changing boldness does affect performance slightly, the satisficing solutions are not overly sensitive to boldness.

3.3.2 Modified Zermelo's Problem

The minimum-time trajectory for Zermelo's problem leads to a set of differential equations that are difficult to solve; the problem is compounded by the lack of explicit boundary conditions on θ. For nonlinear current fields, an optimal solution is not readily obtained. A satisficing solution, however, is readily obtained. To test this assertion, Zermelo's problem was modified to include the nonlinear current field

$$\begin{aligned} c_x &= V\operatorname{sgn}(x)\frac{\min\{|x|, 4\}}{h}\cos(x) \\ c_y &= V\operatorname{sgn}(y)\frac{\min\{|y|, 4\}}{h}\sin(y), \end{aligned}$$

where $\mathrm{sgn}(x)$ is the the sign function

$$\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ 0 & x = 0 \end{cases}.$$

The trajectories for the one-step control horizon solution to this modified Zermelo's problem are given in Figure 3.17 for two values of $\mathbf{x}(0)$. For Set #1 ($x(0) = 4.9h$, $y(0) = 1.66h$) the elapsed time was $\Delta t = 1.26$, and for Set #2 ($x(0) = 2.45h$, $y(0) = -1.66h$) the elapsed time was $\Delta t = 1.81$.

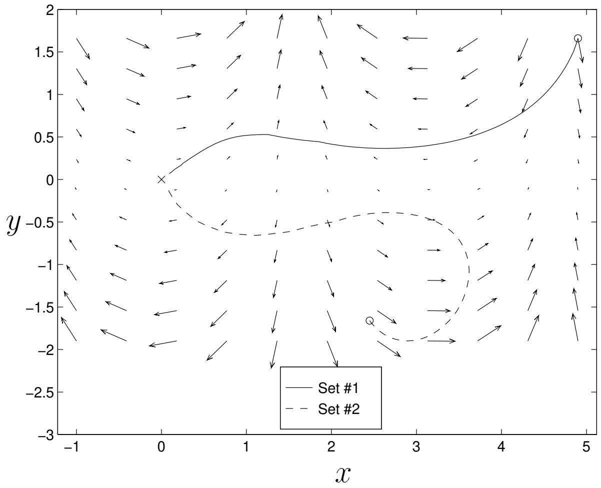


Figure 3.17: 1-step satisficing solutions for modified Zermelo's problem.

Zermelo's problem and modified Zermelo's problem illustrate perhaps the most important property of strongly satisficing control, namely, that it can be applied successfully to problems for which the optimal solution is impractical or even unknown. Furthermore, the satisficing solution not only accomplishes the goal, but does so with reasonable efficiency.

3.4 Some Implementation Details

In the examples in this chapter, controllers have been synthesized for continuous control spaces. Although closed form solutions exist for satisficing linear quadratic regulators, this is not always the case for nonlinear problems. A discretization procedure that exploited the satisficing structure was used. Formally, for the nonlinear regulator problem, the endpoints of the equilibrium set can be calculated in closed form as a function of the current state $\mathbf{x}(t)$ to form the interval (see Theorem 6) (u_*, u^*) as in (3.21)-(3.25). This interval is then discretized into eleven equally spaced points, $U^* = \{u_* = u_1, u_2, \ldots, u_{11} = u^*\}$, where $u_i - u_{i-1} = \frac{u^*(t) - u_*(t)}{10}$. The most discriminating controller from this set is then applied to the system.

Because the most discriminating controller tends to move the system to the origin, the endpoints of the strongly satisficing set tend to converge; i.e., $u^*(t+1) - u_*(t+1) \leq u^*(t) - u_*(t)$. This results in more refined discretization as time progresses. Near the origin, the sampling is very refined. The receding horizon satisficing approach is amenable to this technique because perturbations from the desired trajectory do not destroy the controllers ability to perform. More relevant, perhaps, is that a control which is different from the ideal most discriminating control can be applied without significantly degrading performance, thereby implying that how the control space is discretized is not critical. Thus, the satisficing approach can allow the reduction of computational complexity by permitting flexible discretization schemes.

For Zermelo's problem, the boundaries of the equilibrium set are numerically approximated. The accuracy and rejectability utilities then assign all of their mass to this approximated equilibrium set. The approximated equilibrium set is discretized, and the most discriminating control is applied to the system. Again, because the equilibrium set becomes smaller as the problem approaches the state-space origin, a coarse discretization can be used initially and allowed to refine without increasing the number of samples.

Chapter 4

MULTIPLE AGENT SATISFICING CONTROL: THEORY

The development of a single agent satisficing control theory and applications has established the validity of epistemic utility based satisficing control. Multiple agent control theory is less mature than single agent control theory [56], so it is desirable to to develop multiple agent satisficing control theory and demonstrate its validity. One reason for doing so is that there has been relatively little progress in applying optimal control theory to multi-agent problems. Sometimes this is due to the absence of a well-defined notion of *joint optimality* [37]. In many situations, optimality has different meanings to different agents, and it may be impossible, because of poor communication or ineffective arbitration, for the agents to reach agreement. Even in those cases where an optimality criterion is well defined, it is frequently true that the number, variety, and complexity of possible agent interactions make it difficult to find an optimal solution. Consequently, it is desirable to develop a theory of multiple agent control that can operate when optimality does not or cannot be applied. In this chapter, such a multiple agent satisficing control theory is developed.

4.1 Preliminaries

A multiple agent decision problem consists of two or more decision-makers. The consequences received by each agent are potentially a function the group decision. There are a number of desirable qualities for a multi-agent decision methodology. A methodology should:

1. Allow agents to make decisions when operating under conditions of inter-agent ignorance, independence, or shared information.

2. Provide a mechanism which allows united coalitions to make decisions.

3. Permit changing coalitions and coalitional dynamics.

4. Incorporate the application of strategies (defined as a sequence of decisions).
5. Allow agents to coordinate decisions both in cooperative problems as well as in non-cooperative problems.
6. Degrade gracefully in performance not only when information about other agents is scarce, but also when information about the environment is uncertain.
7. Provide a mechanism for reaching decisions when no clear optimality criterion is available.
8. Allow agents to have their own perspectives which describe their motivations and their relationships to the entire community of agents.

Furthermore, in the control engineering context, the methodology should also

9. Be amenable to a systematic design procedure.

Game theory has been the standard approach to multiple agent decision-making. However, game theory requires restrictive rationality assumptions that may not be satisfied in practice [1]. Consequently, there is a growing perception that progress on multi-agent problems may be more rapid if the requirement for optimal behavior is relaxed in favor of *intelligent* behavior. These controllers have achieved a measure of success [2, 3]. However, some of these techniques are not amenable to systematic design procedures or are restricted to cooperative scenarios [9] thus violating properties 5 and 9. Moore has addressed the problem of allocating resources when information is incomplete [45]. The technique is reasonable for cooperating agents without coalitions, but requires the presence of a centralized policy-maker, thus violating property 1, 2, and 8.

In this chapter, the results of Chapter 2 are generalized by extending satisficing decision and control theory to account for multiple agents. This chapter presents a theory that satisfies properties 1-9 with the exception of property 3. The two limitations imposed in place of property 3 are:

1. An agent may not belong to more than one coalition.

2. Epistemic systems are static. Agents do not learn, and coalitional dynamics are not considered.

Work on coalition formation can be found in [5, 44]. Design examples are presented for agents coordinating in dynamic environments described by differential equations.

A population of M agents may make decisions in many ways: as individuals, as groups, or as a community. These three cases are termed the *individual*, the *coalitional*, and the *communal*, respectively. In this context, the term communal implies that every agent in the community unites in making a decision; i.e., a consensus is intended. With respect to notation, boldface quantities represent vectors: $\mathbf{u} = [u_1, u_2, \dots u_M]^T$ is a vector whose i^{th} component is u_i. The vector made up of the elements of $\mathbf{u}$ with indices in the collection $\mathcal{K} = \{k_1, \dots, k_K\}$ is denoted $\mathbf{u}_{\mathcal{K}} = [u_{k_1}, \dots, u_{k_K}]^T$. Let $\mathbf{u}_{\bar{\mathcal{K}}}$ be the vector formed from $\mathbf{u}$ by removing $\mathbf{u}_{\mathcal{K}}$. The k_ith element of the collection $\mathcal{K}$ is denoted $\mathcal{K}(k_i)$. Quantities pertaining to the jth agent will be identified by a superscript j, as in $P^j_{\mathbf{A}}$ for the accuracy measure. Boldface subscripts are a shorthand notation for K-dimensional measures or densities; e.g., $P^j_{\mathbf{R}} = P_{R^j_1 R^j_2 \cdots R^j_K}$, where K is inferred from the dimension of the argument.

4.2 Satisficing Control

4.2.1 Satisficing Decisions for Agent Coalitions

As defined in Chapter 2, the epistemic system of a single agent acting in isolation consists of a decision space U (with elements denoted u) with associated σ-field $\mathcal{B}$, a state of nature space Θ (with elements ϑ) with associated σ-field $\mathcal{F}$, accuracy and rejectability transition probability densities (measures) $f_A(u;\vartheta)$ and $f_R(u;\vartheta)$ ($P_A(\cdot;\vartheta)$ and $P_R(\cdot;\vartheta)$), respectively, and boldness b. When the agent is in a community of M agents, the joint decision space is denoted $U_{1:M} = U_1 \times \cdots \times U_M$, and the σ-field associated with $U_{1:M}$ is denoted $\mathcal{B}_{1:M}$. Again, nature is denoted by Θ with associated σ-field $\mathcal{F}$. In settings where consequences depend on decisions of all agents in the community, the epistemic system of a single agent, say agent j, includes the M-dimensional rejectability and accuracy transition probability measures (possibly dependent on the state of nature) $P^j_{\mathbf{R}}(\cdot;\vartheta)$ and $P^j_{\mathbf{A}}(\cdot;\vartheta)$, respectively, together with its rejectivity b^j. These measures account for the potential consequences of all decisions made by the community.

If M is large, it may be unusual or impractical for an agent to cooperate with the entire community of agents. More commonly, an agent wishes to cooperate with a subset, or *coalition* $\mathcal{K} = \{k_1, \ldots, k_K\}, K \leq M$ of agents. See [5] for a comparative definition of a coalition. Observe that an agent acting individually corresponds to the special coalition, $\mathcal{K} = \{j\}$, and that agents acting as a community is another special coalition, $\mathcal{K} = \{1, \ldots, M\}$.

When the actions of agents outside $\mathcal{K}$ can affect agent j's decisions, it is beneficial to account for the likely behavior of those agents when forming the set of decisions which obey the satisficing principle. Assume that each coalition can form a subjective model of what group decisions might be made, possibly including dependence upon the state of nature. This dependence on the state of nature is allowed because, for many problems, intelligent behavior depends on external circumstances and these may be uncertain. Let $P^j_{\mathbf{BC}} : \mathcal{B}_{1:M} \times \mathcal{F} \mapsto [0,1]$ represent agent j's subjective belief concerning group decision $G \in \mathcal{B}_{1:M}$ and state of nature $F \in \mathcal{F}$. The exact form of the σ-field $\mathcal{B}_{1:M} \times \mathcal{F}$ is not considered since attention is restricted to the convenient factorization $P^j_{\mathbf{BC}}(G, F) = P^j_{\mathbf{B}|\mathbf{C}}(G|F)P^j_{\mathbf{C}}(F)$. The measure $P^j_{\mathbf{B}|\mathbf{C}}(G|F)$ represents agent j's subjective probability[1] that group decision $G \in \mathcal{B}_{1:M}$ will be made when nature chooses $F \in \mathcal{F}$. The measure $P^j_{\mathbf{C}}(F)$ represents agent j's subjective probability regarding the state of nature.

Accounting for the likely behavior of agents outside of a coalition is a form of coordination since coalition members coordinate their decisions with the likely consequences of the community decision. It is in this sense that satisficing decision theory defines coordination for both cooperative and non-cooperative problems. For coalitional decisions, it is necessary that the σ-field $\mathcal{B}_{1:M}$ be such that it includes the product of the two σ-fields $\mathcal{B}_{\mathcal{K}} \times \mathcal{B}_{\bar{\mathcal{K}}}$; i.e., $\mathcal{B}_{\mathcal{K}} \times \mathcal{B}_{\bar{\mathcal{K}}} \subset \mathcal{B}_{1:M}$. This means that given a set $G_{\mathcal{K}}$ that $\mathcal{K}$ is considering, there must be a set $G \in \mathcal{B}_{1:M}$ such that $G = G_{\mathcal{K}} \times G_{\bar{\mathcal{K}}}$[2] for a potential decision $G_{\bar{\mathcal{K}}}$ made by the agents not in

[1]The game theoretic *max-min* notions are alternatives to the expected behavior method suggested herein. These two methods are not compared in this dissertation.

[2]In this dissertation, $P_{\mathbf{B}|\mathbf{C}}$ will be used on $\mathcal{B}_{\bar{\mathcal{K}}}$ instead of on $\mathcal{B}_{1:M}$. The definition is presented in terms of $\mathcal{B}_{1:M}$ for generality.

$\mathcal{K}$. Define

$$\bar{P}^j_{\mathbf{A}}(G_{\mathcal{K}}) = \int_{U_{\bar{\mathcal{K}}}\times\Theta} P^j_{\mathbf{A}}(G;\vartheta)dP^j_{\mathbf{B}|\mathbf{C}}dP^j_{\mathbf{C}}, \tag{4.1}$$

$$\bar{P}^j_{\mathbf{R}}(G_{\mathcal{K}}) = \int_{U_{\bar{\mathcal{K}}}\times\Theta} P^j_{\mathbf{R}}(G;\vartheta)dP^j_{\mathbf{B}|\mathbf{C}}dP^j_{\mathbf{C}}, \tag{4.2}$$

where the expectation is taken with respect to the likely behavior of the state of nature and the agents in $\bar{\mathcal{K}}$. Define coalition $\mathcal{K}$'s *maximal satisficing set* (from agent j's perspective) to be the element in the σ-field $\mathcal{B}_{\mathcal{K}}$ such that

$$S^j = \arg\sup_{G_{\mathcal{K}}\in\mathcal{B}_{\mathcal{K}}} \left\{\bar{P}^j_{\mathbf{A}}(G_{\mathcal{K}}) - b^j\bar{P}^j_{\mathbf{R}}(G_{\mathcal{K}})\right\}, \tag{4.3}$$

where notational dependence on b and $\mathcal{K}$ has been suppressed for simplicity. Note that this definition incorporates uncertainty about the state and uncertainty about other agent's decisions. Additionally, because of the dependence of the accuracy and rejectability on an uncertain state, the definition implicitly includes uncertainty about the payoffs (or gains) that result from combined actions [4].

Since the Lebesgue Decomposition and the Radon-Nikodym derivative apply for arbitrary probability spaces [53], they apply for $\bar{P}^j_{\mathbf{A}}$ and $\bar{P}^j_{\mathbf{R}}$ defined on $\mathcal{B}_{\mathcal{K}}$. The results from Chapter 2 thereby imply that, when the densities exist and when $\bar{P}_{\mathbf{A}}(S^j) > 0$, the maximal satisficing set is almost everywhere equivalent to

$$S^j = \arg\sup_{G_{\mathcal{K}}\in B_{\mathcal{K}}} \left\{\int_{G_{\mathcal{K}}} \left[\bar{f}^j_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}}) - b^j\bar{f}^j_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}})\right] d\mathbf{u}_{\mathcal{K}}\right\}, \tag{4.4}$$

where

$$\bar{f}^j_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}}) = \int_{U_{\bar{\mathcal{K}}}\times\Theta} f^j_{\mathbf{A}}(\mathbf{u};\vartheta)f^j_{\mathbf{B}|\mathbf{C}}(\mathbf{u};\vartheta)f^j_{\mathbf{C}}(\vartheta)d\mathbf{u}_{\bar{\mathcal{K}}}d\vartheta, \tag{4.5}$$

$$\bar{f}^j_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}}) = \int_{U_{\bar{\mathcal{K}}}\times\Theta} f^j_{\mathbf{R}}(\mathbf{u};\vartheta)f^j_{\mathbf{B}|\mathbf{C}}(\mathbf{u};\vartheta)f^j_{\mathbf{C}}(\vartheta)d\mathbf{u}_{\bar{\mathcal{K}}}d\vartheta. \tag{4.6}$$

If $\bar{P}_{\mathbf{A}}(S^j) = 0$, then

$$S^j = \arg\sup_{G_{\mathcal{K}}\in\mathcal{B}_{\mathcal{K}}} \{\bar{P}_A(G_{\mathcal{K}}) - b\bar{P}_R(G_{\mathcal{K}})\}. \tag{4.7}$$

Because (4.4) and (4.7) are applications of the SLRT applied to multiple decision makers, they are termed Joint SLRT (JSLRT). As in Chapter 2, S^j is nonempty and, if $\bar{f}^j_{\mathbf{A}}$ is concave and $\bar{f}^j_{\mathbf{R}}$ is convex, then S^j is convex.

The maximal satisficing set is the set of all joint coalition decisions for which expected joint accuracy exceeds (rejectivity times) expected joint rejectability, where the expectation is taken with respect to the likely behavior of all agents outside the coalition and with respect to nature.

The nine-tuple

$$\mathcal{M}^j = \{U_{1:M}, \mathcal{B}_{1:M}, \Theta, \mathcal{F}, P^j_{\mathbf{R}}, P^j_{\mathbf{A}}, P^j_{\mathbf{B}|\mathbf{C}}, P^j_{\mathbf{C}}, b^j\} \tag{4.8}$$

is called agent j's M-agent *epistemic system*. The collection of the perspectives of each agent j gives

$$\mathcal{M} = \bigcup_{j \in \mathcal{J}} \mathcal{M}^j = \left\{ U_{1:M}, \mathcal{B}_{1:M}, \Theta, \mathcal{F}, \bigcup_{j \in \mathcal{J}} \left\{ P^j_{\mathbf{A}}, P^j_{\mathbf{R}}, P^j_{\mathbf{B}|\mathbf{C}}, P^j_{\mathbf{C}}, b^j \right\} \right\}. \tag{4.9}$$

Allowing for different perspectives sets the stage for studying coalitional dynamics (changing coalitions). However, coalitional dynamics are not explored in this dissertation. Many cooperative problems can be treated by having all agents share a common perspective and similar utilities. In this case, (4.8) reduces to a simpler form $\mathcal{M} = \left\{U_{1:M}, \mathcal{B}_{1:M}, \Theta, \mathcal{F}, P_{\mathbf{A}}, P_{\mathbf{R}}, P_{\mathbf{B}|\mathbf{C}}, P_{\mathbf{C}}, b\right\}$.

It is frequently desirable to consider the set of decisions acceptable to coalition $\mathcal{K}$ under a constraint condition $\boldsymbol{\zeta}(\mathbf{u}_{\mathcal{K}}) \leq \mathbf{c}$ [8, 13, 65]. Define the set $\mathcal{C} = \{\mathbf{u}_{\mathcal{K}} : \boldsymbol{\zeta}(\mathbf{u}_{\mathcal{K}}) \leq \mathbf{c}\}$, for some constraining function $\boldsymbol{\zeta} : \Re^K \mapsto \Re^L$, $\mathbf{c} \in \Re^L$. Then (4.3) changes to restrict attention to only those sets G in the σ-algebra that are subsets of $\mathcal{C}$

$$S^j = \arg \sup_{G_{\mathcal{K}} \in \mathcal{B}_{\mathcal{K}} : G_{\mathcal{K}} \subset \mathcal{C}} \left[\bar{P}^j_{\mathbf{A}}(G_{\mathcal{K}}) - b\bar{P}^j_{\mathbf{R}}(G_{\mathcal{K}}) \right]. \tag{4.10}$$

In terms of densities and assuming $\bar{P}_{\mathbf{A}}(S_b) \neq 0$, (4.10) becomes

$$S^j = \arg \sup_{G_{\mathcal{K}} \in \mathcal{B}_{\mathcal{K}} : G_{\mathcal{K}} \subset \mathcal{C}} \left\{ \int_{G_{\mathcal{K}}} \left[\bar{f}^j_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}}) - b^j \bar{f}^j_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}}) \right] d\mathbf{u}_{\mathcal{K}} \right\}. \tag{4.11}$$

Though agent j and agent k are in the same coalition, acting in good faith does not guarantee that their joint decision will always be in the satisficing set. This is discussed in the next section.

4.2.2 Consequences of Coalition Involvement

The set S^j consists of all joint decisions (elements of the product space $U_{\mathcal{K}}$) that are satisficing to coalition $\mathcal{K}$. Suppose that agents j and k in $\mathcal{K}$, denoted $\mathcal{K}(j)$ and $\mathcal{K}(k)$, respectively, are disposed to cooperate. Then, from the point of view of agent j, an agent k in the coalition must confine itself to actions in the set

$$G^j_{\mathcal{K}(k)} = \left\{ \mathbf{v} \in U_{\mathcal{K}} : \exists \mathbf{u} \in S^j \text{ with } u_k = v_k \right\}. \tag{4.12}$$

The set $G^j_{\mathcal{K}(k)}$ represents those decisions in $U_{\mathcal{K}}$ for which agent k's decision u_k is compatible with at least one decision $\mathbf{u} \in S^j$. Agent k is said to act *in good faith* with respect to agent j when it confines itself, in the absence of other knowledge, to $G^j_{\mathcal{K}(k)}$.

By extension, community actions outside the set

$$G^j = \bigcap_{k=1}^{K} G^j_{\mathcal{K}(k)} \tag{4.13}$$

cannot be satisficing to agent j. The set G^j consists of all decisions that may be reached when agents in the community act in good faith by restricting attention to components of the satisficing set. Clearly, $U_{\mathcal{K}} \supset G^j \supset S^j$. However, because it is not in general true that $S^j \supset G^j$, agents acting in good faith can produce joint decisions outside S^j. This is easily illustrated for the special case when S^j is convex. In this case G^j is determined by the intervals $({}_*u_k,\ {}^*u_k)$, where ${}_*u_k$ is the kth element of $\inf_{\mathbf{u}\in S^j} \mathbf{u}_k$ and ${}^*u_k$ is the kth element of $\sup_{\mathbf{u}\in S^j} \mathbf{u}_k$. Then

$$G^j = \bigcap_{k=1}^{K} G^j_{\mathcal{K}(k)} = \prod_{k=1}^{K} ({}_*u_k,\ {}^*u_k)$$

is the smallest hyper-rectangle in $U_{\mathcal{K}}$ containing the set S^j. Thus, although agent j and agent k are in the same coalition, even when acting in good faith it is possible for their decision to fall outside the satisficing set.

Because of this, it may be desirable for agents to invoke external criteria to prevent such problems. Suppose agent k wishes to guarantee that its decisions produce joint decisions within S^j, provided other agents act in good faith. To accomplish this desire, agent k must choose a decision in the set

$$H^j_{\mathcal{K}(k)} = \left\{ \mathbf{v} \in U_{1:M} : \text{ for any } \mathbf{u} \in G^j \text{ such that } u_k = v_k \text{ implies } \mathbf{u} \in S^j \right\} \tag{4.14}$$

The intersection of the sets gives

$$H^j = \bigcap_{k=1}^{M} H^j_{\mathcal{K}(k)}. \tag{4.15}$$

Clearly, $H^j \subset S^j$. However, not only can H^j be empty, but each $H^j_{\mathcal{K}(k)}$ can also be empty.

Figure 4.1 illustrates the relationships between these sets from agent 1's perspective for the case $M = 2$ and $\mathcal{K}\{1, 2\}$. The set enclosed by the closed curve in

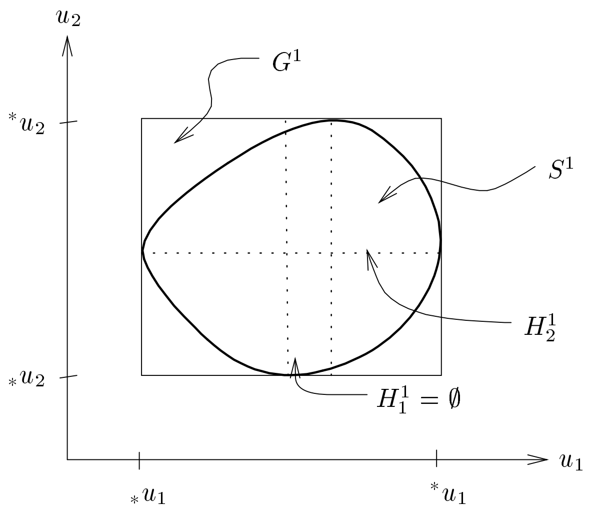


Figure 4.1: The relationships between S^1, $G^1_{\mathcal{K}(k)}$, and $H^1_{\mathcal{K}(k)}$, assuming f^1_A is concave and f^1_R is convex.

boldface represents the satisficing set S^1. The rectangle bounding S^1 represents G^1. The set H^1_2 is represented by the horizontal dotted line and shows that agent 2, by confining its decisions to this set, guarantees that the joint decision will be satisficing for agent 1. However, even if agent 2 confines its decision to G^1, it is still possible for agent 1 to choose a decision such that the joint decision is outside S^1. Hence, $H^1_1 = \emptyset$. Consequently in the absence of certainty about other agents decisions, it is impossible to guarantee that a joint decision will be communally satisficing—even when agents are trying to cooperate.

Thus, the external imposition of sets H^j is rejected in favor of the equilibrium property of the strongly satisficing set. In Sections 4.3 and 4.4 conditions will be established which guarantee that when agents in $\mathcal{K}$ act in good faith the resulting decision is strongly satisficing.

4.2.3 Satisficing Strategies: Coordination over Time

The principles used to generate multi-agent satisficing sets can also be applied to developing a single agent satisficing set for a sequence of decisions. A sequence of decisions over time can be interpreted as a *strategy*, or a coordination with "future selves". In developing a strategy, consider a finite time horizon, and use t as a discrete time variable. The M-step horizon *strategic form* of agent j's epistemic system consists of a joint decision space, $U_{1:M} = U_t \times \cdots \times U_{t+M-1}$, its associated σ-field, $\mathcal{B}_{1:M}$, a strategic joint rejectability, $f^j_{\mathbf{R}}$, a strategic joint accuracy, $f^j_{\mathbf{A}}$, and the rejectivity, b^j. Whereas in multiple agent coordination the joint densities reflect the beliefs and values of several agents, in single-agent strategic coordination these joint probability densities incorporate a single agent's beliefs and values as a function of time; e.g., $f^j_{\mathbf{R}}(\mathbf{u}) = f_{R^j(t)R^j(t+1)\cdots R^j(t+M-1)}(u(t), u(t+1), \ldots, u(t+M-1))$. The marginal probability density reflects an agent's belief that action $u(t+i)$ is the correct action to take at time $t+i$, $i = 0, \ldots, M-1$, independent of actions taken at other times. The joint probability densities include the effects that previous and subsequent actions will have on the action taken at time $t+i$. If the agent's environment is nonstationary, then a model of what an agent's "future self" might do is necessary. This implies a need for a subjective belief $f^j_{\mathbf{B}}$ that incorporates the model parameters[3]. Note that both strategic and multiple agent coordination can be considered simultaneously; the resulting satisficing set retains the form of (4.4).

4.3 Strongly Satisficing Control

As defined in Chapter 2, the strongly satisficing set for single agents consists of those decisions which satisfy a certain equilibrium condition: for a fixed accuracy, they have the smallest rejectability, and for a fixed rejectability, they have the largest accuracy. The construction of the strongly satisficing set for multiple agents parallels that of a single agent, but differs in that it must take into

[3]Whereas extending the control horizon d requires an accurate model of system dynamics, choosing a strategy based on coordinating with a "future self" allows more modeling latitude.

account the actions of several agents.

The set

$$\mathfrak{E}^j = \{\mathbf{u}_\mathcal{K} : \nexists \mathbf{v}_\mathcal{K} \in U_\mathcal{K} \text{ for which } \bar{f}^j_\mathbf{A}(\mathbf{v}_\mathcal{K}) > \bar{f}^j_\mathbf{A}(\mathbf{u}_\mathcal{K}) \text{ and } \bar{f}^j_\mathbf{R}(\mathbf{v}_\mathcal{K}) < \bar{f}^j_\mathbf{R}(\mathbf{u}_\mathcal{K})\} \tag{4.16}$$

is the coalitional *equilibrium set*, where explicit dependence on the coalition $\mathcal{K}$ has been suppressed. Each element in this set enjoys the property that accuracy and rejectability are either both increasing or both decreasing in a neighborhood of the element, and the accuracy is maximized over all decisions that yield the same rejectability. The coalitional *strongly satisficing set* is the intersection of the maximal satisficing set and the coalitional equilibrium set:

$$\mathfrak{S}^j = \mathfrak{E}^j \bigcap S^j. \tag{4.17}$$

That the strongly satisficing set defined above is consistent with the single agent definition follows trivially by considering a coalition with one member $\mathcal{K} = \{j\}$.

The following theorem states that the strongly satisficing set is nonempty.

Theorem 10 *The strongly satisficing set* $\mathfrak{S}^j \neq \emptyset$.

<u>*Proof:*</u> Define the *most discriminating* satisficing control

$$\mathbf{u}_D = \arg\sup_{\mathbf{z}\in S^j} \left\{\bar{f}^j_\mathbf{A}(\mathbf{z}) - b\bar{f}^j_\mathbf{R}(\mathbf{z})\right\}.$$

Then, as in the corresponding proof in Chapter 2, $\mathbf{u}_D \in \mathfrak{E}^j$ and $\mathbf{u}_D \in \mathfrak{S}^j$ which implies that $\mathfrak{S}^j \neq \emptyset$. ■

As in Chapter 2, the most accurate control $\mathbf{u}^j_\mathbf{A} = \arg\max_{\mathbf{v}\in U_\mathcal{K}} \bar{f}^j_\mathbf{A}(\mathbf{v})$ corresponds to the most discriminating control for $b = 0$. Similarly, the least rejectable control $\mathbf{u}^j_\mathbf{R} = \arg\min_{\mathbf{v}\in U_\mathcal{K}} \bar{f}^j_\mathbf{R}(\mathbf{v})$ corresponds to the most discriminating control for $b \to \infty$.

4.4 An Important Special Case

The following two theorems establish desirable properties for the synthesis of controllers. The first theorem states the convexity properties of the satisficing, equilibrium, and strongly satisficing sets for concave accuracy and convex rejectability. The second theorem establishes that the equilibrium set consists of decisions that are maximal for given operating conditions.

Suppose that both accuracy and rejectability are differentiable with gradients given by

$$\nabla \bar{f}_{\mathbf{A}}^{j} = \left[\frac{\partial \bar{f}_{\mathbf{A}}^{j}}{\partial u_{k_1}}, \ldots, \frac{\partial \bar{f}_{\mathbf{A}}^{j}}{\partial u_{k_K}}\right], \quad \nabla \bar{f}_{\mathbf{R}}^{j} = \left[\frac{\partial \bar{f}_{\mathbf{R}}^{j}}{\partial u_{k_1}}, \ldots, \frac{\partial \bar{f}_{\mathbf{R}}^{j}}{\partial u_{k_K}}\right].$$

The derivatives in the direction of a unit vector $\mathbf{v}$ are given by $\nabla \bar{f}_{\mathbf{A}} \cdot \mathbf{v}$, and $\nabla \bar{f}_{\mathbf{R}} \cdot \mathbf{v}$, respectively.

Lemma 11 *For differentiable accuracy and differentiable rejectability,*

$$\mathfrak{L}^{j} \subset \left\{\mathbf{u} \in U_{\mathcal{K}} : \forall \mathbf{v} \ \left(\nabla \bar{f}_{\mathbf{A}}(\mathbf{u}) \cdot \mathbf{v}\right)\left(\nabla \bar{f}_{\mathbf{R}}(\mathbf{u}) \cdot \mathbf{v}\right) \geq 0\right\} \cup \{\mathbf{u}_{\mathbf{R}}, \mathbf{u}_{\mathbf{A}}\}. \tag{4.18}$$

Proof: For any $\mathbf{u} \in \mathfrak{L}^{j}$, accuracy and rejectability are (1) either both increasing or both decreasing in the neighborhood of $\mathbf{u}$ or (2) $\mathbf{u} \in \{\mathbf{u}_{\mathbf{R}}, \mathbf{u}_{\mathbf{A}}\}$. For (1), the accuracy and rejectability directional derivatives in any direction must be the same sign. ■

The condition presented in Lemma 11 can be strengthened when accuracy is both differentiable and concave and when rejectability is both differentiable and convex.

Lemma 12 *For differentiable concave accuracy and differentiable convex rejectability,*

$$\mathfrak{L}^{j} = \left\{\mathbf{u} \in U_{\mathcal{K}} : \forall \mathbf{v} \ \left(\nabla \bar{f}_{\mathbf{A}}(\mathbf{u}) \cdot \mathbf{v}\right)\left(\nabla \bar{f}_{\mathbf{R}}(\mathbf{u}) \cdot \mathbf{v}\right) \geq 0\right\} \cup \{\mathbf{u}_{\mathbf{R}}, \mathbf{u}_{\mathbf{A}}\}. \tag{4.19}$$

Proof: If $\mathbf{u} \in \{\mathbf{u}_{\mathbf{R}}, \mathbf{u}_{\mathbf{A}}\}$ then $\mathbf{u} \in \mathfrak{S}^{j}$ and, hence, $\mathbf{u} \in \mathfrak{L}^{j}$.

Otherwise, suppose there exists a

$$\mathbf{u} \in \left\{\mathbf{w} \in U_{\mathcal{K}} : \forall \mathbf{v} \ \left(\nabla \bar{f}_{\mathbf{A}}(\mathbf{u}) \cdot \mathbf{v}\right)\left(\nabla \bar{f}_{\mathbf{R}}(\mathbf{u}) \cdot \mathbf{v}\right) \geq 0\right\}$$

that is not in $\mathfrak{L}^{j}$. Then, there must exist a $\mathbf{w} \in U_{\mathcal{K}}$ for which $\bar{f}_{\mathbf{A}}^{j}(\mathbf{w}) > \bar{f}_{\mathbf{A}}^{j}(\mathbf{u})$ and $\bar{f}_{\mathbf{R}}^{j}(\mathbf{w}) < \bar{f}_{\mathbf{R}}^{j}(\mathbf{u})$. Since rejectability is concave, the directional derivative in the direction of the vector $\mathbf{v} = \mathbf{w} - \mathbf{u}$, given by $\nabla \bar{f}_{\mathbf{R}}^{j}(\mathbf{u}) \cdot \mathbf{v}$, is negative. However, since accuracy is convex, the directional derivative along $\mathbf{v}$, given by $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u})\cdot\mathbf{v}$, is positive. Thus, there is a $\mathbf{v}$ such that

$$\left(\nabla\bar{f}_{\mathbf{A}}(\mathbf{u})\cdot\mathbf{v}\right)\left(\nabla\bar{f}_{\mathbf{R}}(\mathbf{u})\cdot\mathbf{v}\right) < 0.$$

This is a contradiction which means that $\mathbf{u}\in\mathfrak{E}^j$. ∎

Theorem 13 *Let $U_\mathcal{K}\subset\Re^K$. For $\bar{f}^j_{\mathbf{A}}$ a differentiable concave density function and $\bar{f}^j_{\mathbf{R}}$ a differentiable convex density function over $U_\mathcal{K}$, the maximal satisficing set S^j, the equilibrium set $\mathfrak{E}^j$, and the strongly satisficing set $\mathfrak{S}^j$ are all convex.*

Proof: The proof of Theorem 6 holds for vector $\mathbf{u}_\mathcal{K}$. Hence, S^j is convex.

Let $\mathbf{u}_1,\mathbf{u}_2\in\mathfrak{E}^j$, and let $\mathbf{u}=\alpha\mathbf{u}_1+(1-\alpha)\mathbf{u}_2$. If $\forall\lambda\quad\mathbf{u}\in\{\mathbf{u}_\mathbf{A},\mathbf{u}_\mathbf{R}\}$, then $\mathbf{u}_1=\mathbf{u}_2=\mathbf{u}_\mathbf{A}=\mathbf{u}_\mathbf{R}$ so the set is trivially convex. Otherwise, since the directional derivatives cannot change sign in $\mathfrak{E}^j$ (see proof of Theorem 6), $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_1)\cdot\mathbf{v}$, $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_2)\cdot\mathbf{v}$, $\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_1)\cdot\mathbf{v}$, and $\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_2)\cdot\mathbf{v}$ must have the same sign. Without loss of generality, suppose that $\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_1)\cdot\mathbf{v}\geq\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_2)\cdot\mathbf{v}$ (one must be greater for convex rejectability) and that $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_1)\cdot\mathbf{v}\geq\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_2)\cdot\mathbf{v}$ (one must be greater for concave accuracy). Then $\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_1)\cdot\mathbf{v}\geq\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u})\cdot\mathbf{v}\geq\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_2)\cdot\mathbf{v}$ and $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_1)\cdot\mathbf{v}\geq\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u})\cdot\mathbf{v}\geq\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_2)\cdot\mathbf{v}$ whence $(\bar{f}^j_{\mathbf{A}}(\mathbf{u})\cdot\mathbf{v})(\bar{f}^j_{\mathbf{R}}(\mathbf{u})\cdot\mathbf{v})\geq 0$. By Lemma 12, $\mathbf{u}\in\mathfrak{E}^j$.

Since the intersection of convex sets is convex, $\mathfrak{S}^j$ is convex. ∎

Theorem 14 *Let $U_\mathcal{K}\in\Re^K$, let $\bar{f}^j_{\mathbf{A}}$ be differentiable over the interior of $U_\mathcal{K}$ and let $\bar{f}^j_{\mathbf{R}}$ be differentiable over the interior of $U_\mathcal{K}$. Then for every $\mathbf{u}_\mathcal{K}\in\mathfrak{E}^j$ there exists a rejectivity $b\in[0,\infty)$ such that $\mathbf{u}_\mathcal{K}$ is a most discriminating satisficing control. Furthermore, if $\bar{f}^j_{\mathbf{A}}-b\bar{f}^j_{\mathbf{R}}$ is strictly concave for every $b\in[0,\infty)$, then for any such b there corresponds a unique most discriminating satisficing control.*

Proof: By the definition of $\mathfrak{E}^j$, every $\mathbf{u}_\mathcal{K}\in\mathfrak{E}^j$ must maximize accuracy for a given rejectability. By the Lagrange Multiplier Theorem [38] and Lemma 11, there exists a $\lambda\in[0,\infty)$ such that $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_\mathcal{K})=\lambda\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_\mathcal{K})$. This means that $\nabla\bar{f}^j_{\mathbf{A}}(\mathbf{u}_\mathcal{K})-\lambda\nabla\bar{f}^j_{\mathbf{R}}(\mathbf{u}_\mathcal{K})=0$. However, the discrimination function $\bar{f}^j_{\mathbf{A}}(\mathbf{u}_\mathcal{K})-b\bar{f}^j_{\mathbf{R}}(\mathbf{u}_\mathcal{K})$ is concave since it is the sum of two concave functions ($\bar{f}^j_{\mathbf{A}}(\mathbf{u}_\mathcal{K})$ and $-b\bar{f}^j_{\mathbf{R}}(\mathbf{u}_\mathcal{K})$), it is maximized when $\nabla\bar{f}^j_{\mathbf{A}}-b\nabla\bar{f}^j_{\mathbf{R}}=0$. Setting $b=\lambda$ completes the proof. Furthermore, for $b\in[0,\infty)$, $\mathbf{u}_\mathcal{K}=\arg\sup_{\mathbf{v}_\mathcal{K}\in U_\mathcal{K}}\left\{\bar{f}^j_{\mathbf{A}}(\mathbf{v}_\mathcal{K})-b\bar{f}^j_{\mathbf{R}}(\mathbf{v}_\mathcal{K})\right\}$ exists and is unique provided the discrimination function $\bar{f}^j_{\mathbf{A}}-b\bar{f}^j_{\mathbf{R}}$ is strictly concave. ∎

In consequence of these theorems, when the discrimination function is strictly concave, choosing the maximally discriminating controller for a given rejectivity uniquely chooses the coalition decision. When agents agree on a value for rejectivity, acting in good faith will produce strongly satisficing decisions.

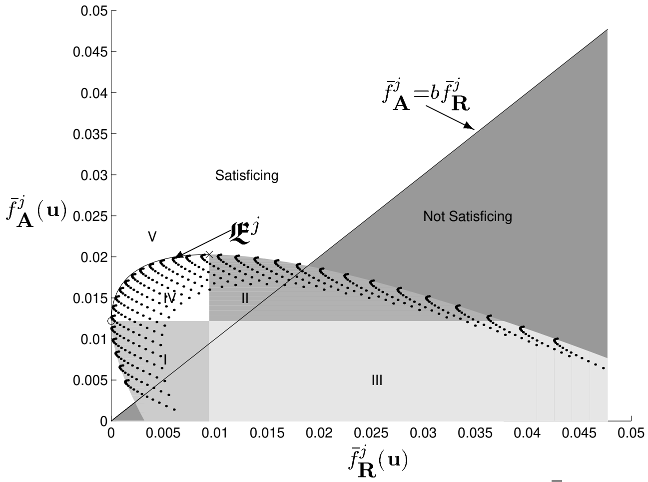


Figure 4.2: Satisficing regions for a concave $\bar{f}_A$ and convex $\bar{f}_R$.

Figure 4.2 illustrates the maximal satisficing and strongly satisficing sets for strictly concave differentiable μ^j. The accuracy and rejectability densities are cross-plotted as parametric functions of $\mathbf{u}$. The "$\cdot$" symbol denotes the pair $(\bar{f}^j_{\mathbf{R}}(\mathbf{u}), \bar{f}^j_{\mathbf{A}}(\mathbf{u}))$ as $\mathbf{u}$ ranges over $U_{\mathcal{K}}$. The "$\times$" symbol represents the point $(\bar{f}^j_{\mathbf{R}}(\mathbf{u}_A), \bar{f}^j_{\mathbf{A}}(\mathbf{u}_A))$, where the accuracy is a maximum. The "$\circ$" symbol represents the $(\bar{f}^j_{\mathbf{R}}(\mathbf{u}_R), \bar{f}^j_{\mathbf{A}}(\mathbf{u}_R))$, where the rejectability is minimum. The points to the left of the $\bar{f}^j_{\mathbf{A}} = b\bar{f}^j_{\mathbf{R}}$ line are the satisficing solutions. The graph may be divided into five regions. In Region I, $\bar{f}^j_{\mathbf{R}}(\mathbf{u}) > \bar{f}^j_{\mathbf{R}}(\mathbf{u}_R)$, but $\bar{f}^j_{\mathbf{A}}(\mathbf{u}) < \bar{f}^j_{\mathbf{A}}(\mathbf{u}_R)$, thus solutions in this region should be rejected since the minimally rejectable solution has both higher accuracy and lower rejectability. In Region II, $\bar{f}^j_{\mathbf{A}}(\mathbf{u}) < \bar{f}^j_{\mathbf{A}}(\mathbf{u}_A)$ but $\bar{f}^j_{\mathbf{R}}(\mathbf{u}) > \bar{f}^j_{\mathbf{R}}(\mathbf{u}_A)$, thus solutions in this region should be rejected since the maximally accurate solution has both higher accuracy and lower rejectability. Region III should also obviously be rejected since both the maximally accurate and minimally rejectable solutions are superior. Region IV is the region for which $\bar{f}^j_{\mathbf{A}}(\mathbf{u}) < \bar{f}^j_{\mathbf{A}}(\mathbf{u}_A)$ and $\bar{f}^j_{\mathbf{R}}(\mathbf{u}) > \bar{f}^j_{\mathbf{R}}(\mathbf{u}_R)$. Solutions in this region should be rejected because they do not maximize accuracy for a given rejectability; i.e., they are not in the equilibrium set. Region V consists of the arc segment between the "$\times$" and the "$\circ$". This region is the region of points that are in the equilibrium set. If the line $\bar{f}^j_{\mathbf{A}} = b\bar{f}^j_{\mathbf{R}}$ intersects $\mathfrak{E}^j$, then only points above the line qualify as members of $\mathfrak{S}^j$ (in Figure 4.2, $\mathfrak{S}^j$ coincides with $\mathfrak{E}^j$).

Chapter 5

MULTIPLE AGENT SATISFICING CONTROL: APPLICATIONS

In this chapter, a receding horizon strategy will be employed to develop a synthesis procedure for satisficing controllers. The synthesis procedure includes both pursuit-evasion and quadratic regulator characteristics. Representative control problems are presented.

5.1 Multiple Agent Controller Synthesis

It is desirable to present a synthesis procedure for multiple agent control problems and to illustrate its application to specific examples. First the principles governing the synthesis of satisficing controllers are reviewed [63]. Then a relatively broad class of multi–agent control problems are described in terms of a particular three–team game. The game includes pursuit-evasion and nonlinear quadratic tracking features, and forms the basis for the applications. Finally, satisficing controllers are designed for two special cases of the game and performance is quantified.

To synthesize a satisficing controller, it is necessary to formulate both an accuracy and a rejectability measure. In Chapter 4, the accuracy utility describes the agent's purpose; it is equivalent to the global objective of the controller. By contrast, the rejectability utility describes the penalty, or cost of control, associated with each action. In multiple agent problems, both measures depend on the actions of other agents. Therefore, the design of a multi-agent satisficing controller may also require a subjective probability measure. Thus, an agent's *purpose*, *penalty*, and *subjective belief* are all explicit in the problem formulation.

As in Chapter 3, when applied to problems for which an optimal control exists, the satisficing procedure should obey the consistency principle. Because of the absence of a well-defined notion of optimality in the capture the flag game presented below, consistency is not shown. However, for the shared resource quadratic regulator, consistency is satisfied.

Assume agent j on coalition $\mathcal{K}$ (denoted by a subscript $\mathcal{K}(j)$) is governed by the dynamics equation

$$\mathbf{x}_{\mathcal{K}(j)}(t+1) = \mathbf{f}_{\mathcal{K}(j)}(\mathbf{x}_{\mathcal{K}(j)}(t), \mathbf{u}_{\mathcal{K}(j)}(t), t).$$

Imagine a game with three teams of agents, denoted $\mathcal{K}_E$, $\mathcal{K}_P$, and $\mathcal{K}_O$. Let K_E, K_P, and K_O be the number of agents on team $\mathcal{K}_E$, $\mathcal{K}_P$, and $\mathcal{K}_O$, respectively. Suppose that team $\mathcal{K}_O$ is trying to track (capture) all members of team $\mathcal{K}_E$ (team *Evader*) while avoiding members of team $\mathcal{K}_P$ (team *Pursuer*). The control problem for team $\mathcal{K}_O$ is then to force a generalized distance between the agent j on $\mathcal{K}_O$ and agent i on team $\mathcal{K}_E$, $\phi(\mathbf{x}_{\mathcal{K}_O(j)}(t), \mathbf{x}_{\mathcal{K}_E(i)}(t))$, to be small, while simultaneously keeping the generalized distance measure between agent j on $\mathcal{K}_O$ and agent k on team $\mathcal{K}_P$, $\psi(\mathbf{x}_{\mathcal{K}_P(k)}(t), \mathbf{x}_{\mathcal{K}_O(j)}(t))$ large and the energy cost $L(\mathbf{x}_{\mathcal{K}_O(j)}(t+1), \mathbf{u}_{\mathcal{K}_O(j)}(t))$ small. Because this game includes both non-cooperative (team $\mathcal{K}_P$ versus team $\mathcal{K}_O$ versus team $\mathcal{K}_E$) and cooperative aspects (on team $\mathcal{K}_O$, agent k_i cooperates with agent k_j), it is a very rich model for multi-agent problems. For agents on team $\mathcal{K}_O$ to make decisions, they require a model of what agents from other teams are doing. The model team $\mathcal{K}_O$ chooses to describe teams $\mathcal{K}_E$ and $\mathcal{K}_P$ dictates the structure of the subjective probability density.

Let Θ denote the states associated with each agent on each team; i.e.,

$$\Theta = \Big\{ \{\mathbf{x}_{\mathcal{K}_E(i)}(t) : i = 1, \dots, K_E\}, \{\mathbf{x}_{\mathcal{K}_O(j)}(t) : j = 1, \dots, K_O\}, \{\mathbf{x}_{\mathcal{K}_P(k)}(t) : k = 1, \dots, K_P\} \Big\}$$

Recall that $U_{\mathcal{K}}$ represents the set of possible group decisions, $\mathbf{u}_{\mathcal{K}}$, made by team $\mathcal{K}$. The community decision space, $U_{1:M}$ can be written as the product space of each team's (=coalition's) decision space:

$$U_{1:M} = \left(\prod_{i=1}^{K_E} U_{\mathcal{K}_E(i)} \right) \times \left(\prod_{j=1}^{K_O} U_{\mathcal{K}_O(j)} \right) \times \left(\prod_{k=1}^{K_O} U_{\mathcal{K}_P(k)} \right) = U_{\mathcal{K}_E} \times U_{\mathcal{K}_O} \times U_{\mathcal{K}_P},$$

where $M = K_O + K_E + K_P$ and $U_{\mathcal{K}(l)}$ is the decision space for agent l on team $\mathcal{K}$. The accuracy and rejectability densities will be formed from two carefully constructed utilities: $g_{\mathbf{A}} : U_{1:M} \mapsto \Re$ and $g_{\mathbf{R}} : U_{1:M} \mapsto \Re$. As in Chapter 2, these utilities will be based on local performance indices.

As in Chapter 2, let $D_{t+\ell}$ denote the explicit influence horizon of control $\mathbf{u}_{\mathcal{K}}$, defined as the maximum number of time increments for which $\mathbf{u}_{\mathcal{K}}(t+\ell)$ has explicit influence on any component of the state. Let $\xi_\ell(t+\tau)$ be the sub-state of $\mathbf{x}(t+\tau)$ that is an explicit function of $\mathbf{u}_{\mathcal{K}}(t+\ell)$, $\tau = 1, \ldots, D_{t+\ell}$. Continuing in this manner, d sequences of sub-states may be generated, denoted the *influence vectors*, of the form

$$\begin{aligned}
\boldsymbol{\xi}_0(\mathbf{u}_{\mathcal{K}}(t)) &= \{\xi_0(t+1), \ldots, \xi_0(t+D_t)\} \\
\boldsymbol{\xi}_1(\mathbf{u}_{\mathcal{K}}(t+1)) &= \{\xi_1(t+2), \ldots, \xi_1(t+1D_{t+1})\} \\
&\vdots \\
\boldsymbol{\xi}_{d-1}(\mathbf{u}_{\mathcal{K}}(t+d-1)) &= \{\xi_{d-1}(t+d), \ldots, \xi_{d-1}(t+d-1+D_{t+d})\}.
\end{aligned} \tag{5.1}$$

The collection $\{\boldsymbol{\xi}_\ell(\mathbf{u}_{\mathcal{K}}(t+\ell)),\ 0 \le \ell \le d-1\}$ then represents the state elements that are explicit functions of $\mathbf{u}(t), \ldots, \mathbf{u}(t+d-1)$, and using these elements *local performance indices*, $\Phi[\boldsymbol{\xi}_0(\mathbf{u}_{\mathcal{K}}(t)), \ldots, \boldsymbol{\xi}_{d-1}(\mathbf{u}_{\mathcal{K}}(t+d-1))]$, $\Psi[\boldsymbol{\xi}_0(\mathbf{u}_{\mathcal{K}}(t)), \ldots, \boldsymbol{\xi}_{d-1}(\mathbf{u}_{\mathcal{K}}(t+d-1))]$, and $\Lambda[\boldsymbol{\xi}_0(\mathbf{u}_{\mathcal{K}}(t)), \ldots, \boldsymbol{\xi}_{d-1}(\mathbf{u}_{\mathcal{K}}(t+d-1))]$ may be formed and used to compute f_A and f_R. These local performance indices are formed according to the same principles that are used to define the global performance indices, ϕ, ψ, and L. Specific examples of these functionals will be provided in subsequent sections.

Consider the case $d = 1$. The accuracy density for team $\mathcal{K}_O$ is based on choosing a control to minimize the local performance index

$$\Phi(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t)), \boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{E(i)}}(t))).$$

Recalling that superscript O means the perspective of team O, taking the sum[1] of these functions over all agents on each team gives the cost of each action set taken by the group

$$J^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_P}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_O}) = \sum_{j=1}^{K_O} \sum_{i=1}^{K_E} \Phi(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t)), \boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{E(i)}}(t))), \tag{5.2}$$

Form the negated function

$$g^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) = \max_{\mathbf{v}_O, \mathbf{v}_E, \mathbf{v}_P} \left\{ J^O_{\mathbf{A}}(\mathbf{v}_{\mathcal{K}_O}, \mathbf{v}_{\mathcal{K}_E}, \mathbf{v}_{\mathcal{K}_P}) \right\} - J^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) + \epsilon, \tag{5.3}$$

[1]A product of the functions could also be used as a model. This would result in a different class of controllers. These alternate classes may be applicable to different problems.

where ϵ is chosen to guarantee that $g_{\mathbf{A}}$ is positive. The accuracy density is then defined as

$$f^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) = \frac{g^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P})}{\int_{U_{1:M}} g^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) d\mathbf{u}}. \tag{5.4}$$

The rejectability is constructed similarly from the local performance indices $\Psi\left(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{P(k)}}(t)), \boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t))\right)$ and $\Lambda\left(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t)), \mathbf{u}_{\mathcal{K}_{O(j)}}(t)\right)$. Form the functions

$$\begin{aligned} J^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) &= \sum_{j=1}^{K_O}\sum_{k=1}^{K_P}\Big[-\Psi(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{P(k)}}(t)), \boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t))) + \\ &\quad \Lambda(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t)), \mathbf{u}_{\mathcal{K}_{O(j)}}(t))\Big] \qquad (5.5) \\ g^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) &= -\min_{\mathbf{v}_M, \mathbf{v}_{\mathcal{K}_E}, \mathbf{v}_{\mathcal{K}_P}}\left\{J^O_{\mathbf{R}}(\mathbf{v}_{\mathcal{K}_O}, \mathbf{v}_{\mathcal{K}_E}, \mathbf{v}_{\mathcal{K}_P})\right\} + \\ &\quad J^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) + \epsilon. \qquad (5.6) \end{aligned}$$

The rejectability density from team O's perspective is then defined as

$$f^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}, \vartheta) = \frac{g^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P})}{\int_{U_{1:M}} g^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}) d\mathbf{u}}. \tag{5.7}$$

After constructing the accuracy and rejectability densities, the subjective belief density $f_{\mathbf{BC}} : U_{1:M} \mapsto [0, 1]$ must be constructed. This subjective belief density is a function of the modeling that an agent invokes in a multi-agent environment. In the following section, some examples of subjective belief densities adopted from different modeling techniques are presented.

The final step in developing the controller is in forming the strongly satisficing set. The resulting controller should allow agents on a coalition to share a common, finite control resource that potentially constrains possible group decisions.

Adopting a strategy (considering several successive moves at one time) requires only a straightforward modification of the cost function followed by shifting and scaling as in (5.2)-(5.7). Let the purpose and penalty functions be applied to a sequence of d time steps, yielding

$$\begin{aligned} J^O_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_P}; t, \ldots, t+d-1) &= \sum_{\ell=0}^{d-1}\sum_{i=1}^{K_O}\sum_{j=1}^{K_E} \Phi(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t+\ell)), \boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{E(i)}}(t+\ell))), \\ J^O_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}_O}, \mathbf{u}_{\mathcal{K}_E}, \mathbf{u}_{\mathcal{K}_E}; t, \ldots, t+d-1) &= \sum_{\ell=0}^{d-1}\sum_{j=1}^{K_O}\sum_{k=1}^{K_P}\Big[-\Psi(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{P(k)}}(t+\ell)), \boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t+\ell))) + \\ &\quad \Lambda(\boldsymbol{\xi}(\mathbf{u}_{\mathcal{K}_{O(j)}}(t+\ell)), \mathbf{u}_{\mathcal{K}_{O(j)}}(t+\ell))\Big]. \end{aligned}$$

5.2 Pursuit-Evasion: Capture the Flag

5.2.1 Problem Formulation

Capture the Flag (CTF) is a special case of the three team game introduced in Section 5.1. In CTF, an *attack* agent attempts to capture a flag while two *guard* agents attempt to prevent the flag's capture. For simplicity, denote the attacker by a, the two guards by g_i, $i = 1, 2$, and the flag by f. From the attacker's perspective, $M = \{f, a, g_1, g_2\}$, team $\mathcal{K}_E = \{f\}$, team $\mathcal{K}_O = \{a\}$, and team $\mathcal{K}_P = \{g_1, g_2\}$; hence $K_E = 1$, $K_O = 1$, and $K_P = 2$. From the guard's perspective, $M = \{a, g_1, g_2\}$ team $\mathcal{K}_E = \{a\}$, team $\mathcal{K}_O = \{g_1, g_2\}$, and team $\mathcal{K}_P = \emptyset$; hence $K_E = 1$, $K_O = 2$, and $K_P = 0$[2]. Depending on the amount of communication allowed, the guards may exploit their coalition or they may operate essentially individually. The agents operate in a two dimensional Cartesian plane with coordinates denoted (x, y). The flag is placed at rest at the origin $(x_f, y_f) = (0, 0)$.

The attacker attempts to "capture" the flag, defined as the condition $(x_a(t) - x_f)^2 = x_a^2(t) \leq W$ and $(y_a(t) - y_f)^2 = y_a^2(t) \leq W$, and wins the game if it can do so for some $0 \leq t \leq t_f$, where t_f is the duration of the game. The guards attempt to protect the flag by capturing the attacker, which occurs if, for either $i = 1$ or 2, the distance $(x_a(t) - x_{g_i}(t))^2 \leq W$ and $(y_a(t) - y_{g_i}(t))^2 \leq W$ for some $0 \leq t \leq t_f$. If the guards do not capture the attacker but prevent the attacker from capturing the flag, the game ends in a draw.

Let the dynamics of each agent be described by discrete-time, linear, time-invariant state equations. Assume that each agent can control its x-directional and y-directional thrust, and that the guards and the attacker have different masses. Applying Newton's second law and Euler integration gives, for each agent, the set of discrete-time dynamical equations

$$\begin{aligned}
\mathbf{x}(t) &= [x(t), y(t), \dot{x}(t), \dot{y}(t)]^T, \\
&= [\mathbf{p}(t), \mathbf{v}(t)]^T,
\end{aligned}$$

[2]Alternately, the guards could model the flag as an agent that they should pursue. This leads to other interesting controllers with $\mathcal{K}_E = \{a, f\}$, $\mathcal{K}_O = \{g_1, g_2\}$, and $\mathcal{K}_P = \emptyset$.

$$\begin{aligned}\mathbf{u}(t) &= [F_x(t), F_y(t)]^T, \\ \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t),\end{aligned}$$

$$A = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{T}{m} & 0 \\ 0 & \frac{T}{m} \end{bmatrix},$$

where $F_x(t)$ and $F_y(t)$ are the x-directional and y-directional thrusts, respectively, $\mathbf{p}(t)$ and $\mathbf{v}(t)$ are the position and velocity sub-state vectors of the state vector $\mathbf{x}(t)$, respectively, and T is the sample time. The attacker and guards require similar constructs and are distinguished by the subscripts a and g_i, respectively.

Using this set of notation, the attacker's dynamical equations are described by

$$\mathbf{x}_a(t+1) = A_a\mathbf{x}_a(t) + B_a\mathbf{u}_a(t), \tag{5.8}$$

where $\mathbf{x}_a(t)$ denotes the attacker's state and $\mathbf{u}_a(t) = [F_{x_a}, F_{y_a}]^T$ denotes the attacker's control at time t. Similarly, the dynamical equations of guard i are denoted by

$$\mathbf{x}_{g_i}(t+1) = A_{g_i}\mathbf{x}_{g_i}(t) + B_{g_i}\mathbf{u}_{g_i}(t), \tag{5.9}$$

where the subscript i denotes guard i, $\mathbf{x}_{g_i}(t)$ is the state of of guard i at time t, and $\mathbf{u}_{g_i}(t) = [F_{x_{g_i}}, F_{y_{g_i}}]^T$ is guard i's control at time t.

Both the attacker and the guards have equal thrust capacity ($F_{\max}$ is the same for the guards and the attacker), but the guards have only half as much mass as the attacker, $m_{g_i} = \frac{1}{2}m_a$. This allows the guards to be more mobile. This extra mobility is granted because of the comparative difficulty of the purpose of the guards versus the attacker. The attacker is required to pursue a stationary point; i.e., the attacker's controller is a regulator. The guards, by contrast, are required to pursue a nonstationary point; i.e., the controllers are trackers. Moreover, the point the guards are required to track is intentionally trying to avoid them. It is because of this difficulty that the guards are allowed to be more mobile. However, when the guards cooperate, it will be shown that this extra mobility may not be needed.

5.2.2 CTF Epistemic System

For the given agent dynamics, the influence vector for the current input $\mathbf{u}(t)$ consists of both the agent velocity at time $t+1$, $\mathbf{v}(t+1)$ and the agent position at time $t+2$, $\mathbf{p}(t+2)$; i.e., $\boldsymbol{\xi}(\mathbf{u}(t)) = [\mathbf{v}(t+1), \mathbf{p}(t+2)]^T$ where, since time is indicated by the argument, the time subscript on $\boldsymbol{\xi}$ has been replaced by a subscript indicating team affiliation. In terms of the dynamics equations, the components of the influence vector are

$$\mathbf{v}(t+1) = \begin{bmatrix} \dot{x}(t+1) \\ \dot{y}(t+1) \end{bmatrix} = \begin{bmatrix} \dot{x}(t) + TF_x(t) \\ \dot{y}(t) + TF_y(t) \end{bmatrix} \tag{5.10}$$

and

$$\mathbf{p}(t+2) = \begin{bmatrix} x(t+2) \\ y(t+2) \end{bmatrix} = \begin{bmatrix} x(t) + 2T\dot{x}(t) + T^2F_x(t) \\ y(t) + 2T\dot{y}(t) + T^2F_y(t) \end{bmatrix}. \tag{5.11}$$

Assume that each agent has an estimate of where other agents are located in state space. Additionally, each agent may also have a model of what other agents can do. Each agent uses its knowledge of where other agents are or where the might be to form an accuracy and a rejectability function. Denote an estimated influence vector by $\hat{\boldsymbol{\xi}}(\mathbf{u}(t))$, and assume that all belief is placed on that estimate; i.e., $f_{\mathbf{BC}} = \delta(\boldsymbol{\xi}(\mathbf{u}(t)) - \hat{\boldsymbol{\xi}}(\mathbf{u}(t)))$.

The attacker's accuracy is based on a desire to capture the flag, i.e., to choose the sequence $\{\mathbf{u}_a(t), \mathbf{u}_a(t+1), \ldots, \mathbf{u}_a(t+t_f)\}$ so that for some $t \leq t_f$ then $x_a^2(t) \leq W$ and $y_a^2(t) \leq W$. The attacker's accuracy can be formed from the influence vectors by considering the attacker's global goal to reach the flag. Thus,

$$\Phi^a(\boldsymbol{\xi}(\mathbf{u}_a(t)), \boldsymbol{\xi}(\mathbf{u}_g(t)), \boldsymbol{\xi}(\mathbf{u}_f(t))) = [\boldsymbol{\xi}_a(\mathbf{u}_a(t)) - \hat{\boldsymbol{\xi}}_f(\mathbf{u}_f(t))]^T P_a [\boldsymbol{\xi}_a(\mathbf{u}_a(t)) - \boldsymbol{\xi}_f(\mathbf{u}_f(t))]. \tag{5.12}$$

Since the flag stays in one position, $\hat{\boldsymbol{\xi}}_f(\mathbf{u}_f(t)) = 0$ for all time. The accuracy is then found using (5.2)-(5.4).

The attacker's rejectability is based on a desire to evade both of the guards and to conserve fuel. The rejectability is therefore defined using

$$\begin{aligned} \Lambda^a(\boldsymbol{\xi}(\mathbf{u}_a(t)), \mathbf{u}_a(t)) &= \mathbf{u}_a^T(t) R_a \mathbf{u}_a(t), \\ \Psi^a(\boldsymbol{\xi}(\mathbf{u}_a(t)), \boldsymbol{\xi}(\mathbf{u}_{g_i}(t)), \boldsymbol{\xi}(\mathbf{u}_f(t))) &= \left[\boldsymbol{\xi}_a(\mathbf{u}_a(t)) - \hat{\boldsymbol{\xi}}_{g_i}(\mathbf{u}_{g_i}(t))\right]^T Q_a \left[\boldsymbol{\xi}_a(\mathbf{u}_a(t)) - \hat{\boldsymbol{\xi}}_{g_i}(\mathbf{u}_{g_i}(t))\right], \end{aligned}$$

where $\hat{\boldsymbol{\xi}}_{g_i}$ represents the estimate of guard i's state at as a function of $\mathbf{u}_{g_i}(t)$. Rejectability is then found using (5.5)-(5.7).

Each guard's accuracy is based on a desire to catch the attacker. When the guards operate independently they have identical accuracy and rejectability measures. The cost function

$$\Phi^{g_i}(\boldsymbol{\xi}(\mathbf{u}_{g_i}(t)), \boldsymbol{\xi}(\mathbf{u}_a(t))) = \left[\boldsymbol{\xi}_{g_i}(\mathbf{u}_{g_i}(t)) - \hat{\boldsymbol{\xi}}_a(\mathbf{u}_a(t))\right]^T P_{g_i} \left[\boldsymbol{\xi}_{g_i}(\mathbf{u}_{g_i}(t)) - \hat{\boldsymbol{\xi}}_a(\mathbf{u}_a(t))\right]$$

measures the distance between guard i and the attacker. When this distance is small, accuracy is large, indicating that the guard should move toward the attacker. The guard coalition's accuracy can then be found using (5.2)-(5.4).

Each guard's rejectability is based on a desire remain close to the flag and to conserve fuel. This translates into a high energy cost when the guard moves away from the flag or when moves require large amounts of fuel. The rejectability cost becomes

$$\begin{aligned}
\Psi^{g_i}(\boldsymbol{\xi}(\mathbf{u}_{g_i}(t)), \boldsymbol{\xi}(\mathbf{u}_a(t))) &= 0 \\
\Lambda^{g_i}(\boldsymbol{\xi}(\mathbf{u}_{g_i}(t)), \mathbf{u}_{g_i}(t)) &= \left[\boldsymbol{\xi}_{g_i}^T(\mathbf{u}_{g_i}(t)) - \hat{\boldsymbol{\xi}}_f\right]^T Q_{g_i} \left[\boldsymbol{\xi}_{g_i}^T(\mathbf{u}_{g_i}(t)) - \hat{\boldsymbol{\xi}}_f\right] + \\
&\quad \mathbf{u}_{g_i}^T(t) R_{g_i} \mathbf{u}_{g_i}(t)
\end{aligned}$$

that is high when an action would move the guard away from the flag or require energy. The guard coalition's rejectability can then be found using (5.5)-(5.7).

Simulation results are given in Figure 5.1. In the figures, a "$\circ$" represents an agent's starting position and the "$\times$" represents the flag. The attacker's trajectory is represented by a solid line, and each guard's trajectory is represented by a dotted line. The simulations reported herein use the following values: $m_a = 4$, $m_g = 2$, $T = 0.01$, $W = 0.2$, $F_{\max} = 5000$, $b_{g_1} = b_{g_2}$, $t_f = 1.5$,

$$2R_a = R_{g_i} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad Q_a = Q_{g_i} = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$P_a = P_{g_i} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In the simulations, the attacker is constrained to start from a position farther from the flag than the most distant guard and in the sector bounded by the lines from the flag to the two guards. Simulations were performed for several rejectivity pairs (b_a, b_{g_i}).

5.2.3 Reactive Modeling

When agents do not try to anticipate future actions of other agents, but base their actions solely on the current positions of other agents, their decisions are termed *reactive*. This form of modeling is the memoryless case of the communication suggested in [2]. A reactive attacker estimates guard positions by $\hat{\boldsymbol{\xi}}(\mathbf{u}_{g_i}(t)) = [0, \mathbf{p}_{g_i}(t)]^T$. The resulting belief function can be interpreted as follows: the attacker believes that the guards will be in their current position and at rest in two time steps. This is reasonable if the attacker is able to estimate the guard's current positions, but unable to predict future positions. Similarly, a reactive guard estimates the positions of the attacker using $\hat{\boldsymbol{\xi}}(\mathbf{u}_a(t)) = [0, \mathbf{p}_a(t)]^T]$. Reactive modeling is equivalent to using the current positions of other agents as the estimate of their future positions.

The results of five-thousand games with random initial starting points are tabulated in Table 5.1. Results are reported from the attacker's perspective (*win* means that the attacker wins, etc.) In the simulations, the attacker wins up to 98% of the games and loses no more than 37%. A typical result is shown in Figure 5.1 (a). For reactive agents, the attacker begins by traveling directly toward the flag. Because it believes the guards will remain in their current positions, it is not aware of the danger presented by the converging guards. The guards begin the game by heading toward the attacker, but they overshoot the attacker and the attacker wins the game.

5.2.4 Strategic Decisions

In this section, the attacker is allowed to use a two-step ($d = 2$) control horizon strategy to improve its performance. The agents still model other agents reactively, but the attacker develops a strategy based on its reactive model of the guards. For the next time step, the attacker assumes that its "future self" at

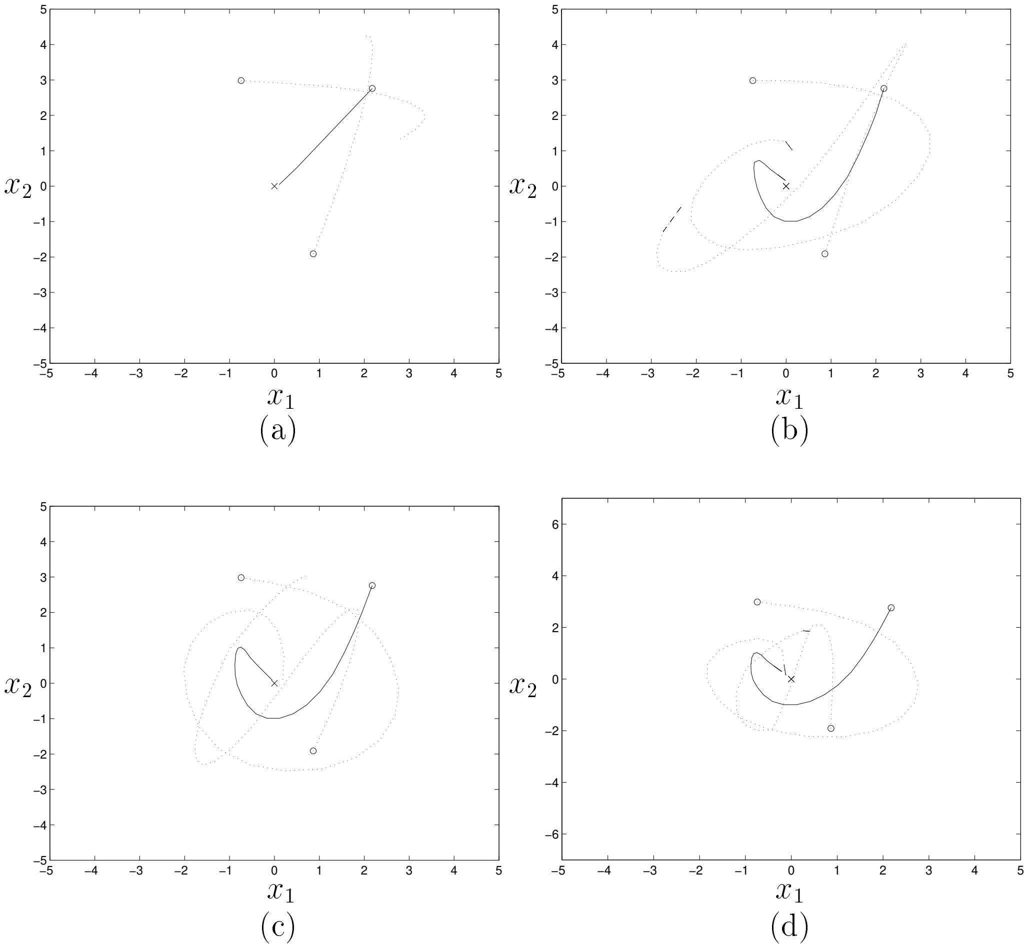


Figure 5.1: Typical results for $(b_a, b_g) = (1.0, 0.5)$ (a) reactive guards and reactive attacker, (b) reactive guards and strategizing attacker, (c) anticipating guards and strategizing attacker, and (d) coordinating anticipating guards and strategizing attacker.

Outcome	(b_a, b_g)		
	(1.0,1.0)	(1.0,0.5)	(1.0,0.0)
% win	76.98	74.66	63.04
% loss	22.98	25.34	36.96
% draw	0.04	0.00	0.00

(a)

Outcome	(b_a, b_g)		
	(0.5,1.0)	(0.5,0.5)	(0.5,0.0)
% win	85.00	85.66	70.98
% loss	13.46	14.32	29.02
% draw	1.54	0.02	0.00

(b)

Outcome	(b_a, b_g)		
	(0.0, 1.0)	(0.0, 0.5)	(0.0, 0.0)
% win	96.16	91.80	98.22
% loss	3.82	3.52	1.62
% draw	0.02	4.68	0.16

(c)

Table 5.1: Simulation results for varying indexes of boldness for attacker and guards using reactive modeling: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$.

time $t+1$ will choose the control action that will move it nearest the flag. This assumption is useful since the game is not played in a stationary environment; the guards move after each time step, changing the attacker's cost functions. By modeling the behavior of its "future self", the attacker effectually invokes a strategy instead of simply increasing the control horizon. The resulting influence vector is $\hat{\boldsymbol{\xi}}(\mathbf{u}_a(t+1)) = \boldsymbol{\xi}(\mathbf{u}_a^*(t+1))$ where $\mathbf{u}_a^*(t+1)$ represents the velocities and positions that obtain under the assumption.

The results of 5000 Monte Carlo simulations are given in Table 5.2. The adoption of a strategy gives the attacker a significant advantage over the reactive guards. The strategizing attacker now wins up to 98% of the games and loses no more than 14%. A typical result for the strategizing attacker is shown in Figure 5.1 (b). The simulation begins with the two guards moving rapidly toward

Outcome	(b_a, b_g)		
	(1.0,1.0)	(1.0,0.5)	(1.0,0.0)
% win	92.08	90.50	89.84
% loss	7.90	9.44	8.82
% draw	0.02	0.06	1.34

(a)

Outcome	(b_a, b_g)		
	(0.5,1.0)	(0.5,0.5)	(0.5,0.0)
% win	94.52	92.18	86.34
% loss	5.36	7.82	13.40
% draw	0.12	0.00	0.26

(b)

Outcome	(b_a, b_g)		
	(0.0, 1.0)	(0.0, 0.5)	(0.0, 0.0)
% win	95.54	95.96	98.32
% loss	4.46	4.04	1.68
% draw	0.00	0.00	0.00

(c)

Table 5.2: Simulation results for varying indexes of boldness for attacker using reactive modeling coupled with strategic decisions, and guards using reactive modeling: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$.

the attacker. The attacker avoids the closer guard and heads toward the flag. One guard trails the attacker, and the second guard overshoots the attacker allowing the attacker to rapidly move toward the flag. When the attacker realizes that it has overrun the flag and that the guards are not an immediate threat, it begins turning toward the origin. This causes the second guard to again overshoot the attacker. When this guard overshoots the attacker, the attacker turns quickly toward the flag and wins the game.

5.2.5 Anticipatory Modeling

Now consider a scenario in which the attacker adopts a strategy based on a reactive model of the guards, but in which the guards attempt to anticipate the attacker's movements. Let each guard know A_a and B_a. Furthermore, let each guard assume that the attacker will choose an action $\mathbf{u}_a(t)$ that will maximize that guard's rejectability, given that the guard chooses $\mathbf{u}_{g_i}(t)$. This strategy is conservative, but not inappropriate. The resulting influence vector is $\hat{\boldsymbol{\xi}}(\mathbf{u}_a(t)) = \boldsymbol{\xi}(\hat{\mathbf{u}}_a(t))$ where $\hat{\mathbf{u}}_a = \arg\max_{\mathbf{v}_a(t)\in U_a} J_{\mathbf{R}}^{g_i}(\mathbf{u}_{g_i}(t), \mathbf{v}_a(t))$. where $J_{\mathbf{R}}^{g_i}$ is given by (5.5).

The simulation results for 500 runs are tabulated in Table 5.3 and indicate that the strategizing attacker is much less successful than before. The attacker wins up to 72% of the games, but loses up to 85%. A typical result is shown in

Outcome	(b_a, b_g)		
	(1.0,1.0)	(1.0,0.5)	(1.0,0.0)
% *win*	46.32	19.88	12.40
% *loss*	53.32	77.78	85.40
% *draw*	0.35	2.34	2.20

(a)

Outcome	(b_a, b_g)		
	(0.5,1.0)	(0.5,0.5)	(0.5,0.0)
% *win*	68.32	51.40	27.08
% *loss*	31.62	48.06	72.24
% *draw*	0.06	0.54	0.68

(b)

Outcome	(b_a, b_g)		
	(0.0, 1.0)	(0.0, 0.5)	(0.0, 0.0)
% *win*	72.44	60.16	28.74
% *loss*	27.28	39.50	70.36
% *draw*	0.28	0.34	0.90

(c)

Table 5.3: Simulation results for varying indexes of boldness for attacker using reactive modeling coupled with strategic decisions, and guards using anticipatory modeling: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$.

Figure 5.1 (c). The importance of this simulation is how the guards do not wildly overshoot the attacker in the first part of the game. Their initial movements are on an intercept course between the attacker and the flag. The attacker wins this game by turning when the guards draw near, but the guards consistently plot intercept courses throughout the game. This ability to anticipate and intercept allows the guards to significantly improve their performance.

5.2.6 Cooperating Guards

Now consider a scenario when the guards use an anticipatory model of the attacker and agree to a policy of coordination based on the relative position of the guards to the attacker. The attacker will compete against the guards using a strategy derived from a reactive model of the guards. Before constructing a cooperation model for the guards, recall that the accuracy and rejectability densities are derived from real-valued utilities defined on the space $U_{1:M} \times \Theta$. Thus far in the development, nature's effect on the accuracy and rejectability densities has only been implicitly addressed. In developing a controller for cooperating guards, the densities are explicitly allowed to change as a function of the state of nature.

Suppose that before the game starts, the guards agree to a coordination policy. In this policy, the guard that is closest to the attacker has the responsibility for pursuing the attacker. The other guard has the responsibility to stay between the attacker and the flag (play zone defense). Formally, let $\| \mathbf{p}_{g_i} - \mathbf{p}_a \|$ denote the Euclidean distance between guard i and the attacker. The two sets

$$\begin{aligned} \Theta_1 &= \{\vartheta \in \Theta : \| \mathbf{p}_{g_1} - \mathbf{p}_a \| \leq \| \mathbf{p}_{g_2} - \mathbf{p}_a \|\} \\ \Theta_2 &= \{\vartheta \in \Theta : \| \mathbf{p}_{g_1} - \mathbf{p}_a \| > \| \mathbf{p}_{g_2} - \mathbf{p}_a \|\} \end{aligned}$$

partition the state of nature. Assume that each guard knows exactly its position and its fellow guard's position, so that Θ_1 and Θ_2 are precisely known.

Define an accuracy density as a function of this partition. For guard 1, if $\vartheta \in \Theta_1$ then $f_{\mathbf{A}}^{g_1'}(\cdot;\vartheta) = f_{\mathbf{A}}^{g_1}(\cdot)$ where the g_1' denotes the modified guard's accuracy and where $f_{\mathbf{A}}^{g_1}(\cdot)$ is defined as in Section 5.2.5. For guard 2, if $\vartheta \in \Theta_2$ then $f_{\mathbf{A}}^{g_2'}(\cdot;\vartheta) = f_{\mathbf{A}}^{g_2}(\cdot)$, where again $f_{\mathbf{A}}^{g_2}(\cdot)$ is defined in Section 5.2.5. This means that the guard closest to the attacker should pursue the attacker.

The other guard (the distant guard) is responsible for staying between the attacker and the flag. The accuracy function can be formulated as minimizing the distance Φ between the guard and a point $\frac{1}{3}$ of the way between the attacker and the flag. Denote a line segment originating from the flag and terminating on the attacker's anticipated position; i.e., the line segment connecting $[0,0]^T$ and the attacker's anticipated position as described in Section 5.2.5. Denote the point of this line segment $\frac{1}{3}$ of the distance from the flag by $\hat{\mathbf{p}}_{\frac{1}{3}}(t)$. The distant guard directly pursues the point $\hat{\mathbf{p}}_{\frac{1}{3}}(t)$. Given that the guard uses a anticipatory model of the attacker, it also implicitly uses an anticipatory model of the $\frac{1}{3}$-point. For $\vartheta \in \Theta_2$, guard 1's accuracy, $f_{\mathbf{A}}^{g_1'}(\cdot;\vartheta)$, is based on $\hat{\boldsymbol{\xi}}_{\frac{1}{3}} = [0, \hat{\mathbf{p}}_{\frac{1}{3}}(t)]^T$. Similarly, for $\vartheta \in \Theta_1$, guard 2's accuracy, $f_{\mathbf{A}}^{g_2'}(\cdot;\vartheta)$.

Formally, for $\vartheta \in \Theta_1$,

$$\begin{aligned}
\Phi^{g_1}(\boldsymbol{\xi}(\mathbf{u}_{g_1}(t)), \boldsymbol{\xi}(\mathbf{u}_a(t))) &= \left[\boldsymbol{\xi}_{g_1}(\mathbf{u}_{g_1}(t)) - \hat{\boldsymbol{\xi}}_{\frac{1}{3}}(\mathbf{u}_{\frac{1}{3}}(t))\right]^T P_{g_1} \left[\boldsymbol{\xi}_{g_1}(\mathbf{u}_{g_1}(t)) - \hat{\boldsymbol{\xi}}_{\frac{1}{3}}(\mathbf{u}_{\frac{1}{3}}(t))\right], \\
\Phi^{g_2}(\boldsymbol{\xi}(\mathbf{u}_{g_2}(t)), \boldsymbol{\xi}(\mathbf{u}_a(t))) &= \left[\boldsymbol{\xi}_{g_2}(\mathbf{u}_{g_2}(t)) - \hat{\boldsymbol{\xi}}_a(\mathbf{u}_a(t))\right]^T P_{g_2} \left[\boldsymbol{\xi}_{g_2}(\mathbf{u}_{g_2}(t)) - \hat{\boldsymbol{\xi}}_a(\mathbf{u}_a(t))\right].
\end{aligned}$$

Similarly, for $\vartheta \in \Theta_2$,

$$\begin{aligned}
\Phi^{g_1}(\boldsymbol{\xi}(\mathbf{u}_{g_1}(t)), \boldsymbol{\xi}(\mathbf{u}_a(t))) &= \left[\boldsymbol{\xi}_{g_1}(\mathbf{u}_{g_1}(t)) - \hat{\boldsymbol{\xi}}_a(\mathbf{u}_a(t))\right]^T P_{g_1} \left[\boldsymbol{\xi}_{g_1}(\mathbf{u}_{g_1}(t)) - \hat{\boldsymbol{\xi}}_a(\mathbf{u}_a(t))\right] \\
\Phi^{g_2}(\boldsymbol{\xi}(\mathbf{u}_{g_2}(t)), \boldsymbol{\xi}(\mathbf{u}_a(t))) &= \left[\boldsymbol{\xi}_{g_2}(\mathbf{u}_{g_2}(t)) - \hat{\boldsymbol{\xi}}_{\frac{1}{3}}(\mathbf{u}_{\frac{1}{3}}(t))\right]^T P_{g_2} \left[\boldsymbol{\xi}_{g_2}(\mathbf{u}_{g_2}(t)) - \hat{\boldsymbol{\xi}}_{\frac{1}{3}}(\mathbf{u}_{\frac{1}{3}}(t))\right].
\end{aligned}$$

Rejectability $f_{\mathbf{R}}^{g_i'}(\cdot;\vartheta)$ stays the same for all $\vartheta \in \Theta$.

Since the state of nature is assumed known, $f_{\mathbf{C}}(\vartheta)$ is a delta function. When the state of nature is precisely known, a change in the state produces a change in the behavior of the guards. This switching between distinct behaviors is similar to the temporal sequencing behaviors suggested in [3].

The results of 500 simulations are tabulated in Table 5.4 and indicate that when the guards coordinate they improve their chances of success by between 12% and 35% with the most significant improvement coming when the guards ignore rejectability. A typical result is shown in Figure 5.1 (d) for a reacting attacker and for coordinating guards. The important thing to notice is how one guard aggressively pursues the attacker as in Figure 5.1 (c), whereas the second guard attempts to position itself between the attacker and the flag. Even when a guard overshoots, the remaining guard is in a good position to prevent the attacker from reaching the flag. This coordination between the two guards distinguishes Figure 5.1 (d) from Figure 5.1 (c) and is the reason that the guards win the game. The improved performance of the guards indicates that the extra mobility granted them may not be necessary when the guards cooperate.

Outcome	(b_a, b_g)		
	(1.0,1.0)	(1.0,0.5)	(1.0,0.0)
% win	29.8	3.2	2.2
% loss	68.4	95.6	97.2
% draw	1.8	1.2	0.6

(a)

Outcome	(b_a, b_g)		
	(0.5,1.0)	(0.5,0.5)	(0.5,0.0)
% win	46.6	15.8	8.8
% loss	51.8	82.4	89.0
% draw	1.6	1.8	2.2

(b)

Outcome	(b_a, b_g)		
	(0.0, 1.0)	(0.0, 0.5)	(0.0, 0.0)
% win	49.4	23.4	9.2
% loss	49.8	73.4	89.2
% draw	0.8	3.2	1.6

(c)

Table 5.4: Simulation results for varying indexes of boldness for strategizing attacker using reactive modeling , and guards coordinating their decisions based on an anticipatory model of the attacker: (a) $b_a = 1.0$, (b) $b_a = 0.5$, and (c) $b_a = 0.0$.

5.2.7 Discussion of CTF

The significance of CTF is in the trend that when modeling, strategy, or cooperation improves, the agents are more successful in reaching their objectives. This trend is plotted in Figure 5.2. When all agent use only reactive modelling, the attacker wins up to 98% of the time time and loses no more than 37%. When the attacker incorporates a strategy, it still wins up to 98% of the time, but loses no more than 13%. This demonstrates that using a strategy improves performance. However, when the guards anticipate the movements of the attacker, the attacker wins up to only 72% of the time and can lose up to 85% of the games. Thus, the ability to anticipate behaviors becomes a strong asset for the guards. Finally, when the two guards coordinate, the attacker wins no more than 49% of the games,

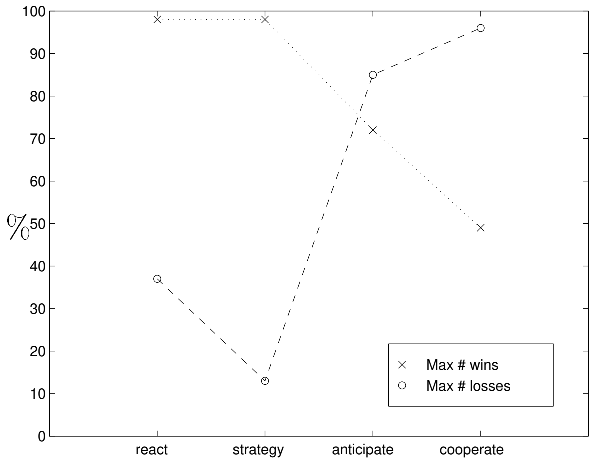


Figure 5.2: Maximum number of wins and losses versus simulation set.

but can lose up to 96%. This indicates that coordination can significantly improve performance. These simple examples demonstrate that the satisficing approach agrees with intuition about how the game should evolve by supporting the idea that performance should improve when strategy is invoked, good modelling is incorporated, and coordination is allowed.

5.3 Shared Resource Quadratic Regulator

The shared resource linear quadratic regulator is presented in this section. The solution is presented as a demonstration of the flexibility of the three team satisficing controller synthesis procedure and compared against the optimal control.

5.3.1 Problem Formulation

As a second example of multi-agent satisficing decision and control theory, consider the problem of regulating M agents who share a bounded resource. Formally, consider M individual LQR problems of the form

Minimize :

$$\begin{aligned} J_i &= \mathbf{x}_i^T(t_f)\mathbf{x}_i(t_f) + \sum_{t=1}^{t_f-1} \Big[\mathbf{x}_i^T(t+1)Q_i(t+1)\mathbf{x}_i(t+1) + \\ &\quad \mathbf{u}_i^T(t)R_{\mathbf{u}_i}(t)\mathbf{u}_i(t)\Big] \end{aligned}$$

$$
\begin{aligned}
\text{Subject to :}& \\
\mathbf{x}_i(t+1) &= A_i(t)\mathbf{x}_i(t) + B_i(t)\mathbf{u}_i(t) \\
\text{Constrained by :}& \\
Z_0 &\geq \sum_{i=1}^{M} \mathbf{u}_i^T(t)\mathbf{u}_i(t).
\end{aligned}
$$

This problem can be structured as a constrained coordination problem.

To construct the coordination problem, define the following vectors:

$$
\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_M(t) \end{bmatrix}, \quad \mathbf{u}(t) = \begin{bmatrix} \mathbf{u}_1(t) \\ \vdots \\ \mathbf{u}_M(t) \end{bmatrix},
$$

$$
\mathsf{A} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_M \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_M \end{bmatrix},
$$

$$
\mathsf{Q} = \begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ 0 & Q_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_M \end{bmatrix}, \quad \mathsf{R}_{\mathbf{u}} = \begin{bmatrix} R_{\mathbf{u}_1} & 0 & \cdots & 0 \\ 0 & R_{\mathbf{u}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{\mathbf{u}_M} \end{bmatrix}.
$$

The joint coordination problem[3] can be constructed as

$$
\begin{aligned}
\text{Minimize :}& \\
\mathcal{J} &= \mathbf{x}^T(t_f)\mathbf{x}(t_f) + \sum_{t=1}^{t_f-1} \Big[\mathbf{x}^T(t+1)\mathsf{Q}(t+1)\mathbf{x}(t+1) + \\
&\qquad \mathbf{u}^T(t)\mathsf{R}_{\mathbf{u}}(t)\mathbf{u}(t)\Big] \\
\text{Subject to :}& \\
\mathbf{x}(t+1) &= \mathsf{A}(t)\mathbf{x}(t) + \mathsf{B}(t)\mathbf{u}(t) \\
\text{Constrained by :}& \\
Z_0 &\geq \mathbf{u}^T(t)\mathbf{u}(t).
\end{aligned}
$$

[3]Because of the structure of the joint coordination problem, it could have been viewed as a vector-valued controller subject to a bound on the control. However, for more complicated constraints and system interrelationships, the optimal solution may not exist whereas the satisficing controller may still be applicable.

In terms of the three-team game introduced in Section 5.1, $\mathcal{K}_E = \{\text{origin}\}$, $\mathcal{K}_P = \emptyset$, and $\mathcal{K}_M = \{\text{M agents}\}$, which means that $K_E = 1$, $K_P = 0$, and $K_O = M$. Consider the one step control horizon solution ($d = 1$). The influence vector is $\boldsymbol{\xi}(\mathbf{u}(t)) = \mathbf{x}(t+1)$. The local performance indices are $\Phi(\boldsymbol{\xi}(\mathbf{u}(t))) = [\mathbf{x}(t+1) - 0]^T[\mathbf{x}(t+1) - 0]$, $\Psi(\boldsymbol{\xi}(\mathbf{u}(t))) = 0$, and $\Lambda(\boldsymbol{\xi}(\mathbf{u}(t)), \mathbf{u}(t)) = \mathbf{x}^T(t+1)\mathsf{Q}(t+1)\mathbf{x}(t+1) + \mathbf{u}^T(t)\mathsf{R}_{\mathbf{u}}(t)\mathbf{u}(t)$. Because all agents share the same perspective, superscript j has been dropped.

Theorem 8 establishes consistency of the nonlinear quadratic regulator for scalar controls. The result is generalized to constrained vector-valued controls in the following theorem.

Theorem 15 (Consistency of constrained vector quadratic regulator) *For the deterministic constrained quadratic regulator problem*

$$
\begin{aligned}
\text{Minimize :} & \\
\mathcal{J} &= \mathbf{x}^T(t_f)\mathbf{x}(t_f) + \sum_{t=1}^{t_f-1}\Big[\mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1) + \\
&\quad \mathbf{u}^T(t)R_{\mathbf{u}}(t)\mathbf{u}(t)\Big] \\
\text{Subject to :} & \\
\mathbf{x}(t+1) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \\
\text{Constrained by :} & \\
Z_0 &\geq \mathbf{u}^T(t)\mathbf{u}(t),
\end{aligned}
$$

if the control horizon spans the full extent of the problem, that is $d = t_f$, then there exists a rejectivity index b such that most discriminating satisficing control is identical to the optimal control.

Proof: The constraint set is $\mathcal{C} = \{\mathbf{z} \in \Re^{n\times t_f} : \mathbf{z}^T(t)\mathbf{z}(t) \leq Z_0\}$, where n is the dimension of the state vector $\mathbf{x}(t)$. The most discriminating control, which is always strongly satisficing, is denoted $\mathbf{u}_D = [\mathbf{u}(0), \ldots, \mathbf{u}(t_f-1)]$, is given by $\mathbf{u}_D = \arg\sup_{\mathbf{z}\in\mathcal{C}}\{\bar{f}_A(\mathbf{z}) - b\bar{f}_R(\mathbf{z})\}$. After some straightforward manipulations,

$$
\begin{aligned}
\mathbf{u}_D &= \arg\sup_{\mathbf{z}\in\mathcal{C}}\{g_A(\mathbf{z}) - b' g_R(\mathbf{z})\} \\
&= \arg\inf_{\mathbf{z}\in\mathcal{C}}\{\Phi[\boldsymbol{\xi}_0(\mathbf{z}(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(\mathbf{z}(t_f-1))] \\
&\quad + b'\Lambda[\boldsymbol{\xi}_0(\mathbf{z}(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(\mathbf{z}(t_f-1))] - b'\Psi[\boldsymbol{\xi}_0(\mathbf{z}(0)), \ldots, \boldsymbol{\xi}_{t_f-1}(\mathbf{z}(t_f-1))]\}
\end{aligned}
\tag{5.13}
$$

where b' is the previously encountered scaled version of b. For the quadratic regulator problem with $d = t_f$, take

$$\begin{aligned}
\Phi[\boldsymbol{\xi}_0(\mathbf{u}(0)),\ldots,\boldsymbol{\xi}_{t_f-}(\mathbf{u}(t_f-1))] &= \mathbf{x}^T(t_f)[P-Q(t_f)]\mathbf{x}(t_f),\\
\Lambda[\boldsymbol{\xi}_0(\mathbf{u}(0)),\ldots,\boldsymbol{\xi}_{t_f-}(\mathbf{u}(t_f-1))] &= \sum_{t=0}^{t_f-1}\mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1)+\mathbf{z}^T(t)R_{\mathbf{u}}\mathbf{z}(t),\\
\Psi[\boldsymbol{\xi}_0(\mathbf{u}(0)),\ldots,\boldsymbol{\xi}_{t_f-}(\mathbf{u}(t_f-1))] &= 0.
\end{aligned}$$

Then,

$$\mathbf{u}_D = \arg\inf_{\mathbf{z}\in\mathcal{C}}\left\{\mathbf{x}^T(t_f)P\mathbf{x}(t_f)+b'\sum_{t=0}^{t_f-1}\mathbf{x}^T(t+1)Q(t+1)\mathbf{x}(t+1)+\mathbf{z}^T(t)R_{\mathbf{u}}(t)\mathbf{z}(t)\right\}, \tag{5.14}$$

where $\mathbf{x}(t)$ is a function of $\mathbf{z}(t)$ via the dynamics equation (3.1). For b chosen so that $b' = 1$, this is exactly the optimal vector-valued constrained quadratic regulator solution. ■

By agreeing to a specific rejectivity index, all agents agree to a specific strongly satisficing solution. Thus, assuming that each agent acts in good faith with respect to the agreed rejectivity (ignoring the incentive compatibility problem [45]), the resulting decision is strongly satisficing for each agent. In the presence of a central policy maker, this condition is trivially satisfied. Additionally, in keeping with the objectives of [45], although the theory and synthesis procedure have been presented for well-defined continuous differentiable utilities, the definitions apply for poorly defined, discontinuous numerical utilities.

5.3.2 Six-Agent Constrained Coordination

In this section, a controller is formulated for the problem where six agents share a resource bounded by an energy function Z_0. Because agents reach a communal agreement, dependence on subjective belief $f_{\mathbf{BC}}$ can be omitted. Figure 5.3 shows the state space trajectories of all six agents for both the constrained optimal control (using saturating state feedback) and the constrained satisficing (b=1.2) control with $Z_0 = 400$. The constrained optimal control is found by first, obtaining the unconstrained optimal control $\tilde{\mathbf{u}}(t)$ and second, if necessary,

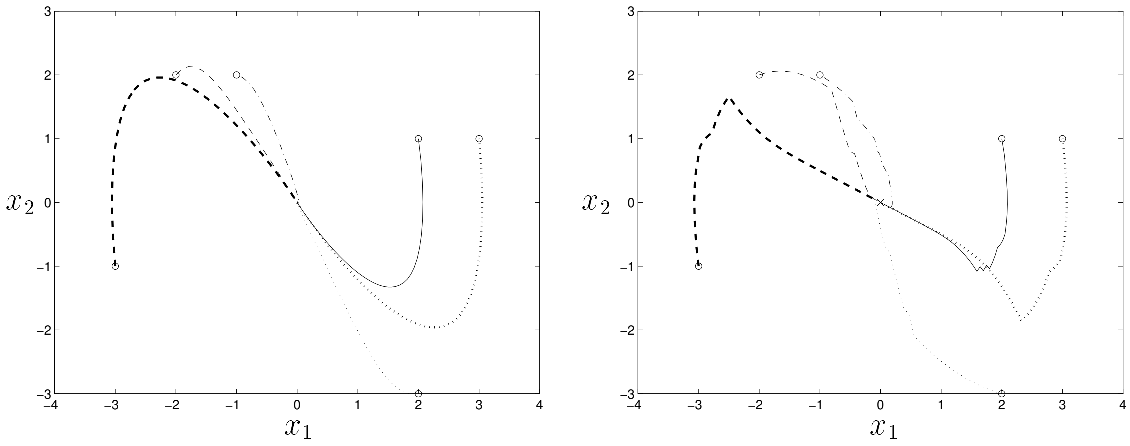


Figure 5.3: Phase plane performance results for the multi-agent shared-resource linear regulator problem: (a) constrained optimal, (b) satisficing (b=1.2).

scaling the optimal control by the energy constraint yielding

$$\mathbf{u}(t) = \begin{cases} \frac{\tilde{\mathbf{u}}(t)}{Z_0} & \tilde{\mathbf{u}}^T(t)\tilde{\mathbf{u}}(t) \geq Z_0 \\ \tilde{\mathbf{u}}(t) & \text{otherwise} \end{cases}.$$

The simulations were conducted for $R_u = [0.05]$, $P = I$,

$$A = \begin{bmatrix} 0.9974 & 0.0539 \\ -0.1078 & 1.1591 \end{bmatrix} \quad B = \begin{bmatrix} 0.0013 \\ 0.0539 \end{bmatrix} \quad Q = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.05 \end{bmatrix}.$$

The agent trajectories originate at the "$\circ$" and terminate at the "$\times$". Note the sharp corners on some of the satisficing trajectories. These sharp corners are produced because agents alternately surrender some of their needed control energy $u_i^2(t)$ so that another agent can have enough and then request sufficient energy to direct their paths toward the origin.

Figure 5.4 shows the control energy as a function of time, $E(t) = \sum_{i=1}^{6} u_i^2(t)$, for the optimal and the satisficing controllers, respectively. Observe that both controllers always keep their energy levels below the constraint, $E(t) \leq Z_0$, and that the satisficing controller requires 10.5% more cost than the constrained optimal. However, when the optimal control is not known, finding the constrained optimal is impossible using the technique presented. However, as demonstrated in the inverted pendulum problem, knowing the optimal control is not necessary

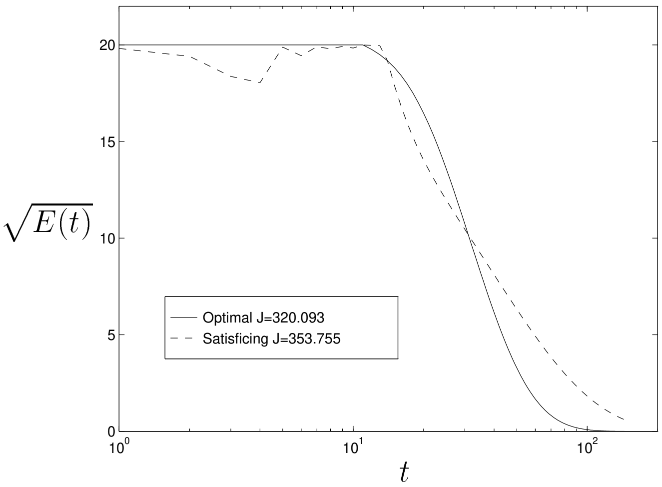


Figure 5.4: Energy consumption for optimal and satisficing controllers.

in developing a satisficing controller. Consequently, a constrained solution can be applied even in the absence of knowing the optimal control.

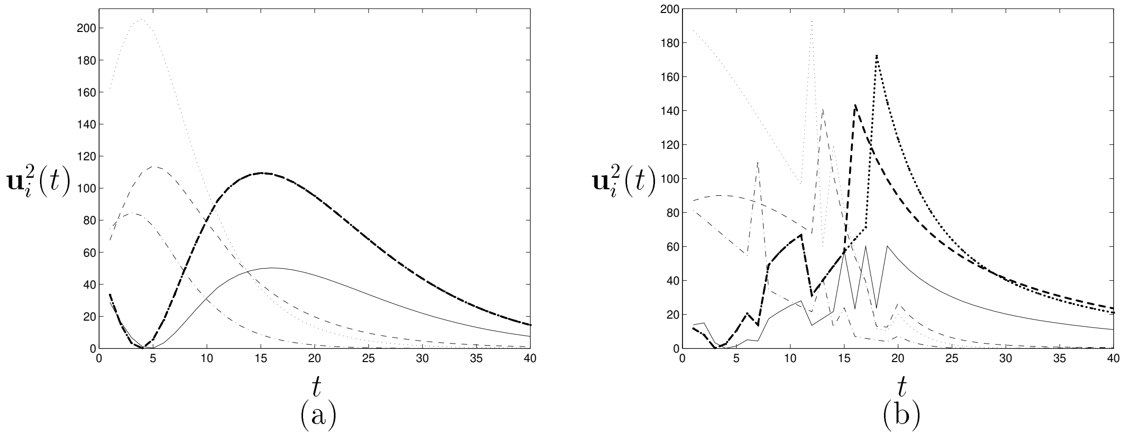


Figure 5.5: Control history results for the multiple-agent shared-resource linear regulator problem: (a) optimal controller: (b) satisficing controller.

Figure 5.5(a) and (b) show the time history of the energy consumption for each agent (the line types correspond to the line types in Figure 5.3) for both the optimal and satisficing controllers. For the satisficing controller, note that the peak energy demand of each agent occurs out of phase with the peaks of other agents. The alternating energy peaks illustrate how the agents can cooperate by switching control authority between agents.

5.3.3 128-Agent Constrained Coordination

For the six-agent linear problem, there exists a closed form solution for the satisficing controller. However, in general this may not be true. Consequently, each potential control must be tested to see if it is satisficing. This can imply a significant degree of complexity for a large number of agents. However, because satisficing controllers do not require perfect models of other agents (recall that subjective probabilities may be used), it is possible to reduce complexity without significantly degrading performance by using simplified agent models $f_{\mathbf{BC}}$.

One simplified satisficing controller was applied to 128 coordinating agents using the same linear models as in the six-agent problem. Under this simplified scheme, agents are placed as leaf nodes in binary tree structure. See Figure 5.6. The parent nodes of the agents do not control an actual plant, but instead represent a coalition of agents. Hence, they are referred to as *virtual agents*.

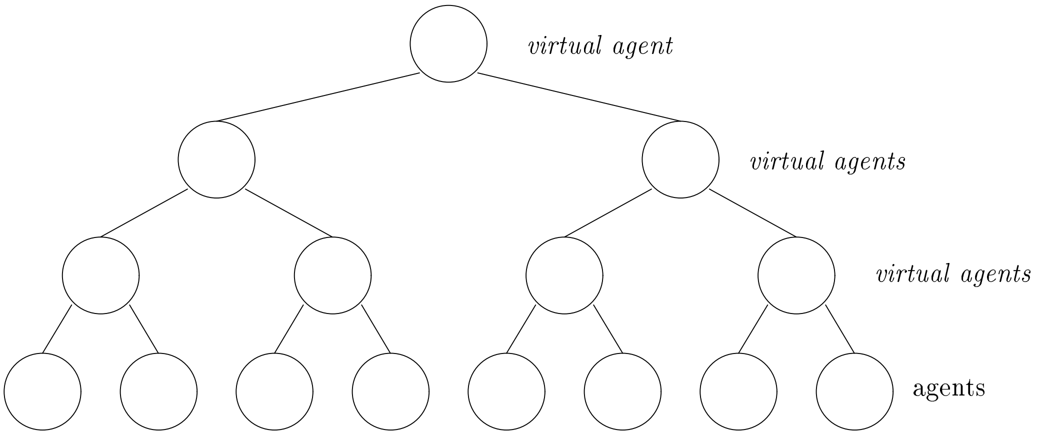


Figure 5.6: Binary tree of coordinating agents.

Under the binary tree structure, the role of each virtual agent is to dispense its allotment of a resource to its two children. Each virtual agent forms an accuracy and rejectability utility based on what it believes the leaf nodes (agents) under it will do with the allotment. Formally, let $\mathcal{K}_i$ represent the coalition of leaf nodes under virtual agent i. Then virtual agent i represents coalition $\mathcal{K}_i$ in seeking resources. Let $Z_i(t)$ denote the portion of the shared resource Z_0 available to virtual agent i at time t. Virtual agent i must allocate this resource to its two children.

To allocate $Z_i(t)$, virtual agent i forms a decision space U_i consisting of ordered pairs (u_1, u_2), where u_1 is the portion $Z_i(t)$ allocated to child 1 and u_2 is the resource allocated to child 2 ($u_2 = \sqrt{Z_i^2(t) - u_1^2}$). The accuracy is then a negated and normalized function of the positions that each leaf node agent can achieve with its allocation. Similarly, rejectability is a normalized function of the cost that each allocation yields. Formally, suppose that virtual agent i has a model for each leaf node agent in $\mathcal{K}_i$. Again, consider $d = 1$ whence $\boldsymbol{\xi}(\mathbf{u}(t)) = \mathbf{x}(t+1)$. Then

$$
\begin{aligned}
J_A^{\mathcal{K}_i}(u_1, u_2) &= \sum_{j \in \mathcal{K}_i} \mathbf{x}_j^T(t+1)\mathbf{x}_j(t+1) \\
&= \sum_{j \in \mathcal{K}_i} [A_j\mathbf{x}_j(t) + B_j\mathbf{u}_j(t)]^T [A_j\mathbf{x}_j(t) + B_j\mathbf{u}_j(t)], \\
J_R^{\mathcal{K}_i}(u_1, u_2) &= \sum_{j \in \mathcal{K}_i} \mathbf{x}_j^T(t+1)Q\mathbf{x}_j(t+1) + \mathbf{u}_j(t)R_j\mathbf{u}_j(t) \\
&= \sum_{j \in \mathcal{K}_i} [A_j\mathbf{x}_j(t) + B_j\mathbf{u}_j(t)]^T Q_j [A_j\mathbf{x}_j(t) + B_j\mathbf{u}_j(t)] + \mathbf{u}_j(t)R_j\mathbf{u}_j(t),
\end{aligned}
$$

where the set of $\mathbf{u}_j(t)$ is constrained by the allotment of resources passed down from virtual agent i.

In general, there are too many combinations to consider every possible allocation of resources to the terminating agents; computational complexity becomes to expensive. However, each virtual agent i can impose a belief function for the agents below it, $\mathcal{K}_i$. This is the source of the computational complexity reduction. Suppose that, in the spirit of constrained optimal control, the virtual agents believe that each agent in $\mathcal{K}_i$ will allocate its resource proportionally to its desired control. This requires that virtual agent i know the desired resource of each child. Let the desired control of terminating agent j that is a descendent of child k be denoted $\tilde{\mathbf{u}}^k_{\mathcal{K}_i(j)}$. When virtual agent i gives u_1 to its first child, and u_2 to its other child, then it assumes that

$$
\hat{\mathbf{u}}^k_{\mathcal{K}_i(j)} = \max\left\{\tilde{\mathbf{u}}^k_{\mathcal{K}_i(j)}, \frac{\tilde{\mathbf{u}}^k_{\mathcal{K}_i(j)}}{\sum_j \tilde{\mathbf{u}}^k_{\mathcal{K}_i(j)}} u_k\right\}.
$$

Then the virtual agent subjective belief density $f^i_{\mathbf{B}}(\mathbf{u}_{\mathcal{K}_i})$ becomes $\delta(\mathbf{u}_{\mathcal{K}_i} - \hat{\mathbf{u}}_{\mathcal{K}_i})$, where $\hat{\mathbf{u}}_{\mathcal{K}_i} = [\hat{\mathbf{u}}^1_{\mathcal{K}_i}, \hat{\mathbf{u}}^2_{\mathcal{K}_i}]^T$. This belief function eliminates the need to consider the effect of any other allocation. Then

$$\bar{J}_A^{\mathcal{K}_i}(u_1, u_2) = \sum_{j \in \mathcal{K}+i} [A_j \mathbf{x}_j(t) + B_j \hat{\mathbf{u}}_j(t)]^T [A_j \mathbf{x}_j(t) + B_j \hat{\mathbf{u}}_j(t)],$$

$$\bar{J}_R^{\mathcal{K}_i}(u_1, u_2) = \sum_{j \in \mathcal{K}_i} [A_j \mathbf{x}_j(t) + B_j \hat{\mathbf{u}}_j(t)]^T Q_j [A_j \mathbf{x}_j(t) + B_j \hat{\mathbf{u}}_j(t)] + \hat{\mathbf{u}}_j(t) R_j \hat{\mathbf{u}}_j(t),$$

where $\hat{\mathbf{u}}_j = (\hat{\mathbf{u}}_{\mathcal{K}_i})_j$ is the j^{th} control pair of the vector $\mathbf{u}_{\mathcal{K}_i}^*$, and the overbar indicates that the expectation is implicitly taken prior to forming the accuracy and rejectability densities. Since the belief function is a delta function, this will not change the resulting expected densities. The accuracy and rejectability densities are then formed from these cost functions as in (5.2)-(5.4) and (5.5)-(5.7). In practice, if agents request more control than they are allocated, then after the resource is distributed among the leaf nodes, any remaining resource is distributed to the agent requesting the greatest additional amount. This improves the allocation efficiency without increasing complexity.

The optimal and satisficing (for $b = 1.6$) phase plane trajectories are shown in Figure 5.7(a) and (b), respectively. The control history plots do not reveal anything not present in Figure 5.4(a), so they are not presented. The controller

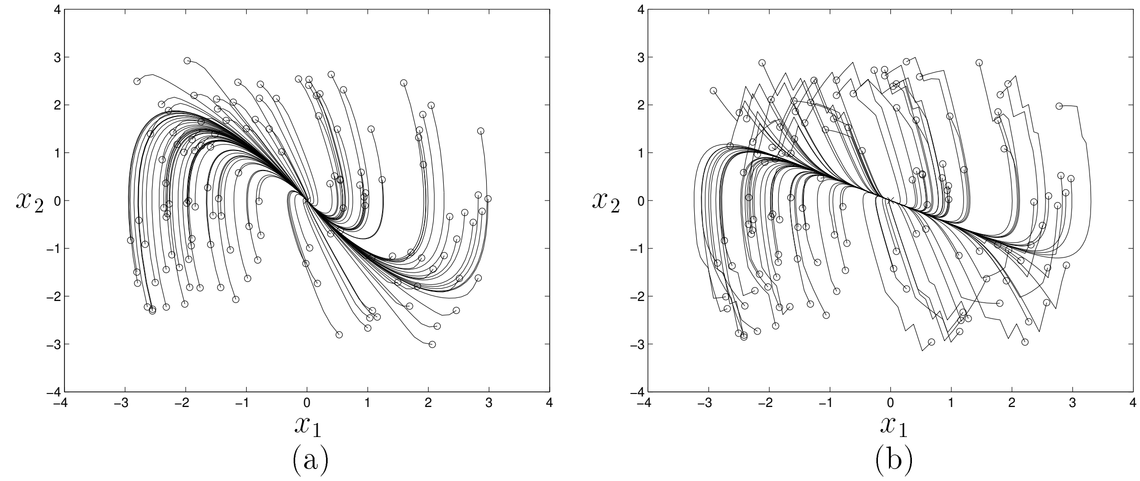


Figure 5.7: Phase plane performance results for the multi-agent shared-resource linear regulator problem: (a) constrained optimal, (b) satisficing for $b = 1.6$.

effectively coordinated the 128 agent's control. The satisficing controller stabilized each agent without violating the constraint of $E_{\text{sat}}(t) \leq 7225 = Z_0$ for all t. The constrained optimal controller yielded $\mathcal{J}_{\text{opt}} = 4490$ and the satisficing controller yielded $\mathcal{J}_{\text{sat}} = 5473$. These numbers translate into a cost $\mathcal{J}$ of -22%. This is worse

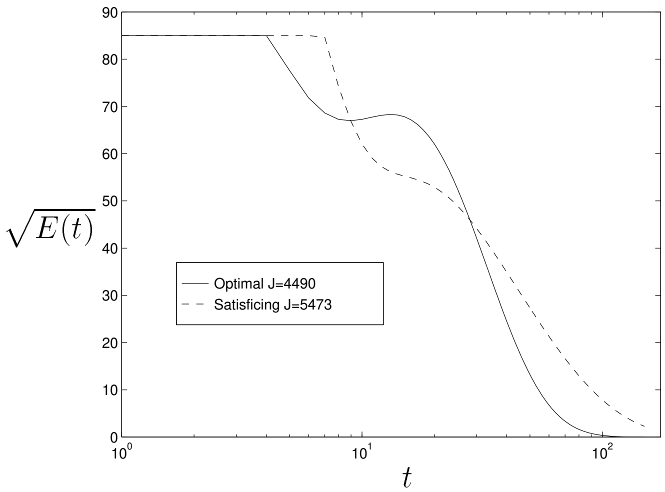


Figure 5.8: Control history for 128 agents sharing a bounded resource: constrained optimal versus constrained satisficing.

than the satisficing controller using a complete model, but acceptable because complexity was significantly reduced. In fact, given that the control space U_i of each agent was discretized into N elements, this simplification reduced complexity from order (N^M) calculations (M is the number of agents, 128 in this problem), to order $(N^{\log M})$ calculations.

5.4 Some Implementation Details

As in the single agent case, the control space was discretized and the most discriminating control from the discrete set was selected. The discretization was performed by restricting the accuracy and rejectability mass to the set of points satisfying the condition

$$U_{\mathcal{K}}(t) = \{\mathbf{u}_{\mathcal{K}} : \nabla \bar{f}_{\mathbf{A}}(\mathbf{u}_{\mathcal{K}}) \nabla \bar{f}_{\mathbf{R}}(\mathbf{u}_{\mathcal{K}}) \geq 0\}. \tag{5.15}$$

This set satisfies one of two conditions required for a control $\mathbf{u}$ to be in the equilibrium set. The control space is then discretized on a uniform grid with 11 sample points. The most discriminating control is selected and applied. Since the most discriminating control tends to move the system toward the origin in state-space, the discretization became more refined as time progresses.

Chapter 6

CONCLUSIONS

6.1 Results

The theory and application of satisficing control are developed in this dissertation. First, the notions of *accurate* and *rejectable* decisions are introduced, and the corresponding *accuracy* and *rejectability* utilities are defined. Using these utilities, the satisficing set is defined via the satisficing likelihood ratio test. These utilities are again used to define the strongly satisficing set as a refinement of the satisficing set. For differentiable concave accuracy and differentiable convex rejectability, the set of strongly satisficing controls are interpreted as maximal for a given rejectivity.

Next, a synthesis procedure is not only presented for single agent satisficing controllers, but also tested on representative control problems. This synthesis procedure, developed for nonlinear systems describable in state-space, is based on a receding horizon formulation and uses cost functions to differentiate between potential consequences of controls. Provided that the control horizon extends the entire duration of the problem, using this synthesis procedure guarantees that the elements of the strongly satisficing set are consistent with the optimal control for both regulator and minimum time problems. Simulation results indicate that, for a single step control horizon, the satisficing control performs within approximately 4% of the optimal control for the linear quadratic regulator problem. For a two step control horizon, the optimal control performs within approximately 15% of the optimal control for Zermelo's problem.

Finally, the single agent theory is extended to multiple agent problems and a synthesis procedure is developed. This synthesis procedure uses a receding control horizon for nonlinear systems describable in state-space, and includes both pursuit-evasion and energy conservation characteristics. For problems requiring only regulator characteristics, the strongly satisficing set is again shown to be consistent with the optimal solution. Simulation results indicate that, for a constrained regulator with a single step control horizon, the satisficing control performed within approximately 10% of the constrained optimal. When a simplified algorithm is used, complexity is significantly reduced and performance remains within approximately 22% of the optimal. For problems where optimality could not be defined, the satisficing solution is seen to be compatible with intuition which says that as knowledge, coordination, and strategy improve, the likelihood of success should improve.

6.2 Discussion

Decision-making is an essential component of an intelligent agent. When forming a decision, two essential criteria emerge: (a) the decision should be credible, and (b) the decision should be beneficial. Judgments of credibility, together with evaluations of propositions with respect to value, importance, or consequences, form the basis of all rational decision-making for both cognitive and practical applications.

Optimal decision theory merges the credibility and beneficiality aspects of a proposition into a single criterion and selects a decision which is best according to this criterion. This methodology, which exploits all the information necessary to provide a global solution, is the preferred method of design in all situations where it can be appropriately applied. When, however, a global solution is too costly or too difficult to implement or when the information necessary to generate an optimal solution is unavailable, applying ad hoc design principles has been the default controller synthesis paradigm. For problems where optimality cannot be applied, satisficing control theory is a justifiable alternative to ad hoc design principles which can provide very good performance. Whereas an optimal control is determined by generating a single performance index which merges credibility and beneficiality, a satisficing control is determined by comparing the credibility (accuracy) of a control against its beneficiality (rejectability). Strongly satisficing then refines the satisficing set by eliminating all controls for which a solution exists which is both more credible and more beneficial.

There is a growing perception that progress for many problems, particularly multi-agent problems, may be more rapid if the requirement for optimal behavior is relaxed in favor of *intelligent* behavior. Intelligent behavior is somewhat loosely defined, but is often understood as behavior which imitates in some meaningful way that of a competent biological organism in similar circumstances. Familiar intelligent control methodologies include fuzzy logic, artificial neural networks, genetic algorithms, and artificial intelligence [54]. These methods focus on various aspects of intelligent behavior; for example, fuzzy logic emphasizes making good decisions in the presence of imprecision, while artificial neural networks exploit a metaphor from biology to learn general patterns from examples. Because satisficing control theory is firmly grounded in sound epistemological principles, it can be considered a form of intelligent control [17]. Additionally, the satisficing principle appears compatible with fuzzy logic [18].

Multiple agent satisficing control theory allows the construction of multiple agent controllers that do not require restrictive rationality or modeling assumptions, but instead choose controls based on the available amount of information, embodied in accuracy, rejectability, and subjective belief functions. Additionally, the controllers allow agents to coordinate not only with agents within their own coalition, but also with agents outside their coalition. Agents can invoke strategies based on expected behavior and they can modify their strategies based on changing situations. The ability to choose controls based on expected behavior means that performance degrades gracefully in the presence of partial information. As models of the environment or other agents improve, performance improves. Performance indices can be used to qualify control choices, and only controls which meet a minimum standard of performance are admissible.

From an academic point of view, it is not sufficient to argue that epistemic utility-based control theory is worthy of attention simply because it appears to perform well for selected examples. Three criticisms may be advanced. First, it does not offer the performance guarantees, such as stability and controllability, provided by traditional control theory and, therefore, simply gives up too much. Second, even if one were to accept the possible usefulness of a satisficing approach to control, there is no unique set of procedures on which to base a satisficing approach and, hence, it is ad hoc. Third, a (strongly) satisficing resolution to a problem might yield inadequate performance. Why should the epistemic utility approach be privileged with respect to other schemes?

To the first criticism, the reply is that theorems regarding stability and controllability seem hard to come by in any context of limited rationality, and that this is not a limitation of epistemic utility *per se*. This limitation is certainly shared with much of contemporary nonlinear control methodology. Although this important question deserves further research efforts, it may be that the many problems simply do not admit such performance guarantees. Meanwhile, to reject good solutions in favor of unknown, and perhaps nonexistent, best solutions seems unreasonable and intransigent. Moreover, the satisficing solutions developed herein are consistent with optimal control theory.

To the second criticism, the reply is that the satisficing approach developed herein may not be unique, but it is not ad hoc. Satisficing control solutions must meet a minimum standard of performance. Epistemic utility provides such a threshold using the satisficing likelihood ratio test, defined in terms of the accuracy and rejectability utility functions. Thus, epistemic utility-based satisficing control is, like traditional optimality, firmly grounded in principles. Ad hoc solutions do not generally enjoy this same grounding, but are typically based on more or less vague notions of desirability.

To the third criticism, the reply is that while it is true that satisficing control might yield inadequate performance, this problem is not unique to satisficing controllers. Much as an optimal controller may give the best performance according to a given criterion but not be satisfactory for every scenario, a satisficing controller may give justifiable results that are not satisfactory for every scenario. However, justifying a control on the basis of a minimum performance standard is a reasonable design principle.

One argument can be advanced specifically against multi-agent satisficing control theory: agents in adversarial situations should base decisions on what other agents are capable of doing instead of what they are likely to do. This is a fundamental tenet of game theory [37]. The response to this argument is two-fold. First, making decisions based on expected behavior is not unreasonable for a large class of problems (such as problems where adequate modeling of other agents is possible). Second, making a decision based on what another agent is capable of doing implicitly invokes a subjective model of that agent. This implicit model indicates that, conditioned on my decisions, another agent will do that which it perceives as most beneficial. Satisficing decision theory makes this modeling process explicit and, consequently, allows an agent to avoid serious errors that may result from applying restrictive rationality assumptions to problems with uncertain models.

The epistemic utility formulation has a number of attractive attributes.

- It uses the same goals and principles to design the accuracy and rejectability density functions that would be used to design an optimal controller.

- It has a formal, straightforward procedure for controller synthesis. Furthermore, strongly satisficing control decisions possess an important equilibrium property: no controls have both lower rejectability and higher accuracy. For convex $\bar{f}_{\mathbf{R}}$ and concave $\bar{f}_{\mathbf{A}}$, the family of strongly satisficing decisions is convex which is a useful in controller synthesis.

- It is robust and degrades gracefully as the amount of information available decreases. Even in the extreme case of temporally instantaneous information ($d = 1$) the satisficing controller can still generate a useful control policy. As more information is made available, it can be incorporated with corresponding performance improvements. In the limit as the horizon increases to the full extent of the problem, the satisficing solution converges, at least for constrained vector-valued quadratic problems and single-agent minimum-time problems, to the optimal solution.

- It does not require the imposition of arbitrary and perhaps objectionable assumptions, such as linearity or time-invariance, to find a solution.

- It allows application of the satisficing principle to both single agent and multiple agent problems using the same basic formalisms. The resulting multiple agent controllers are robust and degrade gracefully in the presence of partial information about other agents.

Until relatively recently, in the absence of finding an optimal decision, ad hoc principles have been the default decision paradigm. Researchers of intelligent controllers have been attempting to establish decision paradigms which are applicable when optimality fails or is inappropriate. Satisficing control theory presents a form a intelligent controller design which can incorporate the performance indices associated with optimal control, but which relaxes some of the restrictive requirements necessitated by optimality. To the best of the author's knowledge, aside from the work developed in this dissertation and the work of Stirling, Frost, Morrell, Moon, and Kenney, epistemic utility based decision theory has not been applied to engineering. The results presented in this dissertation indicate that epistemic utility based satisficing control theory is a feasible controller synthesis methodology which can provide adequate performance in the presence of partial information.

Appendix A

SUGGESTIONS FOR FURTHER WORK

The following ideas originated during the research for and the writing of this dissertation. They are included as suggestions for additional work.

- Use the concepts developed in this dissertation to further develop satisficing estimation theory. Of particular interest to estimation theory might be the concept of a strongly satisficing estimate.

- Characterize the stability properties of the satisficing controllers developed thus far. A Lyapunov approach to characterize the local stability properties of the single agent nonlinear quadratic regulator might facilitate this development.

- Characterize the robustness properties of the satisficing controllers developed thus far. Consider using sensitivity and complementary sensitivity as a basis for forming accuracy and rejectability utilities.

- Instead of requiring accuracy and rejectability to be probability measures, the probability theoretic additivity requirement can probably be replaced by the less restrictive sub-additive or super-additive requirement. This would require the ideas of mutually singular, absolutely continuous, and Radon-Nikodym derivative to be defined in terms of the resulting fuzzy measure [68]. Such a development would provide a theoretical tie not only between fuzzy logic controllers and the satisficing principle, but also between neural networks and the satisficing principle.

- Instead of relying solely upon quadratic cost functions to formulate controllers, a hybrid controller containing both fuzzy logic and traditional optimality components would probably be useful. This hybrid approach would allow the best of both schools of thought to be applied to a single problem.

- Use satisficing decision theory to improve the defuzzification capabilities of fuzzy logic controllers. This would require a natural language description of accuracy and rejectability. The interpretation of fuzziness in [66] could be helpful in this regard.

- Incorporate learning in creating accuracy, rejectability, and subjective probability densities. Possible learning mechanisms include decision-directed neural network training, learning automata, and probabilistic reasoning [55].

- Investigate the relationship between game theory and satisficing multiple agent decision theory.

- Develop a satisficing H_∞ controller based on a two-agent coordination problem.

- Develop algorithms which exploit the ability of satisficing control theory to handle uncertainty to reduce computational complexity.

- Apply multiple agent satisficing control theory to nonholonomic control problems.

- Apply satisficing control theory to discrete event dynamic systems, and specifically to ordinal optimization [22–24].

- Apply information theory to selecting a representative from a convex set of accuracy or rejectability utilities.

- Apply information theory to characterizing the amount of knowledge possessed by agents in multiple agent decision problems.

- Apply random process descriptions to the synthesize accuracy and rejectability utilities as a function of time.

- Use convex sets of model parameters to improve performance robustness.

- Use convex sets of cost functions, and learn the costs that yield best performance.

- Develop a satisficing algorithm that can exploit parallel computing power. Since the determination of a satisficing control (but not strongly satisficing) is based on comparing two independent measures of the control instead of comparing one control against another, an entire set of controls could be tested in parallel to determine if they are satisficing.

Appendix B

TAXONOMY OF MULTIPLE AGENT PROBLEMS

B.1 Introduction

Multiple agent decision problems are often difficult to solve; partly because agents may disagree about what constitutes rational behavior and partly because, even if they agree in principle, the rational solution may be unknown, computationally impractical, or unenforcible. The development of a useful multiple agent decision theory would benefit greatly from a taxonomy of decision-making characteristics. A taxonomy not only allows for clear delineation of the scope of a theory, but also allows comparison of different theories. Moreover, a taxonomy provides a basis for showing theories to be consistent with known solutions to existing problems. In this context, consistency means that a theory incorporates solved problems with known taxonomic characteristics.

A taxonomy that includes elements of group *preference knowledge, coordination,* and *strategy* can be visualized as a three-dimensional space with axes defined by the degree of group knowledge, the degree of group coordination, and the extent of strategy invoked. See Figure B.1. Preference knowledge refers to in-

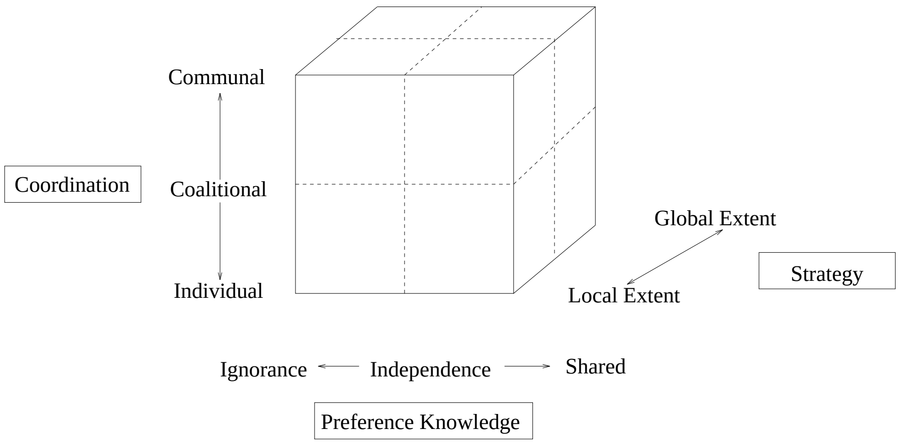


Figure B.1: Taxonomy of coordination types.

formation available about other agents and ranges from *ignorance* to *independence* to *shared* information. Coordination refers to an agents disposition to unitedly make decisions with other agents and includes notions of an *individual*, a *coalition*, and a *community*. Strategy refers to a time sequence of decisions and ranges from *local extent* to *global extent*. Communication and modelling are other important issues, but they deal more with the quality of knowledge and coordination rather than the type. Consequently, no effort is made to include communication as part of the taxonomy.

As an illustrative example of the taxonomy, the game theoretic decision method is classified. As part of this example, the prisoner's dilemma is classified including the characteristics that make this game a dilemma. Finally, a motivation for developing a multiple-agent decision taxonomy in the engineering context is presented. Of particular engineering interest, solutions to a simplified pursuit-evasion game. The game is selected because it typifies many competitive problems are discussed. Also of engineering interest, the game of *robot rugby* is introduced coupled with the necessary and desirable controller characteristics for this game. The solution methods adopted for these games represent different taxonomic classifications.

B.2 Multiple Agent Taxonomic Characteristics

B.2.1 Preference Knowledge

Agent i's decision-making process is a function of how much knowledge it has about the preferences of other agents. At one extreme, agent i may work in complete *ignorance* about other agent preferences. Although this could be problematic since consequences may be determined by community decisions, agent i may have no alternative. However, when agent i knows all other agents preference patterns, but does not have any indication of how its preferences affect other agent preferences, then agent i works *independently* of the other agents. When agent i knows not only the preference patterns of other agents, but also knows how its preferences affect other agent preferences, then agent i can use this *shared* information to improve the likely consequence of its decision. In practice, it is possible for agent i to have more information about the preferences of one group of agents than the preferences of another. This is particularly true in team games.

B.2.2 Coordination

Since agents may make decisions in environments where the consequences they receive depend on group decisions, they may choose to coordinate their decisions with other agents in an attempt to produce consequences that are desirable to them. However, when the number of agents M is large, it is unlikely that every agent in the community will be disposed to cooperate. For example, in two-agent zero-sum games ($M = 2$), although an agent coordinates its decisions with the other agent (consider the minimax solution) the two agents do not cooperate in reaching a group decision $\mathbf{u}$.

This leads to the important distinction between coordination and cooperation. Coordination is based on making decisions using the consequences of group decisions. Examples of coordination types include minimax solutions [37] and expected behavior (like in this dissertation). By contrast, cooperation is based on agents attempting to reach a consensus. Hence, on the taxonomy cube there are different types of coordination depending on the structure of coordinating agents. When agent i participates as an *individual*, it means that it does not attempt to reach a consensus with any other agent in the community. Agents trying to reach a consensus with agents from $\mathcal{K}$ form a *coalition*. When the entire community attempts a consensus, the decision is *communal*. In practice, it is possible for agents to participate as members of more than one coalition in an effort to improve the likelihood of producing a community decision that is beneficial to them.

B.2.3 Strategy

Frequently, decision makers wish to coordinate a sequence of decisions in order to reach a desirable consequence. A strategy is defined as a time sequence of decisions. Frequently, invoking a strategy can improve an agent's performance.

When agent i cannot accurately predict the effect of every decision, it is likely to develop a *partial extent* strategy that considers consequences of one or more decisions in sequence, but does not extend the entire duration of the problem. Such strategies are required by many popular games (such as chess) when the possible decisions grow too quickly to track. At one extreme of local extent strategy, agent i may not be able to or willing to make even a single decision. One scenario in which this could occur is when every possible consequence is catastrophic (such as choosing which move to make under *check-mate*). At the other extreme, when strategies can be invoked for the *global extent* of a problem, then it is desirable to do so.

B.3 Game Theory

Game theory has been the predominant method for making multi-agent decisions. Because of this, game theory is a useful tool for validating the taxonomy presented herein. Loosely, game theory assumes[1] that [37]

1. Players are "rational".

The game theoretic rationality assumption dictates that every agent attempts to maximize a utility over their preference pattern. Utility maximizing is the driving engine for determining acceptable decisions. In addition to the rationality assumption, game theory also assumes that

2. Agent i knows the preference pattern of every other agent. This is referred to as *hperrationality* [1].

3. Agent i makes decisions knowing only the logical consequences of its decisions and without any knowledge about the choices of other agents. This is referred to *indeterminacy* [1].

4. Agent i can form a preference pattern over every possible strategy defined over the entire extent of the problem (the game in normal form).

Finally, depending on how the payoff matrices are defined, agents can cooperate or compete with other agents or coalitions of agents.

The delineation of the game theoretic assumptions dictates the taxonomic characteristics of game theory. Assumptions 2 and 3 indicate that agent i operates with *independent* preference knowledge of other agents. In terms of the geometry, hyperrationality squeezes preference knowledge from the left, and indeterminacy squeezes preference knowledge from the right. Assumption 4 indicates

[1] Not everyone accepts the game theoretic definition of rationality [1, 34].

that agent i invokes a strategy for the *full extent* of the problem. Finally, depending on the structure of the payoff matrix, agent i can coordinate as an *individual* or as part of a *coalition* or as part of the *community*. These characteristics imply that game theory operates at the back of the cube in the center of the horizontal dimension with freedom to occupy any location along the vertical dimension. When critics decry game theory as being overly restrictive, their arguments frequently translate into criticisms of the limited taxonomic characteristics of the decision paradigm. Specifically, hyperrationality and indeterminacy are criticized [1]. An additional criticism of game theory is the restrictive requirement that a global strategy be formed. In many games of practical interest (such as chess) this requirement is prohibitive or impossible.

The difficulties of game theory often arise when it is applied to problems which require taxonomically different characteristics than those of game theory. For example, in the prisoner's dilemma, agents are required to make a decision based on independent preference knowledge and individual decisions when shared knowledge and communal decisions are required. Understanding the taxonomic structure does not aid the prisoners, but it is useful in explaining the difficulty.

B.4 Capture the Flag

Capture the Flag (CTF) is a game wherein an *attack* agent attempts to capture a flag which is protected by two *guard* agents. The attacker wins the game if it reaches the flag without colliding with the guards. The attacker loses the game if it collides with either guard. The game ends in a draw if time expires before the flag is captured and without the guards and attacker colliding. Results of this game are given in Figure 5.1.

In the taxonomic setting, the guards form a coalition which then coordinates its decisions with an adversarial attacker. The attacker coordinates its decisions with the adversarial guards. Strategy is determined by the length of time considered in making a decision. The agents can consider only one time step, or they can invoke more sophisticated strategies based on what they expect to happen. The knowledge available to the agents is a function of the model they adopt for themselves and each other.

B.4.1 Reactive Modeling

Consider a single time-step strategy. When agents do not try to anticipate future actions of other agents, but instead base their actions solely on the current positions of other agents, their decisions may be termed *reactive*. This form of modeling is the memoryless case of the communication suggested in [2]. Since the guards do not know the attacker's preferences nor the attacker the guards' preferences, they act in ignorance of each other. The guards may know each other's preferences, but do not allow this knowledge to affect their own preferences. Hence, they use independent knowledge in their coalition.

A typical result is shown in Figure 5.1 (a). For reactive agents, the attacker begins by traveling directly toward the flag. Because it believes the guards will remain in their current positions, it is not aware of the danger presented by the converging guards. The guards begin the game by heading toward the attacker, but they overshoot the attacker and the attacker wins the game.

B.4.2 Strategic Decisions

The attacker can attempt to improve its performance by using a two-step strategy. The agents still model other agents reactively (preference knowledge and coordination type do not change), but the attacker develops a strategy based on its reactive model of the guards. In forming a strategy, the attacker assumes that its "future self" will choose the control action that will move it toward the flag most quickly. In a sense, developing a strategy in this way can be viewed as forming a coalition with a "future self", sharing preferences, and forming a decision. The assumption that the "future self" will choose a particular action is invoked since the game is not played in a stationary environment; the guards move after each time step, changing the attacker's preferences.

A typical result for the strategizing attacker is shown in Figure 5.1 (b). The simulation begins with the two guards moving rapidly toward the attacker. The attacker avoids the closer guard and heads toward the flag. The second guard overshoots the attacker allowing the attacker to rapidly move toward the flag. When the attacker realizes that it has overrun the flag and that the guards are not an immediate threat, it begins turning toward the origin. This causes the second guard to again overshoot the attacker. When the second guard overshoots, the attacker turns quickly toward the flag and wins the game.

B.4.3 Anticipatory Modeling

Now consider a scenario in which the attacker adopts a strategy based on a reactive model of the guards, but in which the guards attempt to anticipate the attacker's movements. This has the effect of refining the guards preferences, but not the type. The preference knowledge about the attacker is still ignorance since they are assuming what the attacker will do according to their own preferences.

A typical result is shown in Figure 5.1 (c) for a strategizing attacker and for anticipating guards. The important thing to notice in this simulation is how the guards do not wildly overshoot the attacker in the first part of the game. Their initial movements are on an intercept course between the attacker and the flag. The attacker wins this game by turning when the guards draw near, but the guards consistently plot intercept courses throughout the game. This ability to anticipate and intercept allows the guards to significantly improve their performance.

B.4.4 Cooperating Guards

Now consider a scenario when the guards use an anticipatory model of the attacker, but agree to a policy of coordination based on the relative position of the guards to the attacker. In this policy, the guard that is closest to the attacker has the responsibility for energetically pursuing the attacker. The other guard has the responsibility to stay between the attacker and the flag (play zone defense). The guards knowledge about each other now changes from independent, to shared. This change comes because the guards allow a pre-disclosed agreement to dictate how their preferences affect one another. However, they still only have ignorant knowledge of the attacker. The attacker competes against the guards using a strategy based on a reactive model of the guards.

A typical result is shown in Figure 5.1 (d) for a reacting attacker and for coordinating guards. The important thing to notice is how one guard aggressively pursues the attacker as in Figure 5.1 (a), whereas the second guard attempts to position itself between the attacker and the flag. Even when a guard overshoots, the remaining guard is in a good position to prevent the attacker from reaching the flag. The coordination between the two guards distinguishes Figure 5.1 (d) from Figure 5.1 (c) and is the reason that the guards win the game.

B.5 Robot Rugby

Robot rugby is a modification to the CTF game previously introduced. In this modification, there are five coordinating agents: two teams each consisting of two agents, and one rogue agent. Each team has an assigned territory and an

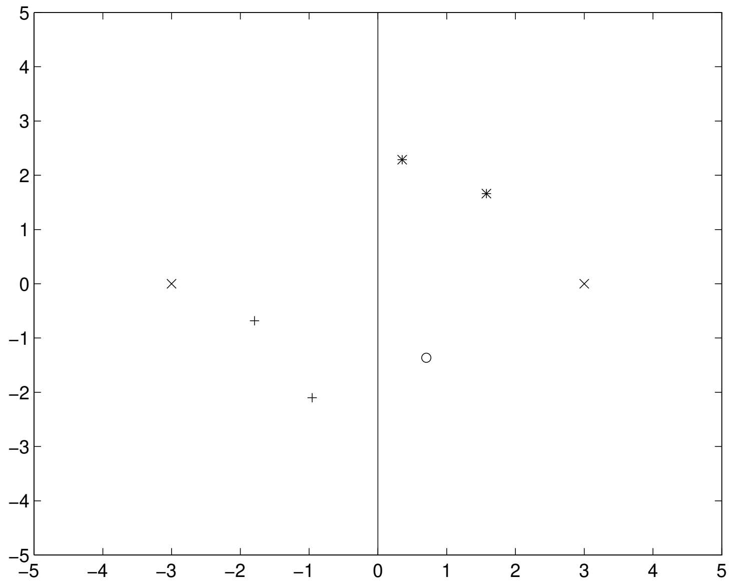


Figure B.2: Robot rugby playing field.

assigned goal (or flag) located in their territory. The playing field is illustrated in Figure B.2. A solid line divides the playing field. Agents on one team are designated with a "+" whereas agents on the other team are designated with a "$*$". The flags are designated with a "$\times$" and the rogue agent is designated with a "$\circ$". The objective is to reach the other teams flag before the other team reaches your flag. If two agents on different teams collide, the agent outside of its territory is removed from further competition. The rogue agent can win by capturing either flag, but if it is collides with any other agent at any time, then it must join that agent's team. The rogue agent can win either as part of a team, or as an individual. Assume that a team does not know if the rogue agent has been captured by the other team.

We will begin our discussion of the desirable taxonomic characteristics of this game with *preference knowledge*. Because it will be helpful for agents on a team to cooperate, they will attempt to share knowledge with each other. However, they will also attempt to prevent the other team from knowing their preferences and strategies. This means that, unless opposing teams intercept plans, they operate in ignorance. However, agents on the same team can share preferences or they may operate with independent preferences. The rogue agent will attempt to share, and glean as much information as it can to reach one of the flags. Thus, the rogue agent will operate somewhere between ignorance and shared preference knowledge. Agents on other teams would be foolish to share all of their information with the rogue agent, but they would be wise to share as much information as would be beneficial. Perhaps they might even give false information in hopes of deceiving the other team.

The amount of strategy invoked by a team of agents can vary. Agents may want to include a sequence of moves in their strategy. Possible strategies include trying to capture the rogue agent to enhance team strength, bluffing with one agent on the team while moving the other team member toward the flag, or even passing false information to the other team via the rogue agent.

Coordination is easily defined for the two teams. An agent is in a coalition with a teammate, but teams act individually with respect to the other team. However, teams may include the rogue agent in a coalition, and the rogue agent can act as a member of a coalition with either team, both teams, or neither team.

B.6 Discussion

The significance of CTF is in the trend that when modeling, strategy, or coordination improves, the agents are more successful in reaching their objectives. This means that as knowledge becomes more effective, strategy is invoked, and shared knowledge improves, agents can perform better. This demonstrates that the taxonomy agrees with intuition about how the game should evolve.

The significance of robot rugby is in the descriptive power of desirable properties of the game. Given a language to describe multiple agent decision problems, agents can make informed decisions. Since robot rugby can be viewed as a prototype for many multiple agent problems, the ability to describe various robot rugby scenarios can be a powerful tool for developing appropriate decision-making methods.

Appendix C

ACCURACY AND REJECTABILITY SETS

Levi's epistemic utility theory did not restrict attention to single utilities only. Instead, convex sets of credal probabilities were considered [34]. Allowing for sets of belief functions instead of single subjective probabilities is a departure to traditional approaches to decision theory [25]. Morrell and Stirling extended this idea in the development of the set-valued Kalman filter [46, 47, 49–51] and in set-valued estimation [26, 42, 43, 48].

In the control engineering context, it is also desirable to address the problem of sets of utilities. After defining the problem, three potential resolutions are presented. A discussion is then included.

C.1 Convex Utilities

In the formation of the satisficing set, four implicit assumptions are made. First and second, that the expected accuracy and expected rejectability utilities are uniquely specified. Third and fourth, that the subjective belief over the states of nature and over the actions of other agents are uniquely specified. If any of these assumptions are violated, then satisficing control theory must be extended.

There are at least six scenarios when expected accuracy and expected rejectability can only be specified to within a convex set.

- More than one objective of a controller is specified. Then it may be appropriate to allow for more than one utility to describe the objectives.
- When system models are not completely specified, then the accuracy and rejectability utilities may depend on which model of the system is used. Given a set of system parameters, the resulting accuracy and rejectability are only known to within a set.

- The subjective belief function of the state of nature may not be precisely known. Perhaps only the variance is known, but the mean of the distribution is specified only to within a set [46]. Then the resulting expected accuracy and rejectability will only be known to within a set.
- The behavior of other agents may not be describable by a unique probability density. Perhaps the model of the agent can only describe the mean of the distribution, but not the variance. The resulting expected accuracy and expected rejectability will only be known to within a set.
- Cost functions may be specified only as elements of particular sets. This would allow learning of cost functions that yield desired performance.
- Dynamic coalition formation may allow for agents to belong to more than one coalition and, consequently, have more than one accuracy and rejectability utility.

The mathematical description of each of these scenarios is found in treating the totality of expected accuracy and expected rejectability densities as a set. As suggested by Levi [34], if any set of accuracy utilities is justified, then the convex closure of this set is also justifiable thereby restricting attention to closed convex sets of utilities. Let the set of accuracies and rejectabilities be denoted as

$$\mathcal{A} = \{\bar{f}_{\mathbf{A}}\}, \tag{C.1}$$

$$\mathcal{R} = \{\bar{f}_{\mathbf{R}}\}. \tag{C.2}$$

These sets may be determined by a convex set of $f_{\mathbf{C}}$, a convex set of $f_{\mathbf{B}|\mathbf{C}}$, a convex set of $f_{\mathbf{A}}$, a convex set of $f_{\mathbf{R}}$, or any combination of these. The next section discusses which controls will be satisficing according to various notions of admissibility.

C.2 Definitions of Admissibility

A rather weak notion of admissibility is one where $\mathbf{u}$ is admissible if and only if it is satisficing for at least one $\bar{f}_{\mathbf{R}} \in \mathcal{R}$ and at least one $\bar{f}_{\mathbf{A}} \in \mathcal{A}$. Omitting coalitional dependence, this *w-admissible* (weakly admissible) set is defined as

$$\mathcal{W} = \left\{\mathbf{u} : \exists \bar{f}_{\mathbf{R}} \in \mathcal{R}, \exists \bar{f}_{\mathbf{A}} \in \mathcal{A} \text{ such that } \mathbf{u} \in S(\bar{f}_{\mathbf{R}}, \bar{f}_{\mathbf{A}})\right\}, \tag{C.3}$$

where $S(\bar{f}_{\mathbf{R}}, \bar{f}_{\mathbf{A}})$ is a functional representation of the satisficing set which makes explicit the dependence upon the expected accuracy and rejectability[1]. Observe that $\mathcal{W}$ may be written as

$$\mathcal{W} = \bigcup_{\mathcal{R},\mathcal{A}} S(\bar{f}_{\mathbf{R}}, \bar{f}_{\mathbf{A}}). \tag{C.4}$$

A much more restrictive notion of admissibility is when actions are admissible if and only if they are satisficing for every possible accuracy and rejectability. Omitting coalitional dependence, this *s-admissible* (strongly admissible) set is defined as

$$\mathcal{S} = \left\{\mathbf{u} : \mathbf{u} \in S(\bar{f}_{\mathbf{R}}, \bar{f}_{\mathbf{A}})\ \forall \bar{f}_{\mathbf{R}} \in \mathcal{R} \text{ and } \forall \bar{f}_{\mathbf{A}} \in \mathcal{A}\right\}. \tag{C.5}$$

Observe that $\mathcal{S}$ may be written as

$$\mathcal{S} = \bigcap_{\mathcal{R},\mathcal{A}} S(\bar{f}_{\mathbf{R}}, \bar{f}_{\mathbf{A}}). \tag{C.6}$$

A third potential definition for admissibility is justifiable. It is based on finding a representative density in the convex set of accuracies and rejectabilities. This method effectually combines the other elements of the set in a weighted average. The weighted average can be seen as a compromise between elements of the set. Possible criteria for selecting a representative density include selecting the density that maximizes entropy, selecting a typical density (in an information theoretic sense), or learning a density that yields the best performance. The *t-admissible* (typically admissible) set is defined as

$$\mathcal{T} = S(\bar{f}_{\mathbf{A}}^{\mathcal{T}}, \bar{f}_{\mathbf{R}}^{\mathcal{T}}), \tag{C.7}$$

where $\bar{f}_{\mathbf{A}}^{\mathcal{T}}$ and $\bar{f}_{\mathbf{R}}^{\mathcal{T}}$ denote the typical accuracy and rejectability, respectively.

C.3 Discussion

Selecting a mechanism for making decisions based on convex sets of accuracy and rejectability is an important problem for multiple agent control theory. At one extreme, a decision in $\mathcal{W}$ is satisficing under very weak conditions. For example, if the convex sets equal the entire probability simplex, then every

[1]could also consider options which are satisficing for at least one $b \in [0, \infty)$. However, this is equivalent to the satisficing set for $b = 0$.

potential control is admissible. At the other extreme, a decision in $\mathcal{S}$ is very restrictive[2]. The option of reaching a compromise or typical density and choosing a decision from $\mathcal{T}$ may be the most useful from an implementation perspective, but depending on who you listen to, it may be the easiest to justify on epistemological principles [25, 52]. Answers to these problems will endow multiple agent satisficing control theory with a great facility for handling ambiguity.

[2]Using the set $\mathcal{S}$ may be a method for connecting game theory to satisficing control theory.

VITAE

Michael A. Goodrich received his M.S. and B.S. (cum laude) degrees from Brigham Young University in 1995 and 1992, respectively. He is a member of Phi Kappa Phi, Eta Kappa Nu, Golden Key, and the Utah Academy of Sciences, Arts, and Letters. His research interests include theoretical decision theory, decision theory applied to intelligent control, multiple agent coordination and decision theory, set-valued estimation theory, set-valued signal processing, and fuzzy logic.

Bibliography

[1] Michael Bacharach. The epistemic structure of a theory of a game. *Theory and Decision*, 37:7–48, 1994.

[2] T. Balch and R. Arkin. Avoiding the past: A simple but effective strategy for reactive navigation. In *Proceedings of the 1993 IEEE International Conference on Robots and Automation*, Atlanta, Georgia, 1993.

[3] T. Balch, G. Boone, T. Collins, H. Forbes, D. MacKenzie, and J. C. Santamaría. Io, Ganymede, and Callisto – a Multiagent Robot Janitorial Team. *AI Magazine*, 1996. To appear.

[4] E. A. Billard. Stability of Adaptive Search in Multi-Level Games under Delayed Information. *IEEE Transactions of Systems, Man, and Cybernetics Part A: Systems and Humans*, 26(2):231–240, March 1996.

[5] E. A. Billard and J. C. Pasquale. Probabilistic Coalition Formation in Distributed Knowledge Environments. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(2):277–286, February 1995.

[6] A. E. Bryson and Y. Ho. *Applied Optimal Control*. Blaisdell, 1969.

[7] R. T. Bupp, D. S. Bernstein, and V. C. Coppola. *A Benchmark Problem for Nonlinear Control Design: Problem Statement, Experimental Testbed, and Passive Nonlinear Compensation*. ACC, 1995. Presented at the 1995 ACC.

[8] E. B. Castelan, J. M. Gomes da Silva Jr., and J. E. R. Cury. A reduced-order framework applied to systems with constrained controls. *IEEE Transactions on Automatic Control*, 41(2):249–255, February 1996.

[9] A. Chaudhury, P.-S. Deng, and S. Rathnam. A computational model of coordination. *IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans*, 26(1):132–141, January 1996.

[10] Chung Choo Chung and John Hauser. Nonlinear Control of a Swinging Pendulum. *Automatica*, 31(6):851–862, 1995.

[11] P. L. Combettes. The Foundations os Set Theoretic Estimation. *Proceedings of the IEEE*, 81(2):181–208, February 1993.

[12] E. Cox. Fuzzy fundamentals. *IEEE Spectrum Magazine*, pages 58–61, October 1992.

[13] C. E. T. Dórea and B. E. A Milani. Design of L-Q Regulators for State Constrained Continuous-Time Systems. *IEEE Transactions on Automatic Control*, 40(3):544–551, March 1995.

[14] R. L. Frost and W. C. Stirling. An Architecture for Intelligent Control Based on Epistemic Utility Theory. In *Proceedings of the 1994 American Control Conference*, pages 2580–2584, June–July 1994.

[15] J. D. Glover and F. C. Schweppe. Control of Linear Dynamic Systems with Set Constrained Disturbances. *IEEE Transactions of Automatic Control*, 16(5):411–423, October 1971.

[16] M. A. Goodrich, W. C. Stirling, and R. L. Frost. A Satisficing Controller for the RTAC. Submitted for review.

[17] M. A. Goodrich, W. C. Stirling, and R. L. Frost. A satisficing approach to intelligent control of nonlinear systems. In *Proceedings of the 1996 IEEE International Symposium on Intelligent Control*, Dearborn, Michigan, USA, September 15-18 1996. To Appear.

[18] M. A. Goodrich, W. C. Stirling, and R. L. Frost. A satisficing fuzzy logic controller. In *Proceedings of the FUZZ-IEEE '96 International Conference on Fuzzy Systems*, New Orleans, Louisana, USA, September 8-11 1996. To Appear.

[19] M. A. Goodrich, W. C. Stirling, R. L. Frost, and J. D. Kenney. A Theory of Satisficing Decision and Control. Submitted for review.

[20] N. B. Hadj-alouane, S. Lafortune, and F. Lin. Variable Lookahead Supervisory Control with State Information. *IEEE Transactions on Automatic Control*, 39(12):2398–2410, 1994.

[21] O. Hernández-Lerma and J. B. Lasserre. Error Bounds for Rolling Horizon Policies in Discrete-Time Markov Control Processes. *IEEE Transactions on Automatic Control*, 35:1118–1124, 1990.

[22] Y. C. Ho. Heuristics, Rules of Thumb, and the 80/20 Proposition. IEEE *Transactions on Automatic Control*, 39(5):1025–1027, 1994.

[23] Y. C. Ho and M. E. Larson. Ordinal Optimization Approach to Rare Event Probability Problems. *Discrete Event Dynamic Systems: Theory and Applications*, 5:281–301, 1995.

[24] Y. C. Ho, R. S. Sreenivas, and P. Vakili. Ordinal Optimization of DEDS. *Discrete Event Dynamic Systems: Theory and Applications*, 2:61–88, 1992.

[25] H. E. Kyburg Jr. and M. Pittarelli. Set-Based Bayesianism. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 26(3):324–339, May 1996.

[26] Venceslav Kafedziski and Darryl Morrel. Epistemic utility estimation with valuational convexity. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(1):190–193, January 1995.

[27] J. D. Kenney. Stochastic Satisficing Control. Ph.D. Dissertation, Brigham Young University. To appear.

[28] J. Kim, Y. Moon, and B. P. Zeigler. Designing Fuzzy Net Controllers Using Genetic Algorithms. *IEEE Control Systems Magazine*, 15(3):66–72, 1995.

[29] D. Kirk. *Optimal Control Theory*. Prentice-Hall, Englewood Cliffs, N.J., 1970.

[30] H. Kwakernaak. *Linear Optimal Control Systems*. Wiley-Interscience, New York, 1972.

[31] S. Lall and K. Glover. Robust performance and adaptation using receding horizon H_∞ control of time varying systems. To appear in Proceedings of the 1995 American Control Conference.

[32] S. Lall and K. Glover. A game theoretic approach to moving horizon control. In *Advances in Model-Based Predictive Control*. Oxford University Press, 1994.

[33] I. Levi. *Gambling with Truth*. MIT Press, Cambridge, Massachusetts, 1967.

[34] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts, 1980.

[35] I. Levi. *Decisions and Revisions*. Cambridge University Press, London, 1984.

[36] F. L. Lewis. *Optimal Control*. Wiley-Interscience, New York, 1976.

[37] R. D. Luce and H. Raiffa. *Games and Decisions*. John Wiley, New York, 1957.

[38] J. E. Marsden and A. J. Tromba. *Vector Calculus*. W. H. Freeman and Company, 3 edition, 1988.

[39] D. Q. Mayne and H. Michalska. Receding Horizon Control of Nonlinear Systems. *IEEE Transactions on Automatic Control*, 35:814–824, 1990.

[40] J. M. Mendel. Fuzzy logic systems for engineering: A tutorial. *Proceedings of the IEEE*, 83(3):345–377, March 1995.

[41] H Michalska and D. Q. Mayne. Moving horizon observers and observer-based control. *IEEE Transactions on Automatic Control*, 40(6):995–1006, June 1995.

[42] T. K. Moon and S. E. Budge. Classification Using Set-Valued Kalman Filtering and Levi's Decision Theory. *IEEE Transactions onf Systems, Man, and Cybernetics*, 24(2):313–319, February 1994.

[43] T. K. Moon, S. E. Budge, W. C. Stirling, and J. B. Thompson. Epistemic Decision Theory Applied to Multiple-Target Tracking. *IEEE Trans on Systems, Man, and Cybernetics*, SMC-24(2):234–245, 1994.

[44] T. K. Moon, R. L. Frost, and W. C. Stirling. An Epistemic Utility Approach to Coordination in the Prisoner's Dilemma. *BioSystems: Journal of Biological and Information Processing Sciences*, 37(1-2):234–245, 1996.

[45] J. C. Moore, H. R. Rao, and A. B. Whinston. Multi-Agent Resource Allocation: An Incomplete Information Perspective. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(4):1208–1219, August 1994.

[46] D. R. Morrell. *A Theory of Set-valued Estimation.* PhD thesis, Brigham Young University, 1988.

[47] D. R. Morrell. Epistemic Utility Estimation. Technical report, Telecommunications Research Center, College of Engineering and Applied Science, Arizona State University, Tempe, AZ, 1992.

[48] D. R. Morrell. Epistemic Utility Estimation. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(1):1–13, January 1993.

[49] D. R. Morrell and W. C. Stirling. Set-valued filtering and smoothing. In *Conference Record of Twenty-second Asilomar Conference on Signals, Systems, and Computers*, pages 236–239, Oct 1988.

[50] D. R. Morrell and W. C. Stirling. Tracking of Marginally Observable Targets. In *Proceedings ICASSP89*, pages 2720–2723, May 1989.

[51] D. R. Morrell and W. C. Stirling. Set-Valued Filtering and Smoothing. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(1):184–193, January/February 1991.

[52] A. Mosleh and V. M. Bier. Uncertainty About Probability: A Reconciliation with the Subjectivist Viewpoint. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 26(3):303–310, May 1995.

[53] J. Neveu. *Mathematical Foundations of the Calculus of Probability.* Holden Day, San Francisco, 1965.

[54] K. M. Passino. Intelligent Control for Autonomous Systems. *IEEE Spectrum*, 32(6):55–62, June 1995.

[55] J. Pearl. *Probablistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, Ca, 1988.

[56] J. F. Rudge. Multi-Agent Linear-Quadratic-Guassian Systems with "Poor" Observations. *SIAM J. Control and Optimization*, 16(5):832–846, September 1978.

[57] F. C. Schweppe. *Uncertain Dynamic Systems*. Prentice-Hall, 1973.

[58] H. A. Simon. A Behavioral Model of Rational Choice. *Quart. J. Economics*, 59:99–118, 1955.

[59] Samuel M. Smith. *A Computational Approach to Fuzzy Logic Controller Design*. PhD thesis, Brigham Young University, Provo, UT, July 1991.

[60] W. C. Stirling. Coordinated Intelligent Control Via Epistemic Utility Theory . IEEE Control Systems Magazine, 13(5):21–29, October 1993.

[61] W. C. Stirling. Decision-making using epistemic utility theory. In Cristiano Castlefranchi and Eric Werner, editors, *Lecture Notes in Artifical Intelligence*, number 830 in Lecture Notes in Artificial Intelligence, pages 164–183. Springer-Verlag, 1994.

[62] W. C. Stirling and R. L. Frost. Making Value-Laden Decisions Under Conflict. In *Proceedings 1994 IEEE Conference on Systems, Man, and Cybernetics*, volume 2, pages 1559–1564, October 1994.

[63] W. C. Stirling, M. A. Goodrich, and R. L. Frost. Procedurally rational decision-making and control. *IEEE Control Systems Magazine*, 1996. To Appear.

[64] W. C. Stirling and D. R. Morrell. Convex Bayes Decision Theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(1):173–183, January/February 1991.

[65] H. J. Sussamnn, E. D. Sontag, and Y. Yang. A General Result on the Stabilization of Linear Systems Using Bounded Controls. *IEEE Transactions on Automatic Control*, 39(12):2411–2425, 1994.

[66] P. Wang. The Interpretation of Fuzziness. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 26(2):321–326, April 1996.

[67] Qifeng Wei, W. P. Dayawansa, and W. S. Levine. Nonlinear Controller for an Inverted Pendulum Having Restricted Travel. *Automatica*, 31(6):841–850, 1995.

[68] Chen yuan Wang. *Fuzzy Measure Theory*. Plenum Press, 1992.

[69] L. A. Zadeh. What is optimal? *IRE Transactions on Information Theory*, 4(1):3, March 1958.