MULTIFINGERED GRASPING: GRASP REFLEXES AND CONTROL CONTEXT

A Dissertation Presented

by

JEFFERSON A. COELHO JR.

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2001

Department of Computer Science

MULTIFINGERED GRASPING: GRASP REFLEXES AND CONTROL CONTEXT

A Dissertation Presented

by

JEFFERSON A. COELHO JR.

Approved as to style and content by:

Roderic A. Grupen, Chair

Andrew G. Barto, Member

Neil E. Berthier, Member

Paul E. Utgoff, Member

Bruce Croft, Department Chair
Department of Computer Science

ACKNOWLEDGMENTS

I am a very fortunate man, and my good fortune put me in contact with wonderful people that helped me in every step in the realization of this work. First of all, this work would not be possible without the vision, the stimulating discussions and arguments, the unconditional and broad support offered by Rod Grupen. Together we embarked in an unrelenting search for the essence of multifingered grasping, while keeping an eye for interesting insights on the onset of motor control and the embodiment of knowledge about the physical world. It was a wild and rewarding ride – thanks, Rod! The thesis committee helped me stay grounded in this journey, and they made significant contributions to the structure and quality of this document. I am very grateful to Andy Barto, Neil Berthier and Paul Utgoff.

Many friends and colleagues also contributed in diverse ways to the success of this endeavor. Kamal Souccar implemented many of the controllers and sensor interfaces I used in my experiments, and showed me how to function in a robotics lab. Manfred Huber, with his unflappable character, wore many hats, as guru, sound box, and friend. Elizeth Araujo, my wife, contributed with insights during all stages of this work, as we discussed every development in detail. She also provided me with the necessary intangibles: love, energy, passion, common-sense, and round-the-clock support, respect and friendship. I would like also to acknowledge the contributions of my family. I have excellent role models, and little could I accomplish without their advice. As I said earlier, I am a very fortunate man.

This work had the support of CNPq - Conselho Nacional de Pesquisa, under the grant 202.107/90.6.

ABSTRACT

MULTIFINGERED GRASPING: GRASP REFLEXES AND CONTROL CONTEXT

SEPTEMBER 2001

JEFFERSON A. COELHO JR.

B.S.E.E., FEDERAL UNIVERSITY OF RIO DE JANEIRO

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Roderic A. Grupen

Grasping and manipulation of objects provide important stimulation toward maturing mental representations in human beings during their development. Both are open-ended activities, which can only be fully characterized in situational terms involving the purpose of the activities, the resources deployed, and the role of each effector. This dissertation describes how representations of grasp activity or *grasp state* may be constructed by an agent as it interacts with a collection of objects. The state representation proposed is based on a set of empirically-derived generative models of the system dynamics. These models are constructed autonomously by the grasping agent as it observes dynamic patterns in the phase portrait induced by a set of feedback grasp controllers. The resulting state representation serves as the basis on which competing grasp strategies are constructed, evaluated, and stored. The results

obtained indicate that it is possible to learn grasp policies that maximizes a user-supplied metric for individual objects even if object geometry, pose and identity are not explicit components of the grasp state vector. Contrary to many other approaches described in the literature, the resulting grasp strategy deals with noisy sensors and actuators, multiple objects, and has been implemented in a real hand-arm system. The techniques contributed in this dissertation can also be applied to the development of general sensorimotor control strategies for agents operating in open environments. They also provide insight into the automated construction of situated representations for sensorimotor activities, the scheduling of motor resources, the basis of perceptual categorization, and the uniform integration of continuous and discrete control primitives, control knowledge, control recovery/recognition into context-dependent strategies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Grasping and manipulation are very complex activities which, along with language, reveal properties of intelligence most uniquely attributed to human beings. Wilson [66] presents convincing arguments and archaeological evidence that points to manual dexterity as the driving force behind brain development. His arguments refute the prevailing view that a powerful brain preceded the mechanisms for manual dexterity. In addition to evolutionary pressure toward cortical development, hands and motor play provide important stimulation toward maturing mental representations in human beings during their development. Educators agree on the importance of block play [41], citing it as the single most stimulating activity a pre-schooler can have, exactly because of its open-ended nature.

Despite the valuable function of human hands and the critical stimulation they provide to the growth of mental representations, researchers in AI have largely overlooked this opportunity to understand the formation of the human mind – primarily because of the thorny issues involving the control of non-linear phenomena, the interpretation of sensory information, and the role of the mechanism (kinematic structure, sensory apparatus and computational constraints) in the formation of mental representations. This dissertation contends that these issues are interrelated, and together determine how autonomous agents perceive their environment. An agent's representation of the current context is dependent upon its sensorimotor apparatus and the history of its interaction with the environment. Autonomous agents refine continuously the perception of their environment, in a process shaped by the agent's perceptual resources and limitations, the agent's limited computational capability, and the kinematic and dynamic constraints imposed by the agent's body.

The broader objective of this dissertation is the study of issues pertaining to the development of sensorimotor control strategies for agents operating in open domains or environments. Specifically, this work addresses the synthesis of robust control strategies for multifingered grasping. Robustness is sought with respect to the agent's limited actuation and sensing abilities, including the monitoring of object properties that affect the grasp operation (e.g., object geometry and pose, friction between object surface and fingers, and object weight), sensory and actuation noise, and non-linearities. Robust control strategies can also be characterized as context-dependent control strategies, in which the current context is defined by the combined effects of the factors listed above.

1.1 Open Environments

Open environments are domains characterized by wide ranges of operating conditions in which the agent must function. In this specification, control strategies must be robust to noise processes affecting sensing and actuation and more importantly to truly distinct operational regimes or *contexts*. Both sources of variability are addressed in the approach developed in this dissertation, respectively, through the adoption of two important design principles.

- Feedback control laws are used as the building blocks of the agent's control strategy in the environment. Feedback controllers constitute the basic mechanism for accommodating a wide range of operating conditions and rejecting noise. Moreover, they can be designed to promote safe operation of the robot in the environment from the first servo cycle onward, a particularly relevant characteristic for any robot control architecture. In this framework, the agent is constrained at all times to adopt a feedback control strategy selected from the task *control basis*, an instruction set for the agent's behavioral programs. Each controller in the control basis is defined by a subset of fingers or grasp resources. In principle there exists an optimal schedule of grasp resources for each context. The overall control architecture seeks the best match between problem instances and grasp resources.

- The many control contexts experienced by the agent as it interacts with the environment can be represented by combinations of empirical models of control dynamics. Control dynamics are modeled empirically and the resulting *generative models* are later used to categorize run-time experience and potentially recover context information previously unavailable to the grasping agent. Transients in the control process reveal this information as discernible variations in the control dynamics. Temporal histories contain information regarding the operational context that is not generally available in instantaneous sensor observation. This technique represents a completely general means if recovering hidden or unobservable state that is applicable to all robot systems. Moreover it eliminates the need to construct complete prior models – an assumption common to the vast majority of existing, open-loop grasp planners that exist today. Given adequate context, the agent may evaluate the context-dependent utility of each control law in the control basis and improve its overall control strategy.

This problem formulation is suitable to any open domain task in which physical agents must interact with non-stationary environments, characterized by many operating modes or contexts. Agents acting in open domains must have a rich repertoire of control strategies and be able to relate the current context to the best line of action in their repertoire, continuously reacting to the history of sensory observations.

1.2 Synthesis of Context-Dependent Grasp Policies

This dissertation breaks down the synthesis of context-dependent grasp policies into three sub-problems: the first problem is the definition of a *control basis*, a set of feedback controllers designed to address locally specific objectives within a task domain. The agent's action repertoire must be *complete*; a set of controllers Π is complete if for each problem instance in a task set $\mathcal{T}$ there exists a sequence of controller activations that addresses it. Guaranteeing the completeness of a set of controllers is a hard task in general, especially in open domains where the required behavior varies significantly. Note that the design of a complete set of controllers can be trivial in some cases; a set of controllers capable of displacing a mobile robot in the x and y directions is complete for the navigation problem domain – yet, it does not make the navigation task any easier. The central issue is, therefore, how to derive controllers that span a functional specification. General answers to this issue are not known at this point, but I claim that the control architecture presented in this work addresses rich problem domains, including the development of sensorimotor strategies and that the architecture itself has been shown to be versatile and quite expressive [10, 12, 23, 27, 28].

The second sub-problem is the description of the agent's experience in the environment in terms of *state* transitions. Before the agent can evaluate the consequence of its control choices, it must be able to express the current context in terms of *state* variables, so that the agent can summarize the effects of past behavior and create expectations about future course of actions. State variables must be inferred from current and past sensor readings.

The third sub-problem is the selection of the best course of action for the current context or state as perceived by the agent. The agent must discriminate between distinct operational contexts, and select which controller $\pi_i \in \Pi$ is optimal for the perceived context or *state*. These three problems are the object of Chapters 3, 4, and 5, respectively. The following sections introduce briefly the strategy proposed to address these problems.

1.2.1 Grasp Control Basis

The components of the control basis Π represent the native control structure of the agent, expressed as $\Pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$. For the multifingered grasp domain, the control basis Π consists of controllers designed to locally optimize the current grasp configuration, by generating differential displacements of the contacts on the object surface to minimize the resulting wrench (force and torque) residual on the object. Wrench minimization is an important grasp control objective, as it reduces the friction forces required for grasp stability.

The controllers in Π are designed to be asymptotically stable under a variety of operational circumstances. More importantly, the controllers in Π generate system dynamics, producing in phase space that later will be used to identify the current control context.

1.2.2 Grasp State

Skillful, context-dependent policies for grasping are based ultimately on estimates of the *state* of the grasping process; the agent will select a controller, $\pi_i \in \Pi$, that maximizes the utility of the perceived state. In this work, state information is conveyed by sets of generative models that describe the grasp dynamics generated by the controllers in the control basis. Generative models have been used in speech analysis [17], scene reconstruction [50], and handwritten digit recognition [51] to describe the intrinsic variability present in these application domains. Implicit in the application of generative models is the assumption that the phenomena described are generated by processes that end up lending structure to the models themselves. For instance in the handwritten digit recognition domain it is assumed that the shape of the digit "2" may be correlated with the shape of the digit "5" if they are present in the same writing sample; recognizing one digit may help in the recognition of the other, as the underlying generative model is the same for both digits.

In the multifingered grasp domain the object itself is the entity the binds the dynamics across different controllers. Although the agent does not have access to object identity, it may indirectly infer it by observing the dynamics generated by the several controllers $\pi_i \in \Pi$. State information is conveyed by the tuple $(\pi_i, \mathbf{q})$, where π_i denotes the current controller and $\mathbf{q}$ is a vector of Boolean predicates that indicates which generative models of system dynamics are compatible run-time observations of the grasping process. A set of models $\mathcal{M}(\pi_i)$ summarize data collected by the agent as it samples the dynamics generated by each controller π_i. Each generative model in $\mathcal{M}(\pi_i)$ is a conditional probability density function describing the distribution of observations in phase space for an operational regime or context under control law π_i.

The models in $\mathcal{M}(\pi_i)$ collectively represent empirical knowledge acquired about a range of operational contexts found in the environment as the agent adopts primitive π_i. I assume that similar contexts benefit from similar control policies so patterns of membership in $\mathcal{M}(\pi_i)$ identify "equivalent" run-time situations that call for "equivalent" behavior. Once $\mathcal{M}(\pi_i)$ is available, Bayesian estimation is used to estimate $\mathbf{q}$. Due to its definition, state estimation is intrinsically situated, because it is grounded in autonomous interaction with the environment. Moreover, it extends the kinds of information conveyed in individual observations. The resulting state space is discrete, even though the original performance metrics (and observables) are real variables. Moreover, one can prove (Section 4.3.1) that partitioning the phase space with m models will result in $O(m^2)$ states.

1.2.3 Context-Dependent Control Policies

The set of generative models $\mathcal{M}(\pi_i)$ allows the agent to re-describe its experience in the environment in terms of state transitions in a discrete state space. State transitions identify changes in the "informational state" of the agent and can be used as decision points where the agent may adopt a new controller to maximize utility.

The determination of the best switching strategy is a sequential decision problem which can be described and solved within the Markov Decision Process (MDP) framework. The set of possible states $(\pi_i, \mathbf{q})$ defines the state space for the MDP problem and the primitives in the control basis Π define the set of candidate control actions. Ideally, the resulting selection policy will maximize a user specified performance metric while at the same time minimizing its variance across several operational contexts.

The resulting control architecture allows the agent to plan the use of its resources, and to influence the quality of the solutions generated. The resulting system is arguably the most complete multifingered manipulation system described in the literature, and a fine platform for the study of several interesting topics such as the formation of perceptual categories, the sensorimotor basis of problem solving strategies, and the influence of resource constraints. This work provides insight into the automated construction of situated representations for sensorimotor activities, the scheduling of motor resources, the basis of perceptual categorization, and the uniform integration of control and policy in the grasping and manipulation domains. The result is progress toward a general purpose architecture in which control and control knowledge are acquired simultaneously through on-going interaction with the world [13].

1.3 Document Organization

The organization of this document reflects the task breakdown proposed in Section 1.2. The next chapter discusses work related to the architecture described in this document. Chapter 3 provides a formal definition of the grasp control basis, and characterizes the performance of the constituent controllers. It also discusses how optimal control principles can be applied in the derivation of the time-optimal composition schedule for the controllers. The composition rule derived can only be applied if a complete geometric model of the object is available.

The characterization of grasp dynamics and the representation of the grasp process in terms of state transitions are explained in Chapter 4. Chapter 5 discusses the synthesis of context-dependent grasp policies, and details the learning algorithm used in their derivation. It also characterizes the performance (in a simulated environment) of the policy derived. The entirety of Chapter 6 discusses the implementation and the results obtained in experiments involving physical manipulators. Chapter 7 reviews the results and insights attained during the investigative work leading to this dissertation, as well as its contributions. The appendix characterizes the dynamics of grasp formation for the special case in which all contacts are positioned on a planar facet; it also introduces the parametric model of grasp dynamics used in this work.

CHAPTER 2
RELATED WORK

This thesis addresses the development of context-sensitive sensorimotor strategies for physical agents in general, and the development of multifingered grasp strategies in particular. Both subjects have been studied extensively, under many formulations and with different motivations and emphases.

2.1 Multifingered Grasp Synthesis

Salisbury [55] employed the grip Jacobian, G, to transform contact forces to object frame wrenches[1]. Since that time, the issue of constructing appropriate contact configurations (and therefore, grip Jacobians) has received a great deal of attention. The grasp synthesis problem may be viewed as a search in contact configuration space for contact geometries that can transmit the right set of wrenches to the object.

2.1.1 Off-line Grasp Synthesis

Most work in the literature addresses *off-line* grasp synthesis, in which the search for contact geometries is performed on a geometric model of the object. Previous work in off-line grasp synthesis may be categorized into two broad categories:

- Optimization-driven grasp synthesis: Several authors cast grasp synthesis as an optimization problem over a bounded domain. The set of all contact configurations over a given object geometry has been evaluated through continuation

[1] Wrench denotes the 6-dimensional vector containing the forces and moments along the x, y, and z axis.

techniques [20], or through a direct search over object geometry [47]. Several metrics have been used to compare candidate grasp configurations; in the work of Faverjon and Nguyen [20, 47], the contact configurations that satisfy the force closure condition are further sorted on the basis of the sensitivity of force closure condition to the precision of contact placement. Sorting may also be based on the maximum force required to counteract a worst case wrench [21]. Other evaluation functions incorporate the manipulability of the resulting grasp configuration [35], its reliance on friction forces [6], and the ability of the grasp to resist external disturbances [26, 31]. In these broader contexts, the complexity of optimization-based approaches can be prohibitive.

- Taxonomy-based grasp synthesis observes that grasp geometry is strongly related to hand morphology. Taking cues from human grasping strategies (described in detail by Napier [46]), these techniques relate hand configuration (and thus contact configuration) to the geometry of the object in the context of the task [16, 29, 63]. This approach is a memory-intensive map from object geometry and task to appropriate hand configuration and pose.

All approaches discussed so far can be distinguished on the basis of object geometries permitted (polygons, parametric surfaces, two and three dimensions), and on the basis of the type (point contact, soft contact) and number of contacts. All depend on the availability of complete geometrical models of the objects, and, in some cases, friction coefficients between fingers and object. Satisfying these modeling requirements is expensive and this thesis discusses how the costs of their satisfaction can be circumvented in a on-line framework.

2.1.2 On-line Grasp Planners

Few researchers have presented on-line, adaptable grasp strategies. Pollard [48] proposed an approach in which grasp prototypes are generalized off-line and indexed according to task and grasp quality. The grasp prototype can be used later to constrain an on-line search of grasp configurations for novel objects. Jameson and Leifer [31] present a grasp strategy in which artificial potential fields are used to express joint range and interference constraints as the manipulator grasps an object. In their formulation, the artificial potential is used to score hand/arm configurations, and the resulting quality metric included stability, hand collisions with the table top, and hand kinematics. Although global geometric models of the object and its surrounding environment are not required by the planning algorithm, the resulting procedure did not address the sensory requirements of grasping task. The planner requires information that cannot be trivially inferred by current sensory technology such as the distance between a finger and the table top/object surfaces.

2.1.2.1 Sensor-based Grasping and Manipulation

Sensor-based approaches to the on-line grasp synthesis problem are mostly vision-based [33, 54]; typically, geometric parameters extracted from the object image are used to predict grasp quality. An open-loop control scheme is used during grasping, and learning procedures are employed to build the map from visual features to expected grasp quality indices.

Tactile sensing has been used mostly for object recognition/modeling [2, 7], monitoring of object manipulation [19], or the control of grasping forces [56, 36]. Allen [2] described how haptic exploratory procedures can be used to acquire shape models of objects. Three exploratory procedures (EPs) – grasping by containment, lateral extent EP, and contour following – were implemented. In each, tactile sensors probed the object surface repeatedly following stereotypical, predefined movement patterns.

Chen *et al.* [7] proposed to explore object surfaces through the use of forward sensor models, that compute differential changes in sensor readings from differential changes in contact parameters (contact normal and position). Given a desired tactile feature or contact parameters, the exploration procedure compares it to the current sensor reading and uses the forward model (or tactile Jacobian) to command incremental displacements of the tactile sensor. The exploration is guided in its entirety by the combination of the desired contact parameters and the tactile data obtained. The work of Chen *et al.* is novel because in it contact information determines contact displacement over the object surface. Although grasping is not their objective, clearly their procedure could be easily adapted to address grasping tasks. The method advanced in this dissertation employs this kind of approach as well.

Farooqi *et al.* [19] described the implementation of two manipulation primitives, namely *object rotation* and *object pivoting*. It is assumed that a geometric model of the object is available, and a combination of ultrasonic sensors and tactile sensors is used to establish the initial object spatial orientation. Tactile sensors are then used to monitor the execution of the manipulation primitives, keeping track of variables such as the distance between the contact point and the closest edge, distance between the center of gravity and the pivot axis, and facet transitions. The execution of the resulting grasp procedure is sensor-based, but the procedure itself is still dependent on the quality of the initial model of object geometry, that is not revised once the grasp process starts. The grasp strategy described in this dissertation is independent of the quality of geometric models, as no model is assumed or built.

For some authors, grasp synthesis corresponds to the search not for grasp configurations, but for a set of contact normal forces capable of guaranteeing grasp equilibrium. Tactile-based grasp procedures often address the stability maintenance problem [36], in which contact forces are modulated to maintain the contact forces within the friction cone. These strategies are complementary to the grasp procedure described in this dissertation, as they can be employed after a locally optimal grasp configuration is defined by the grasp synthesis procedure.

2.1.3 Discussion

The approaches discussed above describe robotic grasping as an optimization problem, where the goal is to derive the grasp configurations that maximize a given metric under controlled circumstances. Object geometry and friction coefficients are known *a priori*, and sensor data acquisition and interpretation play a secondary role – these approaches often contemplate the open-loop execution of the grasping plan. This thesis departs from the traditions in robotic grasping in three important ways:

- it addresses the sensory requirements of the grasping task, by assuming very little prior knowledge about the object geometry, while requiring arguably minimal sensor information for the execution of grasp commands, namely the contacts' positions and respective normal vectors. The grasp formation process models the contact between fingers and object surface as frictionless point contacts, and as a result the convergent grasp configurations will rely minimally on surface friction for grasp equilibrium;

- it provides for online, incremental refinements of existing grasp configurations, based on feedback grasp control laws. The closed-loop grasp controllers reject noise, and can be used for the execution of grasp plans, if such plans are available;

- it deals with open sets of objects by adopting context-dependent control strategies. Control strategies are selected based on past performance under similar circumstances.

These premises are the starting point for the trajectory leading to the synthesis of context-dependent grasp strategies. At the onset of this process, an agent endowed with limited sensorimotor resources is only able to execute fixed feedback control strategies akin to grasp reflexes. Over the course of many grasp trials, the agent explores its control repertoire and the combinations of kinematic and sensory resources available. Its observations lead to the construction of empirical models of grasp dynamics, and at a later stage the agent is able to use the state information inferred to compare sequences of control activations and create context-dependent behavior. This evolutionary trajectory reflects the belief that multifingered grasping and manipulation are critical (perhaps the best) platforms for studying the onset of embodied intelligence.

2.2 Synthesis of Control Strategies

A central premise of this work is the use of local controllers as the building blocks for a global control strategy. This idea has been explored by many researchers [3, 23, 27, 37, 40, 43, 53, 49] when confronted with the problem of synthesizing a complex control law in "open" application domains.

In these cases, it is possible to partition the system's state space into regions of competence and to associate a local control law with each region [53]. The constructive organization of local control laws defines a global control law that attends to user-specified requirements, such as global stability, and robustness with respect to noise or plant variations.

Reinforcement learning-based composition schemes [3, 27, 37, 40, 43] construct control laws by allowing the agent to select which control action to take in each discrete state from a finite set alternatives. This sequential decision problem is solved by allowing the agent to exploit the alternatives and to build local estimates of the utility of each control primitive in each given state. These local estimates, in turn, define a *value function* as the expected discounted sum of future rewards for each state-action pair [59]. The use of controllers or even complete policies as action primitives leads to smaller state and action spaces, allowing for the solution of more complex tasks [3, 27]. It also allows for the construction of hierarchical solutions, involving control actions with distinct time scales [60].

All approaches mentioned presume that state variables are available at all times; in many cases it is further assumed that the set of all current sensor readings encodes system state. The next section reviews the literature related to the state estimation problem.

2.2.1 State Estimation

Constructing/recovering state information for control or forecast tasks has been a goal of several disciplines – operations research, control theory, dynamical system theory, reinforcement learning, and others. The objective is always the same: derive or build enough state information to permit the agent to evaluate control alternatives, ultimately leading to tangible improvements in performance.

The extraction of state information from a sequence of observations is related to the "observability" of the system, a concept used to characterize state-feedback control laws. For linear systems, a system is regarded as *observable* if it is possible to estimate its state at a time t_0 given its output $y(t)$ over a finite time interval $[t_0, t_1]$. It is interesting to note that observability (as well as controllability) is a property of the state representation chosen for the system, rather than a property of the system itself; the lack of an observable state representation does not imply that one does not exist [38].

Within the nonlinear control framework, the concept of observability is closely related to the notion of distinguishability, and it is always defined within a certain neighborhood (c.f. [30]). In intuitive terms, a system is said to be observable if all relevant variables for the current task are incorporated into the system's feedback control laws through estimation or direct sensing. This definition suggests that few systems can be deemed observable; except for trivial tasks, the complete determination of the "relevant variables" may be a hard task in itself.

State estimation methods such as Kalman filtering acknowledge the presence of noise in sensor readings, and provide the means for refining the current state estimate given a statistical characterization of the noise processes and the model of the system dynamics. Kalman estimation has been proved to be optimal for linear systems subject to Gaussian noise; for other noise processes, the Kalman filter is the unbiased linear filter with minimum error variance [38].

State estimates provided by Kalman filters presume a closed-form model of system dynamics. Some researchers have circumvented this requirement by building parametric representations of system dynamics directly from measurements. Tani [61] describes a system in which a mobile robot constructs a description of its environment based on the sensory observations collected by the robot as it navigates in it. Initially the robot explores the environment while avoiding collisions, without specific goal configurations. The continuous stream of data is used to construct a forward model (implemented as a recurrent neural network) describing the evolution of percepts in the navigation task. Model accuracy is monitored by comparing model-based prediction and the observed sensory input. Once the model is available, it can be used to plan optimal paths to arbitrary positions in the environment.

Kalman filters are not applicable to classification tasks, when measurements may come from any element out of a set of distributions. The procedure must know beforehand which model of system dynamics and noise distribution describe the measurements observed. Likewise, parametric model-based estimators will fail if ambiguity is present in the data used to construct the dynamic models. In both cases, system state is considered to be *hidden*; the estimates and predictions are computed based on partial state information. In robotic tasks, limited perception (due to limited sensors, occlusions, and constrained sensory ranges) and computational resources conspire to keep the agent from assessing its current state completely.

2.2.2 Dealing with Partial State Information

Taking on a course of action given partial information has been studied within the partially observable Markov decision problem (POMDP) framework. In it, system state is not observable, and the agent must infer it from a sequence of discrete observations. Exact solutions involve recasting the POMDP as a regular MDP in which the state space consists of vectors representing the probability that the agent is in each state (the belief vector) [32]. If $|S|$ is the cardinality of the state space, then the *belief state* of the agent is a $|S|$-dimensional vector. The resulting representation is continuous, even if the underlying state space is discrete. The solution of the resulting continuous space MDP requires exponential time and memory for its computation. Furthermore, exact solutions are defined only for problems in which the state space is discrete, and require complete models of the environment (including measurement and observation probabilities), typically not available a priori.

The complexity of POMDPs forced researchers to explore for methods more robust with respect to the failure of the Markovian hypothesis [58], methods addressing a certain class of POMDPs [65], and methods that create and explore alternative representations of the state space [39, 9, 42].

Singh et al. [58] propose that stochastic policies may provide better results in terms of the average payoff from each state than any deterministic policy, a result valid for memory-less policies. Their method aims at approximating POMDPs without explicit state estimation, and the main argument against state estimation is the computational cost associated with the estimation procedure.

The Perceptual Distinctions Approach (PDA) [9] breaks away from the convention of estimating state information directly from percepts. Instead, the agent's experience is described in terms of transitions between internal states or *classes*. Initially, the transformation from percepts to classes (or predictive model) is completely random. As the agent interacts with the environment, it tests the validity of the model by testing its ability to predict changes in the sensory input vector. The acquired experience is used to improve the model through a number of tests on the usefulness of partitioning existing classes and introducing new perceptual distinctions. The tests are performed in rounds and distinctions perceived as useful are preserved in future rounds. The current state or class is not directly observable, so the agent must maintain a belief vector in memory containing the probability that the agent is in each class.

Instead of monitoring discrepancies in the perceptual input (a sign of the failure of the Markovian assumption) as proposed by [9], McCallum's U-Tree algorithm [39] searches for discrepancies in the rewards received as the agent takes each action available in each internal state. McCallum's U-Tree algorithm records every (state, action, reinforcement) triple observed in memory. Each triple is also stored in a leaf in a *prefix tree*, and each leaf may contain several triples. The address of the leaf is a concatenation of tests on past values in the observation vector, or tests on past control actions. Each test represents a relevant distinction observed and tested over the course of many trials. Leaf nodes can be split as new distinctions (i.e., distinctions leading to an improvement in the prediction of future rewards) are discovered, increasing the depth of the prefix tree incrementally. Therefore, leaf nodes represent state-action pairs that predict future rewards, and can be used as state in a reinforcement learning problem.

The U-Tree algorithm is still a very expensive algorithm, and simplifications are introduced in order to make it more manageable. First, there is a bound on the "time window" used to collect past observations/actions in order to build distinctions. Second, the order in which the distinctions can be performed must be predetermined by the system designer. Both simplifications may prevent the system from finding the optimal policy. However, McCallum's greatest concern is (in his words) the "chicken and egg problem": the current policy depends on existing distinctions and the distinctions themselves depend on the current policy. Moreover, past distinctions may become irrelevant after changes in the control policy, and the U-Tree algorithm never removes distinctions proven to be relevant in the past – it does not backtrack.

Michaud and Matarić [42] have also proposed a strategy to transform past agent experience into state representation. Their subsumption-based control architecture can improve its performance by learning to deal with non-stationary conditions. Behaviors are split into three classes: task behaviors, maintenance behaviors, and alternative behaviors. Whenever a maintenance behavior is activated (a collision avoidance behavior in their example), the agent has the option of activating one alternative behavior (follow-side, rest, turn-randomly) or keep executing its current behavior. System state is a combination of the history of behavior activation (represented by a task tree) and the expected performance relative to the average performance observed in the recent past. The objective of the optimization procedure is to minimize interference between behaviors during task execution, and the overall time consumed.

2.2.3 Discussion

Like the PDA architecture [9], McCallum's U-tree representation and the "Interaction Models" proposed by Michaud and Matarić, the state representation presented in this dissertation (briefly introduced in Section 1.2.2 and formally described in Chapter 4) is constructed based on information collected over a number of observations. All approaches address the problem of selecting actions given partial state information. Contrary to the assumption considered by Singh et al. [58], all approaches assume that state estimation need not be an expensive operation; it is believed that agents can benefit even if system state is not completely recovered.

The grasp agent constructs state by observing how grasp error evolves over time — in contrast to the PDA architecture, there is no search for relevant features. The resulting representation is similar to the U-tree representation proposed by McCallum, but there are also important differences:

- U-trees are formed as the agent explores the environment through random sequences of atomic control actions, while in this work environment exploration is biased by the controllers in the control basis (briefly introduced in Section 1.2.1 and formally described in Chapter 3). Each controller aims at maximizing a local performance metric, and a local convergent configuration is always reached if the corresponding control law is repeatedly executed. The sequential activation of the controllers in the control basis is not explored during model construction. Exploratory actions can be designed to be safe, and be constrained to search relevant subspaces, speeding up learning.

- U-trees are constructed based on statistics derived from a finite window of observations while the parametric models of system dynamics proposed in this work can encode very long observation sequences that, in principle, can be extended on-line.

- Nodes are added to the U-trees to address observed failures in the Markovian assumption, namely disparities in the distribution of rewards from a given state. The emphasis is placed on the construction of the U-tree, an structures representation of the environment; by contrast, our representation emphasizes behavior organization and performance.

- The representation proposed is more compact and less explicitly related to raw perception than the representations employed by McCallum and Tani [39, 61].

Many other features make the generative models-based approach unique:

- Fast response to context changes: The models of Michaud and Matarić [42] include a "forgetting" mechanism that ultimately will cause the system to adapt its control policy to changes in the environment. The proposed representation emphasizes the selection of an appropriate control policy as opposed to adaptation to changes in the environment.

Within the multifingered grasp domain, there exists a correspondence between control policy selection and the resources invoked and applied to the task. Therefore, context adaptation can be expressed as a resource management task.

- The definition of "state" (as shown in Chapter 4) is based on generative models of system dynamics, obtained as the agent experiments with a set of control strategies. This formulation allows for the refinement of the state space without the need for statistical tests: there is no need to test for independence between any pair of generative models, as no conceptual gain in information or utility is accrued by using the minimal set of generative models. This paradigm also accommodates measurement noise.

- Decision points (i.e., point in time when the agent is allowed to switch its current control strategy) are defined by changes in the informational state of the agent. In the work of Michaud [42], such decision points are a function of the relation between the average performance and the performance expectation - a more complicated formulation. The state representation advanced in this thesis – state defined by the varying probabilistic membership of the observation series to a family of generative model – is independent of the time spent executing the current behavior or controller; moreover, it provides more granularity than formulations relying on convergence events of controllers, while avoiding the extra costs paid by approaches that deal with continuous time [60].

CHAPTER 3

MULTIFINGERED GRASP SYNTHESIS

3.1 Characterization of Grasp Configurations

The contact between fingers and objects has been characterized in terms of the relative motions between finger and object (dynamic contacts [44], for example), and in terms of the wrenches transmitted by the finger to the object (static contacts) [55]. The dynamics of contacts can be further classified as belonging to the class of rolling or sliding contacts, or a combination of both modalities. Static contacts are studied in terms of the wrenches each contact type can transmit to the object surface. Three static contact types are frequently described:

- frictionless point contacts can impart forces only along the normal to the object surface;
- point contacts with friction can cause interaction forces to be in any direction within the friction cone defined by $|\mathbf{f}_t| \leq \mu \mathbf{f}_n$, where $|\mathbf{f}_t|$ is the magnitude of the friction force, $\mathbf{f}_n$ is the magnitude of the normal contact force (always positive), and μ is the static coefficient of friction;
- soft contacts that permit contact torques around the point of contact in addition to normal and tangential (friction) forces. The contact torques $\mathbf{m}$ are modeled similarly to the tangential forces and their magnitudes are limited by the magnitude of the normal forces according to the expression $|\mathbf{m}| \leq \gamma \mathbf{f}_n$, where γ is the coefficient of torsional friction.

For all contact types, the normal force $\mathbf{f}_n$ must always be compressive; the fingers can only push against the object surface.

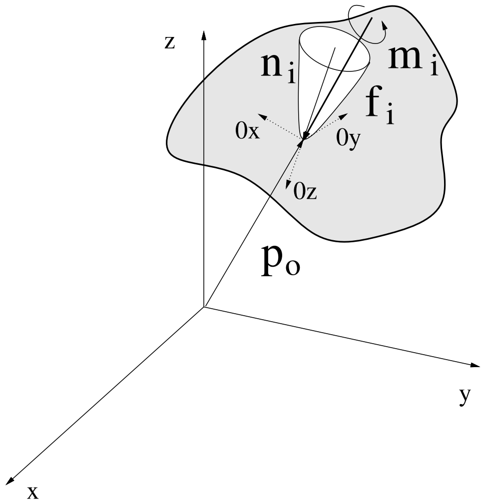

Figure 3.1. Contact wrench for soft contact model.

The analysis of contact systems distinguishes the wrench contributions due to normal forces, friction forces, and torques. Each contact type can be described by a *wrench basis*, $\mathbf{B}$ written in local (contact) coordinates. Table 3.1 lists the contact wrench bases for the contact types discussed. The contact wrench bases are defined with respect to the contact coordinate frame, $\mathbf{O_c} = [\mathbf{o_x} \ \ \mathbf{o_y} \ \ \mathbf{o_z}]^{\mathbf{T}}$, centered at the point of contact and oriented such that the vectors $\mathbf{o_x}$ and $\mathbf{o_y}$ define a plane tangent to the object surface and $\mathbf{o_x} \times \mathbf{o_y} = \mathbf{o_z}$ point inwards along the surface normal – see Figure 3.1.

The wrench in frame $\mathbf{O_c}$ is given by $\boldsymbol{\omega}_c = \mathbf{B}\mathbf{f}_c$, where $\mathbf{B}$ is the wrench basis, a set of unit vectors spanning the wrench space[1] generated by each contact model

[1]As noted by Duffy [18], there are important differences between wrench spaces and the vector spaces defined in Linear Algebra; the orientation angles of wrench vectors is dependent on the units used and they are asymmetric with respect to the magnitude of the force components, as negative contact forces are not defined. These differences do not invalidate the analysis or algorithms presented in this dissertation.

Table 3.1. Wrench bases for simple contact types.

Contact Type	Wrench Basis $\mathbf{B}$
Frictionless point contact	$[0\ 0\ 1\ 0\ 0\ 0]^T$
Point contact with friction	$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$
Soft contact	$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$

(see Table 3.1), and $\mathbf{f}_c$ is the vector formed by the magnitudes of contact forces and moments in the contact frame. Following the formalism introduced by Murray, Li, and Sastry [45], the resulting wrench corresponding to contact i in object coordinates is computed through the transformation

$$\boldsymbol{\omega}_{o_i} = \begin{bmatrix} \mathbf{R}_{oc_i} & \mathbf{0} \\ \mathbf{p}_{oc_i}^T \mathbf{R}_{oc_i} & \mathbf{R}_{oc_i} \end{bmatrix} \mathbf{B}_i \mathbf{f}_{c_i} = \mathbf{G}_i \mathbf{f}_{c_i}. \tag{3.1}$$

In Equation 3.1, $\mathbf{R}_{oc_i}$ is the rotation matrix between the contact coordinate frame and the object coordinate frame, and $\mathbf{p}_{oc_i}$ is the contact position vector in object coordinates.

The matrix $\mathbf{G}_i$ describes how the components of the contact wrench $\mathbf{f}_{c_i}$ are transformed into wrenches in the object coordinate frame. The resulting wrench at the object frame due to k contacts is given by the superposition of the contact wrenches. The sum $\mathbf{W} = \sum_{i=1}^{k} \boldsymbol{\omega}_{o_i}$ is frequently expressed by the product of the matrix $\mathbf{G}$ by the vector $\mathbf{f}_c$ as follows:

$$
\begin{aligned}
\mathbf{W} &= \boldsymbol{\omega}_{o_1} + \boldsymbol{\omega}_{o_2} + \ldots + \boldsymbol{\omega}_{o_k} \\
&= [\mathbf{G}_1 \; \mathbf{G}_2 \; \ldots \; \mathbf{G}_n] \; [\mathbf{f}_{c_1} \; \mathbf{f}_{c_2} \; \ldots \; \mathbf{f}_{c_n}]^T \\
&= \mathbf{G} \; \mathbf{f}_c \qquad (3.2)
\end{aligned}
$$

The grip transform, $\mathbf{G}$ [55], and the friction constraints on contact wrenches completely characterize a static grasp configuration. The quality of a grasp configuration is closely related to the properties of $\mathbf{G}$, especially its span and its null space (assuming that the friction and unisense contact constraints are satisfied). Ideally, the product $\mathbf{G}\ \mathbf{f}_c$ should span $\Re^6$ to suppress general disturbance wrenches applied to the object. This condition is necessary and sufficient for the definition of force closure grasps.

Force closure grasps are also characterized by the existence of internal forces. As demonstrated in [45], a grasp is force closure if and only if $\mathbf{G}$ is surjective (i.e., onto) and there exists a vector of contact forces $\mathbf{f}_{ns}$ in the null space of $\mathbf{G}$ such that $\mathbf{f}_{ns}$ satisfy the friction constraints. Intuitively, internal forces are a necessary condition for force closure because they can be increased (i.e., the object can be squeezed by the fingers) in order to resist disturbance wrenches while applying a zero net wrench to the object.

3.2 Multifingered Grasp Synthesis

In its simplest form, the grasp synthesis problem consists of finding a grasp configuration such that a reference wrench, $\mathbf{W}$, is produced in a specified object coordinate system. Candidate grasp configurations must satisfy bounds on friction wrenches and must employ strictly positive normal forces. In a manner consistent with Equation 3.2, a grasp synthesis algorithm can achieve a reference wrench condition by either modifying the contact configuration, by modifying the contact forces within the noted constraints or by combining these degrees of freedom. In some cases, many configurations satisfy the requirements and a secondary selection criterion may be used to rank the grasp configurations.

Grasp synthesis is frequently posed as a mathematical optimization problem [6, 20, 26] in spite of challenging issues related to the computational complexity of the planning algorithm, and to the required sensory information. The approach in this thesis departs from the traditional formulation in many ways.

- It casts grasp synthesis as an online process in which grasp configurations are fixed points for the control laws applied. Moreover, grasp synthesis is described as a robust control problem with respect to variations in object geometry, friction coefficients, noise and uncertainty;

- It relies directly on sensory information typically available during the grasp formation process, namely contact position and normals;

- The number of contacts employed in the grasp task is a control parameter and may vary during the grasp process. Each combination of contacts or fingers defines a distinct control alternative with unique opportunities to observe the run-time context (the interaction patterns between manipulator kinematics, sensor readings and the unknown object geometry in this thesis). The set of all control laws constitutes a basis for both grasp control and grasp observers.

3.2.1 Closed-Loop Grasp Control

Figure 3.2 depicts the closed-loop control law proposed. The plant consists of a k-contact grasp configuration on an unknown object geometry. All contacts forces are assumed to be unitary and oriented along the inward directed surface normal. During grasp formation, the contacts are modeled as frictionless point contacts. Consequently, the grasp configurations that result are generally composed of uniform contact force distributions and rely minimally on friction. Both conditions lead toward generally robust grasps, not necessarily toward optimal grasps with respect to other performance measures.

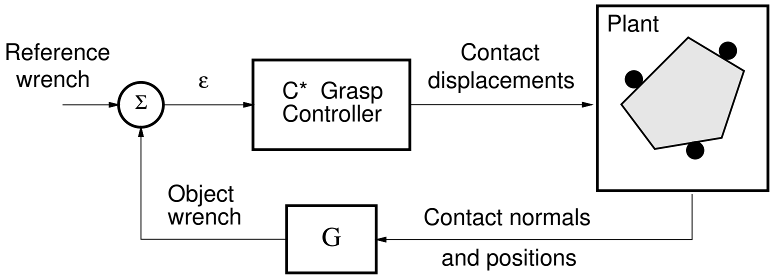

Figure 3.2. Grasp synthesis as a control problem.

The control objective is to bring the current grasp configuration closer to the force closure condition, or in other words to construct a null space in the grip transform $\mathbf{G}$. A grasp configuration is considered statically stabilizable if the grip transform, $\mathbf{G}$, contains a null space consisting of wrenches derived from strictly positive-sense contact forces. If there exists a contact wrench that can be written as a positive linear combination of other contact wrenches, $\boldsymbol{\omega}_i = \sum_{j \neq i}^{k} \boldsymbol{\omega}_j$, then the object may be *squeezed* within a null space of the grip transform without applying a net wrench to the object. It follows that the condition $\sum_{i=1}^{k} \boldsymbol{\omega}_i = 0$ is a necessary condition for stable grasp configurations, and the corresponding control reference is a zero net wrench.

Sensor evidence in the form of contact positions and normals is used to construct the instantaneous grip transform, $\mathbf{G}$, with which contact forces are transformed into object frame wrenches. This local characterization captures the ability of the object surface to carry contact forces and to generate object frame wrenches. Moreover, the local characterization contains enough information to derive differential contact displacements descending a local gradient in a quadratic wrench residual toward the reference wrench.

3.3 Grasp Control Basis

The objective of the grasp controller in Figure 3.2 is to minimize the squared wrench residual ϵ:

$$\begin{aligned} \epsilon &= \boldsymbol{\rho}^T \mathbf{M} \boldsymbol{\rho} \\ &= \left(\frac{1}{k}\sum_{i=1}^{k} \boldsymbol{\omega}_i\right)^T \mathbf{M} \left(\frac{1}{k}\sum_{i=1}^{k} \boldsymbol{\omega}_i\right), \end{aligned} \tag{3.3}$$

where $\boldsymbol{\rho}$ is the wrench residual and expresses the net wrench over k contacts, $\mathbf{M}$ is any normalization matrix that is positive definite, and $\boldsymbol{\omega}_i$ is the object frame wrench resulting from the i^{th} contact force. In the following, $\mathbf{M}$ is defined as the identity matrix $\mathbf{I}$; no weighting is assumed.

The squared residual, ϵ, is minimized by following the negative gradient of ϵ with respect to the contact configuration vector, $\mathbf{q}$. One can compute $\frac{\partial \epsilon}{\partial q_j}$:

$$\begin{aligned} \frac{\partial \epsilon}{\partial q_j} &= 2\left(\frac{1}{k}\sum_{i} \boldsymbol{\omega}_i\right)^T \frac{\partial\left(\frac{1}{k}\sum_i \boldsymbol{\omega}_i\right)}{\partial q_j} \\ &= 2\left(\frac{1}{k}\sum_{i} \boldsymbol{\omega}_i\right)^T \frac{\partial \boldsymbol{\omega}_j}{\partial q_j}\frac{1}{k} \\ \frac{\partial \epsilon}{\partial q_j} &= \frac{2}{k}\left(\frac{\partial \boldsymbol{\omega}_j}{\partial q_j}\right)^T \boldsymbol{\rho}, \end{aligned} \tag{3.4}$$

where $\frac{\partial \boldsymbol{\omega}_j}{\partial q_j}$ is the partial derivative of the j^{th} contact wrench with respect to the coordinate q_j. In vector notation

$$\frac{\partial \epsilon}{\partial \mathbf{q}} = \frac{2}{k}\begin{bmatrix} (\frac{\partial \boldsymbol{\omega}_1}{\partial \mathbf{q}_1})^T \\ (\frac{\partial \boldsymbol{\omega}_2}{\partial \mathbf{q}_2})^T \\ \vdots \\ (\frac{\partial \boldsymbol{\omega}_n}{\partial \mathbf{q}_n})^T \end{bmatrix} \boldsymbol{\rho}$$

$$= \frac{2}{k}\frac{\partial \mathbf{W}^T}{\partial \mathbf{q}}\boldsymbol{\rho}. \tag{3.5}$$

Minima in Equation 3.3 (configurations where $\frac{\partial \epsilon}{\partial \mathbf{q}} = 0$) are potential grasp solutions. These minima can occur when $\boldsymbol{\rho} = 0$ and when $\frac{\partial \mathbf{W}^T}{\partial \mathbf{q}}\boldsymbol{\rho} = 0$. The first case clearly corresponds to a null space in the grip transform and therefore meets our criteria for a statically stabilizable grasp. The second case can arise in a number of circumstances. Some of the equilibrium grasp configurations for which $\boldsymbol{\rho} \neq 0$ can be stabilized by contact force modulation or by friction forces. Still others in the general formulation of Equation 3.5 are truly local minima and must be avoided.

Both $\frac{\partial \mathbf{W}}{\partial \mathbf{q}}$ and $\boldsymbol{\rho}$ in Equation 3.5 require an expression for the contact wrench as a function of contact coordinates. This function depends directly on the geometry of the object. To build controllers applicable to a range of object geometries, two local surfaces types are considered. Force residual control is derived from a model of contact wrenches on a sphere (Section 3.3.1). Moment residual control is derived from a model of contact wrenches on an infinite plane defined by the contact normal (Section 3.3.2). In conjunction with Equation 3.5, these models yield artificial potentials with unique minima and represent two local surface geometries as a constructive basis for approximating a range of object geometries. The next sections define the control formulation and mechanisms for composing them into correct grasp controllers.

3.3.1 Residual Force Control

The residual force controller, $\mathcal{C}_f$, is derived from a model of how object frame contact wrenches vary with contact coordinate on the surface of a sphere as shown in Figure 3.3. If θ and ϕ are the angular coordinates of a single contact on the surface of a sphere, then the corresponding object frame wrench, $\mathbf{w}(\theta, \phi)$, can be written as:

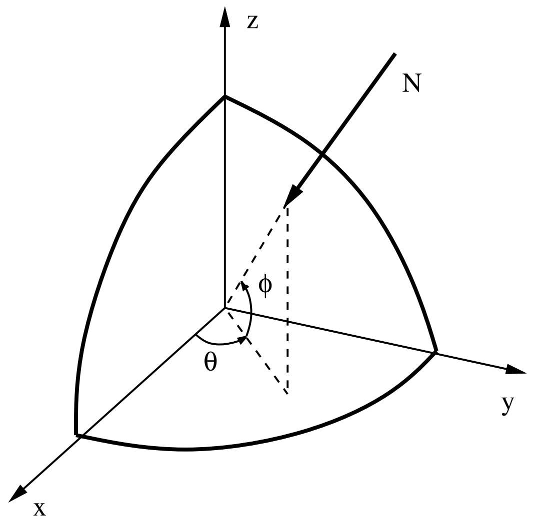

Figure 3.3. Local surface model for the residual force potential.

$$\mathbf{w}(\theta,\phi) = \begin{bmatrix} w_{fx} \\ w_{fy} \\ w_{fz} \\ w_{mx} \\ w_{my} \\ w_{mz} \end{bmatrix} = \begin{bmatrix} -\cos(\theta)\cos(\phi) \\ -\sin(\theta)\cos(\phi) \\ -\sin(\phi) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{3.6}$$

so that

$$\begin{bmatrix} w_{\theta_i} & w_{\phi_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{w}}{\partial \theta_i} & \frac{\partial \mathbf{w}}{\partial \phi_i} \end{bmatrix} = \begin{bmatrix} \sin(\theta_i)\cos(\phi_i) & \cos(\theta_i)\sin(\phi_i) \\ -\cos(\theta_i)\cos(\phi_i) & \sin(\theta_i)\sin(\phi_i) \\ 0 & -\cos(\phi_i) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.7}$$

3.3.2 Residual Moment Control

Figure 3.4 illustrates the geometry from which the residual moment potential is derived. The residual moment controller, $\mathcal{C}_m$, minimizes the moment residual by

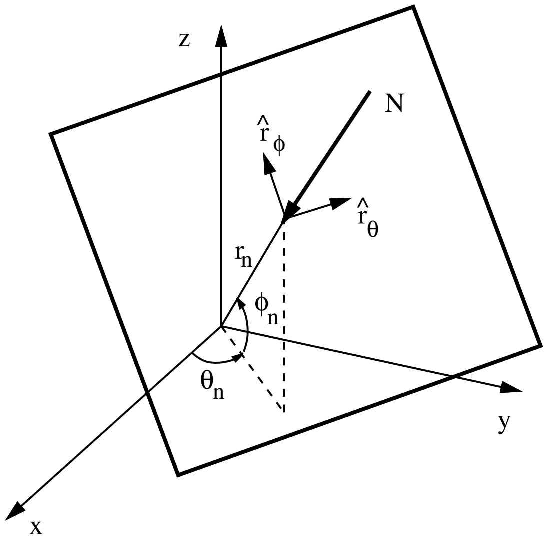

Figure 3.4. Local surface model for the residual moment potential.

controlling the position of the contact (r_θ, r_ϕ) on a locally planar surface defined by parameters (r_n, θ_n, ϕ_n).

Contact forces in the object frame are constant everywhere on the plane and moments, $\mathbf{m} = \mathbf{r} \times \mathbf{f}$, vary linearly in the plane, passing through zero when $r_\theta = 0$ and $r_\phi = 0$. The resulting wrench model is:

$$\mathbf{w}(\theta, \phi) = \begin{bmatrix} w_{fx} \\ w_{fy} \\ w_{fz} \\ w_{mx} \\ w_{my} \\ w_{mz} \end{bmatrix} = \begin{bmatrix} -\cos(\theta_n)\cos(\phi_n) \\ -\sin(\theta_n)\cos(\phi_n) \\ -\sin(\phi_n) \\ -r_\theta \sin(\phi_n)\cos(\theta_n) + r_\phi \sin(\theta_n) \\ -r_\theta \sin(\phi_n)\sin(\theta_n) - r_\phi \cos(\theta_n) \\ r_\theta \cos(\phi_n) \end{bmatrix}. \tag{3.8}$$

This model yields correct instantaneous wrenches (and therefore wrench residuals) for all objects provided that (r_n, θ_n, ϕ_n) are known or measurable. The moment residual controller assumes that these parameters are constant in the neighborhood of contact i, so that

$$\begin{bmatrix} w_{r_{\theta_i}} & w_{r_{\phi_i}} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{w}}{\partial r_{\theta_i}} & \frac{\partial \mathbf{w}}{\partial r_{\phi_i}} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -\cos(\theta_{ni})\sin(\phi_{ni}) & \sin(\theta_{ni}) \\ -\sin(\theta_{ni})\sin(\phi_{ni}) & -\cos(\theta_{ni}) \\ \cos(\phi_{ni}) & 0 \end{bmatrix}. \quad (3.9)$$

The gradient descent procedure resulting from the combination of the above wrench model and Equation 3.5 responds exclusively to moment residuals.

We have shown [11, 25] that artificial potentials defined by Equations 3.6 and 3.8 are unimodal so that gradient descent (Equations 3.7 and 3.9) yield unique grasp distributions in the sphere and the infinite plane, respectively. However, on general closed surfaces there are potentially many local minima that are not grasp solutions. Figure 3.5 shows the squared force and moment residuals for a planar square shape and 2 contacts. Those plots are obtained by fixing the black contact at $\theta_0 = -\pi/6$, and moving the white contact over the range $0 \leq \theta_1 < 2\pi$. Figure 3.5(b) identifies four minima (exactly one per face) for this object in the moment residual and Figure 3.5(c) illustrates the single minimum in the force residual. Figure 3.5 suggests that the globally optimal configuration may be reached by first minimizing the force residual (i.e., performing gradient descent on the residual force potential), and then minimizing the moment residual. For the square shape, at least, this strategy will produce the globally optimal solution. The next section shows that this is true in general for regular, convex polygons.

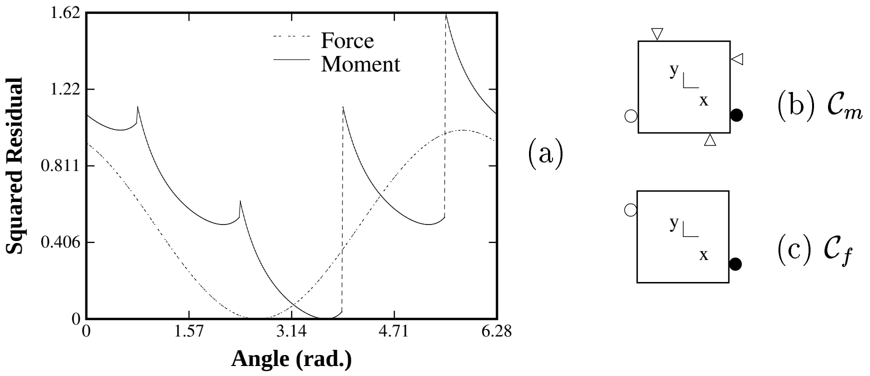

Figure 3.5. (a) The force and moment residuals for the planar square shape. (b) Attractors in the moment residual (one per face). (c) Unique attractor in the force residual.

3.3.3 Composition of $\mathcal{C}_f$ and $\mathcal{C}_m$

A simple composition rule involving $\mathcal{C}_f$ and $\mathcal{C}_m$ yields a grasp controller that is provably optimal for the class of regular convex polygons, for grasp configurations with two and three contacts. The grasp controller $\mathcal{C}^*$ is defined by the rule

$$\mathcal{C}^* = \begin{cases} \mathcal{C}_f & \text{if } (\epsilon_f > \tau) \\ \mathcal{C}_m & \text{otherwise,} \end{cases} \tag{3.10}$$

where ϵ_f is the instantaneous force residual, and τ is an appropriately chosen convergence threshold; for the purposes of the proof detailed in this section $\tau = 0$; notice that the switching event is defined by convergence in $\mathcal{C}_f$. By construction, $\mathcal{C}_f$ yields a unique $\epsilon_f = 0$ grasp configuration in the plane for $k = 2$ and 3 consisting of evenly spaced contacts separated by $2\pi/k$ radians. This result generalizes in a straightforward manner to 3D geometry. When $\mathcal{C}_m$ is activated, $\mathcal{C}_f$ configurations are modified to minimize ϵ_m. The proof that $\mathcal{C}^*$ will always find the two- and three-contact grasp configurations with zero wrench residual for regular convex polygons consists of:

1. verifying that the solutions generated by $\mathcal{C}_f$ are the closest approximation to a zero force residual on regular n-gons;

2. the determination of the range of contact configurations in which $\mathcal{C}_f$ solutions exist;

3. verifying that there exists a unique $\mathcal{C}_m$ solution within that range, and

4. demonstrating that the moment residual controller $\mathcal{C}_m$ does not increase the observed force residual; a non-zero force residual would activate the $\mathcal{C}_f$ controller and potentially disrupt the actions of the moment residual controller. If a $\mathcal{C}_m$ solution is reached, then it is the global minimum in the squared wrench residual since it minimizes both force and moment residuals on the n-gon.

Theorem 3.3.1 (Correctness) *For 2 and 3 contacts on regular, convex polygons, $\mathcal{C}^*$ yields the globally minimum wrench residual.*

Figure 3.6 depicts a generic polygon and defines several useful geometric properties. For convenience, the edges of the polygon are numbered from 0 to $n-1$ and we assume that contact 0 is positioned on edge 0. If we further assume that contact 0 is

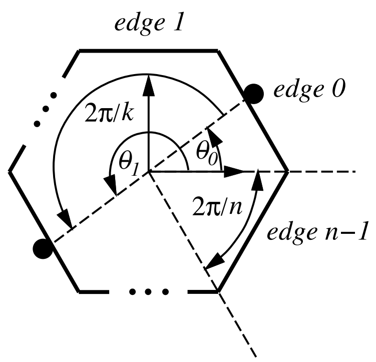

Figure 3.6. 2 contacts ($k = 2$) on an n-sided regular polygon.

positioned at $\theta_0 = 0$, then the coordinate of the i^{th} contact following convergence in $\mathcal{C}_f$ is written:

$$\theta_i = i * \frac{2\pi}{k} \qquad 0 \leq i \leq k-1, \quad for\ k = 2, 3.$$

When k contacts are placed on the perimeter of the regular, n-sided polygon with $\theta_0 = 0$, the remaining $k-1$ contacts occur at discrete, $2\pi/nk$ intervals on their respective edges. Using these conventions, we can establish the conditions necessary for equilibrium in $\mathcal{C}_m$.

Consider the $\phi = 0$ section of a regular polyhedra. In this case, the $\mathcal{C}_m$ control actions are a function of the r_θ coordinates (see Equation 3.5):

$$\frac{\partial \epsilon}{\partial \mathbf{r}_\theta} = -\frac{2}{k}\mathbf{W}_{r_\theta}^T \boldsymbol{\rho}$$

where:

$$\begin{aligned} (\mathbf{W}_{r_\theta}^T)_i &= [0\ 0\ 0\ 0\ 0\ 1] \\ \boldsymbol{\rho} &= \sum_{i=0}^{k-1} [-\cos(\theta_n)\ \ -\sin(\theta_n)\ 0\ 0\ 0\ r_\theta]_i \end{aligned}$$

Under these conditions:

$$\begin{bmatrix} \dfrac{\partial \epsilon}{\partial r_{\theta_0}} \\ \dfrac{\partial \epsilon}{\partial r_{\theta_1}} \\ \vdots \\ \dfrac{\partial \epsilon}{\partial r_{\theta_{(k-1)}}} \end{bmatrix} = \frac{2}{k} \begin{bmatrix} \sum_{i=0}^{k-1} r_{\theta_i} \\ \sum_{i=0}^{k-1} r_{\theta_i} \\ \vdots \\ \sum_{i=0}^{k-1} r_{\theta_i} \end{bmatrix}, \tag{3.11}$$

where r_{θ_i} are the coordinates of contact i on its respective contact plane. The gradient at each contact is identical, so that configuration changes produced by $\mathcal{C}_m$ move all the contacts in the same direction. In other words, the actions generated by $\mathcal{C}_m$ preserve $\mathcal{C}_f$ equilibria ($\epsilon_f = 0$) precisely; $\mathcal{C}_m$ operates strictly within the null space of $\mathcal{C}_f$ [11]. This property guarantees that *limit cycles* cannot exist and therefore no spurious minima can be introduced. Limit cycles are defined as closed trajectories in state space; they characterize the trajectories of dynamically stable systems in which the system visits the same set of states in a bounded region of the state space over and over again (Liapunov stability).

According to Equation 3.11

$$\sum_{i=0}^{k-1} r_{\theta_i} = 0 \tag{3.12}$$

identifies minima in ϵ_m. Two possible conditions arise.

The number of faces n is an integer multiple of k In this case, all the contacts fall on vertices (since $\theta_0 = 0$) and the $\mathcal{C}_f$ grasp configuration may be rotated through the entire angle subtended by an edge, $\frac{2\pi}{n}$, without any of the contacts leaving an edge. In this circumstance,

$$\sum_{i=0}^{k-1} r_{\theta_i} = \sum_{i=0}^{k-1} \tan(-\frac{\pi}{n} + \psi) \Rightarrow \tan(-\frac{\pi}{n} + \psi) = 0$$

where ψ is the angle through which the $\mathcal{C}_f$ configuration is moved by the $\mathcal{C}_m$ controller. It is clear that $\psi = \frac{\pi}{n}$ is a solution and that it remains within the range specified.

The number of faces n is not an integer multiple of k In this case, k contacts are distributed uniformly from $[-\frac{\pi}{n},\ +\frac{\pi}{n} - \frac{2\pi}{kn}]$ at $\frac{2\pi}{kn}$ intervals on their respective edges. In this circumstance,

$$\sum_{i=0}^{k-1} r_{\theta_i} = \sum_{i=0}^{k-1} \tan(-\frac{\pi}{n} + \frac{2\pi i}{kn} + \psi) = 0$$

The *tangent* function is unique and odd over this domain so that this condition is equivalent to requiring that the sum of the relative contact angles be zero:

$$0 \quad = \quad \sum_{i=0}^{k-1} (-\frac{\pi}{n} + \frac{2\pi i}{kn} + \psi)$$

$$
\begin{aligned}
&= k(-\frac{\pi}{n} + \psi) + \frac{2\pi}{kn}(\frac{k(k-1)}{2}) \\
\psi &= \frac{\pi}{kn}.
\end{aligned}
$$

Again, this solution is within the specified range. As in the previous case, this ϵ_m minimum is exactly midway through the range permitted. □

3.4 Control Composition as an Optimal Control Problem

The composition rule expressed by Equation 3.10 is only one correct alternative out of many possibilities for this class of objects. It assumes that only partial information about the object is available: the moment residual controller requires position and normal of each contact, and the force residual controller requires only contact positions. Although this is probably the most accurate scenario for control of real-world complex systems, one can do better if complete or very good models about the problem are available, through the use of optimal control techniques.

In this section, the composition problem is solved optimally, with respect to time. The objective is to demonstrate once again the utility of the grasp controllers discussed previously – because such controllers are available, the optimal control problem can be formulated. Also, portraying the grasping of an object as an optimal control problem serves to illustrate yet another solution strategy to the grasp synthesis problem.

3.4.1 Problem Description

In order to keep the presentation clear, a small and somewhat constrained version of the grasp synthesis problem was chosen to illustrate the composition procedure. There is no impediment to using the technique discussed below to the unrestricted grasping problem, involving grasps with n moving contacts. For the general control composition problem, the only requirement is differentiability of the metric being optimized and the constituent controllers. See [8] for a introduction to Optimal Control Theory.

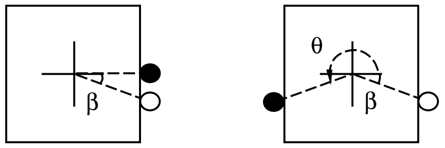

Figure 3.7. (a) Initial configuration (b) Final configuration.

The problem consists in finding out how to change the activation of both the force and moment residual controllers, such as to minimize the number of control steps required to take the system from the initial to the final configuration (see Figure 3.7). For simplicity, we will consider that just one contact moves over the object surface, while the other stays fixed at its initial position. Also, we will assume that the activation for the moment residual controller is constrained to be $(1 - \alpha)$, where $\alpha \in [0, 1]$ is the activation for the force residual controller.

3.4.2 Problem Formulation

Maximize

$$V = \int_0^T -\epsilon_m(\theta)\, dt$$

Subject to

$$\begin{aligned} \dot{\theta}(t) &= -k\left(\alpha \frac{d\,\epsilon_f(\theta)}{d\,\theta} + (1-\alpha)\frac{d\,\epsilon_m(\theta)}{d\,\theta}\right) \\ \theta(0) &= 0.0, \quad \theta(T)\ unknown \\ \alpha &\in [0, 1] \end{aligned}$$

In the formulation above, ϵ_m is the squared wrench residual (Equation 3.3). If the task is to achieve null grasp of a square with two contacts then

$$\epsilon_m(\theta) = E + \frac{(2\tan\theta - 1)^2}{16},$$

where E is a constant corresponding to the force error incurred in positioning the moving contact in the various faces of the square. For the face corresponding to $\theta = 0$, $E = 1$; for faces corresponding to $\theta = \pi/2$ and $\theta = 3\pi/2$, $E = 0.5$, and for the face corresponding to $\theta = \pi$, $E = 0$.

Similarly, ϵ_f is the force residual error, expressed as

$$\epsilon_f(\theta) = 0.5 * (1 + \cos(\theta - \beta)),$$

β being the angle for the fixed contact. In the development that follows, the first and second derivatives of ϵ_m and ϵ_f with respect to θ will be required:

$$\begin{aligned}
\frac{d\,\epsilon_m}{d\,\theta} &= \frac{2\tan\theta - 1}{4\cos^2\theta} \\
\frac{d^2\epsilon_m}{d\,\theta^2} &= \frac{8 + 2(\tan\theta - 1)\sin 2\theta}{16\cos^4\theta} \\
\frac{d\,\epsilon_f}{d\,\theta} &= -0.5\sin(\theta - \beta) \\
\frac{d^2\epsilon_f}{d\,\theta^2} &= -0.5\cos(\theta - \beta)
\end{aligned}$$

The Hamiltonian function for the problem can now be expressed as

$$H(t, \theta, \alpha, \lambda) = -\epsilon_m(\theta) + \lambda(t)\frac{d\,\theta}{d\,t}.$$

The costate variable λ is an auxiliary variable required by the method [8]. The Maximum Principle conditions are

$$H(t, \theta, \alpha^*, \lambda) \;\geq\; H(t, \theta, \alpha, \lambda), \quad \forall t \in [0, T] \tag{3.13}$$

$$\dot{\theta}(t) = \frac{\partial H}{\partial \lambda}$$

$$
\begin{aligned}
&= -k\left(\alpha\frac{d\,\epsilon_f(\theta)}{d\,\theta} + (1-\alpha)\frac{d\,\epsilon_m(\theta)}{d\,\theta}\right) && (3.14)\\
\dot{\lambda}(t) &= -\frac{\partial H}{\partial \theta} = \frac{d\,\epsilon_m}{d\,\theta} \\
&+ \lambda(\alpha\frac{d^2\epsilon_f}{d\,\theta^2} + (1-\alpha)\frac{d^2\epsilon_m}{d\,\theta^2}) && (3.15)\\
\lambda(T) &= 0
\end{aligned}
$$

Equation 3.13 means that the resulting Hamiltonian H is to be maximized over time, and $\alpha^*(t)$ is the activation policy that maximizes it. Notice that selecting α such that $\frac{\partial H}{\partial \alpha} = 0$ does not cover all cases, because of the constrained range of values α can assume — it might be the case that α^* corresponds to one value in the limits of the closed interval $[0, 1]$, for example. And this is the case for the problem we are interested in, because $\frac{\partial H}{\partial \alpha}$ does not depend on α:

$$
\frac{\partial H}{\partial \alpha} = -\lambda(\frac{d\,\epsilon_f}{d\,\theta} - \frac{d\,\epsilon_m}{d\,\theta}).
$$

Therefore, the time optimal policy for this particular problem implies switching α from 0 to 1 and back. This optimal pattern can be easily computed by integrating Equations 3.14 and 3.15 over time:

1. Initialize λ;
2. Compute $\frac{\partial H}{\partial \alpha}$; if $\frac{\partial H}{\partial \alpha} > 0$, select $\alpha = 1$; otherwise, make $\alpha = 0$;
3. Compute $\theta_{t+1} = \dot{\theta}_t \Delta t$;
4. Compute $\lambda_{t+1} = \dot{\lambda}_t \Delta t$;
5. Check for convergence on θ; return to (2) above if not converged.

3.4.3 Results

In all results, we used $\lambda(0) = 5$, and $\Delta t = 0.01$. The method is not sensitive to the values chosen for these parameters; an ample range of values can be used.

However, selecting a low initial value for λ may result in solutions corresponding to local minima. The fixed contact is stationed at $\beta = -0.46365$ radians.

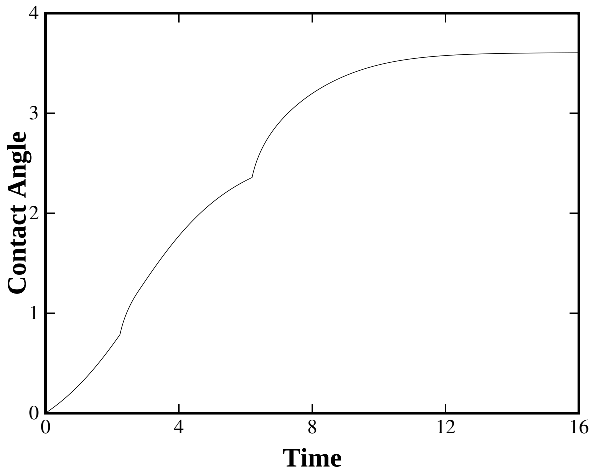

Figure 3.8. Evolution of θ with time.

Figure 3.8 shows how θ varies with time. Starting from $\theta(0) = 0.0$, the contact progresses steadily to the optimal solution $\theta(T) = 3.6052$ radians. Figure 3.9 shows the optimal activation pattern for α as a function of time. Figure 3.9 also displays the corresponding areas on the object surface where the force residual controller was selected; for the white areas, $\alpha = 1$. Gray areas correspond to where in the object the moment residual controller was selected ($\alpha = 0$).

3.4.4 Discussion

Notice that as long as the metric chosen to be optimized is independent of the intermediate activations $\alpha(t)$, the time optimal activation policy will always be a *bang-bang* type solution, no matter the object being grasped.

Another interesting aspect is the fact that this approach also yields an effective grasp controller: just by evaluating $\frac{\partial H}{\partial \alpha}$ one can decide on the next activation α. What remains to be determined is how robust is this controller with respect to object geome-

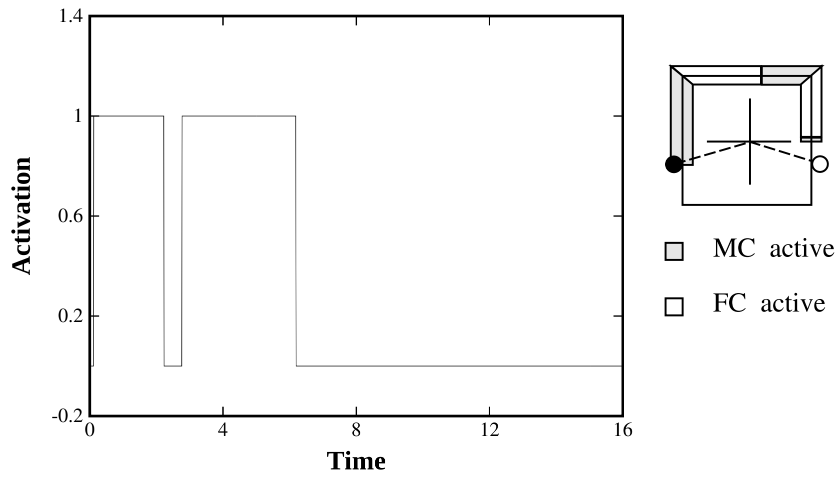

Figure 3.9. Time optimal activation pattern during the grasp. $\alpha = 1$ corresponds to full activation of the force residual controller, while $\alpha = 0$ corresponds to full activation of the moment residual controller. In the inset, gray areas correspond to where in the object's surface $\alpha = 0$; white areas mark the areas where $\alpha = 1$.

tries, and uncertainties on actuation and sensory information. Because this approach essentially involves integration over time, it is not clear how to revise the control history incrementally: presumably, a object model is being continuously or periodically updated, and sensory data may yield discontinuities in the state information. How to accommodate such changes in the control scheme is yet to be determined.

3.5 Experiments with Physical Manipulators

This section documents the experiments realized in the Laboratory for Perceptual Robotics in the time span from roughly 1993 to 1998. These experiments had broad objectives, and they speak to the expressive power of the control basis. The commentary that follows emphasizes the aspects related to grasping. The experiments are described in chronological order.

3.5.1 Integrated Reaching and Grasping

The $\mathcal{C}^*$ controller has been used in the context of pick-and-place tasks, where several control modules were integrated into a system capable of performing autonomous, collision-free reaching and grasping of wood blocks placed over a table at arbitrary positions. Figure 3.10 shows the experimental setup. Grasps are performed by a Utah/MIT hand, attached to a GE P50 robot arm. The objects consist of wood blocks approximately 20 cm in height and with maximum diameters of 5 cm (for the rectangle). All objects are regular and irregular prisms with polygonal bases ranging from triangles to hexagons.

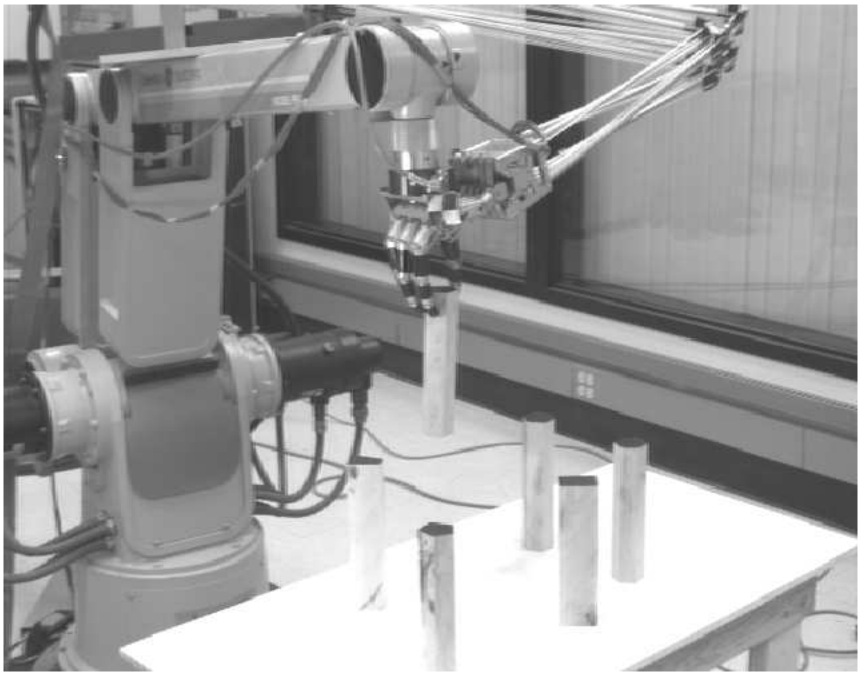

Figure 3.10. Experimental setup, showing the GE P50 robot arm and the Utah/MIT hand carrying an object with a top grasp.

The pick-and-place task consists of grasping a wood block designated by the user in an image from an overhead camera. Following the grasp, the block is moved to another user specified position on the cluttered table. Figure 3.11 depicts the execution of a complete pick-and-place cycle.

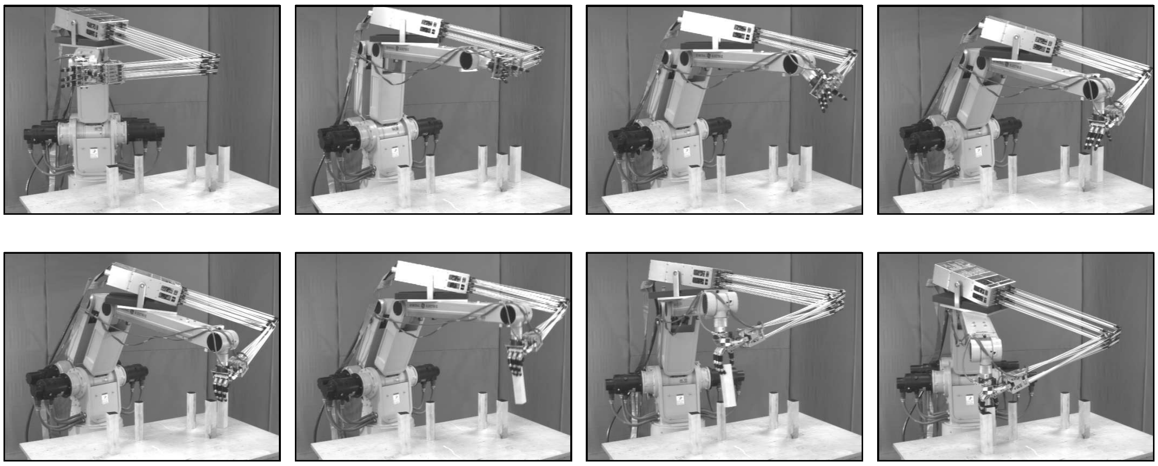

Figure 3.11. The pick-and-place task.

All 21 degrees of freedom in the system are managed autonomously at run time according to the finite state machine depicted in Figure 3.12. In the diagram, double circles denote states in which the system waits for user input which is supplied via a visual interface,i.e. grasp *that* object and put it *there*. Each of the remaining states is labeled $\Phi_i|_\tau^\sigma$, where σ is the input state set for the controller Φ_i, and τ is the output state set including all effectors enlisted by Φ_i. In this task, three basis controllers were enlisted: Φ_0, the harmonic function path controller; Φ_1 the contact configuration controller; and Φ_2 is the kinematic conditioning controller [23]; Φ_1 is an abstraction for the $\mathcal{C}^*$ controller described in this chapter. many instances of these control types are generated by associating unique resource parameterizations of the three basic types.

The pick-and-place cycle starts as the user specifies a target object by clicking on an overhead visual image of the robot's workspace. The selection activates Φ_0, the harmonic function path controller and the hand/arm system navigates through the cluttered configuration space to the neighborhood of the specified object using the

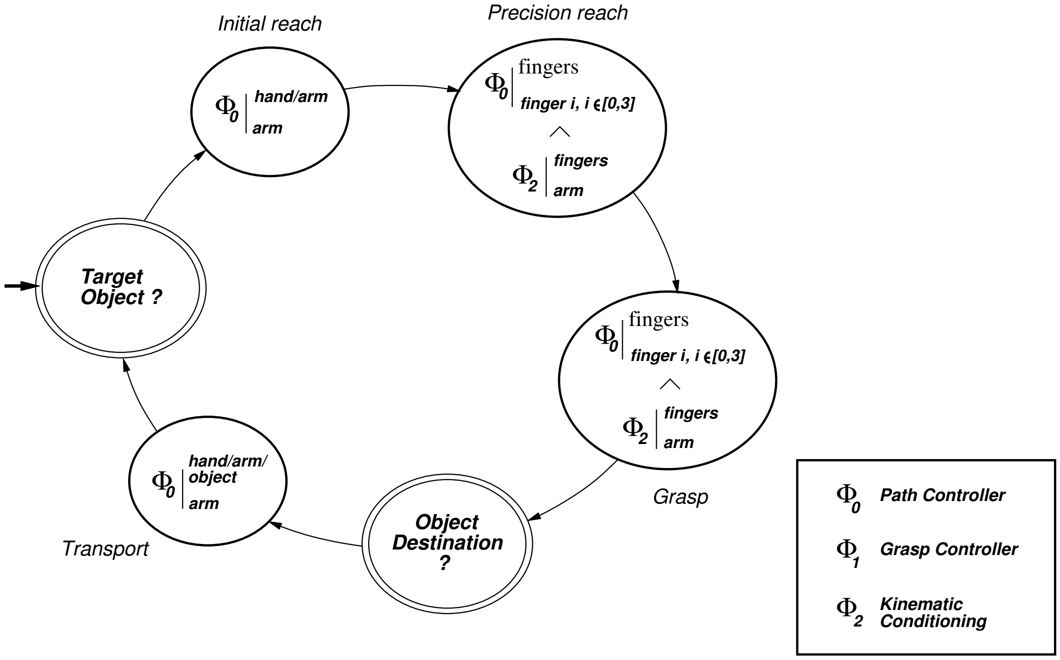

Figure 3.12. Finite state machine for pick-and-place task.

robot arm's five degrees of freedom (DOF) exclusively. The remaining 16 DOFs are held motionless at this stage.

Following convergence, the automaton enters the "Precision reach" (see Figure 3.12) stage, activating multiple instances of Φ_0 (one for each finger) and Φ_2 concurrently, causing the finger tips to approach the object surface while at the same time adjusting the posture of the arm.

At the end of the "Precision reach" stage, the object surface is within the workspace of the fingers and the arm is in a configuration that locally maximizes the kinematic condition of the fingers. The grasp controller, Φ_1, and the kinematic conditioning controller, Φ_2, are then activated concurrently to form the grasp while optimizing the kinematic condition of the hand. The coupled control activation propagates constraints imposed by the grasping task, through the hand posture, to the arm posture. This control sequence concludes with the arm and fingers in a grasp configuration on the object where the fingers apply adequate grasp forces. Figure 3.13 shows examples of convergent contact configurations: a top grasp of an object with pentagonal base and a lateral grasp of an object with rectangular base, from different perspectives.

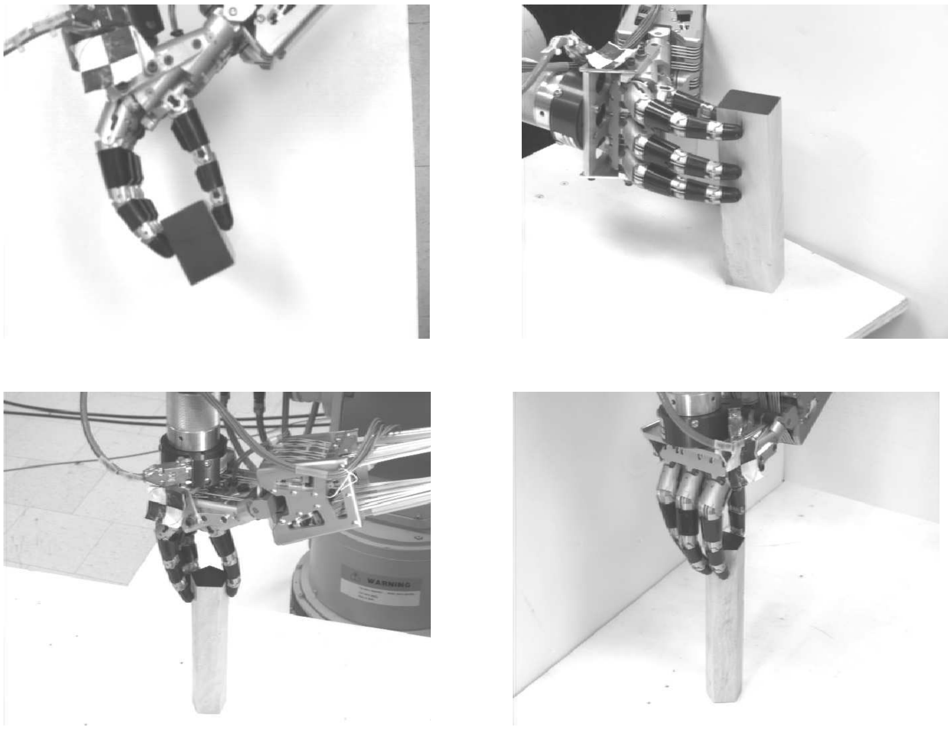

Figure 3.13. Grasps synthesized by system. Top row shows two perspectives of the lateral grasp of a rectangular prism, while bottom row shows the top grasp of a pentagonal prism.

Once the object is grasped, the user designates a target location within the cluttered workspace. The controller associated with the "Transport" stage, Φ_0, is activated and the object is carried to its destination and placed there, where the fingers release it. Overall, system performance is very robust, even in the presence of woes that plague every real world implementation: imprecise object models, discretization errors, calibration drifts, and kinematic model imprecision. Over the course of a number of experiments, no structural problems (i.e., deficits in the grasp control basis) were observed. Failures would result after a sequence of mishaps, typically starting with a bad visual model of the object geometry, followed by an imprecise grasp action, that ultimately would cause the object to wobble and fall when released at the target location. Most of the time, imperfect models would not prevent the system from executing a secure grasp and delivery.

3.5.2 Finger Gaits and Manipulation

The next series of experiments involved a series of finger gaits, designed as manipulation strategies that are robust with respect to sensor noise, portable across manipulators, and designed and implemented rapidly. These manipulation strategies are expressed by sequences of control activations, each one involving a particular set of sensors and manipulation resources. The working hypothesis was that the control basis tacitly encodes domain expertise that will facilitate the planning and execution of manipulation sequences. Moreover, it is proposed that the control specifications are *declarative* in nature, that elements of the control basis reflect domain or context neutral control strategies – largely independent of the task geometry and robot kinematics. A corollary to this proposition holds that policies may be transferred to other "equivalent" resources to accomplish the same kinds of tasks using the same general approach. This is very ambitious hypothesis, but is nevertheless an important experimental objective for intelligent robots.

The best example in this series of experiments is the object rotation task described in [23]. The task objective is to rotate an object within the hand while maintaining object stability at all times. The corresponding control automaton is shown in Figure 3.14. At the outset, the hand holds the object with a stable four-fingered grasp. From this point on, the agent lifts finger i from the object surface and moves it to a configuration such that the grasp configuration formed by fingers $\{i,\ (i+2) \bmod 4,\ (i+3) \bmod 4\}$ is a stable three-fingered grasp configuration, by applying the Φ_1 controller. The resulting grasp positions the fingers in a suboptimal kinematic configuration due to the geometry of the task. This is compensated by the activation of Φ_2, the kinematic conditioning controller. The resulting changes in the arm and the rotational axis of the object lead to the object rotation shown in Figure 3.15. At the next state, the agent increments i (or decrements it, for clockwise rotations) and the process repeats itself.

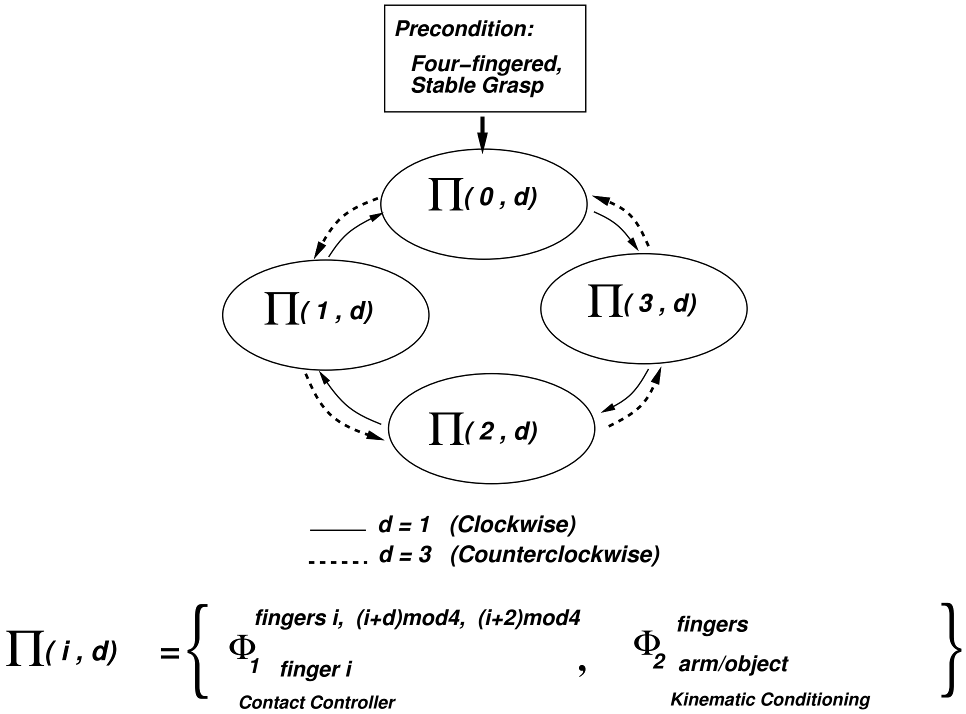

Figure 3.14. Finite State Machine for Rotation Task

This example further demonstrates the reusability of the controllers. The same controllers as in the pick-and-place task are used here to accomplish a conceptually very different task. Moreover, dynamically rebinding this control basis to degrees of freedom results in solutions that are robust with respect to task geometry and robot kinematics [27].

3.6 Summary

Many factors influence the outcome of a grasp operation, some of which cannot be controlled by the grasping agent as for instance object geometry and its exact

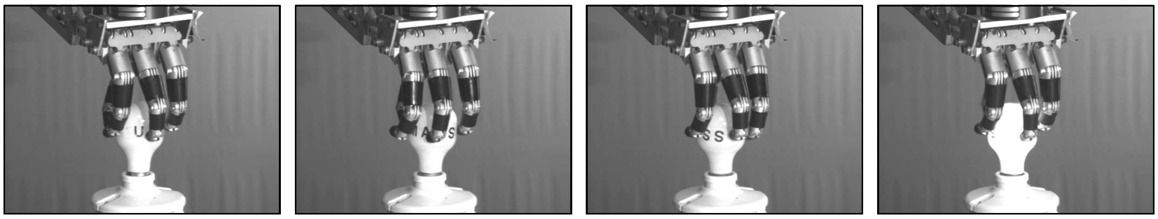

Figure 3.15. Utah/MIT hand Screwing in a Light Bulb

orientation with respect to the manipulator. However, the agent does have control over which sensory and kinematic resources to deploy in each context, as well as the overall control strategy. The overall quality of the solutions can be significantly improved if the grasping agent were able to reject or somehow convert poor solutions into good ones by selecting the best grasp primitive for each context.

In order for the agent to compare the relative merits of alternative control strategies across contexts, it must encode each the grasp context in terms of some representation scheme; moreover, it must be able to retrieve the representation given the history of sensory readings. Three non-trivial problems must be addressed before context-dependent grasp strategies can be created:

- How can one describe the grasp process in terms of state variables?
- How can one retrieve state information given sensory readings?
- How does one build a map between grasp state and control primitives?

The agent must also be able to experiment with the different grasp primitives in each context and select the best control strategy for each context. The representation and retrieval of grasp context information is discussed in the next chapter; Chapter 5 deals with the synthesis of context-dependent grasp strategies.

CHAPTER 4

ON-LINE STATE ESTIMATION FOR GRASP CONTROL DECISIONS

Many factors affect the outcome of a grasp operation: object shape and scale, friction between object surface and fingers, kinematic constraints, sensory information available, and the set of finger surfaces employed in the grasp operation. This work proposes that *grasp state* is effectively expressed in the transient response of the situated grasp controller, $\mathcal{C}^*$.

Each grasp configuration can be associated with the observation $\mathbf{o} = [\epsilon\ \dot{\epsilon}]^T$, where ϵ is the squared wrench residual expressed by Equation 3.3 and $\dot{\epsilon}$ its derivative with respect to time. The sequence of configurations generated by $\mathcal{C}^*$ can be represented by a sequence of t observations $\mathcal{O}= \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t\}$. The evolution of the grasp process captured by the sequence of observations, $\mathcal{O}$, can be plotted in phase space, as shown in Figure 4.1.

The left panel in Figure 4.1 depicts the grasp dynamics for a typical two-contact grasp of an irregular triangle. Initially, the contacts are as shown in configuration (a). $\mathcal{C}^*$ drives the system to an intermediate grasp configuration as shown in (b), and the system converges to configuration (c). The contact configuration (c) at the attractor corresponds to a minimum of the squared wrench residual ϵ, in which the velocity $\dot{\epsilon}$ is approximately zero. Many other paths lead to the same attractor; in fact, the shaded region in Figure 4.1 represents the set of all states leading to the attractor corresponding to configuration (c). This region is termed the *basin of attraction* of the grasp attractor.

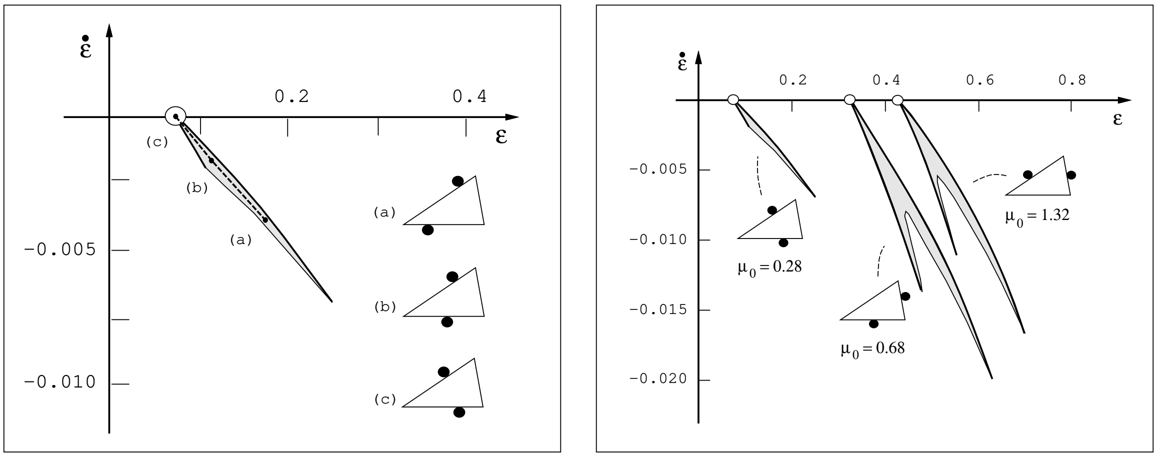

Figure 4.1. Left panel depicts the evolution of a grasp trial: (a) is the initial configuration, (b) an intermediate configuration, and (c) the convergent configuration. The complete phase space plot for the irregular triangle is shown on the right.

The grasp dynamics for a given object and set of resources (contacts) can be represented by a set of trajectories in phase space. Trajectories representing the same control context can be combined to form a *generative model* of prototypical system behavior. An illustration of all generative models for two-contact grasps involving the irregular triangle is depicted in Figure 4.1 (right panel). It has three attractors and basins of attraction, corresponding to the three possible combinations of two contacts and three distinct edges. The convergent grasp configurations and the respective quality indices μ_0 are also shown in Figure 4.1. The index μ_0 is the minimum friction coefficient required for grasp equilibrium, for $\dot{\epsilon} = 0$. It is a quality metric associated with each attractor, and it applies to all states in the basin of attraction.

4.1 State: A Formal Definition

The state representation proposed is based upon two discrete sets: the control basis $\Pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$ and the membership pattern vector $\mathbf{q} = [p_1 \;\; p_2 \;\; \ldots \;\; p_m]^T$, that indicates which models of a set of m generative models are consistent ($p_k = 1$) with the last observation and which are not ($p_k = 0$).

Given the observation vector $\mathbf{o}_t$ collected at time t under primitive π_i, the corresponding system state is denoted by the tuple $s_t = (\pi_i, \mathbf{q})$. The vector $\mathbf{q}$ expresses the models (or operational contexts) that are compatible with the observation $\mathbf{o}_t$ according to the agent's past experience. The design of the tuple $s_t = (\pi_i, \mathbf{q})$ attempts to encompass the information present in the sequence of observations $\mathcal{O}= \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t\}$; the resulting representation defines a discrete state space, that can be used in lieu of the continuous observation variables.

4.1.1 Control Basis

The control basis represents the agent's native control structure, expressed as collection of feedback control laws $\Pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$. Each component controller $\pi_i \in \Pi$ is a stable controller designed to address specific tasks, based on simplified, local models of the interaction between the contacts and object geometry. The controller $\pi_i \in \Pi$ optimizes a time-invariant, locally convex scalar potential $\phi_i(x)$, by driving the robot configuration $\boldsymbol{\theta}$ along the gradient $-\nabla\phi_i(\boldsymbol{\theta})$. The potential $\phi_i(\boldsymbol{\theta})$ measures the utility of the configuration $\boldsymbol{\theta}$ in robotic tasks such as grasping, navigation, posture, or force control. End-to-end tasks can be solved by combining elements of Π and engaging sets of sensorimotor resources, as proposed by Huber and Grupen [27].

The stability of controller π_i ensures that the system will achieve a configuration $\boldsymbol{\theta}$ such that $|\nabla\phi_i(\boldsymbol{\theta})| < \delta$, where $\delta > 0$ is a threshold for equilibrium. The satisfaction of $|\nabla\phi_i(\boldsymbol{\theta})| < \delta$ provides an event defining the termination of π_i. Attractors satisfying the convergence condition correspond to local minima in $\phi_i(\boldsymbol{\theta})$ due to the local nature of the underlying controllers.

Within the multifingered grasp domain, the potential $\phi(\boldsymbol{\theta})$ corresponds to the squared wrench residual ϵ, and $\boldsymbol{\theta}$ corresponds to the contact position and force vectors. The control basis Π is formed by several versions of the $\mathcal{C}^*$ controller corresponding to the many possible combinations of fingers or grasp resources. A hand with three fingers labeled $\{T, 1, 2\}$ (label for the thumb, first finger, and second finger) allows for four distinct contact subsets, assuming that at least two fingers are required to grasp the object:

$$\mathcal{R} = \{(T, 1), \;\; (T, 2), \;\; (1, 2), \;\; (T, 1, 2)\}. \tag{4.1}$$

Each element $r \in \mathcal{R}$ defines a new control law that generates characteristic/signature grasp dynamics and outcomes. The grasp control basis is defined as $\Pi = \{\mathcal{C}_r^* | \; r \in \mathcal{R}\}$.

4.1.2 Generative Models

A generative model describes how sequences of observations can be "generated" within a particular operational context. As the agent executes controller π_i, many distinct control contexts, characterized by distinct dynamics, can be observed. If an exhaustive set of generative models is available, then the agent may infer the set of models compatible with a given sequence of observations by computing and updating the likelihood that each model has generated the data observed. If $P(M_k)$ is the likelihood that model k generates the sequence of observations, and $P(\mathbf{o}|M_k)$ is the likelihood of observing $\mathbf{o}$ given that it is drawn from the distribution expressed by model k, then $P(M_k|\mathbf{o})$ (the likelihood that model k explains the data observed given observation $\mathbf{o}$) can be computed as

$$P(M_k|\mathbf{o}) = \frac{P(\mathbf{o}|M_k)P(M_k)}{\sum_{i=1}^{m} P(\mathbf{o}|M_i)P(M_i)},$$

assuming the existence of m generative models. The prior probability $P(M_k)$ can be determined using the data gathered during construction of the generative models. $P(M_k|\mathbf{o})$ can be interpreted as the belief that model k explains the observation $\mathbf{o}$.

The generative models compute the probability $P(\mathbf{o}|M_k)$; in this work, the corresponding conditional probability distributions have a parametric representation. The cost of acquiring data in real robots favors parametric representations, as they typically require less data than non-parametric representations. The use of parametric representations is common practice in the dynamical systems literature (e.g., the work of Fraser [22]), especially when insight about system dynamics is available.

The appendix investigates the dynamics of the grasp process for the case in which all contacts are localized in a planar facet of the object. It is demonstrated that the temporal rate of change $\dot{\epsilon}$ is a linear function of ϵ, the squared wrench error, according to the following relation:

$$\dot{\epsilon} = -K(\epsilon - \epsilon_0), \tag{4.2}$$

where K is a characteristic gain and ϵ_0 is squared wrench residual for the corresponding attractor. This result suggests that a set of linear models involving ϵ and $\dot{\epsilon}$ can be used as a first order model of grasp dynamics, provided the contacts are not close to discontinuities in the object geometry such as corners or edges. Notice that such discontinuities can be represented as transitions from one set of linear models to another set, constructed by the subsequent observations.

Equation 4.2 describes system dynamics in an idealized situation without sensor noise and other disturbances. The representation proposed here assumes that (1) system dynamics in phase space can be modeled by a set of piecewise linear stochastic models with mean $\bar{\dot{\epsilon}}$ given by Equation 4.2, and (2) noise and uncertainty described by a zero mean Gaussian distribution with variance σ^2. The resulting generative model $M_{\pi_i}(\boldsymbol{\psi}, \mathbf{o})$ follows:

$$\begin{aligned} \bar{\mathbf{o}} &= \begin{bmatrix} \epsilon \\ \dot{\epsilon} \end{bmatrix} = \begin{bmatrix} \epsilon \\ -K\epsilon + \epsilon_0 \end{bmatrix} \\ M_{\pi_i}(\boldsymbol{\psi}, \mathbf{o}) &= \frac{1}{L}\frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-(\mathbf{o}-\bar{\mathbf{o}})^T(\mathbf{o}-\bar{\mathbf{o}})}{2\sigma^2}). \end{aligned} \tag{4.3}$$

The observation vector $\mathbf{o} = [\epsilon \;\; \dot{\epsilon}]^T$ consists of the squared wrench residual ϵ (defined by Equation 3.3) and an estimate of its temporal rate of change $\bar{\dot{\epsilon}}$. The parameter vector $\boldsymbol{\psi}_k = [\sigma^2 \; K \; \epsilon_0]^T$ describes the structure of the probability distribution $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o}) = P(\mathbf{o}|M_k)$. The constant L normalizes the distribution, such that

$$\int_{\mathbf{o}} M_{\pi_i}(\boldsymbol{\psi}, \mathbf{o}) \, \mathrm{d}\mathbf{o} = 1.$$

4.2 Constructing Generative Models of Grasp Dynamics

The representation of the dynamics generated by primitive π_i requires a set of m generative models, expressed as $\mathcal{M}(\pi_i) = \bigcup_{k=1}^{m} M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$. The model $M_{\pi_i}(\boldsymbol{\psi}_1, \mathbf{o})$ is a native model of system dynamics (including fixed points) and it provides a uniform representation for learning grasp policies. Models $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o}) \in \mathcal{M}(\pi_i)$ are constructed using data obtained empirically. As the agents executes the controller π_i, it records a sequence of n observations $\mathcal{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_n\}$. The derivation of model $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$ consists of determining the parameter vector $\boldsymbol{\psi}_k^*$ that maximizes the likelihood that $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$ "generates" $\mathcal{O}$. The log-likelihood $\mathcal{L}(\mathcal{O}, \boldsymbol{\psi}_k)$ associated with sequence $\mathcal{O}$ is defined as

$$\mathcal{L}(\mathcal{O}, \boldsymbol{\psi}_k) = \sum_{j=1}^{n} \log M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o}_j),$$

where n is the length of the sequence of observations $\mathcal{O}$, and $\mathbf{o}_j$ denotes the j^{th} element of $\mathcal{O}$.

Many optimization procedures can be used to derive the parameter vector $\boldsymbol{\psi}_k^*$. The log-likelihood function $\mathcal{L}(\mathcal{O}, \boldsymbol{\psi}_k)$ associated with the model defined by Equation 4.3 is a quadratic function on the parameter vector $\boldsymbol{\psi}_k$, and a closed-form solution for $\boldsymbol{\psi}_k^* = [\sigma^2 \quad K \quad \epsilon_0]^T$ can be easily derived by solving the system of three linear equations defined by $\nabla \mathcal{L}(\mathcal{O}, \boldsymbol{\psi}_k) = \mathbf{0}$:

$$
\begin{aligned}
\sigma^2 &= \frac{\sum_{i=1}^{n}(\dot{\epsilon}_i + K(\epsilon_i - \epsilon_0))^2}{n} \\
K &= \frac{n\sum_{i=1}^{n}\dot{\epsilon}_i\epsilon_i - \sum_{i=1}^{n}\epsilon_i\sum_{i=1}^{n}\dot{\epsilon}_i}{(\sum_{i=1}^{n}\epsilon_i)^2 - n\sum_{i=1}^{n}\epsilon_i^2} \\
\epsilon_0 &= \frac{\sum_{i=1}^{n}\dot{\epsilon}_i\sum_{i=1}^{n}\epsilon_i^2 - \sum_{i=1}^{n}\dot{\epsilon}_i\epsilon_i\sum_{i=1}^{n}\epsilon_i}{n\sum_{i=1}^{n}\dot{\epsilon}_i\epsilon_i - \sum_{i=1}^{n}\epsilon_i\sum_{i=1}^{n}\dot{\epsilon}_i}
\end{aligned}
$$

4.2.1 Elimination of Redundant Models

One generative model is derived for each sequence $\mathcal{O}$ recorded; after n trials, n models $M_{\pi_i}(\boldsymbol{\psi}_j, \mathbf{o})$, $j = 1, \ldots, n$ will have been derived. Many of these n models are redundant, representing a common dynamic pattern in the phase portrait. Two models characterized by the parameter vectors $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}_j$ are said to be redundant if $|\boldsymbol{\psi}_i - \boldsymbol{\psi}_j| < \delta$; therefore, a clustering algorithm in parameter space $\boldsymbol{\psi}$ can be used to identify and purge redundant models.

In the experiments reported in Chapter 5, an agglomerative $k-$means clustering algorithm [52] is used to identify and merge redundant models. Every pair of models $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}_j$ such that $i < j$ is compared and merged if $|\sigma_i^2 - \sigma_j^2| < 0.01$, and $|K_i - K_j| < 0.2$, and $|\epsilon_{0i} - \epsilon_{0j}| < 0.01$. Selected constant thresholds were determined empirically, and they implement a conservative model merging policy; only very similar models are merged. It is worthwhile to mention that the elimination of redundant models is not strictly necessary for the derivation of context-sensitive composition policies; however, it does reduce the computational load associated with the derivation of these policies.

4.2.2 Sampling Dynamics

The construction of $\mathcal{M}(\pi_i)$ requires experience over the whole spectrum of control contexts. The problem of determining how much experience is necessary to construct adequate models of system dynamics is open in general. A more complete set of models is certainly better, but it is important, in practice, to balance the utility of the missing information against the cost of acquiring it. Moreover, model completeness cannot be guaranteed, unless special assumptions are made. The issue of how much experience is necessary is related to exploration-exploitation issues in the reinforcement learning literature – see Thrun [62] for a lucid review of the techniques proposed to address this issue.

The number of epochs required for sampling dynamics provides an upper bound on the total number of models in $\mathcal{M}(\pi_i)$. Its specification can be made within the framework of dynamical systems theory, where the traditional approach is to monitor how much "novelty" each interaction episode brings to the existing set of models[1]. If the exploration policy is such that the all states have the same chance of being selected as the initial state, then the exploration policy is considered uniform, and model acquisition should stop as soon as the "novelty" index for the last n episodes falls below a given threshold τ.

4.3 Derivation of Membership Pattern q

The membership pattern $\mathbf{q} = [p_1\ p_2\ \ldots\ p_m]^T$ is a Boolean vector containing m predicates that indicate which models out of the set of m generative models are consistent with the last observation. The predicate p_k is defined as $p_k = (M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o}) > \kappa)$, where κ is an conveniently chosen threshold that reflects the confidence on the quality of the generative models.

The predicate p_k evaluates TRUE in a region of the domain of $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$ defined by the relation $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o}) \leq \kappa$. The regions defined by the several model boundaries may overlap in phase space, as shown in Figure 4.2. Each region is uniquely labeled by the membership pattern $\mathbf{q}$, and one can re-describe the transitions between regions in terms of a discrete graph, shown in the right panel of Figure 4.2. For instance, the region in which models $M_{\pi_1}(\boldsymbol{\psi}_2, \mathbf{o})$ and $M_{\pi_1}(\boldsymbol{\psi}_3, \mathbf{o})$ (respectively models M_2 and M_3 in Figure 4.2) overlap is identified by the vector $[0\ 1\ 1\ 0\ 0]^T$. The resulting

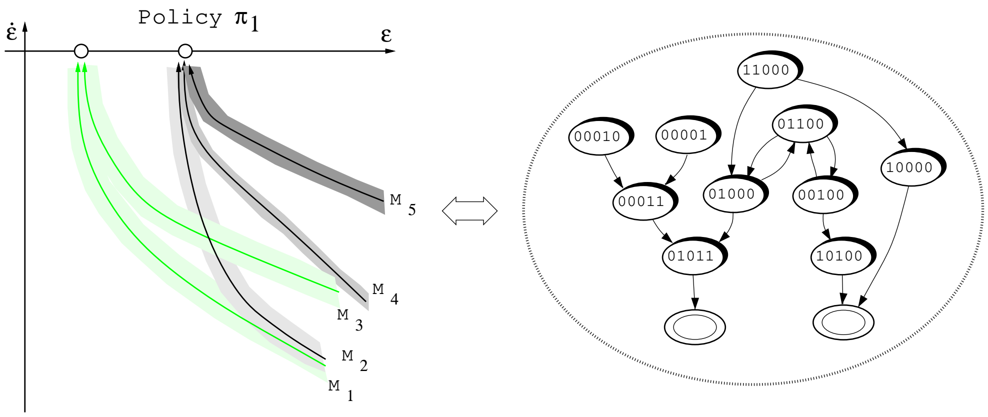

Figure 4.2. Diagrams depict the phase space plot $(\epsilon, \dot{\epsilon})$ for primitive π_1 (left) and all possible state transitions (right).

representation defines a discrete state space, that in principle re-describes the grasp process in terms of transitions between these discrete states.

Notice that the set $\mathcal{M}(\pi)$ of generative models may be incomplete, due to its empirical nature; as the grasp progresses, the system may find itself in a region of the phase space that has not been visited during the model construction stage. When no existing model adequately predicts recent observations, $\mathbf{q}$ evaluates to $\mathbf{0}$. Therefore, state $(\pi, \mathbf{0})$ is visited whenever the agent has no applicable prior knowledge.

The experiments reported in Chapter 5 suggest that substantial performance improvements can still be attained even if only a subset of all possible generative models are available. Moreover, the incompleteness of $\mathcal{M}(\pi)$ can be addressed in direct ways, as suggested in Chapter 7.

4.3.1 Cardinality of State Space

The total number of states (or equivalently, operational contexts) in the representation proposed can be estimated for the case in which model boundaries are described by linear functions. Each model is characterized by the parameter vector $\boldsymbol{\psi}_k = [\sigma^2 \ K \ \epsilon_0]^T$, and represented by a pair of lines in the phase plane as shown in

Figure 4.3(a). Model overlapping in the phase plane (Figure 4.3(b)) divides the phase plane in regions. Each region can be labeled by the subset of models that form it. In the state representation proposed, each region also defines a state; the total number of regions defined in the phase plane is therefore equivalent to the corresponding number of states.

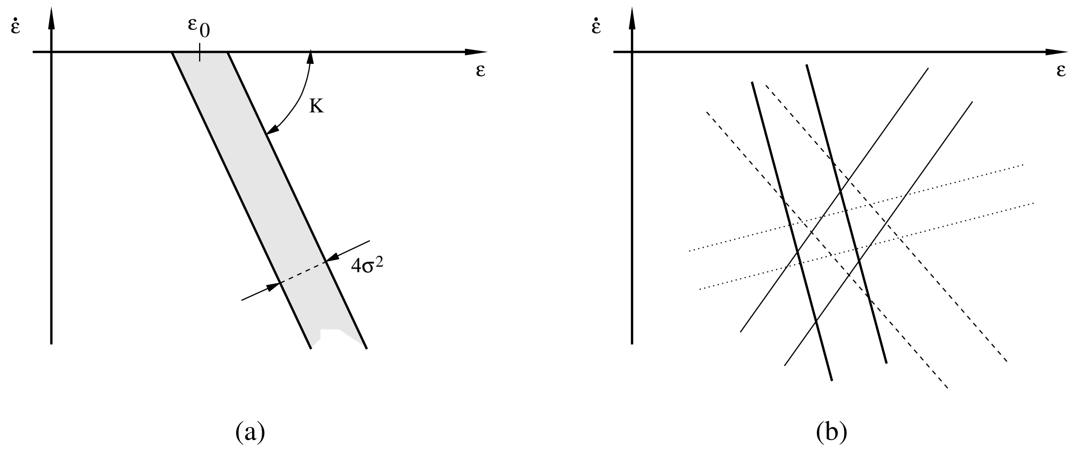

Figure 4.3. (a) Characterization of linear model in the phase plane. (b) Arbitrary arrangement of models in the phase plane; given m models, there exist $O(m^2)$ distinct regions in the phase plane.

Claim 1 *$O(m^2)$ states are defined by the planar arrangement of a set of m linear models in the phase plane.*

Proof: The proof consists of determining an upper bound for the total number of line segments defined by two intersection points, and using it to construct an upper bound for the number of regions.

Each model is represented by two parallel lines in the phase plane; there exist exactly $2m$ lines in the phase plane. Each line is intersected by at most $2m - 2$ lines, as no line is intersected by itself or by the other parallel line that belongs to the same model. The $2m-2$ intersection points divide the line into $2m-1$ line segments; it follows that there exists at most $2m\,(2m-1)$ line segments in the phase plane.

Regions are defined by a minimum of three line segments, and each line segment is in between at most two distinct regions. Assuming that all line segments are between two regions, and assuming that all regions are triangles, it follows that there are at most $\frac{2}{3}\,2m\,(2m-1)$ regions in the phase plane. The upper bound on the number of corresponding states is proved.□

The experimental data obtained for the grasping domain confirms that m^2 is a pessimistic upper bound for the total number of distinct states. Contrary to the assumption in the proof above, empirical models have a bounded domain, and not every model overlap with all other models.

4.4 Temporal Resolution in Control Composition

The control composition process presumes the existence of specific events or *choice points* in which the agent is allowed to choose a controller in order to maximize utility. Two polar approaches for the determination of the choice points have been described in the literature; one allows the agent to change its strategy at every servo cycle (essentially at every clock tick), and at the other end of this continuum, the agent may choose to switch strategies only after a pre-defined termination event, such as for example the convergence of the current control law. Both extremes have drawbacks: as the temporal resolution decreases (from a clock tick to termination events), the agent becomes less reactive but the complexity of building a policy in the resulting MDP decreases. At the other extreme, increased temporal resolution (moving from termination events to clock ticks) results in increased complexity and responsivity.

It is reasonable to assume that the best point in terms of accumulated utility for the agent is often between these extremes. The state representation introduced in this dissertation supports such intermediate solutions. In it, changes in the membership vector, $\mathbf{q}$, are used to segment the agent experience. Changes in $\mathbf{q}$ reflect changes in the information available to the agent; the resulting segmentation is inherently more efficient than clock tick-driven segmentation. At the same time, it brings about higher reactivity than termination event-driven segmentation, as the number of choice points is always larger for the state change-driven segmentation. Moreover, changes in $\mathbf{q}$ are determined autonomously by the agent, contrary to the other alternatives.

CHAPTER 5

SYNTHESIS OF CONTEXT-DEPENDENT GRASP STRATEGIES

The effect of the grasp controllers can be characterized by transitions between discrete states as demonstrated in the previous chapter. The discrete state space allows the system to experiment with sequences of control activations to discover those sequences that lead the system to optimal grasp configurations, as evaluated by a given reward or utility metric.

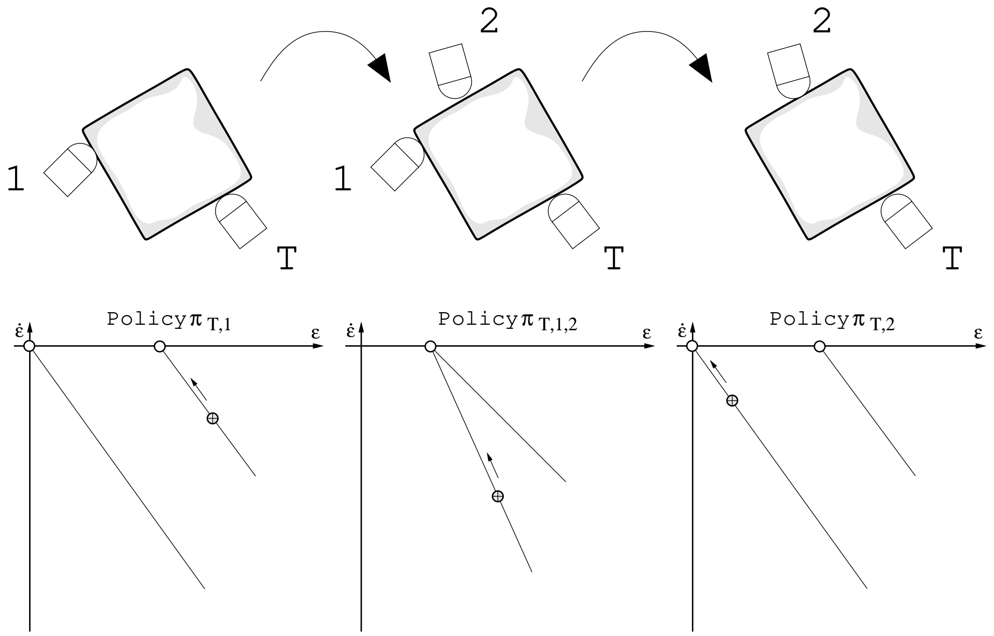

Figure 5.1. Context-dependent grasp of a cube (top view). Controller $\pi_{T,1}$ is initially applied, and a transition to controller $\pi_{T,1,2}$ is found during learning to be advantageous. After convergence of controller $\pi_{T,1,2}$, a transition to controller $\pi_{T,2}$ is made.

Figure 5.1 illustrates how sequential activation of the grasp primitives in the control basis Π may lead to improved performance. Initially, the system adopts controller $\pi_{T,1}$ to grasp an object of unknown geometry and pose. The resulting phase space coordinate corresponds to a state in the discrete graph and the experience accumulated during learning indicates that a switch to controller $\pi_{T,1,2}$ is advantageous from this state. Controller $\pi_{T,1,2}$ is followed until convergence where the system transitions again to controller $\pi_{T,2}$. Controller $\pi_{T,2}$ is executed and the system converges to the attractor that optimizes the objective performance.

The problem of deciding when to adopt a distinct grasp primitive, as well as which one to adopt can be characterized as a sequential decision problem, that can be solved using reinforcement learning methods. The goals of the reinforcement learning task are primarily to address the variability and uncertainty introduced by many possible objects and poses. The resulting *grasp policy* overcomes (to a certain degree) the locality of the component control primitives, π_i. It also expands the range of possible behavior beyond that of the individual controllers, creating grasp artifacts as well as information regarding the run-time context. The final control policy is capable of adapting its strategy to the many observable operational contexts encountered. The construction of such context-dependent control strategies is described next. In this work, "grasp policy" or simply "policy" refers to the control strategy formed by sequencing the grasp controllers, π_i, also referred to as "control primitives" as described in Chapter 3.

5.1 Derivation of the Optimal Switching Policy

From any grasp configuration, there exists a grasp resource allocation schedule that leads the system to a configuration that is optimal. Optimality implies that no other reachable grasp configuration has better quality as measured by the objective function. In this work, resource allocation schedules are equivalent to control switching policies, as there exists an one-to-one correspondence between set of resources and constituent controllers. Identifying the optimal switching policy can be expressed as a sequential decision problem, and reinforcement learning [59] can be used in its resolution. It allows the agent to evaluate the relative merits of each policy incrementally, as the agent interacts with the environment.

Several reinforcement learning algorithms have been proven to converge to the optimal policy under prescribed conditions. The Q-learning algorithm [64] is employed here to derive the optimal policy. The goal of this dissertation is to advance the interactionist state representation introduced in Chapter 4 rather than to make theoretical contributions to reinforcement learning algorithms. In the pursuit of this goal, experiments involving both physical and simulated arm/hand systems and objects were performed. This chapter describes the experiments performed in simulation; the experiments involving physical manipulators are described in Chapter 6.

The complete specification of the learning problem includes a description of the experimental setup, the state space, the action space, the reward structure, and the learning parameters. The following section will provide these details.

5.1.1 Experimental Setup

All grasp trials involved the simulation of a 14 DOF robot arm/hand platform consisting of a five degree-of-freedom (DOF) GE P50 robotic arm and the Stanford/JPL hand. The Stanford/JPL hand has three fingers, each one with three independently actuated joints.

In the beginning of each trial, an object is placed within the workspace of the robotic arm. The location of the object's center of mass is known for simulated experiments, but not its orientation (a random variable uniformly distributed in the interval $(-\pi, \pi])$, as well as its identity: the test object is selected at random from a set containing cylinders, cubes and triangular prisms.

In experiments involving physical manipulators, errors in the kinematic transformations and errors due to the actuation of the fingers disturb the measurements of contact location and normals so that the same object presented at the same position and orientation will produce different tactile signatures across trials. No two grasp trials are performed on exactly the same object, at least as perceived by the system. For the experiments performed in simulation, a visual model of the object geometry is constructed with the sole purpose of generating simulated contact information. Cubes and triangular prisms are described by their height h, the edge distance to each object's reference frame r, and the corresponding edge angle θ; cylinders are described by the height and radius parameter. The measurement noise observed in the real manipulator is modeled through variations in the parameters r and θ, as for most grasp configurations the contacts are placed on the side of the objects and are therefore not affected by changes in object height. The distributions for r and θ were constructed based on data collected over thirty-two models of the cube and the triangular prism derived from images of real objects. Figure 5.2) depicts the top view of nine cubes generated by sampling the r and θ distributions repeatedly.

In the case of the cube, the model geometry is described by four parameter pairs $P = \{(r, \theta)_i,\ i = 1, \ldots, 4\}$. The first three parameter pairs are determined by sampling the distributions for r and θ repeatedly; the last parameter pair is computed so that the resulting object has its center of mass centered at the origin; see Figure 5.2 (right panel) for examples of objects constructed according to this procedure. The models for triangular prisms are generated similarly, and models for cylinders are an exception, not being subject to variations in their geometry.

Once object location is estimated, the robot arm reaches to the object and the hand is positioned over the object in a configuration such that the object surface is within the workspace of the participating fingers. With the hand positioned over the object, the system is ready for interacting with the object.

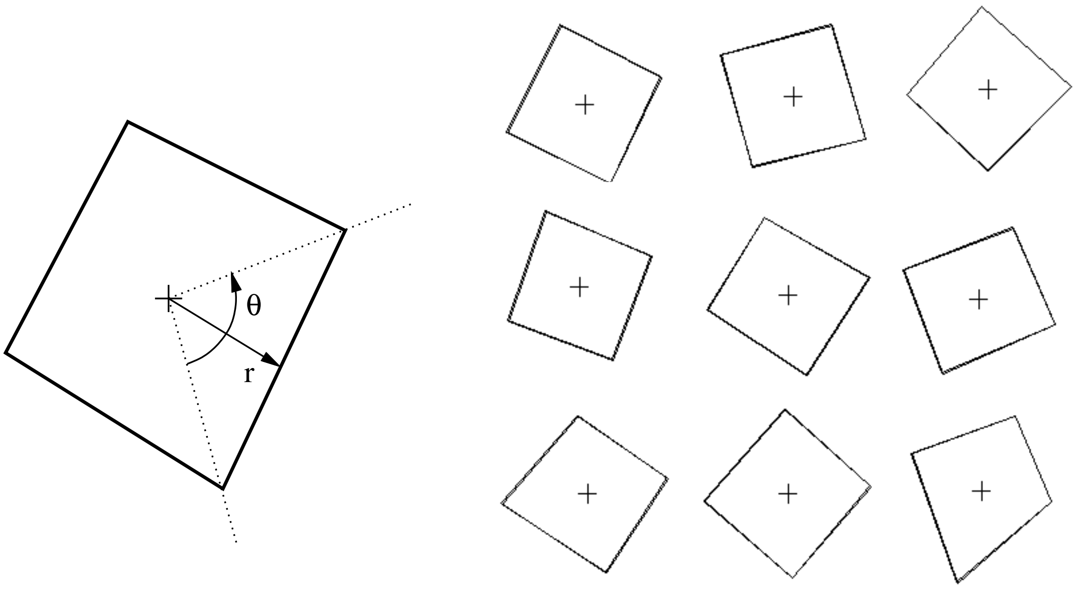

Figure 5.2. (Left) Model geometry is described by parameters r, θ and the height h (not shown). (Right) Models constructed by sampling the distributions of r and θ.

5.1.2 State Space

The state space associated with the grasp task is the object of Chapter 4. State representation is based on a set of generative models of grasp dynamics, as described in Section 4.1. Such generative models can be constructed as the agent learns the optimal switching policy, provided there is a bound on the total number of generative models. For the experiments described in this section, model construction and policy learning are performed in two different stages. This strategy avoids the overhead incurred as the agent keeps track of the model construction and learning processes.

The collection of models describing the grasp dynamics generated by controller π_i is denoted by $\mathcal{M}(\pi_i) = \bigcup_{k=0}^{m} M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$ (see Section 4.2). The data required for the construction of each set $\mathcal{M}(\pi_i)$ was collected over the course of 35 grasp trials for each object type. No switching takes place during this stage: the same control primitive is used from the beginning of the trial until its conclusion. A total of 4 (control options for π_i) $*$ 3 (object types) $*$ 35 (trials) $=$ 420 grasp trials were performed in simulation, resulting in 420 generative models of grasp dynamics. After the elimination of the redundant models (see Section 4.2.1), a total of 61 models resulted. Figures 5.3 and 5.4 depict the generative models for each object type and for 2- and 3-contact grasp controllers, respectively. Each model, $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$, is represented by a polygon that delimits the region of phase space in which $p_k = 1$ (see Section 4.3), corresponding to two standard deviations Models with an associated prior probability below 5% are colored light gray.

The lower right plot in both figures shows all models. Each region defined by the two or more overlapping models is uniquely identified by a membership pattern vector $\mathbf{q}$, as defined in Section 4.3. The total number of states for a given policy is equal to the number of distinct regions the overlapping models define in the phase portrait. As discussed in Section 4.3.1, m models define a state space with $O(m^2)$ states.

5.1.3 Action Space: Primitive Grasping Actions

During learning the agent can adopt any control primitive out of the grasp control basis Π at every choice point. For the Salisbury hand, the agent has four choices of controllers, as indicated in Equation 4.1:

$$\Pi = \{\pi_{(T,1)}, \quad \pi_{(T,2)}, \quad \pi_{(1,2)}, \quad \pi_{(T,1,2)}\},$$

where the subscript indicates the set of resources enlisted by each grasp controller. For instance, $\pi_{(T,2)}$ denotes the grasp primitive that engages both the thumb (T) and the second finger (2).

During every learning trial the first control primitive is selected randomly from the control basis Π. The corresponding set of fingers is enabled and the first tactile probe is performed, resulting in the first observation $\mathbf{o}_1 = (\epsilon, \dot{\epsilon})_1$. Given $\mathbf{o}_1$, the list of generative models is scanned to identify those models explaining $\mathbf{o}_1$ and to

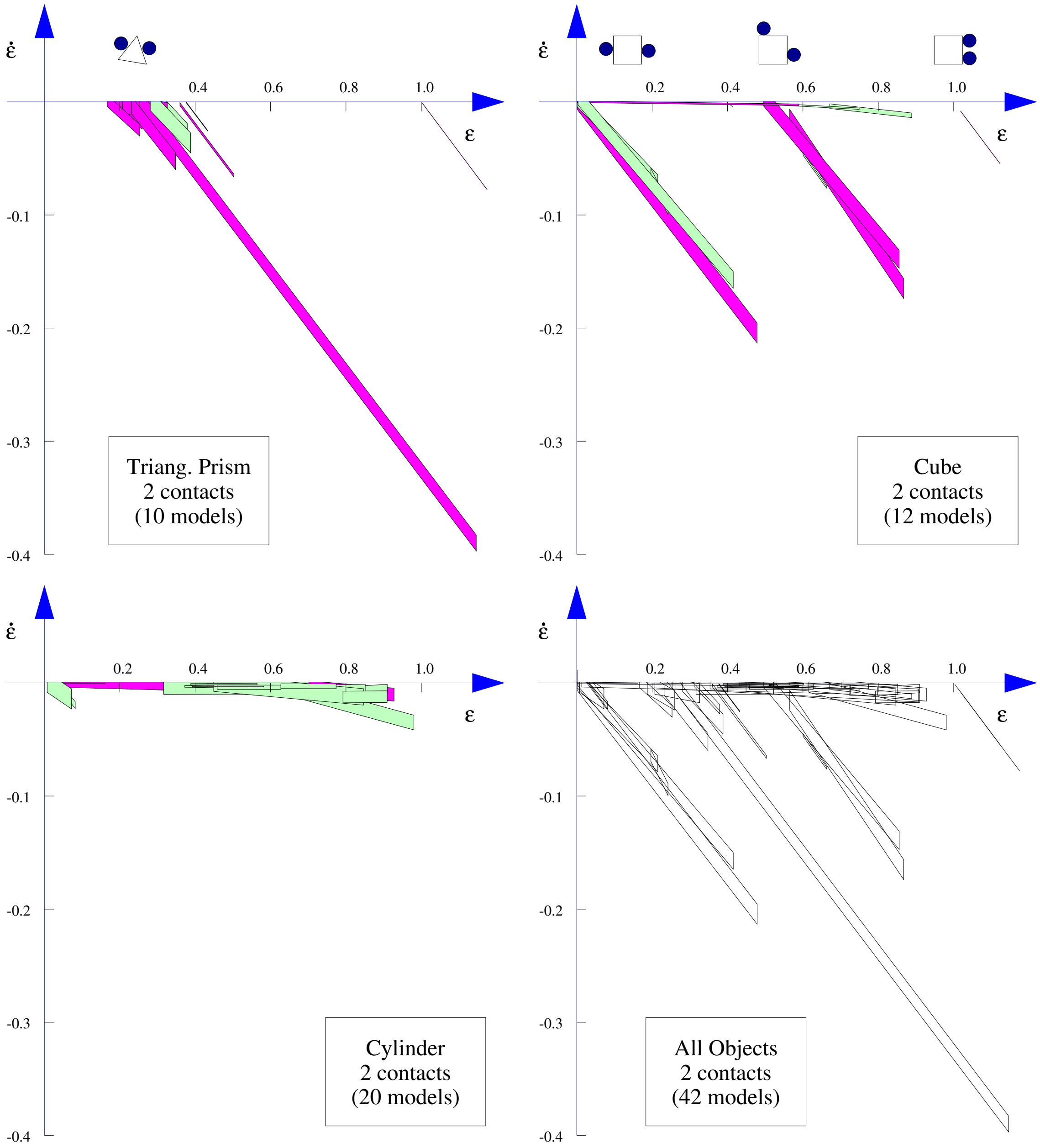

Figure 5.3. Generative models for 2 contact grasp controllers, for each object type. Each model, $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$, is represented by a polygon that delimits the region of phase space in which $p_k = 1$.

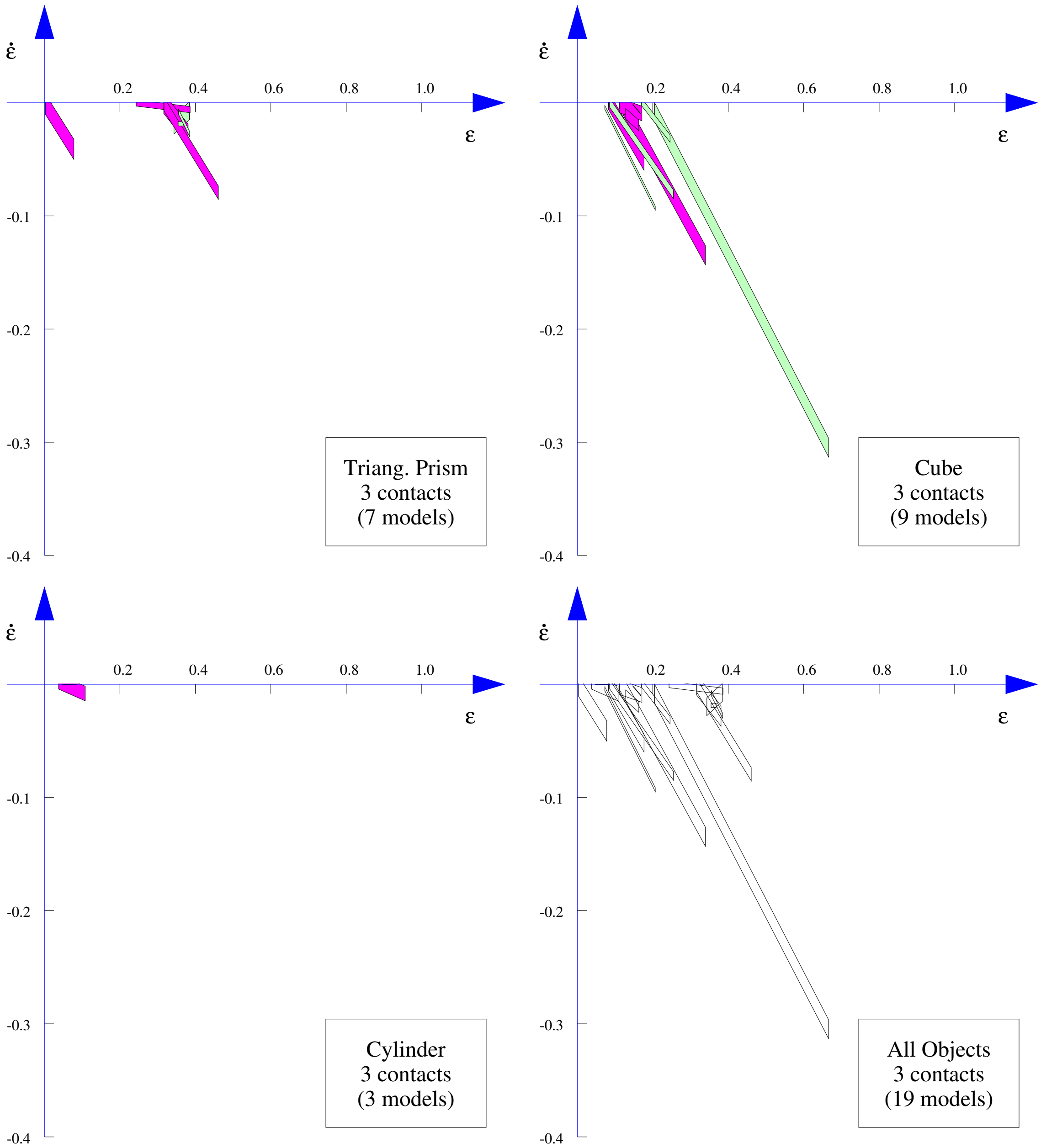

Figure 5.4. Generative models for 3 contact grasp controllers, for each object type.

form the membership pattern vector, $\mathbf{q}_1$. The state $(\pi, \mathbf{q}_1)$ points to an entry in the state-action value table and the corresponding best action is selected.

A control primitive is executed until a change in the membership pattern $\mathbf{q}$ is detected; at this *choice point* the best action for the new state is selected. Many tactile probes may be executed under the same controller before a change in the membership vector, $\mathbf{q}$, occurs. As the current control law reaches a convergent configuration, the agent reaches a special choice point signaled by the change in predicate p_1, the first component of the membership pattern $\mathbf{q}$. In this special choice point, the agent may continue the grasp trial by activating another control law, or it may terminate the grasp trial by selecting the control law that has just converged. In other words, the convergence of the current control law does not necessarily imply termination of the grasp process. This allows the agent to use suboptimal, convergent configurations as way points to the optimal configurations and to proceed with the grasp trial even if the initial configuration is at, or very close to, a suboptimal grasp configuration. The trial may also end if the limit of 50 tactile probes is reached. Once the grasp episode terminates, the corresponding grasp configuration receives a score of $(1 - \mu_0)$, as detailed below.

5.1.4 Reward Structure

The reward structure is task specific, as it defines the agent's objectives through the specification of the relative merits/costs of taking each action in each state. Within the grasp synthesis research community, there is no consensus as to which single grasp metric best captures the multiple dimensions along which a grasp configuration can be evaluated, and many grasp metrics have been proposed [21, 47].

The grasp metric optimized in the experiments described in this document is proportional to the negative of μ_0, the minimum friction coefficient required for grasp stability. Specifically, grasp configurations are scored corresponding to $(1 - \mu_0)$; a score of 1 ($\mu_0 = 0$) is best; it implies that friction forces will not be required to construct a force closure grasp. There are no costs or penalties associated with the number of tactile probes conducted *en route* to a convergent configuration.

The computation of μ_0 requires the resolution of a mathematical programming problem subject to the following constraints on normal and tangential forces [34, 24]:

1. $|\mathbf{f}_n| \geq 1$ — normal forces are compressive with magnitude greater than or equal to 1,

2. $|\mathbf{f}_t| \leq \mu|\mathbf{f}_n|$ — the net contact force must remain within the friction cone, and

3. $\sum_{i=1}^{k} \mathbf{f}_i = \sum_{i=1}^{k}(\mathbf{r}_i \times \mathbf{f}_i) = \mathbf{0}$.

In general, infinitely many force distributions satisfy these constraints. The solution selected corresponds to the one with the minimum combined magnitude $M = \sum_{i=1}^{k} |\mathbf{f}_{t_i}| + |\mathbf{f}_{n_i}|$. The solution for μ_0 is computed from the resulting force distribution. A more efficient algorithm for computing μ_0 is available [57]; the approach reported here was chosen for its simplicity.

5.1.5 Learning Parameters

The Q-learning algorithm computes the discounted sum of future rewards for each state-action pair (denoted as $Q(s,a)$), for a given control policy. The control policy specifies which action, a, is to be selected from every state, s. At the outset, before a policy exists, action selection is random and the algorithm is said to be stochastically exploring control alternatives. The rewards obtained are consolidated in the corresponding entries of $Q(s,a)$. As the learning process progresses, the control policy becomes increasingly focused on high-quality actions, even as $Q(s,a)$ changes due to changes in the control policy.

The values of $Q(s, a)$ are updated as the system transitions from state s_{t-1} to state s_t according to the equation

$$Q(s_{t-1}, a_{t-1}) = (1 - \alpha)Q(s_{t-1}, a_{t-1}) + \alpha(r_t + \gamma \max_a Q(s_t, a_t)),$$

where r_t is the reward received at time t, α is the learning rate, and γ is the discounting factor. The learning rate was set to $\alpha = 0.20$ and discounting factor was set to $\gamma = 0.98$. For the experiments described in this chapter, the selection of which grasp primitive π_r to adopt (action a) was regulated by a Boltzmann distribution. The probability $p_s(a)$ that the agent selects action a while in state s is described by:

$$p_s(a) = \frac{\exp \frac{Q(s,a)}{T}}{\sum_{b \in \mathcal{A}} \exp \frac{Q(s,b)}{t}},$$

where $\mathcal{A}$ is the set of all possible control choices or actions and T (for temperature) is the parameter that controls the amount of exploration (the probability of executing actions other than the one with the highest Q-value). If T is high, or if Q-values are all the same, this will pick a random action. If T is low and Q-values are different, it will tend to pick the action with the highest Q-value.

Parameter tuning was not necessary; the (relatively high) learning rate reflects the assumption that the state space is approximately Markovian. Under certain conditions [64], this iterative process has been shown to converge to an optimal policy and $Q(s, a)$ accurately reflects the value obtained from state s under the optimal control policy.

5.2 Results

Figure 5.5 depicts a typical learning curve (curve labeled $\triangle\square\bigcirc$, top left panel). The learning curve was smoothed with a sliding window ten data points wide. Each point corresponds to the grasp score associated with the convergent grasp configuration, for the randomly chosen object. The data corresponding to each object can be separated into three classes corresponding to the three object types.

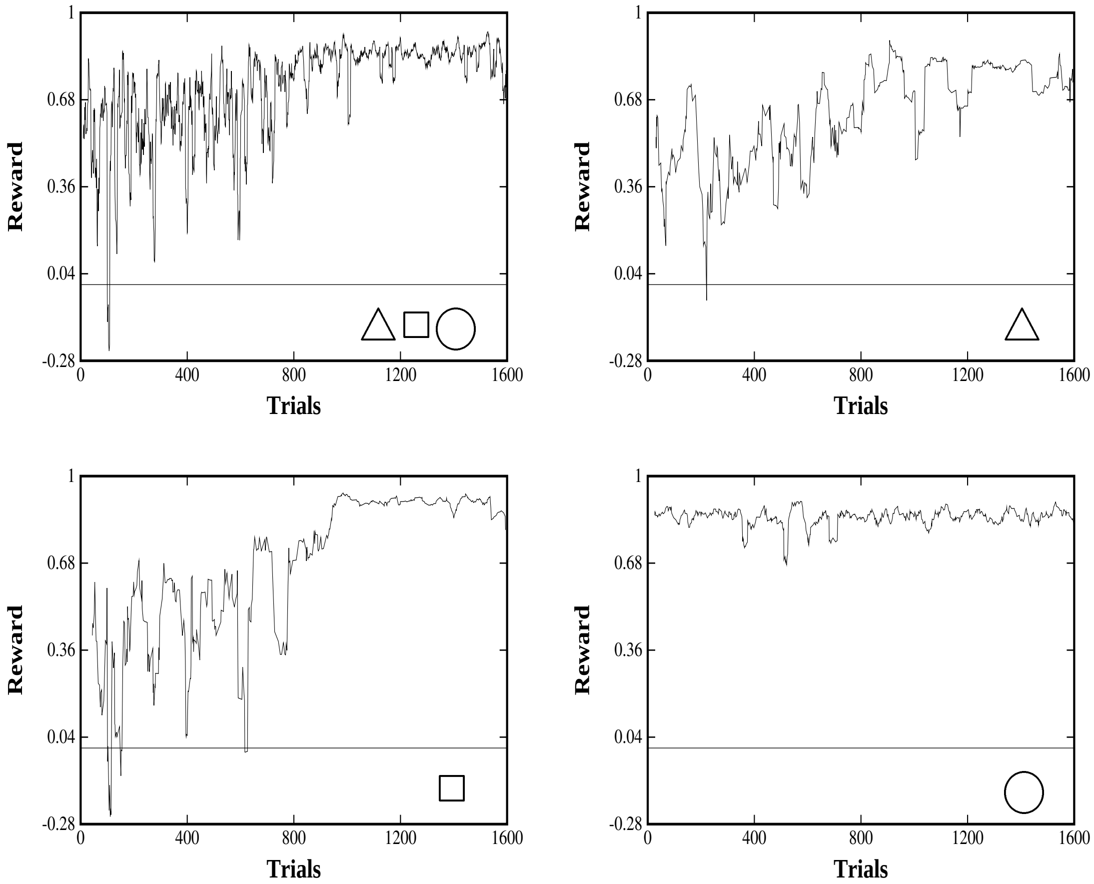

Figure 5.5. △□◯ Typical learning curve for all objects. The learning curves corresponding to the individual objects are labeled △, □, ◯. Vertical axes represent the reward $(1-\mu_0)$ received after convergence, and horizontal axes indicate the trial number.

As indicated in Figure 5.5, the learning algorithm is able to derive a switching policy that improves system performance for two of the three object types. The exception is the learning curve for the cylinder, where no noticeable improvement is indicated. As it turns out, any controller in the control basis is able to achieve a high quality grasp configuration on the cylinder; variations in performance are due solely to the precision of the convergent grasp configuration.

Note that many factors contribute to make an average score of 1.00 unattainable:

- The objects are not perfectly symmetrical; for instance, if two contacts are placed on opposing facets of a cube, some friction will be necessary for equilibrium, unless the corresponding planes are perfectly parallel to each other (this observation does not apply to the cylinder);

- Even if the facets were parallel to each other, control execution is halted when $\dot{\epsilon} < \delta$ and $\delta > 0$; the corresponding wrench residual will have to be canceled with additional friction forces; variations in performance are due solely to the precision of the convergent grasp configuration.

- Some of the conditions required for the convergence of the Q-learning algorithm may be violated, including non-Markovian state space, deficiencies in the exploration policy, and insufficient number of trials.

5.2.1 Baseline Performance

The performance of the composition policy will be compared to the performance of the best individual controller of the control basis in each context. In each trial, all controllers in Π are executed until convergence, starting from the same initial configuration, characterized by object position and orientation, as well as manipulator and finger configurations. The best score obtained by the set of control primitives is then recorded. The distribution of scores reflects the absolute best performance the agent can achieve, as if it had an "oracle" capable of selecting the best grasp primitive for each context in the beginning of each trial. Figure 5.6 shows the corresponding performance distributions obtained after eight independent runs of 1600 trials. The error bars in each bin indicate the range corresponding to 95% of all samples, assuming a normal distribution.

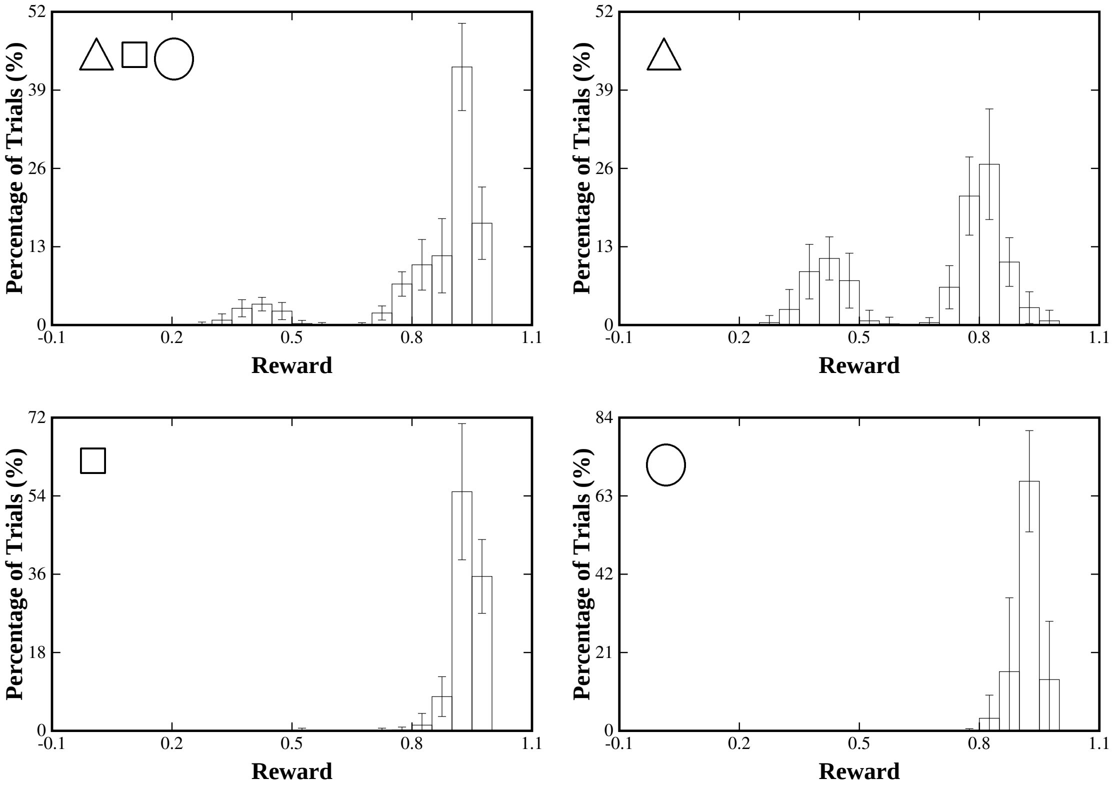

Figure 5.6. △□◯ Reward distribution for the oracle-based grasp controller, for all objects. The distributions for individual object types (labeled △, □, ◯) are also shown.

A secondary performance metric of interest is the number of tactile probes required for convergence. The probing operation is expensive: the tactile sensor must be moved along a planned trajectory slowly towards the object surface, while monitoring for contact in order not to disturb the object during grasp formation. Even though the minimization of the number of tactile probes is not the objective of the learning algorithm, it is important to investigate whether the learning-based grasp policy requires an impractical number of tactile probes. Section 6.1.1 discusses the implementation of the tactile probe operation in the physical robot.

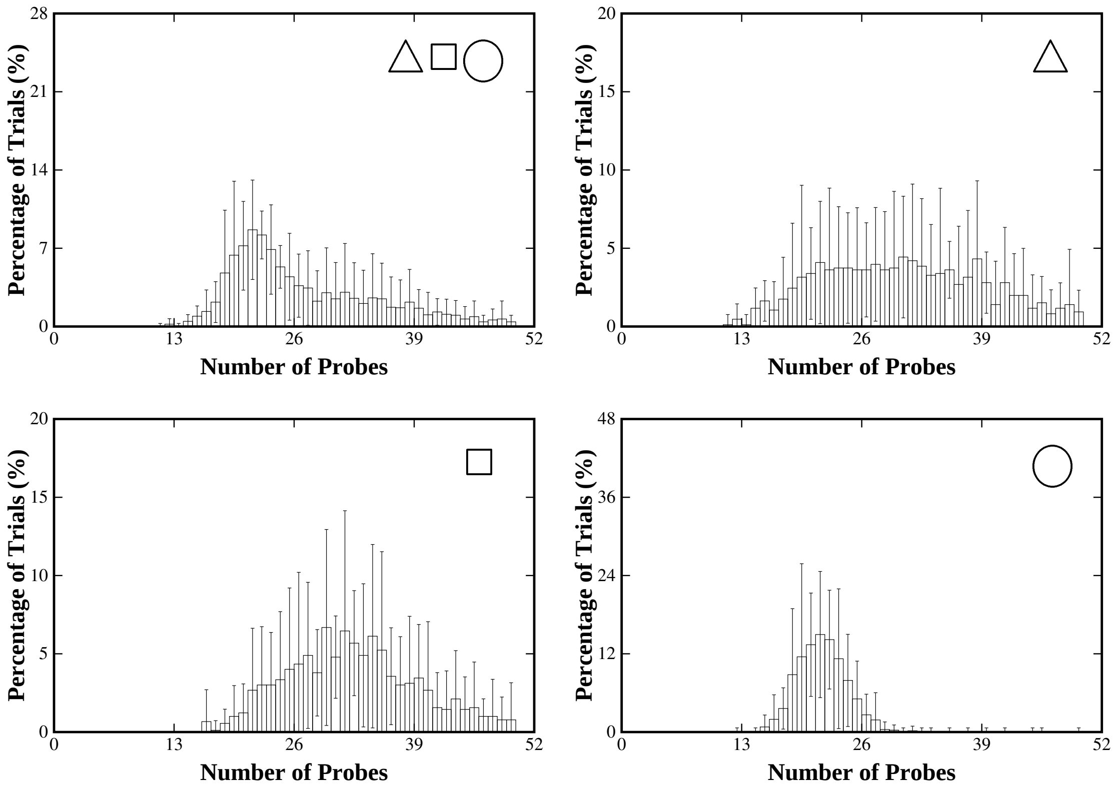

Figure 5.7. △□◯ Distribution of the number of tactile probes required until convergence for the oracle-based grasp procedure, for all objects. The distributions for individual object types (labeled △, □, ◯) are also shown.

Figure 5.7 presents the distributions for the number of tactile probes for all objects (top left, labeled △□◯), for the triangular prism (top right, labeled △), for the cube (bottom left, □), and for the cylinder (bottom right, ◯). The average count obtained in each bin is indicated as a percentage of the total number of trials (vertical axis).

5.2.2 Results for the Learned Composition Policy

System performance is evaluated by testing 15 switching policies (corresponding to 15 independent runs of the learning algorithm) over 1600 simulated grasp trials. Each grasp trial is scored in terms of the metric $(1-\mu_0)$ at convergent grasp configurations. Figure 5.8 depicts the performance histogram for all object types (top left, labeled △□◯), for the triangular prism (top right, labeled △), for the cube (bottom left, □), and for the cylinder (bottom right, ◯). The score obtained (horizontal axis) is segmented into 21 bins; negative scores are counted in the leftmost bin, and scores in the range $[0, 1]$ are recorded in the remaining 20 bins of equal width. As before, the average count obtained in each bin is indicated as a percentage of the total number of trials (vertical axis).

As suggested by Figures 5.8 and 5.9, most grasps elicited by the composition policy result in high quality grasp configurations, with scores in the range $[0.8, 1.0]$ for all objects. They also require a small number of tactile probes (the median number of probes for the composition policy is 4, as shown in Figure 5.9).

Note that about 4% (on average) of all trials involving the triangular prism (△) result in negative scores, meaning that a friction coefficient of at least 1.0 is required for force closure. A review of the data indicates that, in most cases, the negative score was due to the interruption of the grasp trial when the limit on the number of tactile probes (50) was reached. The analysis also shows that three factors explain this failure mode in part:

1. **Insufficient exploration:** In many cases, the composition policy leads the system to very good grasp configurations that nevertheless are mapped to states

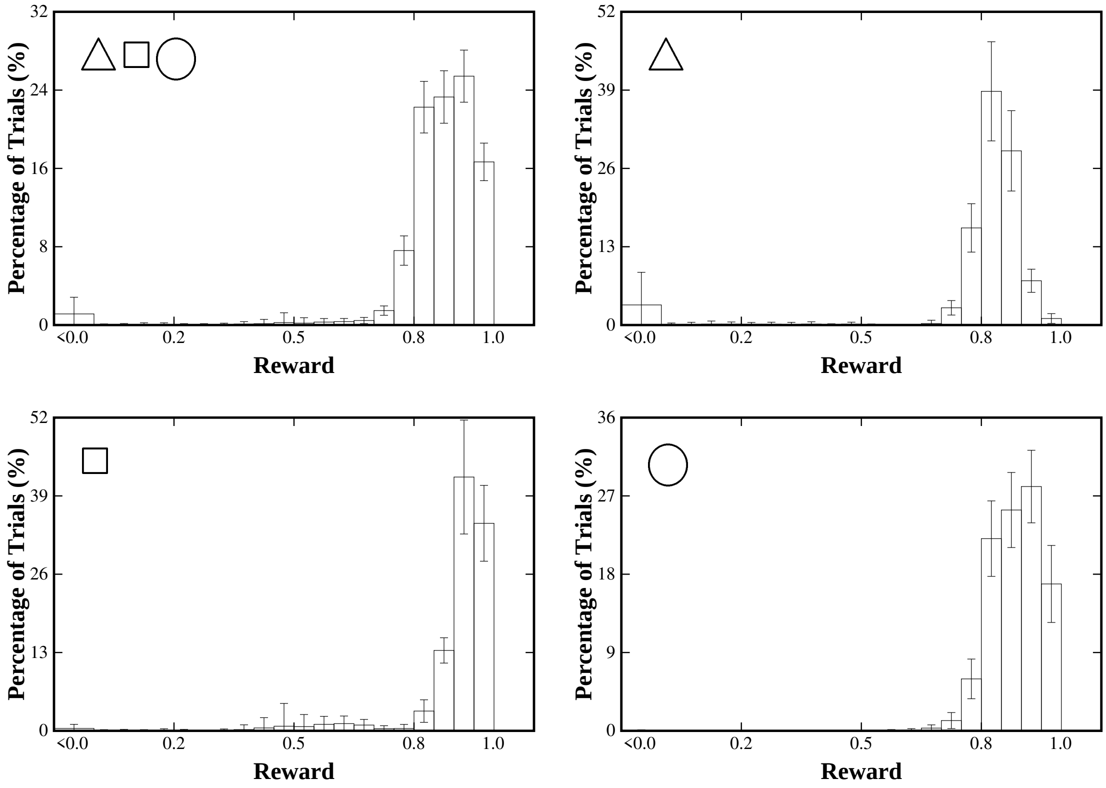

Figure 5.8. △□◯ Score distribution observed for 15 independent switching policies, for all objects. The distributions for individual object types (labeled △, □, ◯) are also shown.

that were never visited during learning. In these cases, the choice of which grasp primitive to adopt is random.

2. **Incomplete set of generative models:** Whenever the observation $\mathbf{o} = (\epsilon, \dot{\epsilon})$ is not supported by any model, the controller transitions to a special state depending on the current control policy (see Section 4.3). These states are labeled $s = 0$, $s = 400$, $s = 800$, and $s = 1200$, for the controllers $(T, 1)$, $(T, 2)$, $(1, 2)$, and $(T, 1, 2)$, respectively. Such special states are rarely visited in typical trials, but visits to these states are associated with a large number of tactile probes.

3. **Limit cycles:** In some trials, the system makes virtually no progress (i.e. contact configurations are not being displaced) over a number of tactile probes as the system keeps on alternating between two grasp primitives. The combination of normals and contact positions in both configurations is such that the control gradient associated with the second configuration cancel the control gradient for the first configuration almost exactly. This situation is not captured by the state representation proposed; in most cases the alternation between these controllers cause the grasp configurations to evolve until a favorable state is reached. One observation about this behavior is in order: such near complete gradient cancellation over consecutive grasp primitives is extremely rare for real systems; in a way, this is an artifact of the simulation of system behavior.

Together, these factors suggest that the shape of the objects associated with interrupted trials is quite unusual, and the composition policy failed to discriminate these particular object geometries adequately.

Figure 5.9 presents the distributions for the number of tactile probes for all objects (top left, labeled △□◯), for the triangular prism (top right, labeled △), for the cube (bottom left, □), and for the cylinder (bottom right, ◯). Again, the average count obtained in each bin is indicated as a percentage of the total number of trials (vertical axis).

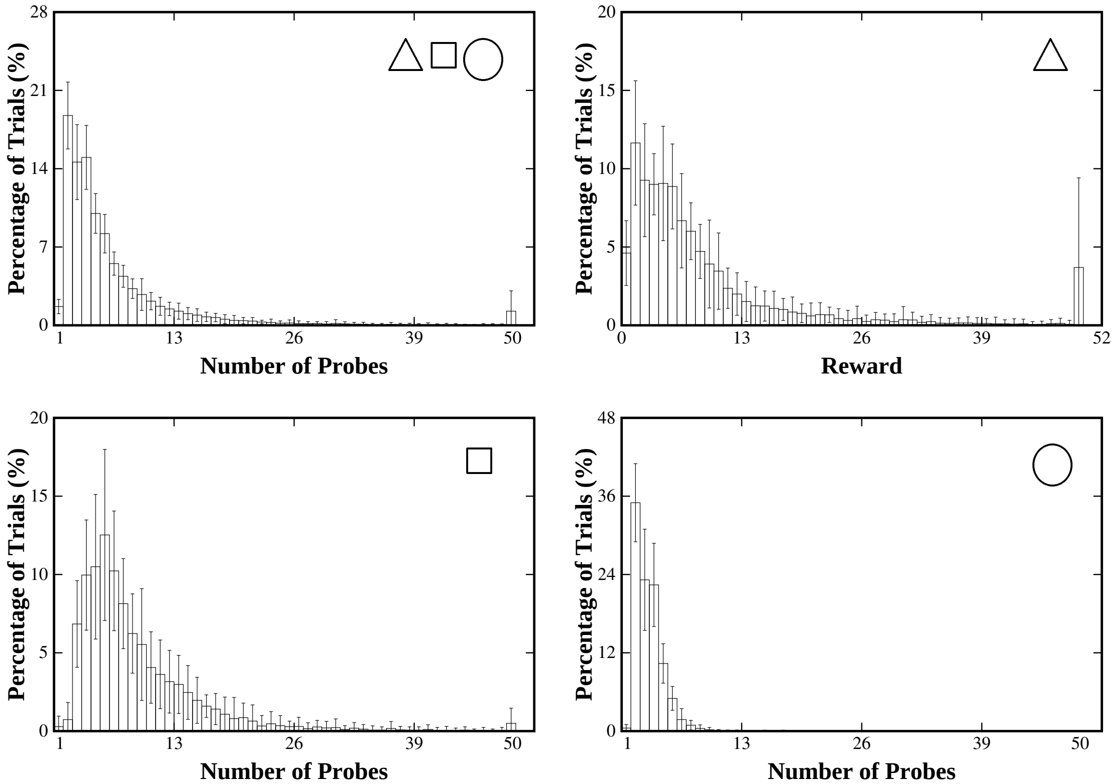

Figure 5.9. △□◯ Distribution of the number of tactile probes required until convergence for 15 independent switching policies, for all objects. The distributions for individual object types (labeled △, □, ◯) are also shown.

As noted earlier, about 4% of all trials involving triangular prisms are terminated as the limit of the number of tactile probes is reached – the effects of the termination are reflected in the performance distribution (see Figure 5.8).

The distributions shown in Figures 5.8 and 5.6 indicate that the learned grasp policy compares favorably to the oracle-based grasp procedure if the grasp involves triangular objects. The oracle-based grasp procedure cannot generate grasp configurations (either with two or three fingers) with scores above 0.5 for a number of grasp contexts involving the triangular object. The initial conditions are such that the agent is unable to position a finger in each of the facets of the triangular prism by executing the three-finger grasp primitive and attain grasp scores in the range between 0.8 and 1.0. The performance of the learned grasp controller and the oracle-based grasp procedure is comparable if the grasp involves cylindrical objects, and the oracle-based grasp procedure has the edge if the grasp involve cubes.

As shown in Figures 5.9 and 5.7, the context-based grasp policy requires fewer tactile probes than the oracle-based grasp procedure. The disparity in terms of the number of tactile probes reflects the greater efficiency of the context-based grasp policy in the use of information: the sequence of tactile probes establishes control context, that in turn informs control selection. It also reflects the nature of the oracle-based grasp procedure, in essence an off-line search strategy. The grasp policy finds positive marginal utility, in this case, in modeling run-time dynamics which produces higher average quality at lower cost.

5.2.3 Qualitative Behavior of the Learned Composition Policy

There are many ways of characterizing the composition policy derived by the learning process. The composition policy can be represented by a transition matrix expressing the probability associated with the transition for each pair of states. However, the matrix representation does not provide insight into the policy acquired, not expressing much of the skillful balance between contact control and information gathering.

Table 5.1 provides a first view of the distinct activation patterns elicited by each object. In it, the average frequency of activation of each grasp primitive (as a percentage of the total number of actions) is shown for each object type. It also shows that all grasp primitives are necessary (i.e., there is not a single controller in the control basis capable of addressing all problem instances).

Another view on the process can be constructed by merging states into classes (state supersets) and describe the grasp process by transitions between classes of

Table 5.1. Frequency of control law selection as a percentage of total number of actions, for each object.

Objects	Controllers			
	(T, 1)	(T, 2)	(1, 2)	(T, 1, 2)
△	12.71%	12.27%	9.72%	65.30%
□	46.16%	28.07%	9.55%	16.22%
○	15.71%	9.19%	7.15%	67.94%

states. This characterization is valid only in a very general sense and for particular objects; nonetheless, it provides a kind of intuition regarding the knowledge encoded in the policy and the expressive power of the interaction-based representation.

Figure 5.10 depicts a particular state arrangement constructed with the sole purpose of allowing external observers visualize and better understand the convergent composition policy. It is based on the transitions observed as the convergent composition policy was applied over 1600 grasp trials. States and transitions were merged and pruned according to three rules:

- Figure 5.10 depicts only the relevant transitions, defined as the top 90% transitions of all observed transitions.

- Transitions are segmented according to the object that it is being grasped; in practice, this yields finite state machines that are projections of the composition control policy into the different object geometries. It is important to stress that this segmentation is being done by an external observer; the control policy does not use this representation in any way, shape, or form.

- States are grouped according to the contact distribution over the object facets (for the triangular prism and cube objects) or according to grasp primitive used, in the case of the cylinder. For example, all three-contact grasp configurations over the triangle in which the contacts are placed in distinct facets are merged into one state; similarly for the all configurations in which two contacts are placed in one facet and the last contact is placed in a distinct facet. Again, such grouping can only be done by an external observer; in the case of the triangular prism and the cube, the agent has no concept of "object facet."

Figure 5.10 also shows that the interaction between the grasping agent and each particular object geometry (as commanded by the composition control policy) can be characterized by specific activation patterns. These activation patterns define object identity, as far as the grasping agent is concerned. The patterns in Figure 5.10 suggest that the grasping agent will start the grasp trial by selecting a two-fingered configuration. In the case of the cylinder (Figure 5.10, bottom diagram), the agent transitions almost in all cases to a three-fingered configuration, and applies the corresponding grasp primitive until convergence. The activation pattern for cubes (Figure 5.10, middle diagram) suggests that the agent has the goal of reaching a two-fingered grasp configuration in which the contacts are placed in opposite facets of the cube. Once in this configuration, the agent activates the primitive repeatedly until convergence. The activation pattern for triangular prism (Figure 5.10, top diagram) reveals the same goal oriented behavior observed in activation pattern for the cubes; in this case however the agent searches for a favorable three-fingered configuration before committing to the grasp primitive that will be used until the conclusion of the grasp trial.

5.2.4 Number of States

During the learning procedure, states are recorded as they are visited. By virtue of their definition, the number of distinct states is limited, so the number of states actually visited provides some insight to the number of learning trials necessary for the derivation of the optimal composition policy. Table 5.2 indicates the average and standard deviation of the number of states visited during learning (top table) and during testing (bottom table). The number of states visited is further broken down

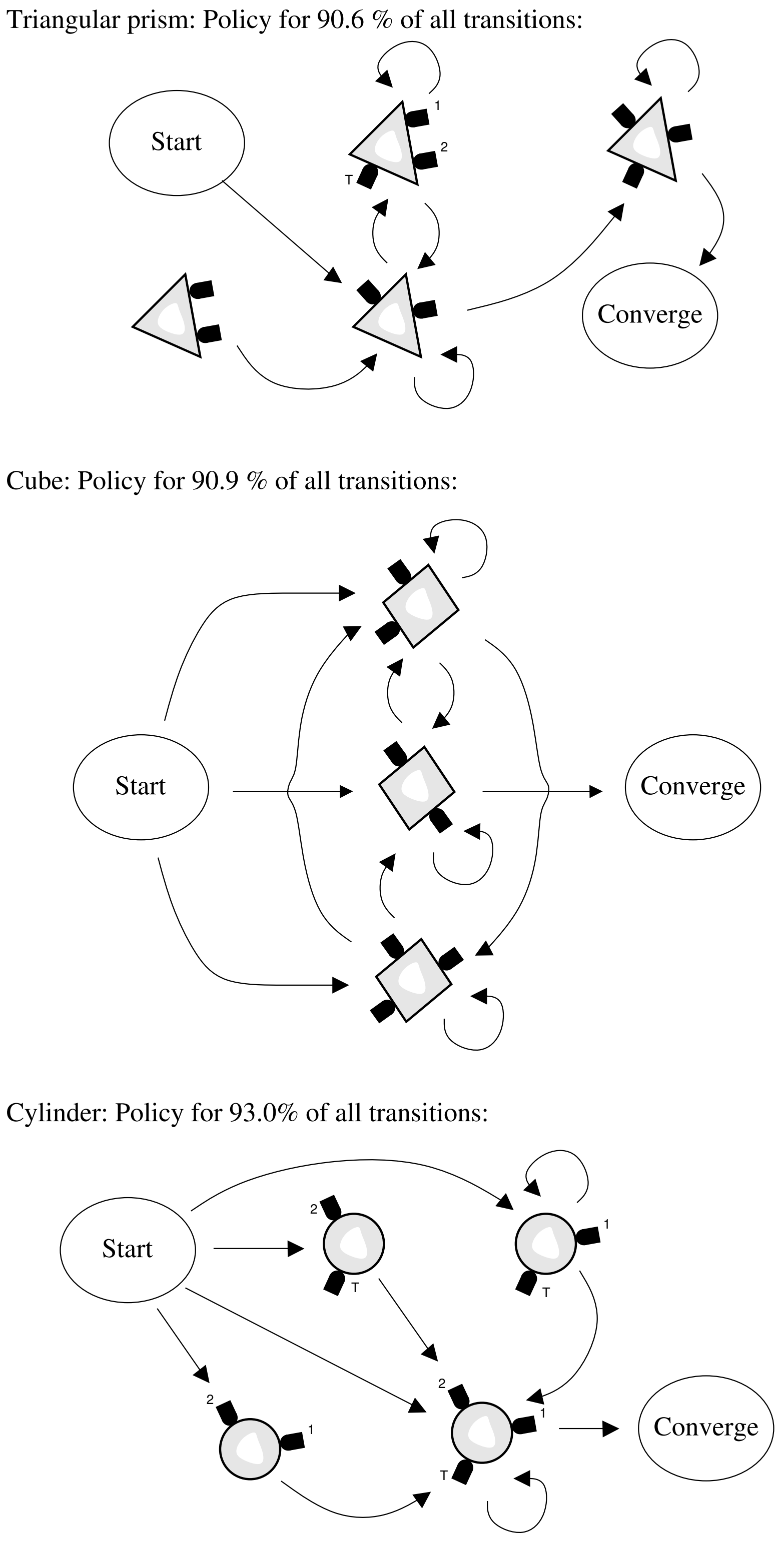

Figure 5.10. Transition diagrams for the triangular prisms (top), cubes (middle) and cylinders (bottom). The state machines for the triangular prism and the cube are projections of the learned composition policy into the particular object geometry.

into the number of states visited before ($\dot{\epsilon} \geq 0.01$ and $p_1 = 0$) and after ($\dot{\epsilon} < 0.01$ and $p_1 = 1$) convergence. Table 5.2 also lists the number of states visited for the top 90% most popular transitions, during learning and testing.

Table 5.2. Average and standard deviation (in parenthesis) of the number of states visited, (a) for all transitions, before convergence, (b) for the top 90% transitions, before convergence, (c) and (d) same, after convergence. Each line corresponds to one control primitive, as indicated. Top table shows variables observed during learning, while bottom table shows the same variables, as observed during test of learned policies.

Controllers	Average number of states visited during exploration			
	$\dot{\epsilon} \geq 0.01$		$\dot{\epsilon} < 0.01$	
	All trans. (a)	Top 90% (b)	All trans. (c)	Top 90% (d)
(T,1)	206.6 (7.86)	87.7 (3.03)	162.9 (7.93)	54.5 (3.04)
(T,2)	202.9 (6.36)	85.9 (2.49)	158.5 (6.31)	56.6 (3.11)
(1,2)	223.7 (9.47)	89.9 (3.64)	186.7 (7.22)	97.3 (4.62)
(T,1,2)	72.9 (3.20)	24.6 (1.50)	54.8 (2.81)	12.1 (0.64)

Controllers	Average number of states visited for learned policies			
	$\dot{\epsilon} \geq 0.01$		$\dot{\epsilon} < 0.01$	
	All trans. (a)	Top 90% (b)	All trans. (c)	Top 90% (d)
(T,1)	149.8 (24.6)	65.3 (10.8)	94.8 (29.1)	30.8 (11.2)
(T,2)	134.4 (30.6)	60.0 (11.5)	81.6 (30.6)	26.9 (12.3)
(1,2)	125.3 (39.7)	46.2 (18.8)	67.2 (49.1)	38.9 (24.8)
(T,1,2)	62.7 (3.90)	23.9 (1.81)	43.3 (5.47)	6.33 (2.67)

The upper bound on the total number of states given in Section 4.3.1 evaluates to 3220 states, corresponding to 35 models. As remarked earlier, the theoretical upper bound assumes that the models have unbounded domain, and each models may overlap with every other model. As expected, the number of visited states decreases dramatically when empirical models are used: in this case, the data shown in Table 5.2 indicates that in average a total of 1269.0 states are visited during learning, while only 759.1 states are visited during testing. When only the relevant transitions are considered, these numbers are further reduced to 508.6 and 298.33, respectively. Not surprisingly, the standard deviation for the number of states visited is smaller in the learning stage; action selection is mostly random and distinct runs will visit about the same number of states. Once a composition policy is available, fewer states are visited; the learned policy forces transitions between generic states into states associated with more positive outcomes. This funneling action of the learned policy ends up restricting the total number of states visited.

CHAPTER 6

LEARNING TO GRASP WITH PHYSICAL MANIPULATORS

The development of autonomous reaching and grasping systems is a challenging endeavor. This chapter describes grasp experiments involving the Stanford/JPL hand and its force/torque fingertip sensors. In these experiments, actual tactile information is used to implement the grasp controllers described earlier.

6.1 Haptic Models of Grasp Dynamics

The $\mathcal{C}^*$ controller is predicated on the assumption that contact position and normal are available, ideally through tactile feedback. In the pick and place experiments described in Section 3.5, no tactile feedback was available; instead, vision and proprioceptive information were used to derive contact normal and position estimates.

The experiments to be described shortly make use of tactile feedback. Brock sensors mounted at the fingertips of the JPL/Stanford hand were employed to deliver run-time contact feedback. The integration of the tactile sensors into the system proved to be a challenging task. It requires the calibration of the tactile sensors, the development of procedures for driving the tactile sensor to the object surface, procedures for gravity compensation as well as procedures for noise filtering. A detailed account of the development process can be found in [14]; however a brief review of the procedure used to manage the tactile probe operation is reviewed here for completeness. The system described in this dissertation is the first online grasp system based entirely on tactile information – most of its time is spent performing probing operations so facile tactile probing is critical and is the first challenge I undertook.

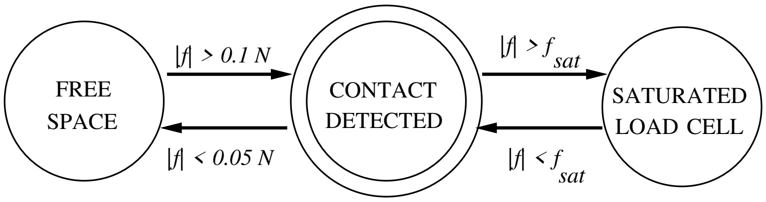

Figure 6.1. Sensory events during contact formation.

6.1.1 Tactile Probing Operation

Tactile probing takes the form of a sequence of contact and release operations. Contact forces must be managed for the duration of these operations in order to minimize unwanted displacements of the object and to improve the precision of contact position and normal estimates.

The quality of the sensor signal during contact formation and during manipulation is affected by several factors, including control parameters, actuator performance, and tendon dynamics. Moreover, the fingertip sensors may go through several sensory events before reaching a configuration in the grasp state may be measured. The sequence of sensory events is described by Figure 6.1. To guarantee good tracking characteristics while employing an existing PID tendon tension controller, the approach trajectory of each fingertip is resolved into relatively small incremental displacements and every differential movement is monitored. A contact event, with estimated positions and normals, is only generated if the *contact detected* state is achieved and maintained in Figure 6.1.

Detecting a contact load greater than $0.1N$ causes the transition from *free space* to the *contact detected* state. Transitions to the *saturated load cell* state are triggered whenever at least one of the eight strain gage channels of the Brock sensor becomes saturated. In the saturated condition, the contact parameters derived are not reliable. From the *contact detected* state, the system may also transition back to *free space* if the magnitude of the contact force decreases below $0.05N$. A valid grasp state can be observed only when all contacts are in the *contact detected* state simultaneously. Since the contact forces are coupled through the rigid object, a very delicate balance must be struck. In all cases, recursive filters are used over several measurements on the load cell to minimize noise when estimating force loads.

The approach trajectory of each fingertip sensor to the object surface proved to be as important as the management of the contact forces. In the solution adopted, each fingertip approaches the projected point of contact along an arc of a circle, computed so that the last displacement along the arc would cause the fingertip sensor to approach the object surface along its estimated normal. Both the projected point of contact and the normal to the object surface are computed based on the last tactile observation. For the first probe, the system assumes that the object is a cylindrical volume with a radius of roughly 5 cm. This simple approach is consistent with more complete path control treatments and allows the focus of this work to be on grasp synthesis.

The computational overhead of contact force monitoring and path control during contact formation is an additional burden on the grasp formation process. More complete approaches to path control may be incorporated into the platform. Moreover, skillful contact formation control can take tremendous advantage of passive compliance properties in the hand. Several designers are examining this issue in the form of new actuators in series with elastic members. Such technology promises to increase the dynamic range of sensing technology and to improve contact formation performance.

6.1.2 Construction of Generative Models of Grasp Dynamics

The first step in the implementation of the context-dependent grasp controller is the construction of generative models of grasp dynamics (see Section 4.2) on our experimental platform. The setup is similar to the setup for the pick-and-place task

Table 6.1. Summary of the empirical data acquired for model construction.

Object	# contacts	# trials	# probes	probes/trial
Triangular	2	20	207	10.4
Prism	3	27	352	13.0
Cube	2	20	148	7.4
	3	30	167	5.6
Cylinder	2	12	82	6.8
	3	20	213	10.7

described in Section 3.5, with the Utah/MIT hand replaced by the Stanford/JPL hand. The controllers for reaching in the pick-and-place task are employed in this experiment as well, following the adaptation of the kinematics to describe the Stanford/JPL hand.

In order to build the empirical models of grasp dynamics, a number of grasp trials were executed for each controller (involving two and three fingers) and each object type (cylinder, cube, and triangular prisms) – see Table 6.1. The data acquired in the grasp trials generate many models, some of them redundant. Figure 6.2 depicts the models for each object type, for the two-fingered version of $\mathcal{C}^*$. As in Figure 5.3, each model $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$ is represented by a polygon that delimits the region of phase space in which $p_k = 1$ (see Section 4.3). Models with an associated prior probability below 5% are colored light gray.

Figure 6.3 depicts the acquired models for the three-contact version of the grasp controller $\mathcal{C}^*$. The lower right plot in both figures shows the overlap of all models. A direct comparison between the models derived based on simulated data (Figures 5.3 and 5.4) and the models derived based on grasp trials involved real objects (Figures 6.2 and 6.3) reveals interesting contrasts.

- Models derived from physical grasp trials are wider than their counterparts derived in simulation; this reflects the noise levels in excess of that modeled in simulation;

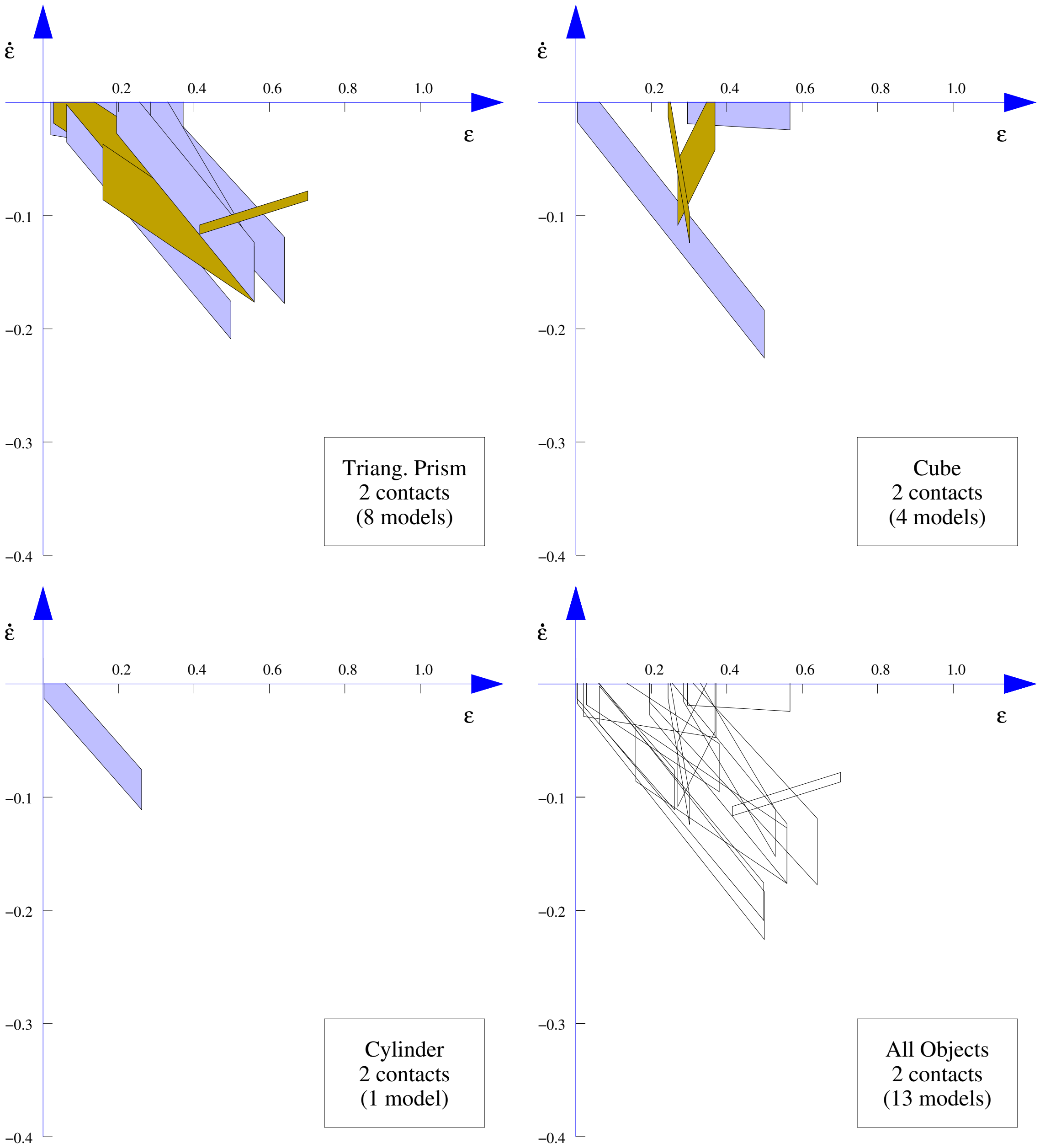

Figure 6.2. Generative models for 2 contact grasp controllers, for each object type. Each model $M_{\pi_i}(\boldsymbol{\psi}_k, \mathbf{o})$ is represented by a polygon that delimits the region of phase space in which $p_k = 1$.

- The scope of simulation-based models (measured in terms of the range of values for both ϵ and $\dot{\epsilon}$) is larger than for real models. This is a reflection of the extreme situations to which grasp controllers are submitted in simulation, especially for two-fingered grasps. Such extreme cases are rare whenever the grasps involve real objects and real hand configurations.

- There is agreement between the models for both cases, especially if only the relevant models (those with a prior probability greater than 5% and colored in dark grey in Figures 5.3, 5.4, 6.2 and 6.3) are considered. All relevant models obtained in simulation have roughly the same linear and angular coefficients as the corresponding models derived from physical grasps.

6.2 Context-Dependent Grasp Policies

The general agreement between the models constructed in simulation and in actual experiments suggests that it might be possible to implement the grasp policy acquired in simulation on the real robot hand/arm system. All experiments were performed on the platform consisting of a five degree-of-freedom (DOF) GE P50 robotic arm and the Stanford/JPL hand as shown in Figure 6.4.

The same control structure depicted in Figure 3.12 was used in the experiments. The object is placed in the robot's reachable workspace at a random orientation. Its position is estimated from an overhead image of the workspace and the robot arm reaches to the object using a harmonic function-based path planner [15]. After the reach operation, the robot hand is in a configuration such that the object surface is within the workspace of the participating fingers. The first tactile probe returns the first state observation, and the learned composition policy determines the next grasp controller to be applied. The trial proceeds until the grasp policy terminates or the limit of 50 tactile probes is reached.

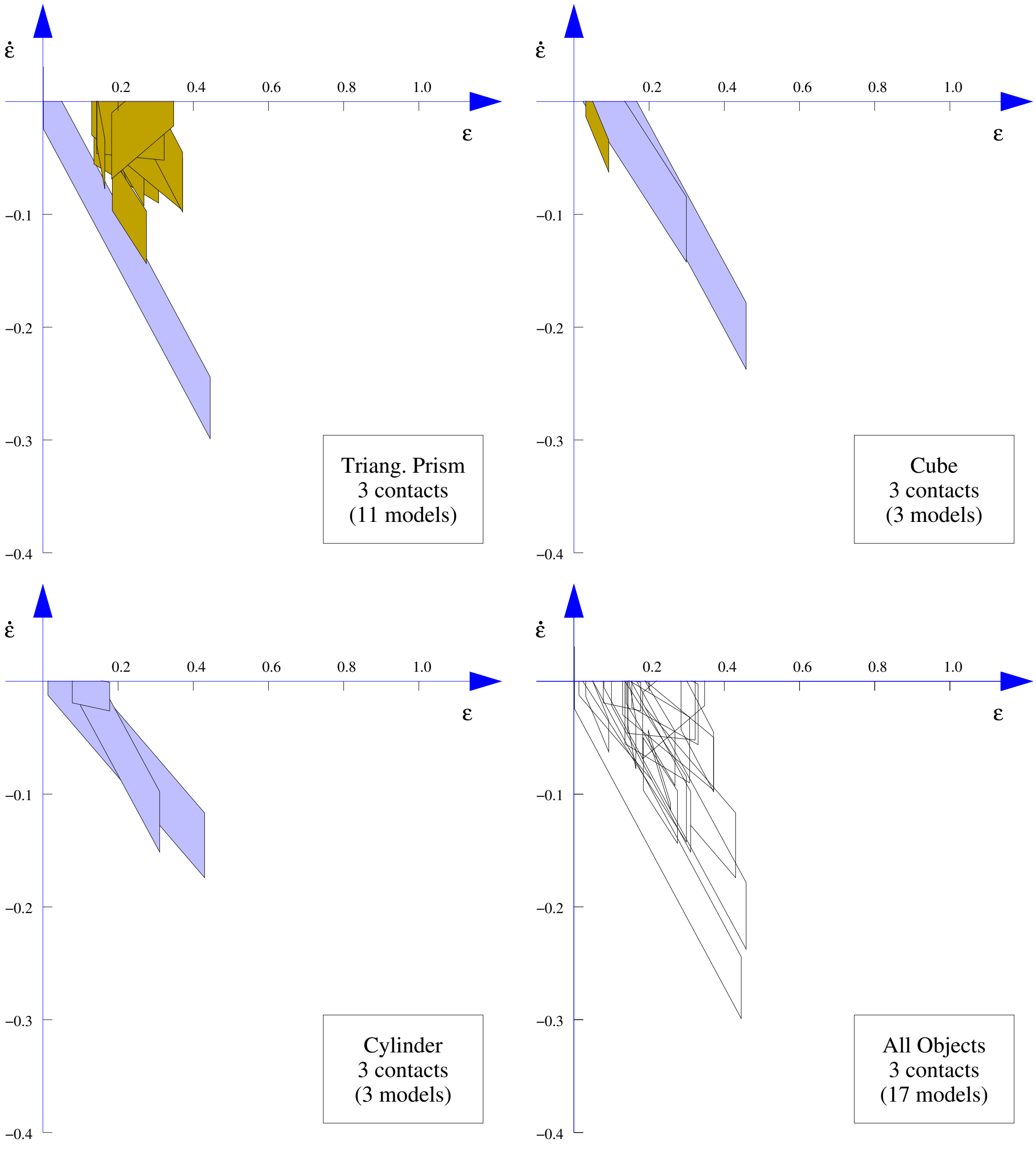

Figure 6.3. Generative models for 3 contact grasp controllers, for each object type.

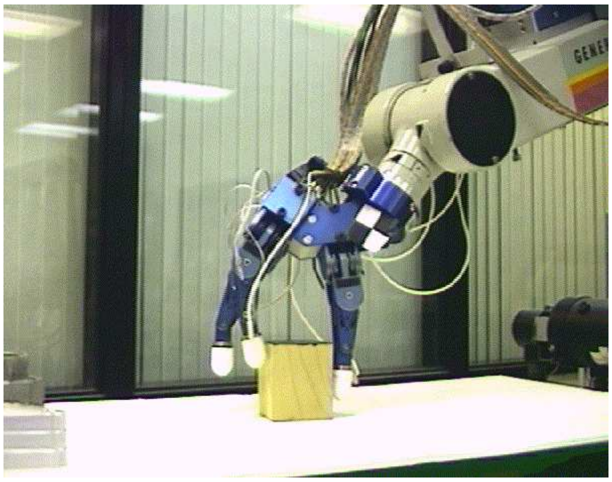

Figure 6.4. Setup for the grasp task.

6.2.1 Stanford/JPL Hand

The Stanford/JPL hand was designed by Ken Salisbury [55]. It has three fingers, each one with three degrees of freedom. Our hand is equipped with Brock tactile sensors that consist of a six axis force-torque fingertip sensor with a sensor geometry that allows the location of a single point contact and the respective contact normal to be estimated as described by Bicchi [5]. Contact normals and locations are transformed from the fingertip coordinate frame to the object's coordinate frame as required by the parametric $\mathcal{C}^*$ controller.

6.2.2 Objects

The test object is selected at random from a set containing a cylinder, a cube, and a triangular prism. The objects are wood blocks with heights of 10 cm (cube and triangular prism) and 20 cm (cylinder). As mentioned earlier, the noise associated with the tactile sensing operation as well as non-linearities in the actuation of the finger joint angles cause the tactile signatures of the same object to differ from trial to trial.

6.2.3 Observed Behavior

A number of grasp trials were performed involving all objects in the test set. The emphasis in the experimental effort was placed on trials involving the cube; unlike cylinders, cube-like objects afford interesting grasp contexts in which the careful selection of grasp primitives yields high utility.

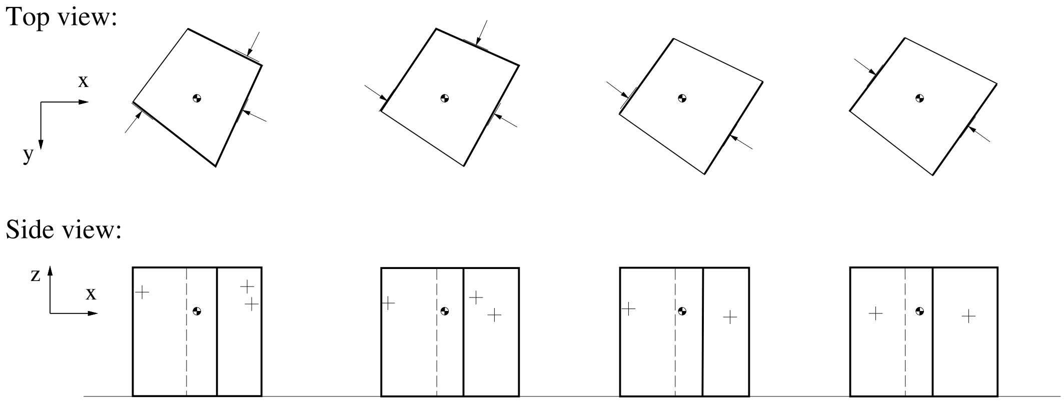

Figure 6.5. Grasp sequence for a cube.

Figure 6.5 depicts a grasp sequence observed for the cube, consisting of four tactile probes. The grasp sequence illustrated is typical for successful grasps which generally require a small number of tactile probes and end up deploying two fingers in the final grasp configuration for a cube. The tactile readings (contact position and normal) for the first tactile probe are usually noisier than subsequent probes. This is due to the prior assumption that the object geometry is a blob of cylindrical shape. The resulting inaccuracy in the finger trajectory and consequent contact load on the sensor translates into excess friction forces, affecting the accuracy of the tactile sensor. Notice that the perceived contact geometry varies from probe to probe; this is caused by measurement noise and by small displacements induced by the probing operation.

Table 6.2 lists the number of probes and the score $(1-\mu_0)$ for a set of six additional grasp trials in which the learned composition policy is employed in the physical hand-arm system. The cube is the object grasped in all trials; in all cases, the system

Table 6.2. Performance characterization for the learned composition policy.

Trial number	# probes	Grasp score
1	2	0.81
2	4	0.85
3	4	0.75
4	8	0.74
5	3	0.78
6	2	0.73

concludes the trial in a two-fingered grasp configuration. The system performed well under conditions of moderate noise; unfortunately, each experiment requires a good deal of preparation, including the careful homing and fine tuning of joint angle control parameters.

The complexity of the tactile probing operation is not conveyed in Figure 6.5. First off, before a clean data set can be obtained, it is necessary to home the finger joints. The quality of contact observations is sensitive to the accuracy of the spatial transforms that translate contact frame wrenches to object frame wrenches. The Stanford/JPL hand does not measure joint angles directly; instead the angular displacements of the motors and the tension on the tendons is used to infer a hand configuration. This design is consistent with the apparent lack of direct joint angle feedback in the human hand. Contact normals and positions can only be measured if one knows the joint angles at some point in time precisely. Thus, particular care had to be exercised at the beginning of each trial to deliver the hand to a known configuration. Moreover, during tactile probing the fingertips must touch the object surface very lightly in order to minimize disturbances on the object. Unfortunately, stiction in the tendons caused the finger joints to exercise jerky movements. Even though the grasp controllers are closed-loop and the resulting strategy can deal with disturbances, if the fingers displace the object during every probe, the whole interaction can be unnecessarily prolonged. Finally, internal forces in the grasp couple the fingers so that the placement of the second finger will press the object against the first finger and sometimes causes the saturation of its tactile sensor (see Figure 6.1). Contact forces are reduced by moving the fingers away from the object surface normal. If this load adjustment is too aggressive, contact may be lost and no useful observations are gathered.

Under unfavorable conditions of noise or uncertainty, progress (measured as an incremental decrease in the wrench residual, ϵ) is not necessarily monotonic over consecutive tactile probes. Moreover, very noisy measurements may not find support in the set of dynamic models and the corresponding novelty as reflected in the predicate vector ($\mathbf{q} = \mathbf{0}$, more often). In these cases, the composition policy will adopt the controller found to be the most effective in situations of no information – this may disrupt the grasp process and lead to an increase in the number of probes required. The solutions adopted aimed only at ameliorating bad contact state feedback; additional effort on the hardware are required including the re-design of actuators or the deployment of other tactile sensing technology. The limited number of trials performed were insufficient to characterize the noise conditions that could compromise system performance.

CHAPTER 7
CONCLUSION

The results reported in this document demonstrate that it is possible to approximate optimal grasp policies for individual objects, even if factors such as object identity, object orientation, and sensor noise are not explicitly part of the state vector. In spite of the progress reported in this thesis, many aspects of the grasping task remain to be addressed. Broader problems include:

Prospective Control: In principle, the knowledge structures acquired within a task domain (e.g., grasping an object) can be propagated backwards in time to previous tasks (e.g., reaching for an object with the intent of grasping it). In the grasp domain, for example, system performance can be incremented through the use of reaching as an anticipatory control action; similarly grasp configurations can be selected in such a way that the fingers do not interfere with possible assembly operations to be executed later. Anticipatory control actions could position the system upstream of a good solution, and are similar in spirit to backward chaining problem solving strategies studied within Artificial Intelligence. These actions are also observed in the sensorimotor development of infants, and are described in the Psychology literature as *prospective behavior* [4], the ability to select actions while accounting for constraints that will arise in the future.

Use of Multimodal Information: Context, as described in this dissertation, is derived essentially from haptic information. In principle, "visual contexts" can also be defined, and they could impact system performance a great deal by enriching the information available to the agent. The integration between visual and haptic contexts is an intriguing research topic in itself.

Flexible Scheduling of Grasp Resources: There are many other grasp resources besides robotic fingers – in fact, any robot surface qualifies as a grasp resource. Object scale plays a role in determining the appropriate set of resources to use; a good example is the grasp of a beach ball, that requires the use of two arms. "Scale" in geometry and force suggests variations in prehensile strategy that are independent of "shape". As a sphere geometry progresses in scale between a ball bearing and a beach ball, there are discrete resource combinations that are likewise recommended. The ball bearing is pinched between the thumb and the index fingertips, the grapefruit may begin to enlist proximal phalanges, the basketball, two hands, and the beach ball (particularly if it is heavy) may employ forearm and torso. The models employed in the dissertation refer to force and moment residuals and their rate of change by virtue of control; as such they may be applicable to classes of tasks and somewhat independent of the resources. Another important dimension to the scheduling problem is the interference introduced by constraints such as previous resource commitments, sensing and kinematic constraints. The development of a uniform mechanism for the scheduling of grasp resources has yet to be addressed.

The control architecture described in this thesis can also be improved upon in specific ways, especially in regard to the representation of system dynamics. More flexible parametric models of system dynamics, capable of expressing how the uncertainty varies with respect to $\dot{\epsilon}$, could remove some of the ambiguity present in the system and yield better performance. In the same vein, the system could accommodate a growing number of generative models, simply by padding the membership predicate vector $\mathbf{q}$ with extra entries and by allowing three logic levels in each one of its entries:

- $p_k = 0$: observation not supported by model k;
- $p_k = 1$: observation supported by model k;
- $p_k = X$: model k not in use.

7.1 Contributions

The contributions of this dissertation can be categorized by two major themes: the representation of continuous sensorimotor processes in term of discrete and computationally tractable state estimates, and the synthesis of robust multifingered grasp strategies on this principled representation.

7.1.1 State Estimation for Continuous Sensorimotor Processes

The state representation is innovative in using feedback control primitives to elicit patterns of interaction with the environment and in using the dynamic models to construct state information. Many of the properties of the state representation are inherited from the control laws from which it is derived:

- The state representation is suitable for implementations involving learning on real robots; the elements of the control basis Π can be designed to maintain safety at all times, by constraining the set of configurations the robot may adopt.
- It supports the re-use of existing knowledge expressed in the form of feedback control laws.
- It accommodates sensor noise, as it is based on stochastic models of system behavior.

- The state distinctions formed are independent of the subsequent learning process; therefore, the state representation compares favorably against the representation proposed by McCallum [39], where the distinctions formed are dependent upon the exploration policy used.

- It permits the segmentation of experience in terms of state transitions; the temporal granularity obtained is an intermediate solution between the extremes proposed in the literature, namely the use of control convergence events or clock ticks as decision points where the agent may change its control strategy. This work advocates using asynchronous events characterized by the availability of new information as choice points instead.

7.1.2 Multifingered Grasping

The $\mathcal{C}^*$ controller described in [11] is the first closed-loop grasp controller described in the literature; it is also the only grasp controller based exclusively on haptic information. Furthermore, the $\mathcal{C}^*$ grasp controller is provably optimal for regular, convex objects, as it will always reach a grasp configuration requiring the least frictional forces for grasp stability, over the whole set of grasp configurations. It is also one of very few manipulation controllers that has been fully implemented [14]. It is robust with respect to the manipulation device, and has been implemented on the Utah/MIT hand, the Stanford/JPL robot hand, and Thing (UMass/LPR's 4-legged walking platform). It has been used in grasping [14], manipulation [23] and walking [27] tasks. It advances the state-of-the-art in terms of multifingered grasp synthesis in several tangible ways.

- It generates grasp strategies that are robust with respect to sensory/actuation noise, due to their formulation as feedback controllers. Moreover, their sensory-driven formulation makes incremental grasp refinements robust with respect to global object geometry.

- Use of realistic sensors: the grasp agent should rely on realistic sensors that can be mounted on actual manipulators. Current research typically assumes the existence of sensors for object identity or geometry. In some extremes, an algebraic description of object geometry is assumed.

- It relies on local models describing the dependence of wrench residuals on contact displacement, yielding a uniform formulation for 2D and 3D grasp tasks, and permitting a variable number of point contacts.

- The resulting prototypical solutions incorporate sensory limitations and kinematic constraints, and can be used as an effective abstraction for the grasping process in planning architectures.

7.2 Insights

Expert-level performance of sensorimotor activities requires fine coordination of sensors, actuators and computational resources. This dissertation provides insights into how an agent with a limited action repertoire, sensor noise, and limited memory could solve some of the problems encountered during its sensorimotor development. In particular, the following observations can be made regarding this study.

- **Situated representations for sensorimotor activities:** Chapter 4 describes how an agent can construct a representation of its interactions with physical objects. The representation described accommodates noisy measurements and it can be acquired automatically by the agent as it interacts with the environment.

- **Integration of information gathering and control actions:** A fundamental relationship between "effecting" a grasp and observing the efficacy of the grasping strategy is exploited in this dissertation. This relationship is formalized in the interaction-based state estimator, wherein hidden state estimation is acquired by observing temporal histories of the interaction and used to inform control decisions about the critical features of rewarding grasp policies. This is a new and powerful aspect of learning with a potentially profound impact on the way we view control and perception – it is also a particularly satisfying aspect of the study.

- **Temporal scheduling of computational resources:** The rate at which control decisions are made must balance responsiveness to novel situations and the overhead incurred in re-evaluating the current context too often. Control decisions should be located at points in time when feedback regarding utility is available and not simply at the tick of a system clock or at the termination of a previous action. The first commits too litle to the governing dynamics and the latter is too often blind to critical information. The state representation introduced in Chapter 4 allows the agent to use changes in the available information as the triggering events for the re-evaluation of the current context.

- **Perceptual categorization:** Autonomous agents must be able to identify and reason about features in the environment that predict its success in carrying out certain tasks; however, categorization is rarely a goal in itself. External observers may infer that the system described in this dissertation classifies objects, as it displays characteristic grasp strategies for each object. Nevertheless, object identity is not a control parameter, and the system does not attempt to classify/categorize objects during the grasp process. Categories are formed to distinguish contexts in available observations that affect utility – there is no prior commitment to the dimensions or purpose of categorical distinction. This approach proposed a principled and uniform theory to cognitive organization that has not been considered before.

The contributions of this dissertation support a rich framework constructed to study the sensorimotor development of humanoid robots [13]. The main objective of this investigation is to understand how interaction leads to representation and how representation leads to problem solving behavior, an important aspect that (maybe due to its complexity) remains to be addressed by the intelligent system research community. This dissertation offers answers on how haptic categories can be formed and represented, within a control architecture that also addresses issues pertinent to the operation of real robots.

STATE EVOLUTION IN PHASE SPACE

The evolution of the observation $(\epsilon, \dot{\epsilon})$ in phase space can be solved in closed form, if the (local) object geometry is known. In this appendix, the case corresponding to three contacts interacting with three planar facets is solved. It is assumed that the contacts do not leave the facets in the course of the grasp operation; each contact is equipped with tactile sensors capable of reporting the contact normal and position, without any measurement noise.

Each contact i elicits a wrench $\mathbf{w_i}$ at the object's reference frame, where $\mathbf{w_i} = (\mathbf{f_i}, \mathbf{r_i} \times \mathbf{f_i})$ is a six dimensional vector containing the force and moments components induced by contact i. It is convenient to express the wrench vector in terms of the facet coordinates, indicated as $\hat{\mathbf{r}}$ in Figure 1. $\hat{\mathbf{r}}_\mathbf{i}$ can be easily computed as the difference vector between the original vector $\mathbf{r_i}$ and its projection along the normal $\mathbf{f_i}$:

$$\hat{\mathbf{r}}_\mathbf{i} = \mathbf{r_i} - (\mathbf{r_i} \cdot \mathbf{f_i})\mathbf{f_i}$$

The origin $\mathbf{r_{0i}} = (\mathbf{r_i} \cdot \mathbf{f_i})\mathbf{f_i}$ of the facet coordinate system is defined as the point for which the resulting moment is zero $(\mathbf{r_i} \times \mathbf{f_i} = \mathbf{r_{0i}} \times \mathbf{f_i} = \mathbf{0})$. It follows that $\mathbf{w_i} = (\mathbf{f_i}, (\mathbf{r_{0i}} + \hat{\mathbf{r}}_\mathbf{i}) \times \mathbf{f_i}) = (\mathbf{f_i}, \hat{\mathbf{r}}_\mathbf{i} \times \mathbf{f_i})$.

For n contacts, the squared residual wrench is defined as

$$\epsilon = (\frac{1}{n}\sum_{i=1}^{n} \mathbf{w_i})^T (\frac{1}{n}\sum_{i=1}^{n} \mathbf{w_i}).$$

Therefore, the residual ϵ is a function of the position of the n contacts. Since the contacts are constrained to be located at the object surface, ϵ is a function of $2n$ variables, or $(\hat{\mathbf{r}}_\theta, \hat{\mathbf{r}}_\phi)_\mathbf{i}$, for $i = 1, \ldots, n$.

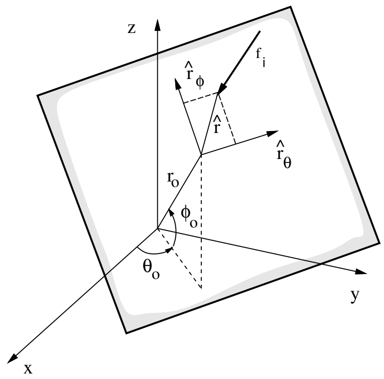

Figure 1. Local coordinate system for each contact.

The objective of the grasp control actions is to minimize the squared residual ϵ, and the contact displacements are made proportional to the gradient $\nabla\epsilon$:

$$\left\{ \begin{array}{rcl} \frac{\mathrm{d}\hat{\mathbf{r}}_{\theta_\mathbf{i}}}{\mathrm{d}t} & = & -\alpha\frac{\partial\epsilon}{\partial\hat{\mathbf{r}}_{\theta_\mathbf{i}}} \\ \frac{\mathrm{d}\hat{\mathbf{r}}_{\phi_\mathbf{i}}}{\mathrm{d}t} & = & -\alpha\frac{\partial\epsilon}{\partial\hat{\mathbf{r}}_{\phi_\mathbf{i}}} \end{array} \right.$$

In the implementation of the grasp controller, a distinct (but equivalent) formulation was employed. This formulation transforms a unimodal (but potentially non-convex) error metric into a convex metric, and yields better numerical estability for the case in which the magnitude of one gradient component is much larger than the other components:

$$\left\{ \begin{array}{rcl} \frac{\mathrm{d}\hat{\mathbf{r}}_{\theta_\mathbf{i}}}{\mathrm{d}t} & = & -\alpha\frac{(\epsilon-\epsilon^*)}{\sum_{i=1}^n(\hat{\mathbf{r}}_{\theta_\mathbf{i}}-\hat{\mathbf{r}}^*_{\theta_\mathbf{i}})^2+(\hat{\mathbf{r}}_{\phi_\mathbf{i}}-\hat{\mathbf{r}}^*_{\phi_\mathbf{i}})^2}(\hat{\mathbf{r}}_{\theta_\mathbf{i}}-\hat{\mathbf{r}}^*_{\theta_\mathbf{i}}) \\ \frac{\mathrm{d}\hat{\mathbf{r}}_{\theta_\mathbf{i}}}{\mathrm{d}t} & = & -\alpha\frac{(\epsilon-\epsilon^*)}{\sum_{i=1}^n(\hat{\mathbf{r}}_{\theta_\mathbf{i}}-\hat{\mathbf{r}}^*_{\theta_\mathbf{i}})^2+(\hat{\mathbf{r}}_{\phi_\mathbf{i}}-\hat{\mathbf{r}}^*_{\phi_\mathbf{i}})^2}(\hat{\mathbf{r}}_{\phi_\mathbf{i}}-\hat{\mathbf{r}}^*_{\phi_\mathbf{i}}), \end{array} \right. \tag{1}$$

where ϵ^* is the minimum value achievable for ϵ, and $(\hat{\mathbf{r}}^*_{\theta_\mathbf{i}}, \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})$ are the corresponding coordinates for $\epsilon = \epsilon^*$. For the case of n contacts interacting with planar facets, the calculation of ϵ^* is straightforward: given that the contacts do no leave the planes defined by $\hat{\mathbf{r}}_\mathbf{i} \cdot \mathbf{f}_\mathbf{i} = \mathbf{0}$, the minimum value for ϵ is $\epsilon = (\sum_{i=1}^n \mathbf{f}_\mathbf{i}/n)^T(\sum_{i=1}^n \mathbf{f}_\mathbf{i}/n)$, that is, ϵ^* corresponds to the squared wrench residual when the residual moment is zero.

1 Determination of Coordinates $(\hat{\mathbf{r}}^*_{\theta_\mathbf{i}}, \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})$

The coordinates $(\hat{\mathbf{r}}^*_{\theta_\mathbf{i}}, \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})$ correspond to the situation in which $\sum_{i=1}^{n} \hat{\mathbf{r}}_\mathbf{i} \times \mathbf{f}_\mathbf{i} = \mathbf{0}$. This equation establishes three constraints $(m_x = m_y = m_z = 0)$ that must be satisfied for $\epsilon = \epsilon^*$. The total number of independent control variables is $2n$, and therefore the calculation of $(\hat{\mathbf{r}}^*_{\theta_\mathbf{i}}, \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})$ requires the determination of the remaining $2n-3$ variables. The determination of these variables can be accomplished in many ways; it is in fact an opportunity to address secondary control objectives, such as kinematic conditioning of hand or manipulator, or others. In this appendix, it is assumed that the secondary control objective is the minimization of the squared total distance traveled by the contacts during grasping, defined as:

$$D = \sum_{i=1}^{n} (\hat{\mathbf{r}}^{\mathbf{0}}_{\theta_\mathbf{i}} - \hat{\mathbf{r}}^*_{\theta_\mathbf{i}})^\mathbf{2} + (\hat{\mathbf{r}}^{\mathbf{0}}_{\phi_\mathbf{i}} - \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})^\mathbf{2},$$

where $(\hat{\mathbf{r}}^{\mathbf{0}}_{\theta_\mathbf{i}}, \hat{\mathbf{r}}^{\mathbf{0}}_{\phi_\mathbf{i}})$ are the initial coordinates for the contact i. D is a function of $2n-3$ variables, and reaches its minimum when $\nabla\mathbf{D} = \mathbf{0}$. Therefore, the determination of the coordinate variables $(\hat{\mathbf{r}}^*_{\theta_\mathbf{i}}, \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})$ requires the solution of a linear system of $2n-3$ equations.

2 Integration of Dynamic Differential Equations

In the integration of the system of differential equations 1 it will be initially assumed that the factor

$$k = \alpha \frac{(\epsilon - \epsilon^*)}{\sum_{i=1}^{n} (\hat{\mathbf{r}}_{\theta_\mathbf{i}} - \hat{\mathbf{r}}^*_{\theta_\mathbf{i}})^\mathbf{2} + (\hat{\mathbf{r}}_{\phi_\mathbf{i}} - \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})^2} \qquad (2)$$

is a constant. Later, this assumption will have to be verified.

If

$$\begin{cases} \frac{\mathrm{d}\hat{\mathbf{r}}_{\theta_\mathbf{i}}}{\mathrm{d}t} & = & -k(\hat{\mathbf{r}}_{\theta_\mathbf{i}} - \hat{\mathbf{r}}^*_{\theta_\mathbf{i}}) \\ \frac{\mathrm{d}\hat{\mathbf{r}}_{\theta_\mathbf{i}}}{\mathrm{d}t} & = & -k(\hat{\mathbf{r}}_{\phi_\mathbf{i}} - \hat{\mathbf{r}}^*_{\phi_\mathbf{i}}), \end{cases}$$

then it follows that

$$\begin{cases} \hat{\mathbf{r}}_{\theta_\mathbf{i}}(\mathbf{t}) & = & \hat{\mathbf{r}}^*_{\theta_\mathbf{i}} + (\hat{\mathbf{r}}^\mathbf{0}_{\theta_\mathbf{i}} - \hat{\mathbf{r}}^*_{\theta_\mathbf{i}})e^{-kt} \\ \hat{\mathbf{r}}_{\phi_\mathbf{i}}(\mathbf{t}) & = & \hat{\mathbf{r}}^*_{\phi_\mathbf{i}} + (\hat{\mathbf{r}}^\mathbf{0}_{\phi_\mathbf{i}} - \hat{\mathbf{r}}^*_{\phi_\mathbf{i}})e^{-kt}. \end{cases}$$

One can now express $\hat{\mathbf{r}}_\mathbf{i}$ as a function of time:

$$\hat{\mathbf{r}}_\mathbf{i}(\mathbf{t}) \quad = \quad \hat{r}_{\theta_i}(t)\hat{\theta}_i + \hat{r}_{\phi_i}(t)\hat{\phi}_i, \tag{3}$$

where $\hat{\theta}_\mathbf{i}$ and $\hat{\phi}_\mathbf{i}$ are the unit vectors for the surface coordinate system. If $||r_0||$ is the magnitude of r_{0_i}, and $||r_{0xy}||$ the magnitude of the vector $(r_{0x}, r_{0y}, 0)$, then

$$\begin{aligned} \hat{\theta}_\mathbf{i} &= \left(\frac{-r_{0y}}{||r_{0xy}||}, \frac{r_{0x}}{||r_{0xy}||}, 0\right) \\ \hat{\phi}_\mathbf{i} &= \left(\frac{-r_{0z}r_{0x}}{||r_{0xy}||\,||r_0||}, \frac{-r_{0z}r_{0y}}{||r_{0xy}||\,||r_0||}, \frac{||r_{0xy}||}{||r_0||}\right). \end{aligned}$$

Replacing $\hat{\theta}_\mathbf{i}$ and $\hat{\phi}_\mathbf{i}$ into Equation 3, $\hat{\mathbf{r}}_\mathbf{i}(\mathbf{t})$ results

$$\hat{\mathbf{r}}_\mathbf{i}(\mathbf{t}) = \begin{bmatrix} \hat{\mathbf{r}}^*_{\mathbf{x_i}} + (\hat{\mathbf{r}}^\mathbf{0}_{\mathbf{x_i}} - \hat{\mathbf{r}}^*_{\mathbf{x_i}})e^{-kt} \\ \hat{\mathbf{r}}^*_{\mathbf{y_i}} + (\hat{\mathbf{r}}^\mathbf{0}_{\mathbf{y_i}} - \hat{\mathbf{r}}^*_{\mathbf{y_i}})e^{-kt} \\ \hat{\mathbf{r}}^*_{\mathbf{z_i}} + (\hat{\mathbf{r}}^\mathbf{0}_{\mathbf{z_i}} - \hat{\mathbf{r}}^*_{\mathbf{z_i}})e^{-kt} \end{bmatrix}.$$

3 Calculation of $\epsilon(t)$

The moments induced by contact i can be computed by the cross product $\mathbf{m_i} = \hat{\mathbf{r}}_\mathbf{i}(\mathbf{t}) \times \mathbf{f_i}$:

$$\mathbf{m_i}(\mathbf{t}) = \begin{bmatrix} m_{x_i} \\ m_{y_i} \\ m_{z_i} \end{bmatrix} = \begin{bmatrix} \hat{r}_{y_i}f_{z_i} - \hat{r}_{z_i}f_{y_i} \\ \hat{r}_{z_i}f_{x_i} - \hat{r}_{x_i}f_{z_i} \\ \hat{r}_{x_i}f_{y_i} - \hat{r}_{y_i}f_{x_i} \end{bmatrix} = \begin{bmatrix} m^*_{x_i} + (m^0_{x_i} - m^*_{x_i})e^{-kt} \\ m^*_{y_i} + (m^0_{y_i} - m^*_{y_i})e^{-kt} \\ m^*_{z_i} + (m^0_{z_i} - m^*_{z_i})e^{-kt} \end{bmatrix}.$$

The squared wrench residual is defined as

$$\begin{aligned}
\epsilon(t) &= (\frac{1}{n}\sum_{i=1}^{n}\mathbf{w_i})^{\mathbf{T}}(\frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{w_i}) \\
&= \frac{1}{n^2}\left((\sum_{i=1}^{n} f_{x_i})^2 + (\sum_{i=1}^{n} f_{y_i})^2 + (\sum_{i=1}^{n} f_{z_i})^2 + (\sum_{i=1}^{n} m_{x_i})^2 + (\sum_{i=1}^{n} m_{y_i})^2 + (\sum_{i=1}^{n} m_{z_i})^2\right) \\
&= \epsilon^* + \frac{1}{n^2}\left((\sum_{i=1}^{n} m_{x_i})^2 + (\sum_{i=1}^{n} m_{y_i})^2 + (\sum_{i=1}^{n} m_{z_i})^2\right).
\end{aligned}$$

By definition $\sum_{i=1}^{n} m_{x_i}^* = \sum_{i=1}^{n} m_{y_i}^* = \sum_{i=1}^{n} m_{z_i}^* = 0$, and $\epsilon(t)$ can be simplified to

$$\epsilon(t) = \epsilon^* + \frac{1}{n^2}\left((\sum_{i=1}^{n} m_{x_i}^0)^2 + (\sum_{i=1}^{n} m_{y_i}^0)^2 + (\sum_{i=1}^{n} m_{z_i}^0)^2\right) e^{-2kt} \qquad (4)$$

One can now show by inspection that k is indeed a constant. All terms in the denominator and numerator of Equation 2 have a common multiplicative factor, e^{-2kt}. With the elimination of this common factor, all remaining terms in k are constants:

$$k = \frac{\alpha}{n^2}\frac{(\sum_{i=1}^{n} m_{x_i}^0)^2 + (\sum_{i=1}^{n} m_{y_i}^0)^2 + (\sum_{i=1}^{n} m_{z_i}^0)^2}{\sum_{i=1}^{n}(\mathbf{\hat{r}^0_{\theta_i}} - \mathbf{\hat{r}^*_{\theta_i}})^{\mathbf{2}} + (\mathbf{\hat{r}^0_{\phi_i}} - \mathbf{\hat{r}^*_{\phi_i}})^{\mathbf{2}}} \quad \square$$

4 Evolution of $(\epsilon, \dot{\epsilon})$ in Phase Space

The closed form solution for ϵ can be used to derive the path in phase space corresponding to the grasp process, from the initial contact configuration to the convergent configuration. If ϵ is given by Equation 4, then

$$\dot{\epsilon} = \frac{-2k}{n^2}\left((\sum_{i=1}^{n} m_{x_i}^0)^2 + (\sum_{i=1}^{n} m_{y_i}^0)^2 + (\sum_{i=1}^{n} m_{z_i}^0)^2\right) e^{-2kt}$$

Eliminating the time variable from the expressions for ϵ and $\dot{\epsilon}$, one gets:

$$\dot{\epsilon} = -2k(\epsilon - \epsilon^*).$$

Therefore, under the assumptions of this appendix, the observation $(\epsilon, \dot{\epsilon})$ will evolve in phase space along a line, with slope $-2k$ and linear coefficient $2k\epsilon^*$.

BIBLIOGRAPHY

[1] Aeyels, Dirk. On the dynamic behaviour of the novelty detector and the novelty filter. In *Analysis of Controlled Dynamical Systems* (1990), B. Bonnard, B. Bride, J.P. Gauthier, and I. Kupka, Eds., pp. 1 – 10.

[2] Allen, P.K. Mapping haptic exploratory procedures to multiple shape representations. In *Proceedings of the 1990 Conference on Robotics and Automation* (Cincinnati, OH, May 1990), vol. 3, IEEE, pp. 1679–1684.

[3] Araujo, Elizeth G., and Grupen, Roderic A. Learning control composition in a complex environment. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (Cape Cod, MA, September 1996), Pattie Maes, Maja Mataric, Jean-Arcady Meyer, Jordan Pollack, and Stewart W. Wilson, Eds., MIT Press/Bradford Books, pp. 333–342.

[4] Bertenthal, B.I., and Clifton, R.K. Perception and action. In *Handbook of Child Psychology - Cognition, Perception, and Language*, D. Kuhn and Siegler R., Eds., vol. 2. Wiley, New York, NY, 1997.

[5] Bicchi, A., Salisbury, J.K., and Brock, D.L. Contact sensing from force measurements. *International Journal of Robotics Research 12*, 3 (1993), 249–262.

[6] Chen, I., and Burdick, J.W. Finding antipodal point grasps on irregularly shaped objects. In *Proc. 1992 IEEE Int. Conf. Robotics Automat.* (Nice, FRANCE, May 1992), vol. 3, pp. 2278–2283.

[7] Chen, N.N., Zhang, H., and Rink, R.E. Touch-driven robot control using a tactile jacobian. In *Proc. 1997 IEEE Int. Conf. Robotics Automat.* (Albuquerque, NM, April 1997), vol. 3, pp. 1737–1742.

[8] Chiang, A.C. *Elements of Dynamic Optimization.* McGraw-Hill, Inc., New York, NY, 1992.

[9] Chrisman, L. Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In *Proceedings of the Tenth National Conference on Artificial Intelligence* (Cambridge, MA, 1992), AAAI/MIT Press.

[10] Coelho Jr., J.A., Connolly, C.I., Grupen, R.A., Huber, M., and Souccar, K. Experiments in autonomous reaching and grasping. In *Video Proc. 1994 IEEE Int. Conf. Robotics Automat.* (San Diego, CA, May 1994).

[11] Coelho Jr., J.A., and Grupen, R.A. A control basis for learning multifingered grasps. *Journal of Robotic Systems 14*, 7 (1997), 545–557.

[12] Coelho Jr., J.A., Grupen, R.A., Huber, M., MacDonald, W., and Souccar, K. Robust control basis for coordinating multiple manipulators. In *Video Proc. 1995 IEEE Int. Conf. Robotics Automat.* (Nagoya, JAPAN, 1995).

[13] Coelho Jr., J.A., Piater, J.H., and Grupen, R.A. Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robot. *To appear in the Journal of Robotic and Autonomous Systems* (2001).

[14] Coelho Jr., J.A., Souccar, K., and Grupen, R.A. A control basis for haptically-guided grasping and manipulation. Tech. Rep. CMPSCI Technical Report 98-46, Dept. Computer Science, University of Massachusetts, Sept. 1998.

[15] Connolly, C. I., and Grupen, R.A. The applications of harmonic functions to robotics. *Journal of Robotic Systems 10*, 7 (Oct. 1993), 931–946.

[16] Cutkosky, M.R. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Trans. Robotics Automat. 5*, 3 (June 1989).

[17] Deng, L. Switching dynamic-system models for speech articulation and acoustics. In *IMA Workshop on Mathematical Foundations of Speech Processing and Recognition* (Minneapolis, MN, Sept. 2000).

[18] Duffy, J. The fallacy of modern hybrid control theory that is based on orthogonal complements of twist and wrench spaces. *Journal of Robotic Systems* (April 1990).

[19] Farooqi, M.A., Tanaka, T., Ikezawa, Y., Omata, T., and Nagata, K. Sensor based control for the execution of regrasping primitives on a multifingered robot hand. In *Proc. 1999 IEEE Int. Conf. Robotics Automat.* (Detroit, MI, May 1999), vol. 4, IEEE, pp. 3217–3223.

[20] Faverjon, B., and Ponce, J. On computing two-finger force-closure graps of curved 2d objects. In *Proc. 1991 IEEE Int. Conf. Robotics Automat.* (Sacramento, CA, May 1991), vol. 1, pp. 424–429.

[21] Ferrari, C., and Canny, J. Planning optimal grasps. In *Proc. 1992 IEEE Int. Conf. Robotics Automat.* (Nice, FRANCE, May 1992), vol. 3, pp. 2290–2295.

[22] Fraser, A.M., and Dimitriadis, A. Forecasting probability densities by using hidden markov models with mixed states. In *Time Series Prediction: Forecasting the Future and Understanding the Past*, A.S. Weigend and N.A. Gershenfeld, Eds. Santa Fe Institute/Addison-Wesley, 1993, pp. 265–281.

[23] Grupen, R., Huber, M., Coelho Jr., J. A., and Souccar, K. Distributed control representation for manipulation tasks. *IEEE Expert 10* (Apr. 1995).

[24] Grupen, R.A., Biggers, K., Henderson, T.C., and Meek, S. Task defined internal grasp wrenches. Tech. Rep. UUCS-88-001, Department of Computer Science, University of Utah, 1988.

[25] Grupen, R.A., Coelho Jr., J.A., and Souccar, K.X. On-line grasp estimator: A partioned state space approach. Tech. Rep. COINS Technical Report 92-75, COINS Department, University of Massachusetts, Oct. 1992.

[26] Guo, G., Gruver, W.A., and Jin, K. Grasp planning for multifingered robot hands. In *Proc. 1992 IEEE Int. Conf. Robotics Automat.* (Nice, FRANCE, May 1992), vol. 3, pp. 2284–2289.

[27] Huber, Manfred, and Grupen, Roderic A. A feedback control structure for on-line learning tasks. *Robotics and Autonomous Systems 22*, 3-4 (December 1997), 303–315.

[28] Huber, Manfred, MacDonald, Willard S., and Grupen, Roderic A. A control basis for multilegged walking. In *Proc. 1996 IEEE Int. Conf. Robotics Automat.* (Minneapolis, MN, April 1996), IEEE, pp. Vol.4 2988–2993.

[29] Iberall, T. Grasp planning for human prehension. In *Proc. IJCAI* (Milan, ITALY, 1987), pp. 1153–1156.

[30] Isidori, A. *Nonlinear control systems*, 3 ed. Springer-Verlag, Berlin, 1995.

[31] Jameson, J.W., and Leifer, L.J. Automatic grasping: An optimization approach. *IEEE Trans. Syst., Man, Cybern. SMC-17*, 5 (Sept./Oct. 1987).

[32] Kaelbling, L.P., Littman, M.L., and Cassandra, A.R. Planning and acting in partially observable stochastic domains. *Artificial Intelligence 101*, 1 (1998).

[33] Kamon, I., Flash, T., and Edelman, S. Learning to grasp using visual information. In *Proc. 1996 IEEE Int. Conf. Robotics Automat.* (Minneapolis, MN, April 1996), vol. 3, IEEE, pp. 2470–2476.

[34] Kerr, J., and Roth, B. Analysis of multifingered hands. *Journal of Robotics Research 4*, 4 (Winter 1986), 3–17.

[35] Li, Z., and Sastry, S. Task-oriented optimal grasping by multifingered robot hands. *IEEE Trans. Robotics Automat. 4*, 1 (Feb. 1988).

[36] Maekawa, H., Tanie, K., and Komoriya, K. Dynamic grasping force control using tactile feedback for grasp of multifingered hand. In *Proc. 1996 IEEE Int. Conf. Robotics Automat.* (Minneapolis, MN, April 1996), vol. 3, IEEE, pp. 2462–2469.

[37] Matarić, Maja J. Reinforcement learning in the multi-robot domain. *Autonomous Robots 4*, 1 (1997), 73–83.

[38] Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, vol. 141-1 of *Mathematics in Science and Engineering.* Academic Press, Inc., New York, NY, 1979.

[39] McCallum, Andrew K. *Reinforcement Learning with Selective Perception and Hidden State.* PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, 1996.

[40] Meila, M., and Jordan, Michael I. Markov mixtures of experts. In *Multiple Model Approaches to Modelling and Control*, R. Murray-Smith and T.A. Johansen, Eds. Taylor and Francis, 1997.

[41] Meltz, B. F. Blockbusters: Blocks are just the single most important toy a child can have. Published on the Boston Globe, 2/24/2000.

[42] Michaud, F., and Matarić, M.J. Representation of behavioral history for learning in nonstationary conditions. *Robotics and Autonomous Systems 29* (1999), 187–200.

[43] Michie, D., and Chambers, R.A. Boxes: An experiment in adaptive control. In *Machine Intelligence 2*, E. Dale and D. Michie, Eds. Edingburgh, 1968.

[44] Montana, D.J. The kinematics of contact and grasp. *International Journal of Robotics Research 7*, 3 (1988), 17–32.

[45] Murray, R.M., Li, Z., and Sastry, S.S. *A Mathematical Introduction to Robotic Manipulation.* CRC Press, Boca Raton, CA, 1994.

[46] Napier, J. The prehensile movements of the human hand. *Journal of Bone and Joint Surgery 38B*, 4 (1956).

[47] Nguyen, V.D. Constructing stable grasps. *The Int. J. Robotics Res. 8*, 1 (1989), 26–37.

[48] Pollard, N. Synthesizing grasps from generalized prototypes. In *Proc. 1996 IEEE Int. Conf. Robotics Automat.* (Minneapolis, MN, April 1996), vol. 3, IEEE, pp. 2124–2130.

[49] Raibert, M.H. *Legged Robots that Balance.* The MIT Press, Cambridge, MA, 1986.

[50] Ramamoorthi, R., and Arvo, J. Creating generative models from range images. In *SIGGRAPH 99 Proceedings* (Los Angeles, CA, August 1999).

[51] Revow, M., Williams, C.K.I., and Hinton, G.E. Using generative models for handwritten digit recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence 18*, 6 (June 1996).

[52] Ripley, B. D. *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, GB, 1996.

[53] Rizzi, A.A., Whitcomb, L.L., and Koditschek, D.E. Distributed real-time control of a spatial robot juggler. *IEEE Computer Magazine 25*, 5 (May 1992).

[54] Salganicoff, M., and Bajcsy, R.K. Robotic sensorimotor learning in continuous domains. In *Proc. 1992 IEEE Int. Conf. Robotics Automat.* (Nice, FRANCE, May 1992), vol. 3, pp. 2045–2050.

[55] Salisbury, J. Kenneth. *Kinematic and Force Analysis of Articulated Hands.* PhD thesis, Department of Mechanical Engineering, Stanford University, Stanford, CA, May 1982.

[56] Schlegl, T., and Buss, M. Hybrid closed-loop control of robotic hand regrasping. In *Proc. 1998 IEEE Int. Conf. Robotics Automat.* (Leuven, BELGIUM, May 1998), IEEE, pp. 3026–3031.

[57] Schwartz, J.T., and Sharir, M. Finding effective "force targets" for two-dimensional multifinger frictional grips. Tech. rep., Robotics and Computer Science Vision Laboratory, Courant Institute of Mathematical Science, 1986.

[58] Singh, S.P., Jaakkola, T., and Jordan, M.I. Learning without state-estimation in partially observable markovian decision problems. In *Proc. Eleventh Int. Conf. on Machine Learning*, W.W. Cohen and H. Hirsch, Eds. Morgan Kaufmann Publishers, San Francisco, CA, 1994, pp. 284–292.

[59] Sutton, R.S., and Barto, A.G. *Reinforcement Learning: an Introduction.* The MIT Press, Cambridge, MA, 1998.

[60] Sutton, R.S., Precup, D., and Singh, S. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence 112* (1999), 181–211.

[61] Tani, J. Model-based learning for mobile robot navigation from the dynamical systems perspective. *IEEE Trans. Syst., Man, Cybern. 26*, 3 (June 1996), 421–436.

[62] Thrun, S.B. The role of exploration in learning control. In *Handbook of Intelligent Control: Neural Fuzzy, and Adaptive Approaches*, D.A. White and D.A. Sofge, Eds. Van Nostrand Reinhold, New York, 1992, pp. 527–559.

[63] Tomovic, R., Bekey, G., and Karplus, W. A strategy for grasp synthesis with multifingered robot hands. In *Proc. 1987 IEEE Int. Conf. Robotics Automat.* (Raleigh, NC, April 1987), IEEE, pp. 83–89.

[64] Watkins, C. J. C. H. *Learning from Delayed Rewards.* PhD thesis, Cambridge University, Cambridge, England, 1989.

[65] Wiering, M., and Schmidhuber, J. Hq-learning. *Adaptive Behavior 6*, 2 (1997), 219–246.

[66] Wilson, F.R. *Hand : How Its Use Shapes the Brain, Language, and Human Culture*, first ed. Vintage Books, New York, NY, 1999.