Efficient Planar Sample-Based Grasp Planning with Uncertainty Using Budgeted Multi-Armed Bandit Models for [v3 Jan 5, 2015]

Michael Laskey[1],Jeff Mahler[1], Zoe McCarthy[1], Florian T. Pokorny[3], Sachin Patil[1],
Jur Van Den Berg[4], Danica Kragic[3], Pieter Abbeel[1], Ken Goldberg[2]

Abstract—[TODO: NOT SURE IF WE NEED TO CHANGE THE AUTHORS LIST OR NOT]Sampling perturbations in shape, state, and control can facilitate grasp planning in the presence of uncertainty arising from noise, occlusions, and surface properties such as transparency and specularities. Monte-Carlo sampling is computationally demanding, even for planar models. We consider an alternative based on the multi-armed bandit (MAB) model for making sequential decisions, which can apply to a variety of uncertainty models. We formulate grasp planning as a "budgeted multi-armed bandit model" (BMAB) with finite stopping time to minimize "simple regret", the difference between the expected quality of the best grasp and the expected quality of the grasp evaluated at the stopping time. To evaluate MAB-based sampling, we compare it with Monte-Carlo sampling for grasping an uncertain planar object with shape uncertainty defined by a Gaussian process implicit surface (GPIS), but the method is also applicable to other models of uncertainty. We derive distributions on contact points, surface normal, and center of mass under shape uncertainty and use these to formulate the associated MAB model, finding that it computes grasps of similar quality to Monte-Carlo sampling and can reduce computation time by an order of magnitude. This suggests a number of new research questions about how MAB can be applied to other models of uncertainty and how different MAB solution techniques can be applied to further reduce computation.

Note to Practitioners—Planning for a grasp in an unknown environment can be difficult due to uncertainties. For example, a given object may have transparency which makes modern kinect-like sensor unable to accurately determine shape or an object may have a built up of mildew or dust, which makes the friction coefficient unknown. To handle all the uncertainties and plan for grasps using a grasp metric, one must perform monte-carlo integration and exhaustively evaluate samples. We show that by using simple algorithms from the Multi-Armed bandit literature, you can intelligently decided which grasp to sample from next and quickly determine the best grasp in a given set of proposed grasps, thus helping to alleviate the computational burden that is associated to incorporating uncertainty.
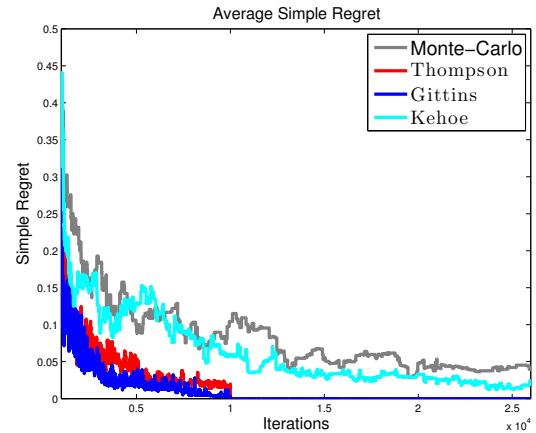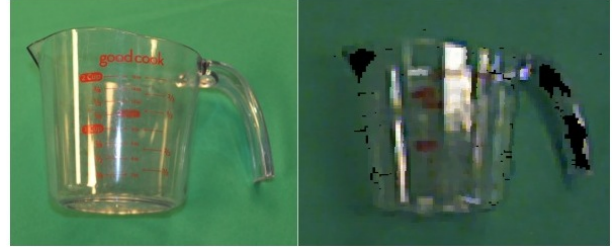
Fig. 1: [TODO: NEW TEASER PHOTO, NOT SURE IF WE SHOULD HAVE SOME SORT OF REAL OBJECT OR NOT](Top Left) Image of a common household measuring cup. (Top Right) Image from a PR2 Primesense camera that reflects the uncertainty that can be induced in objects from transparency. (Bottom) Comparison of Multi-Armed Bandit Techniques (Bayes UCB, Thompson, Gittins) vs. Monte-Carlo sampling to determine the best grasp in a set of 1000 grasps. As you can see the bandit techniques converge in simple regret Eq. 2 a magnitude faster than the traditional approach of Monte-Carlo sampling to determine the highest quality grasp.

I. Introduction

Consider a robot processing orders in a warehouse , where it frequently encounter new consumer products and need to process them quickly. The robot may need to rapidly plan grasps for these objects without prior knowledge of their shape, pose and material properties like friction coefficient or center of mass. Furthermore, the robot may not be able to measure these quantities exactly due to sensor noise and missing data resulting from occlusions and object properties such as transparency or reflective surfaces .

Grasp quality metrics have been developed to determine if a grasp will be successful or not prior to actually executing the grasp and how much force it needs to exert to resist an opposing force, however the ones that we consider evaluate a grasp assuming all the parameters are known [12]. Recent work [19], [26] motivated using knowledge of uncertainty to select grasps, but most methods for evaluating grasp quality in the presence of uncertainty use Monte Carlo sampling over

[1]Department of Electrical Engineering and Computer Sciences; {mdlaskey, zmccarthy, jmahler, sachinpatil, pabbeel}@berkeley.edu

[2]Department of Industrial Engineering and Operations Research and Department of Electrical Engineering and Computer Sciences; goldberg@berkeley.edu

[1−2] University of California, Berkeley; Berkeley, CA 94720, USA

[3]Computer Vision and Active Perception Lab, Centre for Autonomous Systems, School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden {fpokorny, dani}@kth.se

[4]Google; Amphitheatre Parkway, Mountain View, CA 94043, USA jurvandenberg@gmail.com

the possible values of the uncertain quantity [21], [40].

To select a grasp with high quality this evaluation is usually performed for a large set of perturbations in state, shape, and material properties, which can be very time-consuming. However, when evaluating a set of grasps we may be able to determine the difference of quality between proposed grasps with only a few samples and adaptively choose what grasp to evaluate next. We can concentrate grasp quality evaluation on the grasps that are most likely to have the highest quality based on the evaluation done so far.

The multi-armed bandit (MAB) model for sequential decision making problems [5], [23], [?] provides a way to reason about selecting the next option to evaluate. The goal in a MAB model is to make a sequence of decisions over a set of possible options, or "arms", such that a measure of expected reward is maximized. Solutions to the MAB model are particularly useful in applications where it is too expensive to fully evaluate a set of options; for example, in optimal design of clinical trials [35], market pricing [34], and choosing strategies for games [37]. The budgeted multi-armed bandit model [25] is a specialization of the MAB model where at a given "stopping time" the agent must choose the arm that currently has the highest expected reward based on its observations. The objective is to maximize the expected reward of the decision made at the stopping time, or equivalently to minimize "simple regret", which is the difference between the true expected reward of an optimal arm and the true expected reward of the arm pulled at the stopping time.

Our main contribution is formulating the problem of planning grasps according to some quality metric in the presence of uncertainty as a budgeted multi-armed bandit model. We use this formulation to rank a set of potential grasps by the probability of force closure [10], [20] under uncertainty in pose, shape, motion and friction coefficient. We choose the model of shape uncertainty to be a Gaussian process implicit surface (GPIS), a Bayesian representation of shape uncertainty that has been used in various robotic applications [11], [15]. Uncertainty in pose is represented as a normal distributions around the orientation and translation of the object. Uncertainty in motion is represented as a normal distribution around the end point of a planned gripper trajectory and uncertainty in friction coefficient is a normal distribution around an expected friction coefficient. Our approach of treating grasp selection as a BMAB requires only the ability to sample from a given representation of uncertainty and is not restricted to aforementioned representations.

We also show how to estimate distributions on the contact points and surface normals and center of mass using GPIS representation for shape uncertainty. Our experiments demonstrate that using the MAB sampling method improves the time to find the best grasp in a set of 1000 grasps by 5x over the baseline Monte-Carlo approach. [TODO: GOING TO GET A TABLE WITH RESULTS]

II. RELATED WORK

Past work on grasping under uncertainty has considered shape uncertainty [14], [38], uncertainty in contact locations with an object [42], uncertainty in object pose [10], [40], [21]. The effect of uncertainty in object geometry on grasp selection has been studied for spline representations of objects [10], extruded polygonal mesh models [19], [20], and point clouds [16].

One method of evaluating the expected grasp quality under uncertainty is to rank a set of random grasps on an object using samples on shapes, pose or parameters to evaluate a quality measure [10], [19], [20]. Monte-Carlo sampling involves drawing random samples from a distribution to approximate an expected value[8], which can be slow when the distribution is high-dimensional, such as for distributions on possible shapes.

To address this, Kehoe et al. [19] demonstrated a procedure for finding a minimum bound on expected grasp quality given shape uncertainty, which reduced the number of samples needed in Monte-Carlo sampling to choose the highest quality grasps. However, the proposed adaptive sampling approach pruned grasps using only the sample mean and did not utilize any estimates of how accurate the current sample mean is, which in practice could lead to good grasps being thrown away. Laaksonen et al. [22] used Markov Chain Monte-Carlo (MCMC) sampling to estimate grasp quality and object pose under shape and pose uncertainty when the robot is able to obtain new information via tactile sensor. MCMC provides a Bayesian framework for inference with a hidden state, but it can be slow to converge to the correct distribution due to burn in and mixing conditions [2].

We chose to study our MAB sampling method on distributions for friction coefficient, pose, motion and shape. For shape uncertainty we decided to use a Gaussian process implicit surface representation. Our decision to use this uncertainty model is based on GPIS's ability to combine different modes of noise observations such as tactile, laser and visual [31], [41], [11] and its recent use in modeling uncertainty for a number of robotic applications. Hollinger et al. used GPIS as a model of uncertainty to perform active sensing on the hulls in underwater boats [15]. Dragiev et al. showed how GPIS can enable a grasp controller on the continuous signed distance function [11]. Mahler et al. used the GPIS representation to find locally optimal anti-podal grasps by framing grasp planning as an optimization problem [26]. However, this relied on an approximation to grasp quality without guarantees on accuracy. We propose an adaptive sampling approach known as the Multi-Armed Bandit Model. This paper is a substantially revised and expanded version on [?] and [26].

III. PRELIMINARIES AND PROBLEM DEFINITION

Before we present the problem definition, we introduce a way to evaluate the quality of a grasp and our grasping model, the line of action. We also introduce a graphical model to represent the distributions on friction coefficient, motion, pose and shape.

A. Grasp Metric

Ferrari and Canny [12] demonstrated a method to rank grasps by considering points of contact with an object and the object surface normals at those contacts. The magnitude

of the metric allows one to rank grasps by their ability to resist external wrenches (forces and torques) applied to the grasped object. The Ferrari-Canny metric has wide spread use in grasp packages like GraspIT[28], OpenGrasp[24] and Simox [39], which motivates studying its effect with uncertainties.

Let Q be the L^1 version of the metric depends on the contact points $\mathbf{c}_1, ..., \mathbf{c}_m \in \mathcal{R}^2$, surface normals $\mathbf{n}_1, ..., \mathbf{n}_m \in \mathcal{R}^2$, center of mass $\mathbf{z}$ and friction coefficient μ. The metric is evaluated by constructing a convex hull around the wrenches made up of those parameters and finding the radius of the largest unit ball centered at the origin in wrench space. If the convex hull encloses the origin then the grasp is in "force-closure," meaning the grasp can resist any external wrenches if enough force is used. A grasp can be parameterized by the following tuple $g = (\mathbf{c}_1, ..., \mathbf{c}_m, \mathbf{n}_1, ..., \mathbf{n}_m, \mu, \mathbf{z})$[30].

In this work we use the probability of achieving force closure, or $P(Q > 0)$, [10][20], to rank grasps . $P(Q > 0)$ may be computed by sampling from distributions on pose (rotation and translation of the object), shape and material properties (friction coefficient) and averaging the qualities that are computed.

B. Line of action

In an uncertain environment, a grasp planner may not know the true grasp parameters or g exactly due to sensing imprecision. Thus we propose to work with the trajectory of the gripper when analyzing grasps. Similar to the work of [10], we assume that each gripper finger approaches along a *line of action*, a 1D curve $\gamma(t)$ with endpoints a and b as seen in Fig. 2. A gripper finger starts at point a and moves towards b, we assume a is far enough away to be collision free of the object. Each gripper contact is defined by a line of action, so we assume the following tuple is provided $\Gamma = (\gamma_1(\cdot), ..., \gamma_m(\cdot))$, which designates a proposed *grasp plan*.
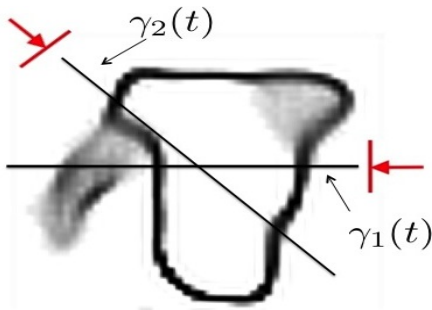

Fig. 2: Illustration of a grasp plan Γ composed of two lines of action, $\gamma_1(t)$ and $\gamma_2(t)$

C. Types of Uncertainty

In this work we consider the following types of uncertainty shape, pose, movement, and friction coefficient. The property of force closure is not affected by the center of mass because it measures if a wrench can be resisted not how large the force is need to resist. Fig. 3 shows the probabilistic relationship of the types of uncertainty considered in this work. We model each one of these with a corresponding distribution and demonstrate how MAB algorithms can handle the types of uncertainty. We

note though that our choice of distributions does not effect the ability to use the MAB algorithms and that the reader should choose distribution that effect their needs.

1) Distribution on Shape

Shape uncertainty can occur from transparency, occlusions and sensor noise [?]. To model this uncertainty we use a Gaussian Process Implicit Surface (GPIS) representation, which is explained in detail in Appendix A. A GPIS is a Gaussian distribution on a signed distance function, sd, that describes the workspace. A signed distance function is greater than 0 outside the surface of an object, 0 at the surface and less than 0 beneath. Signed distance functions can be sampled from $N(\mu(x), \Sigma(x))$, where $\mu(x)$ and $\Sigma(x)$ are the mean and covariance functions of the GPIS. For our application, we set x to discretized points along the workspace $\mathcal{W}$ and for each sample the points that correspond to zero, $sd(x) = 0$, are the contours of the shape. For convenience we will represent the $\mu(x)$ and $\Sigma(x)$ as the tuple $\theta = (\mu(x), \Sigma(x))$.

2) Distribution on Pose

In typical robotics applications the pose of objects in the envinroment is determined by registering the object frame-of-reference to the control frame-of-reference used for grasp execution. Therefore pose uncertainty may come from two primary sources a) uncertainty about the registration of the robot's grasping frame-of-reference to its sensing frame-of-reference and b) uncertainty about the pose of known object models in the robot's sensor data. The effects of pose uncertainty on robotic grasping has been studied by [].

The pose of an object T is a member of the Lie Algebra $SE(3)$ in 3-dimensional space (something analogous holds for 2D). This matrix is defined by 3 rotation angles $\mathbf{w} = (\alpha, \beta, \gamma)$ and 3 translation coordinates $\mathbf{t} = (t_x, t_y, t_z)$, summarized in parameter vector $\xi = (\mathbf{w}, \mathbf{t})^T \in \mathbb{R}^6$. One challenge with pose is that uncertainty in the pose matrix T used to apply transformations is needed in practice, but uncertainty is mathematically more easily to quantify in terms of the pose parameters ξ. Thus given a mean pose matrix $\bar{T} \in SE(3)$ and zero-mean uncertainty on the pose parameters $\xi \sim \mathcal{N}(\mathbf{0}, \Sigma)$ we define the pose random variable T as

$$T = \exp(\xi^\wedge) \bar{T}$$

where the $\wedge$ operator is defined as in [4].

3) Distribution on Motion

In practice a robot may not be able to execute a desired grasp plan Γ exactly due to errors in trajectory following or registration to the object [19]. To handle this uncertainty, we sample the angle of approach ρ of the proposed grasp plan Γ from a zero mean one-dimensional Guassian $\rho \sim N(0, \sigma_{mot}^2)$. In practice σ_{mot}^2 might be set from repeatibility measurements for a robot [?].

4) Distribution on Friction Coefficient

As shown in [42], uncertainty in friction coefficient can play a large role in grasp quality evaluation. The expected friction coefficient $E(\mu)$ can be derived by means of object classification and a look up table. However, because their could be material in between the objects surface and the robot gripper (i.e. dust, water, moisture), we purpose sampling

the friction coefficient from a Gaussian around the expected friction coefficient $\mu \sim N(E(\mu), \sigma_\mu^2)$.

To estimate $P(Q(\Gamma) > 0)$, we sample from the graphical model shown in Fig. 3. To sample from $p(Q(\Gamma) > 0)$, we need to sample from the distributions associated with a line of action $p(\mathbf{n}_i, \mathbf{c}_i | \gamma_i(t), \xi, \theta, \rho)$. Using Bayes rule we can rewrite this as

$$p(\mathbf{n}_i, \mathbf{c}_i | \gamma_i(t), \theta, \xi, \rho) = $$
$$p(\mathbf{n}_i | \mathbf{c}_i, \theta) p(\mathbf{c}_i | \gamma_i(t), \theta, \rho, \xi)$$

Appendix A, describes how to draw shape sample from a GPIS model, which is used to compute $p(\mathbf{c}_i | \gamma_i(t), \theta, \rho, \xi)$ along with the other sampled distribution on pose (ξ) and motion (ρ). Appendix C, describes how to sample from $p(\mathbf{n}_i | \mathbf{c}_i, \theta)$ and presents a novel visualization technique for the distribution on surface normals. Appendix D, describes a way to calculate the expected center of mass assuming a uniform mass distribution.
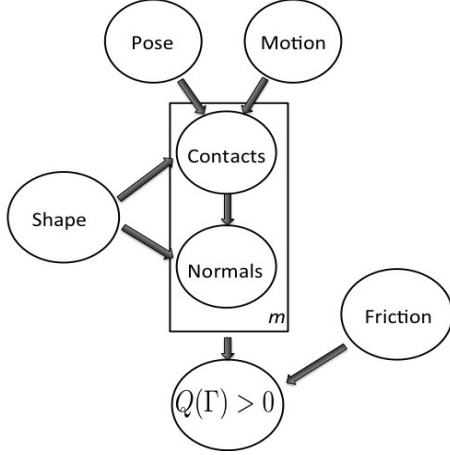

Fig. 3: A graphical model that illustrates the relationship between the different types of uncertainty in an object. Center of Mass uncertainty is dependent on the pose and shape of the object, however friction coefficient is independent of all other types. [TODO: SHOULD WE USE WORDS OR SYMBOLS FOR THE GRAPHICAL MODEL?]

D. Problem Definition

We assume we are given a 2-D workspace $\mathcal{W}$ with an unknown object represented as a trained GPIS model, described in Section A, and set of possible grasp plans G which are generated either uniformly randomly or with a heuristic as in [26]. We are interested in determining Eq. 1 with respect to the probllaliltiy of achieving force closure or $P(Q(\Gamma) > 0)$. [TODO: NEED TO FIGURE OUT IF WE CAN USE OPTIMAL OR NOT]

$$\Gamma^* \in \underset{\Gamma \in G}{\arg\max} P(Q(\Gamma) > 0) \tag{1}$$

IV. MULTI-ARMED BANDITS FOR GRASP SELECTION

While a standard approach to solving the problem in Eq. 1 would be to perform Monte-Carlo integration on each Γ_i and compute the probability of force closure, we propose treating the problem as a multi-armed bandit model and

forming a policy for selecting which grasp to sample. The multi-armed bandit model, originally described by Robbins [33], is a statistical model of an agent attempting to make a sequence of correct decisions while concurrently gathering information about each possible decision. Solutions to the multi-armed bandit model have been used in applications for which evaluating all possible options is expensive or impossible, such as the optimal design of clinical trials [35], market pricing [34], and choosing strategies for games [37].

The traditional setting of a multi-armed bandit model is a gambler that has K independent slot machine arms and decides what machines to play, how many times to play each one, what order to play them in. A successful gambler would want to exploit the machine that currently yields the highest reward and explore new arms to see if they give better rewards. Developing a policy that successfully trades between exploration and exploitation has been the focus of extensive research since the problem formulation [7], [33], [6].

In our setting, we have a probabilistic representation and would like to evaluate many potential grasps on the samples drawn from the graphical model in Fig. 3. Motivated by limited computational resources we allocate sampling resources to efficiently find the best arm. Here each arm corresponds to a different grasp plan and pulling the arm is sampling from the graphical model in Fig. 3 and evaluating the arm's grasp plan on the sample.

A common measure of success in MAB problems is *regret*, the difference between the expected optimal reward and the expected reward of the selected arm on a single pull. Traditional bandit algorithms minimize cumulative regret, the sum of regret over the entire sequence of arm choices. Lai and Robbins showed that an optimal solution to the bandit problem is bounded by a logarithmic function [23]. They presented an algorithm called (Upper Confidence Bound) UCB that obtains this bound asymptotically. Many variants of UCB have thus been proposed. The algorithm maintains a confidence bound on the distribution of reward based on prior observations and pulls the arm with the highest upper confidence bound.

In the context of finding a high quality grasp, one only cares about the regret when the grasp is actually executed in real life. The agent shouldn't be penalized for exploring different grasps, but only for the grasp that they suggest at the termination of the algorithm. Hence, the exploration and exploitation stage are decoupled. Our goal is to identify the best arm in as few decisions as possible. Given a set of arms $\{1, ... K\}$ with respective mean rewards $\mu_1, ..., \mu_K$ and the optimal arm $\mu^* = \underset{k \in \{1,...,K\}}{\max} \mu_k$. The simple regret at time t is given by

$$r_t = \mu^* - \mu_t \tag{2}$$

where μ_t is the estimate of the best arm at time t from the previous observations.

Best arm identification has a wide variety of literature that largely falls into two camps: one where the algorithm terminates once a fixed confidence interval around the best arm is met and the Budgeted Multi-Armed Bandit (BMAB)

model, in which the algorithm must make a decision at the end of a fixed "budget" number of arm pulls.

A. Fixed Confidence

In the fixed confidence setting the forecaster seeks to minimize the simple regret until a fixed confidence threshold is met at which point it terminates. Originally the problem was solved with 'racing' algorithms, which used Hoeffding inequalities or the empirical Bernstein inequality to prune arms that were likely to be suboptimal and used uniform allocation to explore the remaining set [27] [29]. These methods were later extended to return the top m arms instead of only the best arm[13].

The speed of termination is affected by the hardness of the problem, which relates to the how close the other arms expected reward is to the expected reward top arm in the set [3]. In the grasping context this means that achieving a fixed confidence for two similar grasps would require exhaustive sampling from each distribution.

B. Fixed Budget

In the Budgeted Multi-Armed Bandit (BMAB) setting the algorithm is given a stopping time and needs to return the current best arm at that stopping time this can be thought of as an anytime algorithm. Audibert el al. demonstrated an algorithm called Successive Rejects that divides up the total budget into successively shorter phases and discards the worst arm left at the end of each phase. This algorithm can return the best arm with near-optimal probability depending on the hardness of the problem [3]. In addition, UCB-like methods have been proposed that measure a confidence gap and then pull the arm with the highest confidence interval [13]. For example in [7], they showed a link between simple regret and cumulative regret that allowed for the analysis of the existing bandit algorithms like UCB1.

In this work, we consider a setting in which a robot has a fixed computational time limit to select a grasp and thus the fixed budget setting is more appropriate. The time limit can be chosen via repeatability experiments on a desire task. In Section V, we will discuss some of the specific budgeted multi-armed bandit algorithms that are used in practice.

V. BANDIT ALGORITHMS FOR BEST ARM IDENTIFICATION

While, UCB type algorithms adopt a frequenist perspective for distributions with bounded rewards in their empirical estimate of an upper bound. A growing interest though in the MAB community is in bandit algorithms that have Bayesian approach [18] [1]. Here a prior distribution is placed on the distribution to be estimated and utilizing the prior enables a trade off between exploration and exploitation. Theoretical results have shown that these methods are capable of achieving the lower bound described by Lai and Robbin [1] [18]. Furthermore in an empirical study Bayesian methods have been shown to outperform the UCB family [9].

In the context of grasping, the reward from sampling from the grasp distribution is 1 if the sampled grasp is in force closure and 0 if the grasp is not in force closure. The distribution that describes whether an event is $\{0, 1\}$ is known as a Bernoulli distribution and can be described by the parameter θ or the probability an event occurs. In the Bayesian setting one would treat θ as belonging to a distribution. A convenient and common choice for such a distribution is the Beta distribution. Beta distributions are specified by shape parameters α and β, where ($\alpha > 0$ and $\beta > 0$). The mean of the Beta distribution is given by $\alpha/(\alpha + \beta)$. To update the prior Beta distribution one simply adds the count of observed successes of the event to α and the count of the observed failures to β. The default $\alpha = 1$ and $\beta = 1$ corresponds to a uniform distribution on θ.

Given a proposed grasp plan Γ, we draw samples from the shape distribution $P(\theta)$, the distribution on pose $P(\xi)$, distribution on motion $p(\rho)$ and the distribution on friction coefficient $P(\mu)$. The distribution on force closure can then be estimated as Beta- Bernoulli Process with shape parameters α and β. Thus, we can write the expected probability of force closure as follows

$$P(Q(\Gamma|\theta, \mu, \xi, \rho) > 0) = \frac{\alpha}{\alpha + \beta} \qquad (3)$$

Where $Q(\Gamma|\theta, \xi, \mu, z)$ is the grasp quality that is computed on a shape sample drawn from $p(\theta), p(\xi), p(\rho)$ and $p(\mu)$. Whats interesting in the context of a Bayesian BMAB problem our graphical model in Fig. 3, is now equivalent in terms of inference to 4 and we only need to estimate α and β to determine grasp quality. Regardless of the distributions in Fig. 3, we are still able to maintain the theoretical guarantees given by the Bayesian MAB algorithms because the probability of force closure is a Beta-Bernoulli Process [1] [18] [?].
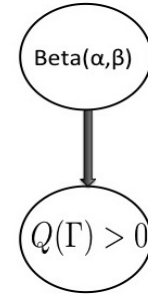

Fig. 4: A graphical model that illustrates the relationship between the Bernoulli distribution of the probability force closure and its conjugate prior Beta distribution that has two shape parameters α and β

A. Bayes-UCB

Bayes UCB, detailed in Algorithm 1, is a version of the UCB algorithm that uses a Bayesian posterior update to get a confidence bound instead of the frequentist concentration inequality approach. To get an upper confidence bound from the posterior, the Bayes UCB algorithm uses the quantile of the Beta distribution up to a specified probability which depends on the timestep and horizon. TAt each time step the arm with the highest quantile is chosen and the Beta distribution is updated based on the observed reward. For binary rewards (i.e. Bernoulli distributions) the expected number of pulls of suboptimal arms is bounded and experimentally this algorithm has been shown to outperform UCB for binary rewards [18].

Algorithm 1: Bayes-UCB for Beta-Bernoulli Process

Result: Current Best Arm, Γ^*

For Beta(1,1) prior, Stopping Horizon n:

for *t=1,2,...,n* **do**

    for *j = 1,...,K* **do**

        Compute: $q_j(t) = Q\left(1 - \frac{1}{t}, p_j^{t-1}\right)$

    Draw arm$I_t = \text{argmax}_{j=1...K} q_j(t)$

    Observe reward $X_{I_t,t} \in \{0,1\}$

    Update posterior:

    Set $S_{I_t,t+1} = S_{I_t,t} + X_{I_t,t}$

    Set $F_{I_t,t+1} = F_{I_t,t} + 1 - X_{I_t,t}$

    Set $\theta_j^t \sim \text{Beta}(S_{I_t,t+1}, F_{I_t,t+1})$

B. Thompson Sampling

Thompson Sampling is a Bayesian method for the multi-armed bandit problem. We will describe it now in detail for the Beta-Bernoulli process. All arms are initialized with a prior Beta distributions, which is normally Beta($\alpha = 1, \beta = 1$) to reflect a uniform prior on the θ of the Bernoulli distribution. Then for each arm draw $\theta_{j,t} \sim \text{Beta}(\alpha, \beta)$ and pull the arm with the highest $\theta_{j,t}$ drawn. The reward, $X_{i,t}$ is observed from that arm, j, and the corresponding Beta distribution is updated. This is repeated until a stopping time is reached. The full algorithm is shown in Algorithm 1.

The randomness of Thompson sampling allows for it to quickly explore the more arms then Bayes-UCB, which makes conservative arm pulls based on confidence bounds. It is also less prone to local solutions because of its stochastic nature. Thus, for cases where exploration and exploitation are decoupled Thompson sampling can find a better arm faster. Thompson sampling has recently been shown to approach the Lai and Robbins bound [1] and has empirically been shown to outperform frequentist methods like UCB in certain settings [9]. Variants of it are even used commercially in products like Microsoft's adPredictor, which is used by Bing, the search engine, [**?**].

Algorithm 2: Thompson Sampling for Beta-Bernoulli Process

Result: Current Best Arm, Γ^*

For Beta(1,1) prior:

for *t=1,2,...* **do**

    Draw $\theta_{j,t} \sim \text{Beta}(S_{j,t} + 1, F_{j,t} + 1)$ for $j = 1,...,k$

    Play $I_t = j$ for j with maximum $p_{j,t}$

    Observe reward $X_{I_t,t} \in \{0,1\}$

    Update posterior:

    Set $S_{I_t,t+1} = S_{I_t,t} + X_{I_t,t}$

    Set $F_{I_t,t+1} = F_{I_t,t} + 1 - X_{I_t,t}$

C. The Gittins Index Method

[TODO: LET ME KNOW HOW CLEAR THIS IS, GITTINS CAN BE HARD TO DESCRIBE IN GENERAL] One possible solution to solve the MAB problem is to treat it as an Markov Decision Process (MDP) and use Markov Decision theory. This solution makes a lot of sense when the distribution is known because the all elements in the standard MDP tuple, $\{S, A, T, R, \gamma\}$, would be known and it is optimal with respect γ [**?**].

However, the curse of dimensionality effects performance because if you have K arms, a finite horizon of T and a Beta-Bernoulli distribution on your arms then your state space is on the order of T^{2*K}. Hence the complexity of solving MAB using Markov Decision theory increases exponentially with the number of bandit processes. A key insight though was given by Gittins, who showed that instead of solving the k-dimensional MDP one can instead solve k 1-dimensional optimization problems: for each arm i, $i = 1,...,k$, and for each state of $x^i = \{\alpha_0 + S_t, \beta_0 + F_t\}^i$, where S_t and F_t correspond to the number of success and failures at pull t.

$$v^i(x^i) = \max_{\tau > 0} \frac{\mathcal{E}[\sum_{t=0}^{\tau} \gamma^t r^i(X_t^i)|X_0^i = x_i]}{\mathcal{E}[\sum_{t=0}^{\tau} \gamma^t | X_0^i = x_i]} \quad (4)$$

The indices can be considered as a computation of the value in choosing an arm conditioned on the fact that you will give up an choose another arm at some point. Once you know the state of your k arms, the algorithm is to select the one with the highest index. For Best Arm Identification you want your discount factor γ to approach 1, since you should never stop pulling the best arm. Generally the computation of the Gittins indices is too expensive, however in the Beta-Bernoulli case it is actually possible [18]. We computed the Gittins indices offline using the restart method proposed by Katehakis et al. [17].

Algorithm 3: The Gittins Index Method for Beta-Bernoulli Process

Result: Current Best Arm, Γ^*

For Beta(1,1) prior, Table of Indices v, Discount Factor γ:

for *t=1,2,...* **do**

    Pull arm $k = \underset{x_k \in X}{\text{argmax}} v(x_k)$

    Observe reward $R_{I_t,t} \in \{0,1\}$

    Update posterior:

    Set $S_{I_t,t+1} = S_{I_t,t} + R_{I_t,t}$

    Set $F_{I_t,t+1} = F_{I_t,t} + 1 - R_{I_t,t}$

    Set $x_k = \{1 + SI_t, t + 1, 1 + F_{I_t,t+1}\}$

.

VI. Experiments

For the experiments we used the Brown Vision Lab 2D dataset, the same used in [10]. We downsampled the image by a factor of 2 to create a 40 x 40 occupancy map, which holds 1 if the point cloud was observed and 0 if it was not observed, and a measurement noise map, which holds the variance 0-mean noise added to the SDF values. The parameters of the GPIS were selected using maximum likelihood on a held-out set of validation shapes. The noise of the motion, position and friction coefficient was set to the following variances $\sigma_{mu} = 0.4$, $\sigma_{rot} = 0.3$ rads, $\sigma_{trans} = 3$. Our visualization technique

follows the approach of [26] and consisted of drawing many shape samples from the distribution and blurring accordingly to a histogram equalization scheme.

We did experiments for the case of two hard contacts in 2-D, however our methods are not limited to this implementation. We drew random lines of actions $\gamma_1(t)$ and $\gamma_2(t)$ by sampling around a circle with radius $\sqrt{2}n$ and sampling the circles origin, then projecting onto the largest inscribing circle in the workspace.

A. Multi-Armed Bandit Experiments

[TODO: ADD BAYES AND KEHOE RESULTS] We consider the problem of selecting the best grasp plan, Γ^* out of a set G. For our experiments we look at selecting the best grasp out of a size of $|G| = 1000$. In Fig. 5, we plotted the simple regret averaged over 100 randomly drawn shapes in our data set and compare the different methods (UCB, Thompson, Gittens and the naive random allocation). We initialize both the Monte-Carlo and bandit technique by sampling each grasp 1 time. We draw samples from our calculated distributions $p(g)$. Interestingly, Gittens and Thompson converge much faster than random and UCB. In Fig. 6, you can see that Gittens and Thompson allocate grasp samples to only the grasps of high quality, thus they are quickly able to ignore the low quality grasps. UCB takes a more conservative approach to sample allocation, which leads to poor performance in the best arm identification problem [7].

B. Sensitivity Analysis

We now will show how well the top two algorithms Thompson Sampling and the Gittens Index Method perform under a variation in noise from friction coefficient uncertainty, shape uncertainty, rotational pose and translation pose. The experiments are performed with the same setup as before but now we increase the variance parameters across a set range for each parameter to simulate low, medium and high levels of noise. All experiments were averaged across 10 shapes with a set size $|G| = 1000$.

For friction coefficient we varied the variance across the following values $\sigma_\mu = \{0.05, 0.2, 0.4$. As illustrated in Fig. 8, the performance of the bandit algorithm remains largely unchanged, with typical convergence to zero in simple regret less than 2000 evaluations.

For rotational uncertainty in pose, we varied σ_{rot} over the set of $\{0.03, 0.12, 0.24\}$ radians. As you can see in Fig. 9, the performance of the bandit algorithms is effected by the change in rotation, increase in variance to 0.24 radians or 13^{deg} causes the convergence in simple regret to not be reached until 5500 samples or an average of 5.5 samples per grasp. This can be explained because such a large variance causes a a drop in quality across all grasps and makes it harder to separate the outliers [?]. The quality of the best grasp along with the grasp for each round is shown in 10.

For translational uncertainty in pose, we varied σ_{trans} in the range of $\{3, 12, 24\}$ units (on a 40 x 40 unit workspace). As you can see in Fig. 9, the performance of the bandit algorithms is effected by the change in rotation, increase noise

of $\sigma_{trans} = 24$ causes the convergence to not be reached until around 5000 samples for the Gittens Method and 8200 evaluations for Thompson Sampling.
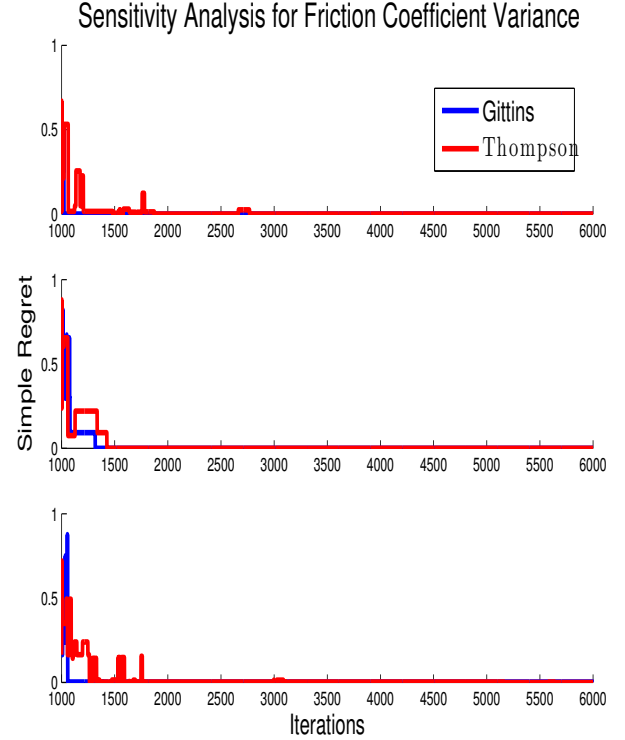

Fig. 8: Sensitivity Analysis for Thompson Sampling and the Gittens Index Method under friction coefficient uncertainty $\sigma_{fric} = \{0.05, 0.2, 0.4\}$ from top to bottom on a 40 x 40 unit workspace averaged over 10 shapes from the Brown Vision Lab Data set. The increase in noise has little effect on the convergence of the two algorithms in simple regret.

VII. LIMITATIONS

Our budgeted multi-armed bandit approach appears promising, but we still do not know how well it will perform on 3D shapes and large scale grids. Future work will be building an efficient construction of GPIS to scale to 3D and test the bandit method there.

The methods we showed (UCB, Thompson and Gittens) are guranteed to find the best grasp as the stopping time approaches infinity [18], [?], [?] but when do you terminate the algorithm is still an open question. Fixed confidence methods do exist that terminate when a certain confidence interval is reached [27] [29]. However, a known problem is that if two grasps have very similar quality it could greatly increase the time for needed to reach the statistical confidence interval [3]. We proposed treating the BMAB as anytime algorithm and have demonstrated empirically that the grasps at a given stopping time are better than prior methods of Monte-Carlo sampling or the approach of Kehoe et al. However, this has a hyper parameter that has to be user-defined, which is not ideal.

Another problem that was revealed in our analysis was that our current grasp metric, probability of force closure, is not dependent on the center of mass [12]. It only measure the probability that a grasp controller can resist any force provided
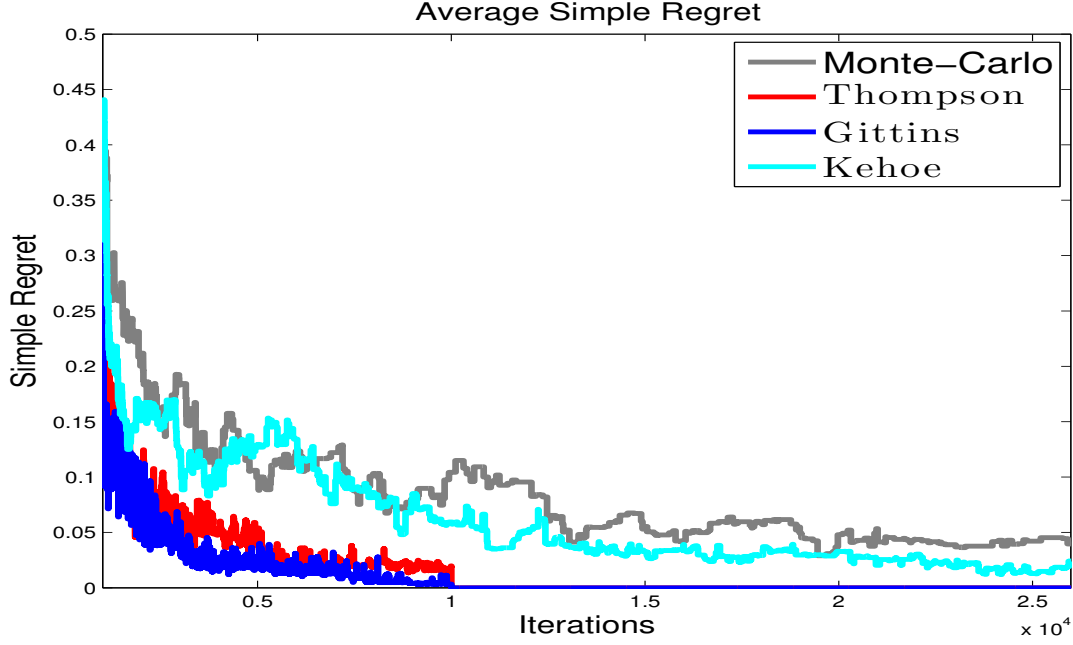
Fig. 5: Comparison of Simple Regret convergence for the four sequential decision methods (Monte-Carlo, Bayes -UCB, Thompson, Gittins). Graph is averaged over 100 shapes from the Brown Sillohoute Dataset [**?**] with a set $|G| = 1000$ for each shape. As you can see the BMAB methods converge almost a magnitude faster than random allocation. It is worth noting that Gittins outperform the other two algorithms, which is useful when choosing which one to implement
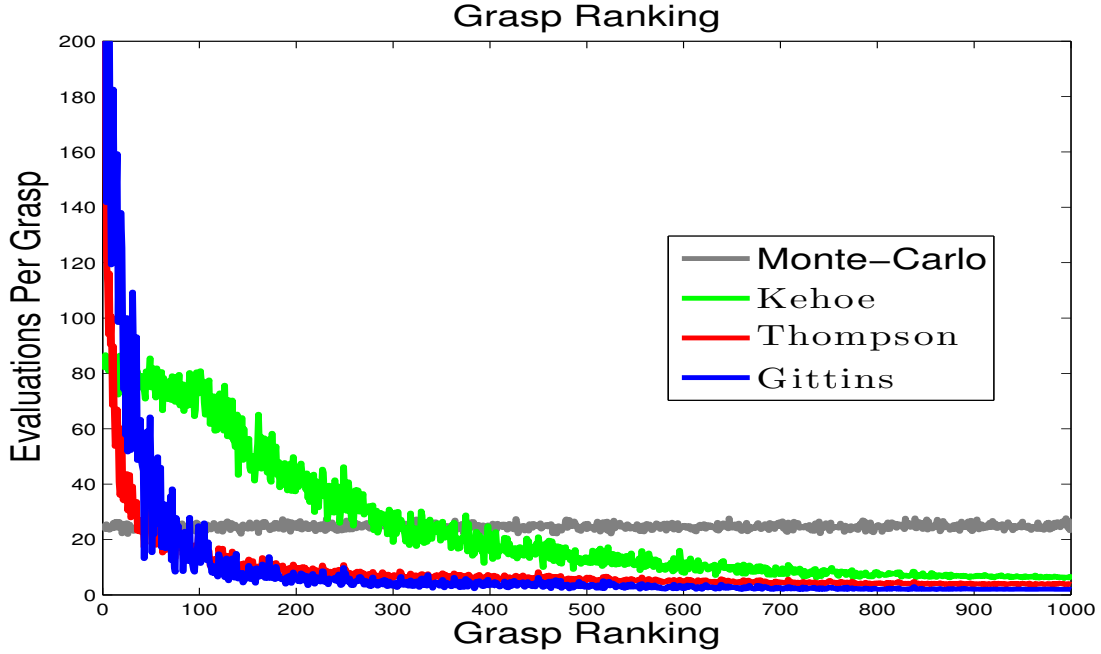

Fig. 6: Comparison of sample per grasp for the four sequential decision methods (Random, Bayes - UCB, Thompson, Gittins). Graph is averaged over 100 shapes from the Brown Silhouette Dataset [**?**] with a set $|G| = 1000$ for each shape. The best grasps are ranked 1 and worst are 1000. As you can see the MAB algorithm intelligently allocate samples towards high quality grasps based on past observations, where Monte-Carlo Integration takes a uniform approach to allocation. [**?**]

it can exert an infinite force. One can assume that the grasp controller on a robot hand is powerful enough to apply the proper resistance, but that assumption might be invalid in some applications. A similar metric that is still from Beta-Bernoulli and takes center of mass into account, would be ideal for both accurate grasp quality prediction under uncertainty and the utilization of MAB algorithms. Recent work by Kim et

al. developed a physics based simulator that could potentially achieve this goal [21].

VIII. CONCLUSION

Assessing grasp quality under uncertainty is computationally expensive as it often requires repeated evaluations of the grasp metric over many random samples. In this work,
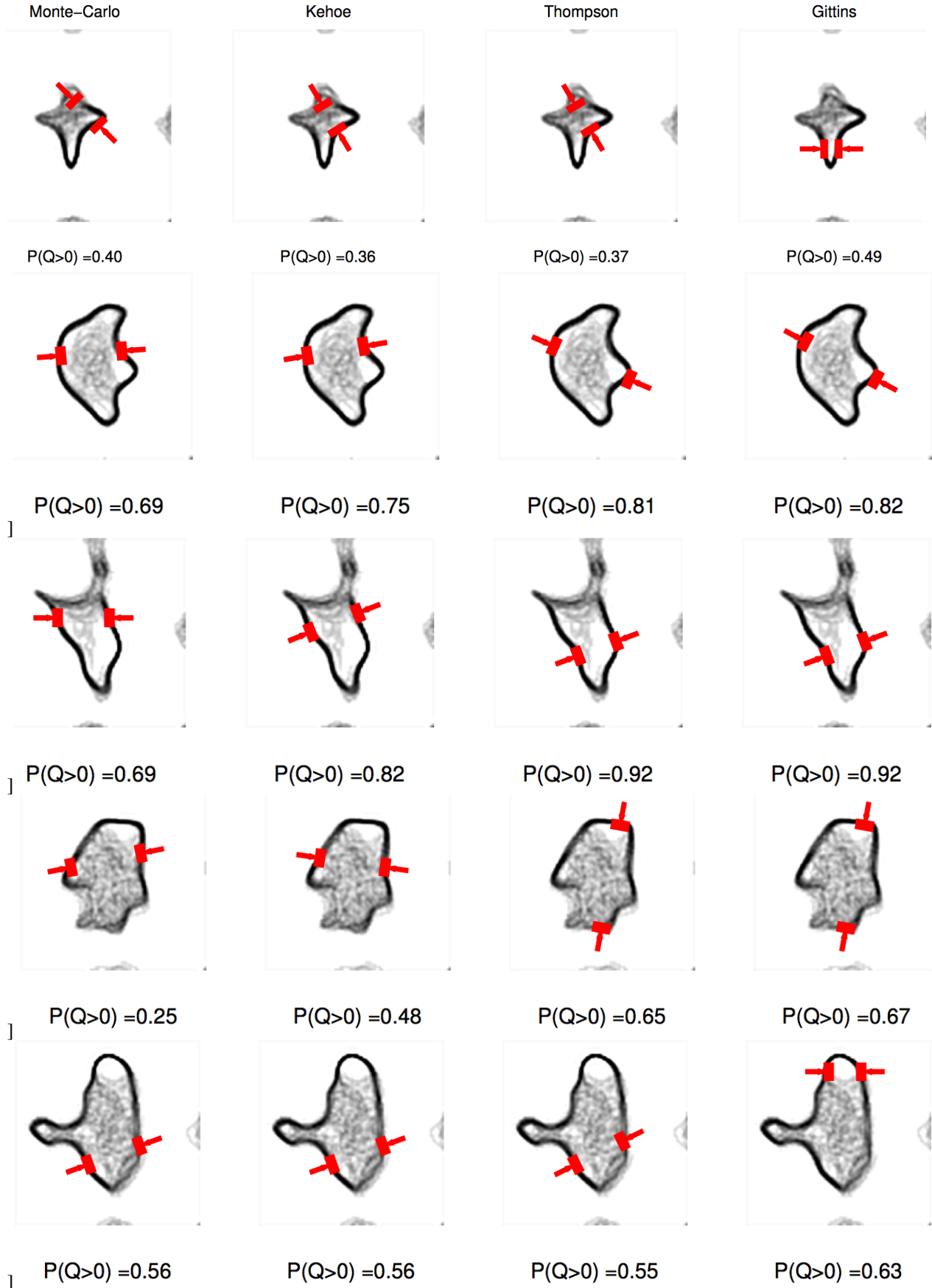
Fig. 7: Five shapes shown from the Brown Visual Lab Dataset with induced shape uncertainity and visualized according to the method described in [26]. The four methods (Monte-Carlo, Kehoe's, Thompson and Gittins) were all run until a stopping time of 9000 evaluations with a uniformly initialized grasp set of $|G| = 1000$. The grasps and the quality each one found is shown above. The Gittins Index Method in all cases shown found the best grasp among the methods.
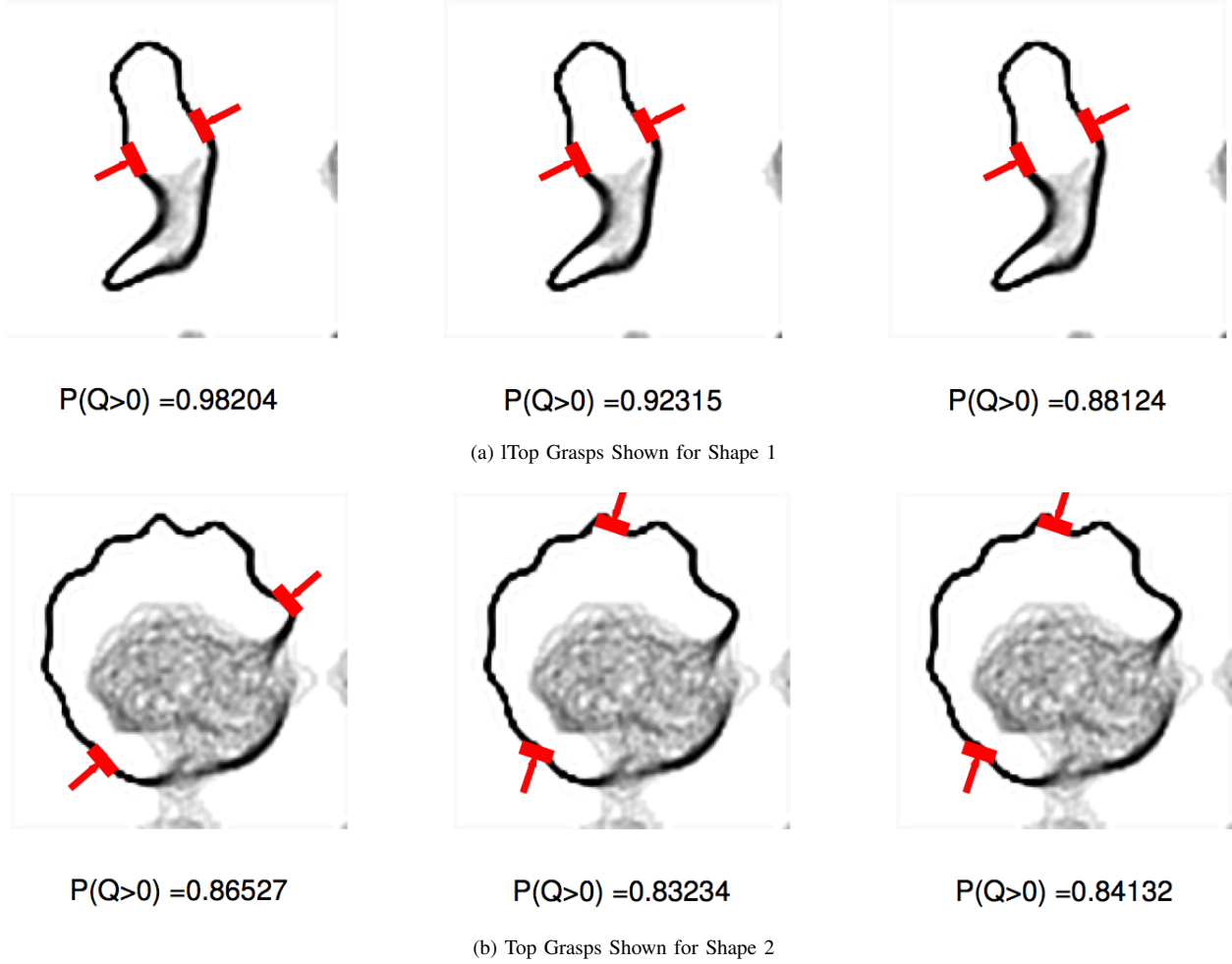
P(Q>0) =0.98204   P(Q>0) =0.92315   P(Q>0) =0.88124

(a) lTop Grasps Shown for Shape 1

P(Q>0) =0.86527   P(Q>0) =0.83234   P(Q>0) =0.84132

(b) Top Grasps Shown for Shape 2

Fig. 10: Two shapes shown from the Brown Visual Lab Dataset,from Left to Right the variance on rotation σ_{rot} is increased 0.03, 0.12 to 0.24 radians. As you can see the overall variance increase effects the quality of the top grasp in the set of possible grasps. Furthermore for Shape 2, the grasp with low rotational variance is different than that for higher variance because the original grasp is more likely to touch the area of higher shape uncertainty when subjected to high variance in rotation. [TODO: FIND BETTER SHAPES IN THE BROWN DATASET]

we proposed a multi-armed bandit approach to efficiently identify high-quality grasps under uncertainty in shape, pose, friction coeffiecient and motion. A key insight from our work is that uniformly allocating samples to grasps is inefficient, and we found that a MAB approach prioritizes evaluation of high-quality grasps while quickly pruning-out obviously poor grasps. A pre-requisite for applying a bandit approach is to formulate a representation of how uncertainty affects grasp parameters and thus grasp quality. We purpose treating this as a graphical model and use model the parameters as stochastic noise. Our choice of distributions though is not the focus of the paper. The MAB algorithm is applicable in any context of bounded reward distributions and all the theoretical results we mentioned will transfer to the Beta-Bernoulli case, or the estimation of probability of force closure.

Our anytime BMAB approach is guaranteed to find the best grasp in a given proposal set in the limit of an infinite time and has empirically been shown to outperform the methods of prior work Monte-Carlo and the method purposed by Kehoe et al. [19].

IX. FUTURE WORK

Our results are promising and they suggest many avenues of future work. By utilizing the BMAB model, we can encode uncertainty in the grasp parameters and then leverage the existing algorithms to efficiently find the best grasp.

In principle, our method can be applied to other representations of shape uncertainty such as perturbations on polygonal vertices [19] or splines [10]. It can further be applied to other grasp quality metrics or simulation based evaluation methods [24].

Future work will also consider applying BMAB approach to grasp planners like GraspIt! [28] to see if our method can handle uncertainty while working under the time constraints needed for most real time applications. While our results are promising, it remains to be seen how well it deals with the increased complexity of 3D models over 2D models and larger scale experiments. However, the BMAB model has a large amount of literature to draw from as we encounter new and more challenging problems [6].
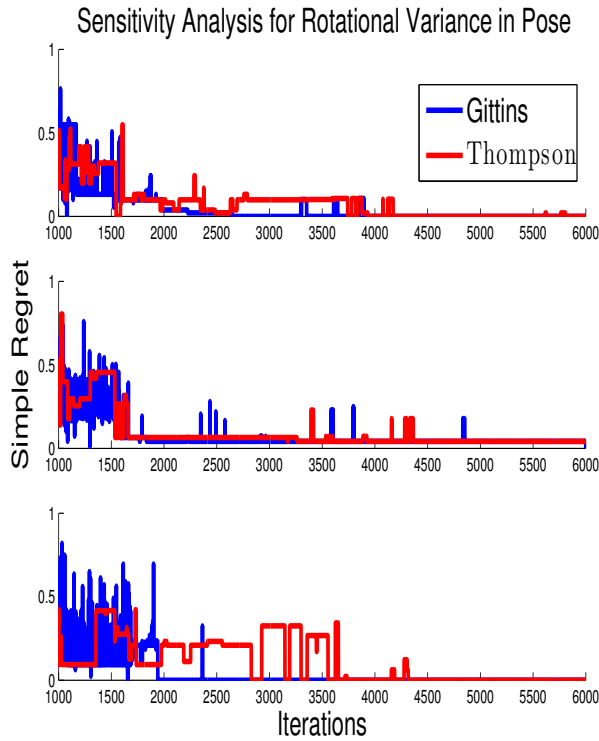
Fig. 9: Sensitivity Analysis for Thompson Sampling and the Gittens Index Method under coefficient of friction uncertainty $\sigma_{rot} = \{0.03, 0.12, 0.24\}$ radians from top to bottom on a 40 x 40 unit workspace averaged over 10 shapes from the Brown Vision Lab Data set. The increase in noise has little effect on the convergence of the two algorithms in simple regret. [TODO: GOING TO RERUN WITH MORE SHAPES TO BE SURE ABOUT THIS]

Fig. 11: Sensitivity Analysis for Thompson Sampling under translation uncertainty $\sigma_{trans} = \{3, 12, 24\}$ units from top to bottom on a 40 x 40 unit workspace. As you can see the increase in noise effects performance, however the 5000 samples needed for Gittins to converge at the the highest level of noise (which is a variance of over half the workspace) is much less that the samples needed for uniform allocation to converge in Fig. 5

REFERENCES

[1] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," *arXiv preprint arXiv:1111.1797*, 2011.

[2] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to mcmc for machine learning," *Machine learning*, vol. 50, no. 1-2, pp. 5–43, 2003.

[3] J.-Y. Audibert, S. Bubeck *et al.*, "Best arm identification in multi-armed bandits," *COLT 2010-Proceedings*, 2010.

[4] T. Barfoot and P. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *Robotics, IEEE Transactions on*, vol. 30, no. 3, pp. 679–693, June 2014.

[5] A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 1998.

[6] D. Bergemann and J. Välimäki, "Bandit problems," Cowles Foundation for Research in Economics, Yale University, Tech. Rep., 2006.

[7] S. Bubeck, R. Munos, and G. Stoltz, "Pure exploration in multi-armed bandits problems," in *Algorithmic Learning Theory*. Springer, 2009, pp. 23–37.

[8] R. E. Caflisch, "Monte carlo and quasi-monte carlo methods," *Acta numerica*, vol. 7, pp. 1–49, 1998.

[9] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," in *Advances in Neural Information Processing Systems*, 2011, pp. 2249–2257.

[10] V. N. Christopoulos and P. Schrater, "Handling shape and contact location uncertainty in grasping two-dimensional planar objects," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, 2007, pp. 1557–1563.

[11] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2011, pp. 2845–2850.

[12] C. Ferrari and J. Canny, "Planning optimal grasps," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 1992, pp. 2290–2295.

[13] V. Gabillon, M. Ghavamzadeh, and A. Lazaric, "Best arm identification: A unified approach to fixed budget and fixed confidence," in *Advances in Neural Information Processing Systems*, 2012, pp. 3212–3220.
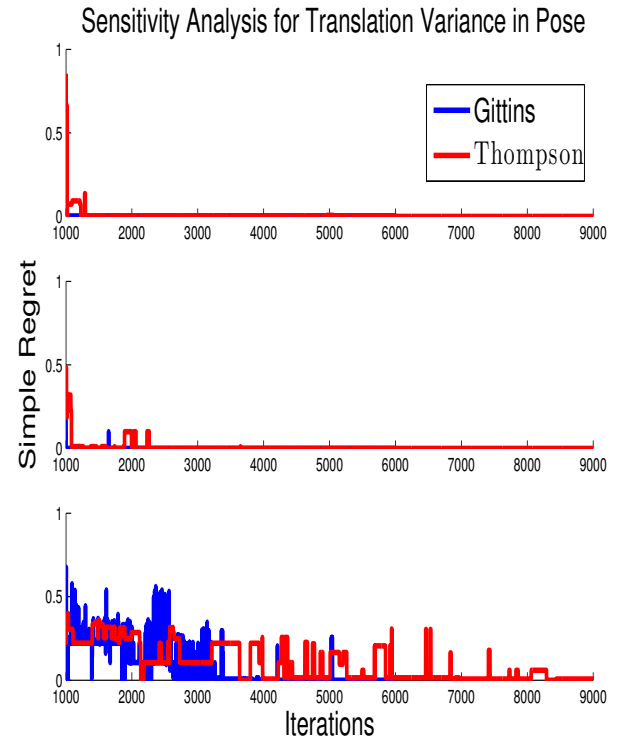
[14] K. Y. Goldberg and M. T. Mason, "Bayesian grasping," in *Robotics and Automation, 1990. Proceedings., 1990 IEEE International Conference on*. IEEE, 1990, pp. 1264–1269.

[15] G. A. Hollinger, B. Englot, F. S. Hover, U. Mitra, and G. S. Sukhatme, "Active planning for underwater inspection and the benefit of adaptivity," *Int. J. Robotics Research (IJRR)*, vol. 32, no. 1, pp. 3–18, 2013.

[16] K. Hsiao, M. Ciocarlie, and P. Brook, "Bayesian grasp planning," in *ICRA 2011 Workshop on Mobile Manipulation: Integrating Perception and Manipulation*, 2011.

[17] M. N. Katehakis and A. F. Veinott Jr, "The multi-armed bandit problem: decomposition and computation," *Mathematics of Operations Research*, vol. 12, no. 2, pp. 262–268, 1987.

[18] E. Kaufmann, O. Cappé, and A. Garivier, "On bayesian upper confidence bounds for bandit problems," in *International Conference on Artificial Intelligence and Statistics*, 2012, pp. 592–600.

[19] B. Kehoe, D. Berenson, and K. Goldberg, "Estimating part tolerance bounds based on adaptive cloud-based grasp planning with slip," in *Automation Science and Engineering (CASE), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1106–1113.

[20] ——, "Toward cloud-based grasping with uncertainty in shape: Estimating lower bounds on achieving force closure with zero-slip push grasps," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 576–583.

[21] J. Kim, K. Iwamoto, J. J. Kuffner, Y. Ota, and N. S. Pollard, "Physically-based grasp quality evaluation under uncertainty," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3258–3263.

[22] J. Laaksonen, E. Nikandrova, and V. Kyrki, "Probabilistic sensor-based grasping," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 2019–2026.

[23] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[24] B. León, S. Ulbrich, R. Diankov, G. Puche, M. Przybylski, A. Morales, T. Asfour, S. Moisio, J. Bohg, J. Kuffner, and R. Dillmann, *OpenGRASP: A Toolkit for Robot Grasping Simulation*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2010, vol. 6472, pp. 109–120.

[25] O. Madani, D. J. Lizotte, and R. Greiner, "The budgeted multi-armed bandit problem," in *Learning Theory*. Springer, 2004, pp. 643–645.

[26] J. Mahler, S. Patil, B. Kehoe, J. van den Berg, M. Ciocarlie, P. Abbeel, and K. Goldberg, "Gp-gpis-opt: Grasp planning under shape uncertainty using gaussian process implicit surfaces and sequential convex programming."

[27] O. Maron and A. W. Moore, "Hoeffding races: Accelerating model selection search for classification and function approximation," *Robotics Institute*, p. 263, 1993.

[28] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *Robotics & Automation Magazine, IEEE*, vol. 11, no. 4, pp. 110–122, 2004.

[29] V. Mnih, C. Szepesvári, and J.-Y. Audibert, "Empirical bernstein stopping," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 672–679.

[30] F. T. Pokorny and D. Kragic, "Classical grasp quality evaluation: New algorithms and theory," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 3493–3500.

[31] C. Rasmussen and C. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[32] C. E. Rasmussen and H. Nickisch, "Gaussian processes for machine learning (gpml) toolbox," *The Journal of Machine Learning Research*, vol. 9999, pp. 3011–3015, 2010.

[33] H. Robbins, "Some aspects of the sequential design of experiments," in *Herbert Robbins Selected Papers*. Springer, 1985, pp. 169–177.

[34] M. Rothschild, "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, vol. 9, no. 2, pp. 185–202, 1974.

[35] R. Simon, "Optimal two-stage designs for phase ii clinical trials," *Controlled clinical trials*, vol. 10, no. 1, pp. 1–10, 1989.

[36] E. Solak, R. Murray-Smith, W. E. Leithead, D. J. Leith, and C. E. Rasmussen, "Derivative observations in gaussian process models of dynamic systems," 2003.

[37] D. L. St-Pierre, Q. Louveaux, and O. Teytaud, "Online sparse bandit for card games," in *Advances in Computer Games*. Springer, 2012, pp. 295–305.

[38] F. Stulp, E. Theodorou, J. Buchli, and S. Schaal, "Learning to grasp under uncertainty," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5703–5708.

[39] N. Vahrenkamp, T. Asfour, and R. Dillmann, "Simo: A simulation and motion planning toolbox for c++."

[40] J. Weisz and P. K. Allen, "Pose error robust grasping from contact wrench space metrics," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 557–562.

[41] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," *Gaussian Proc. in Practice*, 2007.

[42] Y. Zheng and W.-H. Qian, "Coping with the grasping uncertainties in force-closure analysis," *Int. J. Robotics Research (IJRR)*, vol. 24, no. 4, pp. 311–327, 2005.

APPENDIX
GAUSSIAN PROCESS IMPLICIT SURFACE FOR REPRESENTING SHAPE UNCERTAINTY

In order to solve our problem definition, we must estimate $P(Q(\Gamma) > 0)$ for a given grasp . We will first discuss how the GPIS is constructed, then which grasp metric Q we chose and lastly proceed into evaluating $P(Q(\Gamma) > 0)$ efficiently.

A. Gaussian Process (GP) Background

We refer the reader to [**?**] for a more detailed explanation of the GP construction, which we summarize here. Given training data $\mathcal{D} = \{\mathcal{X}, \mathbf{y}\}$ and covariance function $k(\cdot, \cdot)$, the posterior density $p(sd_*|\mathbf{x}_*, \mathcal{D})$, or the distribution on signed distance field, at a test point $\mathbf{x}_*$ is shown to be [32]:

$$p(sd_*|\mathbf{x}_*, \mathcal{D}) \sim \mathcal{N}\big(\mu(\mathbf{x}_*), \Sigma(\mathbf{x}_*)\big)$$
$$\mu(\mathbf{x}_*) = k(\mathcal{X}, \mathbf{x}_*)^\intercal (K + \sigma^2 I)^{-1} \mathbf{y}$$
$$\Sigma(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathcal{X}, \mathbf{x}_*)^\intercal (K + \sigma^2 I)^{-1} k(\mathcal{X}, \mathbf{x}_*)$$

where $K \in \mathbb{R}^{l \times l}$ is a matrix with entries $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\mathcal{X}, \mathbf{x}_*) = [k(\mathbf{x}_1, \mathbf{x}_*), \dots, k(\mathbf{x}_l, \mathbf{x}_*)]^\intercal$. This derivation can

also be used to predict the mean and variance of the function gradient by extending the kernel matrices using the identities [36]:

$$\text{cov}\left(sd(\mathbf{x}_i), sd(\mathbf{x}_j)\right) = k(\mathbf{x}_i, \mathbf{x}_j) \tag{5}$$

$$\text{cov}\left(\frac{\partial sd(\mathbf{x}_i)}{\partial x_k}, sd(\mathbf{x}_j)\right) = \frac{\partial}{\partial x_k} k(\mathbf{x}_i, \mathbf{x}_j) \tag{6}$$

$$\text{cov}\left(\frac{\partial sd(\mathbf{x}_i)}{\partial x_k}, \frac{\partial sd(\mathbf{x}_j)}{\partial x_l}\right) = \frac{\partial^2}{\partial x_k \partial x_l} k(\mathbf{x}_i, \mathbf{x}_j) \tag{7}$$

For our kernel choice we decided to use the square exponential kernel, similar to [11]. Other kernels relevant to GPIS are the thin-plate splines kernel and the Matern kernel [41].

We construct a GPIS by learning a Gaussian process to fit measurements of a signed distance field of an unknown object. Precisely, $x_i \in \mathbb{R}^2$ in 2D and $x_i \in \mathbb{R}^3$ in 3D, and $y_i \in \mathbb{R}$ is a noisy signed distance measurement to the unknown object at x_i.

B. Sampling Shape from GPIS Distribution

To compute the above distribution we must draw samples from $p(\theta)$. In order to draw shape samples from a GPIS, one needs to sample from signed distance function, sd, over the joint on all points in the workspace $\mathcal{W}$ or $p(sd(\mathcal{W}))$. Since this is a GPIS, we know the following

$$p(S) = p(sd(\mathcal{W})) \sim N(\mu(\mathcal{W}), \Sigma(\mathcal{W})) \tag{8}$$

Thus if the workspace is an $n \times n$ grid, the joint distribution is an n^2 multi-variate Gaussian, due to $sd : \mathbb{R}^2 \to \mathbb{R}$. Shape samples drawn from the distribution appear in Fig. 12.
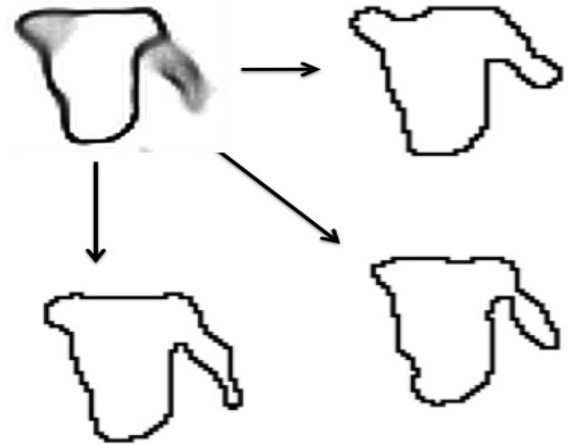

Fig. 12: Shape samples drawn from Eq. 8 on the object in the upper left corner. Given a shape sample we highlight the zero-crossing of the level set in black

C. Distribution on Surface Normals

Using Eq. 6 and Eq. 7, we can compute the mean of the gradient $\mu_\nabla(x)$ and the covariance of the gradient $\Sigma_\nabla(x)$ respectively. Thus we can compute the distribution around the surface normal for a given point in $\mathcal{W}$. We can now write

One interesting effect of this technique is that we can now marginalize out the line of action model and visual what the

surface normal distribution is along a given line of action. To our knowledge this is the first attempt to visual surface normals along a grasp plan. Marginalization can be performed as follows:

$$p(\mathbf{n}_i) = \int_a^b p\big(\mathbf{n}_i = \mathbf{v}|\mathbf{c}_i = \gamma(t)\big)p\big(\mathbf{c}_i = \gamma(t)\big)dt \quad (9)$$

Grasp metrics such as Ferrari-Canny require $\mathbf{n}_i$ be normalized, or, equivalently, a member of the sphere $\mathcal{S}^{d-1}$ [12]. To account for this we densely sample from the distribution $p\big(\mathbf{n}_i)\big)$ and project onto $\mathcal{S}^{d-1}$. In Fig.**??**, we visualize the distribution on $\mathbf{n}_i$ calculated for a given GPIS and approach line of action.

D. Expected Center of Mass

We recall the quantity $P(sd(x) < 0) = \int_{-\infty}^0 p(sd(x) = s \mid \mu(x), \Sigma(x))ds$ is equal to the probability that x is interior to the surface under the current observations. We assume that the object has uniform mass density and then $P(sd(x) < 0)$ is the expected mass density at x. Then we can find the expected center of mass as:

$$\bar{z} = \frac{\int_{\mathcal{W}} xP(sd(x) < 0)dx}{\int_{\mathcal{W}} P(sd(x) < 0)dx} \quad (10)$$

which can be approximated by sampling $\mathcal{W}$ in a grid and approximating the spatial integral by a sum. Since this operation involves the entire SDF, one would want to use a low resolution grid for computational efficiency. We show the computed density and calculated expected center of mass for a marker in Fig. 13.
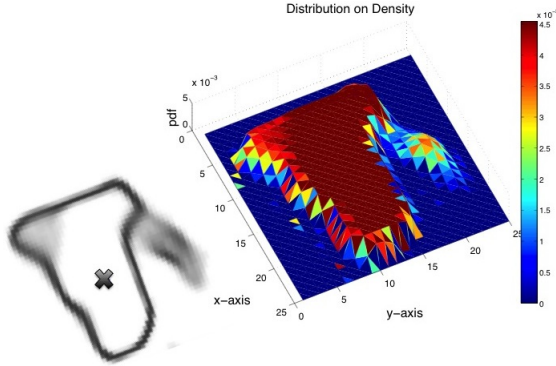

Fig. 13: Left: A surface with GPIS construction and expected center of mass (black X) Right: The distribution on the density of each point assuming uniform density