Budgeted Multi-Armed Bandit Models for Sample-Based Grasp Planning in the Presence of Uncertainty

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

***Abstract*— 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 defined by a Gaussian process implicit surface (GPIS), but the method is applicable to other models of uncertainty. We derive distributions on contact points, surface normal, and center of mass and use these solve the associated MAB model, finding that it computes grasps of similar quality 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.**

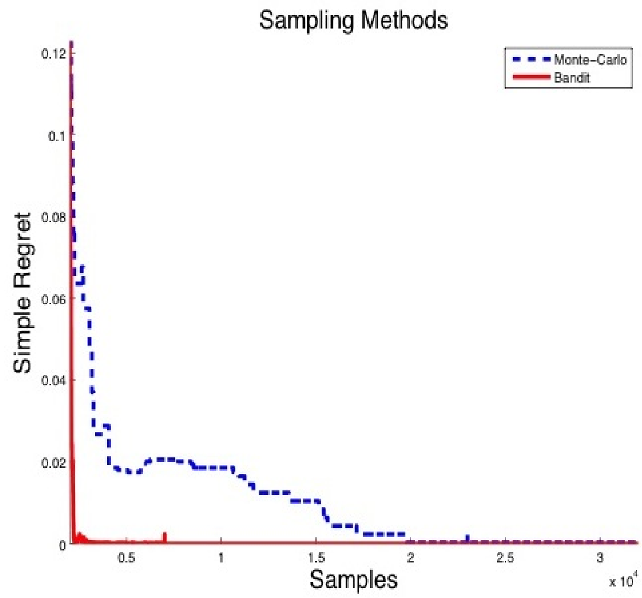

Fig. 1: Convergance in regret of the bandit sampling method (red), compared to the traditional Monte-Carlo method. The fast convergence of the bandit method is due to its ability to intelligently pick what grasp to sample next in a given set of proposed grasps on an object with shape uncertainty

I. Introduction

Consider a robot packing boxes in a shipping warehouse environment, where it may 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 even material properties like friction coefficient or center of mass. Due to sensor noise and missing data due to partial visibility and object properties such as transparency, the robot may not be able to measure these quantities exactly. Grasp planners that assume prior knowledge of object geometry or an exact measurement of pose may fail in this environment.

Grasp quality metrics have been developed to determine if a grasp will be successful or not and how much force it needs to exert to resist an opposing force, however most of them evaulate a grasp assuming all the parameters are known [9]. This motivates using knowledge of uncertainty to select grasps, but most methods for evaluating the quality of a single grasp in the presence of uncertainty require use of an exhaustive sampling over the possible values of the uncertain quantity [15], [36]. To select a grasp with high quality this evaluation is often performed for a large set of potential grasps, which can be very time-consuming. However, when evaluating a set of grasps we may be able to determine the difference of quality between grasps with only a few samples and throw away grasps that are likely to be suboptimal [13]. Thus, we can adaptively 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 [3], [17], [27] provides a way to reason about selecting the next grasp to evaluate and the grasps to discard from consideration. The goal in a MAB model is to make a sequence of decisions over a set of possible options such that a measure of the expected reward of such decisions 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 [30], market pricing [29], and choosing strategies for games [32]. The budgeted multi-

[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

armed bandit model [20] is a specialization of the MAB model with a finite stopping time where 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 in this paper is formulating the problem of planning grasps with a high expected quality in the presence of uncertainty as a budgeted multi-armed bandit model. We use this formulation to rank a set of potential grasps by expected Ferrari-Canny quality [9] under shape uncertainty. We use a budgeted multi-armed bandit model since we would like to execute only one grasp plan after evaluating the expected quality of many potential grasp plans. We choose the model of uncertainty to be a Gaussian process implicit surface (GPIS), a Bayesian representation of shape uncertainty that seen recent use in various robotic applications [8], [11], however the method applies to any model of uncertainty that can be sampled. We also show how to estimate distributions on the contact points and surface normals and center of mass using a GPIS shape representation. We then leverage these distributions to reduce the computational complexity of sampling from the GPIS model from $O(n^6)$ to $O(n^3)$, where the workspace is discretized as a $n \times n$ grid. Our experiments demonstrate that using the MAB sampling method improves the time to rank a set of 1000 grasps by 10x, an order of magnitude improvement, over the baseline Monte-Carlo approach. These promising results suggest that our MAB approach could be used on other types of uncertainty, such as pose, friction coefficent or center of mass and provide significant speed improvements.

II. Related Work

Past work on grasping under uncertainty has considered state uncertainty [10], [33], uncertainty in contact locations with an object [38], uncertainty in object pose [7], [36], [15]. The effect of uncertainty in object geometry on grasp selection has been studied for spline representations of objects [7], extruded polygonal mesh models [13], [14], and point clouds [12].

Currently, the most common 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 [7], [13], [14]. Monte-Carlo sampling involves drawing random samples from a distribution to approximate an expected value[6], which can be slow when the distribution is high-dimensional, such as for distributions on possible shapes. To address this, Kehoe et al. [13] demonstrated a procedure for finding a minimum bound on expected grasp quality given shape uncertainty, which reduced the number of terms needed in Monte-Carlo sampling in order to choose the highest quality grasps. The adaptive sampling pruned grasps using only the sample mean and did not utilize any estimates of how accurate the current sample mean is. Laaksonen et al. [16] used Markov Chain Monte-Carlo (MCMC) sampling to estimate grasp quality and object pose under shape and pose uncertainty. MCMC simplified sampling from a complicated joint distribution on pose and shape, but it can be slow to converge to the correct distribution [2].

We chose to study our MAB sampling method for shape uncertainty using a Gaussian process implicit surface representation. Our decision to use this uncertainty model is based on GPIS's ability to combine various modes of noise observations such as tactile, laser and visual [25], [37], [8] and its recent use in modeling uncertainty for a number of robotic applications. Hollinger et al. used GPIS as a model of uncertainty and performed active sensing on the hulls in underwater boats [11]. Dragiev et al. showed how GPIS can enable a grasp controller on the continuous signed distance function [8]. Mahler et al. used the GPIS representation to find locally optimal anti-podal grasps by framing grasp planning as an optimization problem [21].

III. Multi-Armed Bandit Model

The multi-armed bandit model, originally described by Robbins [28], is a statistical decision model of an agent trying to make correct decisions, while gathering information at the same time. 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 [5], [27], [4]. 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 [30], market pricing [29], and choosing strategies for games [32].

There are a number of algorithms for developing policies to balance exploration and exploitation. One algorithm is $\epsilon-$greedy, which is the idea of choosing the arm with the highest empirical expected reward with $1-\epsilon$ probability and choosing a random arm with probability ϵ [3]. A class of algorithms that have stronger theoretical guarantees are from the Upper Confidence Bound (UCB) family. UCB algorithms maintain the empirical expected reward based off of pulling each arm multiple times, while also estimating an upper bound on the true expected reward using assumptions on the probability distribution of rewards and number of times each arm has been sampled. The algorithm chooses the next option based on Thompson sampling [1] or the Gittins index policy [35]. When the rewards come from exponential family distribution, UCB minimizes cumulative regret, which is the sum over each sub-optimal arm of the number of times that arm is pulled times the difference between the true expected reward of an optimal arm and the true expected reward of that sub-optimal arm.

In the grasping context, one is only interesting in choosing the best grasp at the end of an exploration cycle. We can formulate our problem as a "budgeted multi-armed bandit model" [20]. The budgeted multi-armed bandit model focuses on efficiently exploring the arms until a finite stopping time is reached. The objective of the budgeted multi-armed bandit model is to minimize the "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. Thus the budgeted multi-armed bandit model reduces to exploration until the very last time step, at which time one exploitation step is taken.

IV. 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.

A. Grasp Metric

In their work over two decades ago Ferrari and Canny [9], demonstrated a method to rank grasps by considering their contact points and surface normals. Importantly the magnitude of Q yields a measurement that allows one to rank grasps by their physical stability and evaluate the property of force-closure. Furthermore, it has wide spread use in grasp packages like GraspIT[22], OpenGrasp[18] and Simox [34], which motivates studying its effect with uncertainties.

The L^1 version of the metric works by taking as input the contact points $\mathbf{c}_1, ..., \mathbf{c}_m$, surface normals $\mathbf{n}_1, ..., \mathbf{n}_m$, center of mass z and friction coefficient μ. Then 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. A wrench is defined as concatenation of a force and torque vector. If the convex hull doesn't enclose the origin, the grasp is not in force-closure. Thus a grasp can be parameterized by the following tuple $g = (\mathbf{c}_1, ..., \mathbf{c}_m, \mathbf{n}_1, ..., \mathbf{n}_m, \mu, \mathbf{z})$, our method is applicable to all grasp metrics that represent a grasp as the tuple g, such as [7], [19].

Since we are in an uncertain environment, we are interested in calculating the expected quality, or $E(Q)$. $E(Q)$ is computed by sampling from our distributions on pose, shape or material properties (friction coefficient or density) and averaging the qualities that are computed. Reducing the time to do this for a large number of grasps is the primary focus of the paper.

B. Line of action

In an uncertain environment one wouldn't actually know the true g, thus we have to work with the trajectory of the gripper. Similar to the work of [7], 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*. While we currently assume the gripper moves free of noise, this approach would be applicable to high precision robots such as industrial based on. Future work will look at how to efficiently sample when the approach trajectory has noise.

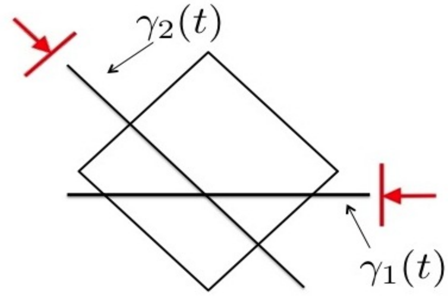

Fig. 2: Illustration of a grasp plan Γ composed of two lines of action, $\gamma_1(t)$ and $\gamma_2(t)$

C. Problem Definition

Given a 2-D workspace $\mathcal{W}$, with an unknown object represented as a trained GPIS model, which we describe in Section VI-A and set of possible grasp plans G. We are interested in determining

$$\Gamma^* \in \text{argmax}_{\Gamma \in G} E(Q(\Gamma)) \tag{1}$$

with respect to a chosen grasp metric Q.

V. Multi-Armed Bandits for Grasp Selection

While a standard approach to solving the problem in Eq. 1 would be to simply perform Monte-Carlo integration on each Γ_i and compute the expected grasp quality, we propose treating the problem as a multi-armed bandit model and forming a policy for selecting which grasp to sample. In our setting, we have a probabilistic shape representation and would like to evaluate many potential grasps on that shape model. Motivated by limited computational resources we are interested in how to intelligently allocate sampling resources to efficiently find the best grasp plan Γ^*. Here each arm corresponds to a different grasp plan and pulling the arm is sampling the shape representation and evaluating the arm's grasp plan on the sampled shape representation. The reward for pulling an arm is the grasp quality of the resulting grasp on the sampled shape. We have a policy for exploration of different grasp plans (i.e. choosing which one to sample next) and at a given stopping time we choose to execute the grasp plan with the highest expected quality based on the samples received so far, which would correspond to actually performing the grasp. The number of samples needed before the simple regret reaches zero, determines how effective an exploration policy is for grasp evaluation.

In solving Eq. 1, we want to pick the grasp plan with the highest $E(Q(\Gamma))$ determined via Monte-Carlo Integration. Due to the non-linear nature of our grasp metric, we currently do not have a way to represent the distribution on grasp quality of $p(Q(\Gamma))$ in closed form, which limits the bandit algorithms available to us.

For our exploration policy, we chose to use a method called *successive elimination*. We maintain a subset of the possible grasp plans that are statistically indistinguishable from the best grasp plan. During exploration, we sample uniformly from this subset. This has been shown to have good performance when the number of evaluations is large [5]. For each grasp plan, we maintain a 95% confidence interval around the expected quality of the grasp plan. The width of the confidence interval for a grasp plan Γ_i that has been sampled n_i times and the samples have standard deviation σ_i is:

$$C_i := \frac{1.96\sigma_i}{\sqrt{n_i}}. \tag{2}$$

Let $\hat{\mu_i}$ be the sample mean of the grasp quality for grasp plan i. Then the confidence interval for grasp plan i is $[\hat{\mu_i} - C_i, \hat{\mu_i} + C_i]$ [6]. After a sample from a grasp plan, we check to see if the confidence interval of the sampled grasp intersects with the confidence interval of the current best grasp plan, and if it does not, we prune that grasp plan from the current active set of grasp plans. Since we prune grasps that our below a confidence interval we have statistical guarantees unlike earlier approaches to this problem[14]. In future work, we will look at using other exploration strategies.

VI. Evaluating a Grasp on Gaussian Process Implicit Surface

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

A. Gaussian Process (GP) Background

We refer the reader to [21] for a more detailed explanation of the GP construction, which we summarize here. Given the 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 [26]:

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

where $K \in \mathbb{R}^{l\times l}$ is a matrix with entries $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\mathcal{X}, \mathbf{x}_*) = [k(\mathbf{x}_1, \mathbf{x}_*), \ldots, k(\mathbf{x}_l, \mathbf{x}_*)]^\intercal$. This derivation can also be used to predict the mean and variance of the function gradient by extending the kernel matrices using the identities [31]:

$$\operatorname{cov}\left(sd(\mathbf{x}_i), sd(\mathbf{x}_j)\right) = k(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

$$\operatorname{cov}\left(\frac{\partial sd(\mathbf{x}_i)}{\partial x_k}, sd(\mathbf{x}_j)\right) = \frac{\partial}{\partial x_k} k(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

$$\operatorname{cov}\left(\frac{\partial sd(\mathbf{x}_i)}{\partial x_k}, \frac{\partial sd(\mathbf{x}_j)}{\partial x_l}\right) = \frac{\partial^2}{\partial x_k \partial x_l} k(\mathbf{x}_i, \mathbf{x}_j) \tag{5}$$

For our kernel choice we decided to use the square exponential kernel, similar to [8]. Other kernels relevant to GPIS are the thin-plate splines kernel and the Matern kernel [37].

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. Calculating the Expected Grasp Quality

Given a proposed grasp plan Γ, the expected grasp quality can be evaluated as follows:

$$E(Q(\Gamma)) = \int Q(g|S,\Gamma)p(S)dS \tag{6}$$

Where $Q(g|S,\Gamma)$ is the grasp quality that is computed on a shape sample drawn from $p(S)$. To compute this we intersect the zero crossing of the level set with the propose grasp plan Γ and determine the parameters g, this has been the approach taken in previous work [13], [14], [7]. See Fig. 3 for an example of what samples drawn from $p(S)$ induced by GPIS look like. For computational reasons we approximate the integral via Monte-Carlo Integration. We use importance sampling to draw from the distribution induced by GPIS and calculate the following

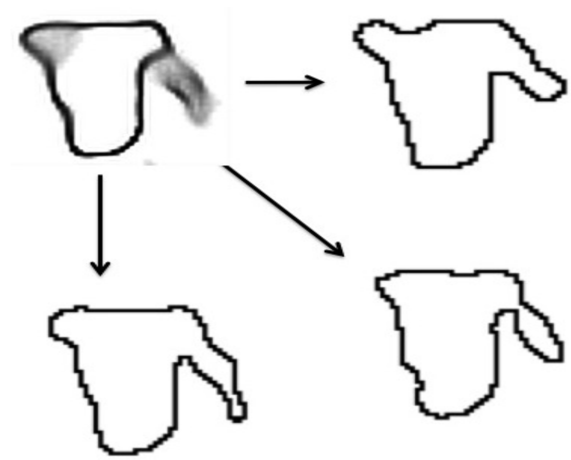

Fig. 3: Shape samples drawn from Eq. 7 on the object in the upper left corner. Given a shape sample we highlight the zero-crossing of the level set in black

$$E(Q(\Gamma)) \approx \frac{1}{N}\sum_{i=1}^{N} Q(g|S_i,\Gamma), \quad S_i \sim p(S)$$

To compute the above distribution we must draw samples from $p(S)$. 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

$$p(S) = p(sd(\mathcal{W})) \sim N(\mu(\mathcal{W}), \Sigma(\mathcal{W})) \tag{7}$$

Thus if the workspace is an $n\times n$ grid, the joint distribution is an n^2 multi-variate Gaussian, due to $sd : \mathbb{R}^2 \to \mathbb{R}$. Sampling from a Gaussian involves inverting the covariance matrix and inversion is in the naive way $O(n^3)$ [24]. Thus the complexity of this operation is $O(n^6)$ in 2D and $O(n^9)$ in 3D.

To reduce complexity we propose sampling not from the shape distributions, but instead from the distributions on the grasps parameters themselves. We recall that a grasp according to our metric is defined as the tuple $g = (\mathbf{c}_1, ..., \mathbf{c}_m, \mathbf{n}_1, ..., \mathbf{n}_m, \mu, z)$. We are thus interested in calculating $p(g|\Gamma, \mu(x), \Sigma(x))$. The distribution on a grasp is defined then as:

$$p(g) = p\big(\mathbf{c}_1, ..., \mathbf{c}_m, \mathbf{n}_1, ..., \mathbf{n}_m | \Gamma, \mu(x), \Sigma(x)\big) \tag{8}$$

We note here that we currently use the friction coefficient μ and the expected center of mass $\bar{z}$ as deterministic values. For grippers that do not approach along the same line of action (i.e. non-parallel jaw grippers) we make the assumption that each contact and normal pair is independent, or

$$p(g) = \prod_{i=1}^{m} p\big(\mathbf{c}_i, \mathbf{n}_i | \gamma_i(t), \mu(x), \Sigma(x)\big) \tag{9}$$

We compute the expected grasp quality now as follows:

$$E(Q(\Gamma)) = \frac{1}{N}\sum_{i=1}^{N} Q(g_i), \quad g_i \sim p(g) \tag{10}$$

We will now show how these distributions can be computed and how the computational complexity for sampling form a grasp plan is reduced from $O(n^6)$ to $O(n^3)$ for a GPIS model.

VII. Distribution of Grasp Parameters

To sample from $p(g)$, we need to sample from the distributions associated with a line of action $p(\mathbf{n}_i, \mathbf{c}_i | \gamma_i(t), \mu(x), \Sigma(x))$. Using Bayes rule we can rewrite this as

$$p(\mathbf{n}_i, \mathbf{c}_i | \gamma_i(t), \mu(x), \Sigma(x)) =$$
$$p(\mathbf{n}_i | \mathbf{c}_i, \gamma_i(t), \mu(x), \Sigma(x)) p(c_i | \gamma_i(t), \mu(x), \Sigma(x))$$

In section VII-A, we look at how to sample from $p(\mathbf{c}_i | \gamma_i(t), \mu(x), \Sigma(x))$. Then in section VII-B, we look at how to sample from $p(\mathbf{n}_i | \mathbf{c}_i, \gamma_i(t), \mu(x), \Sigma(x))$ and present a novel visualization technique for the distribution on surface normals. Lastly in section VII-C, we show a way to calculate the expected center of mass assuming a uniform mass distribution. .

A. Distribution on Contact Points

We would like to find the distribution on contact point $\mathbf{c}_i$. A contact point in terms of the GPIS and line of action model can be defined as the point, t, when the signed distance function is zero and no points before said point along the line have touched the surface. We express this as the following conditions:

$$sd(\gamma(t)) = 0 \tag{11}$$

$$sd(\gamma(\tau)) > 0, \; \forall \tau \in [a, t) \tag{12}$$

We will now demonstrate how to efficiently sample from $p\big(\mathbf{c}_i | \mu x, \Sigma x, \gamma_i t\big)$

The probability distribution along the line $\gamma(t)$ is given by:

$$p\big(sd(\gamma(t)); \mu(t), \Sigma(t)\big) \; \forall t \in [a, b] \tag{13}$$

This gives the signed distance function distributions along the entire line of action in the workspace as a multivariate Gaussian. One could think of this as a marginalization of all other points in signed distance field except the line of action. To sample contact points, one can simply draw samples from Eq. 2 and iterate from a to b until they reach a point that satisfies Eq. 11 and Eq. 12.

B. Distribution on Surface Normals

Using Eq. 4 and Eq. 5, 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

$$p\big(\mathbf{n}_i | \mathbf{c}_i = \gamma(t)\big) = p\big(\mathbf{n}_i | \mu(\gamma(t)), \Sigma(\gamma(t))\big)$$

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:

$$p(\mathbf{n}_i) = \int_a^b p\big(\mathbf{n}_i = \mathbf{v} | \mathbf{c}_i = \gamma(t)\big) p\big(\mathbf{c}_i = \gamma(t)\big) dt \tag{14}$$

Grasp metrics such as Ferrari-Canny require $\mathbf{n}_i$ be normalized, or, equivalently, a member of the sphere $\mathcal{S}^{d-1}$ [9]. 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.4, we visualize the distribution on $\mathbf{n}_i$ calculated for a given GPIS and approach line of action.

C. 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:

$$\bar{z} = \frac{\int_{\mathcal{W}} x P(sd(x) < 0) dx}{\int_{\mathcal{W}} P(sd(x) < 0) dx} \tag{15}$$

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. 5.

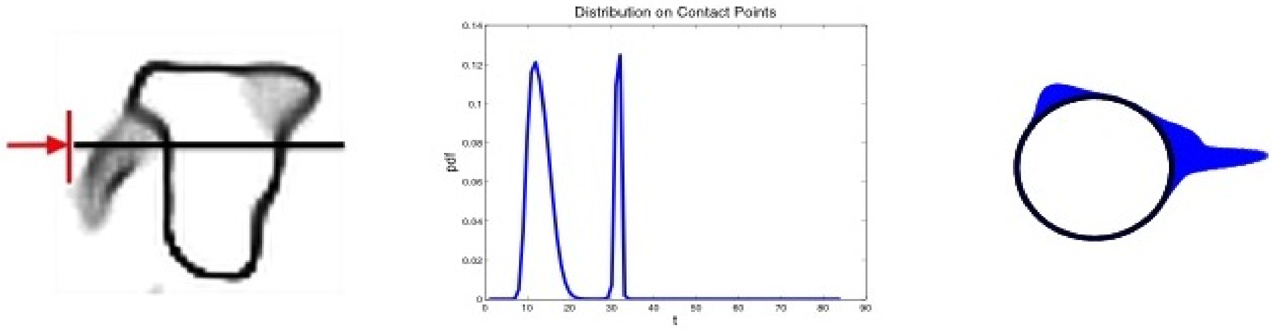

Fig. 4: (Left to Right): Line of action for a given gripper on an uncertain surface representing a measuring cup. Visualization technique comes from [21]. Distribution $p(c)$ as a function of t, the position along the line of action $\gamma(t)$. The two modes correspond to the different potential contact points, either the handle or the base of the cup. Lastly, the distribution on the surface normals (inward pointing) along $\gamma(t)$ described by equation 14. We plotted the probability mass along the unit circle, to reflect the normalization of the surface normals.

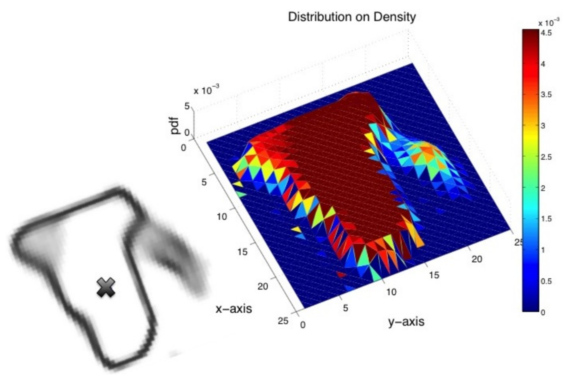

Fig. 5: Left: A surface with GPIS construction and expected center of mass (black X) Right: The distribution on the density of each point assuming uniform density

D. Complexity Analysis on Sampling From Grasps Distributions

After formally deriving each distribution, we can now sample along each line of action $\gamma_i(t)$ from the joint distribution $p(\mathbf{c}_i, \mathbf{n}_i | \gamma_i(t))$, due to our independence assumption Eq. 9. This can be done by drawing samples from the GP for signed distance and normals simultaneously and using our projection technique for the normal distribution.

Having a distribution along a line of action model allows us to sample from those instead of the joint distribution $p(sd(\mathcal{W}))$. Assuming the line of action is on the order of n, sampling from this distribution for a single grasp is $O(n^3)$. However, each proposed grasp plan Γ requires the distribution to be computed, so if we have $T = |G|$ then the complexity is $O(Tn^3)$. In practice, this should be much smaller than $O(n^6)$.

VIII. Experiments

For the experiments below we used common household objects. The objects used can be found at `http://rll.berkeley.edu/grasping/` We manually created a 25 x 25 grid, by tracing a pointcloud of the object on a table taken with a Primesense Carmine depth sensor. To accompany the SDF, we created an 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. Our visualization technique follows the approach of [21] 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

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. 6, we plotted the simple regret for three of the shapes in our data set averaged over 100 runs and compare it to the Monte-Carlo method that randomly chooses a grasp plan to draw a sample from. We initialize both the Monte-Carlo and bandit technique by sampling each grasp 2 times this is to achieve the standard deviation. We draw samples from our calculated distributions $p(g)$. The most interesting thing is that the regret is minimized at least an order of magnitude faster than the traditional approach for the shapes we considered, thus motivating the use of including observations as you select samples.

B. Sampling from Grasps Vs. Shape

We first tested 1000 grasp plans and sampled each one 5000 times and measured the RMS error between converged expected grasp plan qualities for sampling shape Eq. 6 vs. grasps Eq. 10 was 0.004. After confirming the distributions converged close to the same value, we show the computational complexity in Fig. 7 of the two methods for evaluating 100 grasps on an $n \times n$ grid.

IX. 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

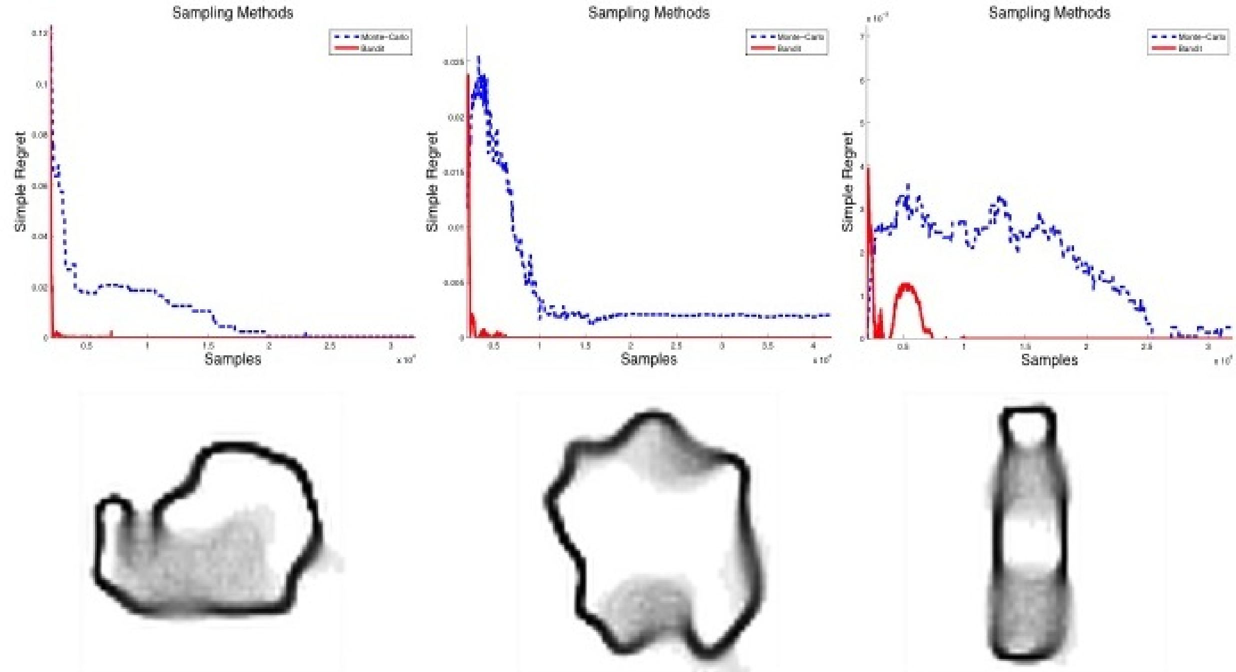

Fig. 6: Top: Comparison of two sampling methods. Red is the multi-armed bandit method that actively chooses which one to sample and Blue is the naive Monte-Carlo method. We measure their ability to converge in terms of simple regret averaged over 100 runs. We had them determine the best grasp when $|G| = 1000$. In all cases the bandit method converges a least a magnitude faster than the pure Monte-Carlo integration. Bottom: Three objects from our data set that we tested on (Tape, Loofa, Water Bottle), we used a visualization technique described in [21]

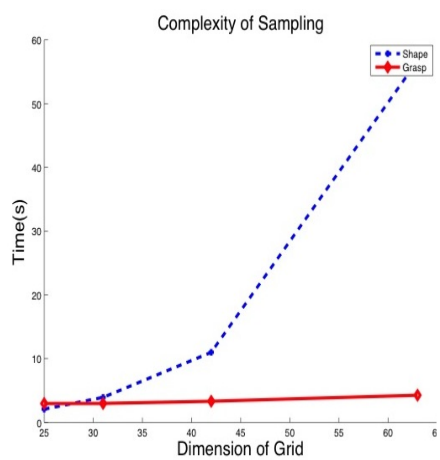

Fig. 7: Time it took to sample from 100 grasp distributions for a given resulution of the workspace. Blue line is sampling from $p(sd(\mathcal{R}))$ or shapes and Red is sampling from $p(g)$ or the calculated distribution on grasps. As you can see sampling from the calculated distributions scales much better.

and test the bandit method there. While we have empirically seen the bandit method always converge to the correct value, our method only has a statistical guarantee of doing so. A low quality grasp plan could be found, albeit with a low probability.

Sampling from our distribution $p(g)$ over $p(S)$ yields a reduction in computationally complexity, but only if the number of grasps one wants to evaluate remains small relative to n^3, techniques to ensure this could be to find locally optimal potential grasps using optimization approaches [21].

An additional problem is that we only have an expected center of mass and not a distribution on the center of mass. This might prove to be to expensive to compute, however recent work by Panahi et al. showed a way to bound the center of mass for convex parts. Extension of his work to implicit surfaces could be of possible interest [23].

X. Conclusion

Assessing grasp quality under shape uncertainty is computationally expensive as it often requires repeated evaluations of the grasp metric over many random samples. In this work, we proposed a multi-armed bandit approach to efficiently identify high-quality grasps under shape uncertainty. A key insight from our work is that uniformly allocating samples to grasps is inefficient, and we found that the Successive Elimination multi-armed bandit 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 an efficient representation of how shape uncertainty affects grasp parameters and thus grasp quality. We modeled uncertainty with Gaussian process implicit surfaces (GPIS) and derived the distribution of grasp parameters when a nominal grasp is applied to the GPIS. As a result, we were able to more efficiently sample from a distribution of grasps executions rather than the shape; leading to a complexity improvement of n^3 in the resolution of the representation. We evaluated this theoretical model on a dataset of common objects and confirmed that: (1) the bandits approach always converged to the best grasp in the candidate set, (2) it performs on average a magnitude faster than a naive uniform sampling approach in our experiments, and (3) sampling from the grasp contacts is 12x faster than sampling from shapes for a 64x64 grid.

XI. FUTURE WORK

Our results are promising and they suggest many avenues of future work. By utilizing the BMAB model, we can simply 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 [13] or splines [7]. It can further be applied to other grasp quality metrics or even simulation based evaluation methods [18].

Future work, will also consider applying BMAB approach to grasp planners like GraspIt! [22] to see if we can handle uncertainty and still maintain the time constraints needed for most real time applications. We currently have only tested one bandit algorithm, successive elimination. While our results were impressive, it remains to be seen how well it scales to the more complicated 3D models. However, the BMAB model has a large amount of literature to draw from as we encounter new and more challenging problems [4].

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

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

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

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

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

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

[13] 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.

[14] ——, "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.

[15] 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.

[16] 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.

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

[18] 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.

[19] Z. Li and S. S. Sastry, "Task-oriented optimal grasping by multifingered robot hands," *Robotics and Automation, IEEE Journal of*, vol. 4, no. 1, pp. 32–44, 1988.

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

[21] 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," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015.

[22] 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.

[23] F. Panahi and A. F. van der Stappen, "Bounding the locus of the center of mass for a part with shape variation," *Computational Geometry*, vol. 47, no. 8, pp. 847–855, 2014.

[24] K. B. Petersen, "The matrix cookbook."

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

[26] 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.

[27] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, vol. 58, pp. 527–535, 1952.

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

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

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

[31] 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.

[32] 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.

[33] 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.

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

[35] R. Weber *et al.*, "On the gittins index for multiarmed bandits," *The Annals of Applied Probability*, vol. 2, no. 4, pp. 1024–1033, 1992.

[36] 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.

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

[38] 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.