Multi-Armed Bandit Models for Sample-Based Grasp Planning in the Presence of Uncertainty [v-13 09-30-2014]

Michael Laskey, Zoe McCarthy, Jeff Mahler, Florian T. Pokorny, Sachin Patil,
Jur Van Den Berg, Danica Kragic, Pieter Abbeel, and Ken Goldberg

Abstract— In new robotic environments, uncertainty in pose and shape can arise from sensor noise, occlusions, and surface properties such as transparency and specularity. We show that framing grasp quality evaluation as a multi-armed bandit problem provides a principled way to reason about evaluating multiple grasps on an uncertain shape quickly. We demonstrate this method for planning grasps in the presence of shape uncertainty encoded by a Gaussian process implicit surface (GPIS). We utilize successive elimination to solve a multi-armed bandit problem and show a speedup in sample complexity versus evaluating grasps using Monte-Carlo integration, a popular method.

I. Introduction

[TODO: GET HIGH RES GPIS VISUALIZATIONS, INCORPORATE NEXT ROUND OF FEEDBACK]

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 plan grasps for these objects without any prior information other than its sensor data. As seen in Fig. 1, it is unavoidable that sensors will have missing or noisy data due to surface properties, such as specularity and transparency. Grasp planners based on an exact representation of shape may fail. This motivates using a representation of shape uncertainty to select grasps, but most current methods for evaluating quality under these representations use a brute-force sampling of shapes to evaluate grasp quality, which can be slow.

Often when evaluating grasps in this manner, there are a number of grasps that can be quickly shown to be inferior. When allocating resources to determine the best grasp, this information should be used in order to spend less time evaluating bad grasps. Multi-armed bandits provide a principled way to reason about evaluating multiple grasps and discarding grasps that are likely to be inferior.

We apply our multi-arm bandit approach to the Gaussian process implicit surface (GPIS), an emerging representation for shapes that gives a formal way to incorporate uncertain measurements from multiple sensors. GPIS has been recently used in a number of robotic applications, such as a feedback grasp controller, finding locally optimal grasps, and active exploration [6], [?], [7].

Our contributions are the following:

1) Framing the selection of a grasp as from a given set as a multi-arm bandit problem.
2) Demonstrating empirically that successive elimination, a multi-armed bandit algorithm, reduces the sampling
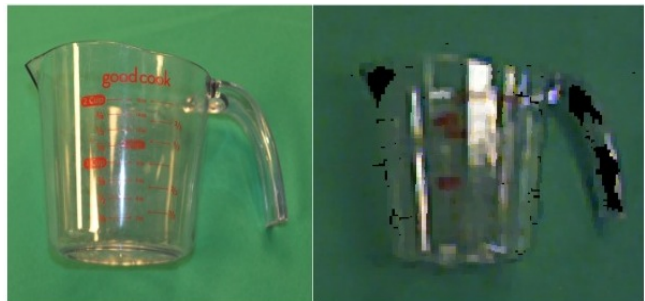

Fig. 1: Left: HD image of a common measuring cup from in a household environment. Right: Image taken from PR2's Primesense Carmine sensor, large parts of the measuring cup are not detected by the depth sensor and are not visible in the RGB image.

complexity to find a grasp with high expected Ferrari-Canny quality from a set of candidate grasps.

II. Related Work

Past work on grasping under uncertainty has considered state uncertainty [10], [11], uncertainty in object pose [2], [12], [13], or uncertainty in contact locations with an object [3]. The effect of uncertainty in object geometry on grasp selectiong has been studied for spline representations of objects [2], extruded polygonal mesh models [1], [8], and point clouds [14].

Currently, the most common method of evaluating grasp quality under shape and pose uncertainty is to rank a set of random grasps on an object using Monte-Carlo integration on shape and / or pose samples to evaluate a quality measure [2], [1], [8]. Monte-Carlo integration involves drawing random samples from a distribution to approximate an integral [15], which can be slow when the distribution is high-dimensional, such as for distributions on possible shape. To address this, Kehoe et al. [1], [8] demonstrated a procedure for finding a minumum bound on grasp quality given shape uncertainty, which reduced the number of terms needed in Monte-Carlo integration in order to rank grasps, but the method could potentially remove high quality grasps. Laaksonen et al. [16] used Markov Chain Monte Carlo (MCMC) sampling to estimate grasp stability and object pose online under shape and pose uncertainty. MCMC simplified sampling from complicated distributions on pose and shape, but it can be slow to converge to the correct distribution [17].

The multi-armed bandit problem [18], [19], [20] is to maximize the total expected reward over a sequence of possible choices over a set of competing options, each of which has a probablistic reward function. The optimal

sequence of choices if the rewards are known is to always pick the option with the maximum expected reward over the set of possible options. However, to discover which choice is optimal requires a balance between exploration, or choosing options that are not well known, and exploitation, or choosing the current 'best' option [18]. Solutions to this problem, called *policies*, include choosing the next option based on Thompson sampling [21], the Gittins index policy [22], and by using an upper confidence bound (UCB) for the expected reward of each option [23]. The UCB can be computed from the sample mean [24], Kullback-Leibler divergence [25], or a prior on the rewards [26]. Solutions to the multi-armed bandit problem are well suited to applications for which it is expensive to exhaustively evaluate all possible options, and therefore these methods have been applied to optimal design of clinical trials [27], market pricing [28], and choosing strategies for games [29].

In this paper we consider the "budgeted multi-arm bandit" problem, where exploration and exploitation are decoupled and the agent can only explore for a finite time before deciding which arm to exploit [30].

We study using Gaussian process implicit surfaces (GPIS) to represent shape uncertainty, based on its ability to combine various modes of nois observations such as tactile, laser and visual. GPIS representations model a distribution over implicit surface functions, which are becoming increasingly popular in robotics due to their success in real-time surface modeling from range images [5], [4]. GPIS representations have recently been used in a number of robotic appications. Hollinger et al. used GPIS to perform active sensing on the hulls in underwater boats [7]. Dragiev et al. showed how GPIS can represent shapes for grasps and used as a grasp controller on the continuous signed distance function [6]. Mahler et al. used GPIS representation to find locally optimal anti-podal grasps by framing grasp planning as an optimization problem [?]. Sampling from GPIS for grasp quality evaluation is computationally intensive, however, because it involves the Cholesky decomposition of an $m \times m$ matrix which takes $O(m^6)$ time, where m is the dimension for a 2D grid over and object [31].

III. PRELIMINARIES AND PROBLEM STATEMENT

In this section we describe the multi-arm bandit problem and how it relates to grasping.

A. Multi-Arm Bandit Problem

The multi-arm bandit problem originally described by Robbins [20] 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-arm bandit problem is a gambler that has K independent slot machine arms and tries to infer which one will yield the highest reward. 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 [32], [20], [33].

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 [18]. A class of algorithms that have theoretical guarantees are Upper Confidence Bound (UCB). 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 has been sampled for each arm. The algorithm chooses the arm at each step which has the highest upper bound on expected reward, since that is the most potentially promising arm. When the rewards come from exponential family distribution, UCB minimizes cumulative regret, which is the number of times a sub-optimal 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 practice, when the distribution on the rewards of arms is not known, the empirical methods such as $\epsilon-$greedy have shown to have better performance [?].

B. Budgeted Multi-Arm Bandit Problem and Grasping

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^*. We can formulate this problem as a "budgeted multi-arm bandit problem" [30]. 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.

We measured performance of our polices in terms of simple regret, which is the difference of the true expected value of the actual best grasp plan and the true expected value of the selected grasp plan at the given stopping time [32]. The number of samples needed before the simple regret reaches zero, determines how effective an exploration policy is for grasp evaluation.

[TODO: ADD MULTI ARM BANDIT PRELIMINARY MAKE IT LONG]

C. Gaussian Process (GP) Background

[TODO: MASSIVE TRIM DOWN] GPs provide an infinite dimensional analogue of Multivariate Gaussian Distributions which allow us to model distributions over functions. Gaussian processes (GPs) are widely used in machine learning as

a nonparametric regression method for estimating continuous functions from sparse and noisy data [35]. In a GP, a training set consists of input vectors $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathbb{R}^d$, and corresponding observations $\mathbf{y} = \{y_1, \ldots, y_n\}$. The observations are assumed to be noisy measurements from the unknown target function f:

$$y_i = f(\mathbf{x}_i) + \epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise in the observations. A zero-mean Gaussian process is completely specified by a covariance function $k(\cdot, \cdot)$, also referred to as a kernel. Given the training data $\mathcal{D} = \{\mathcal{X}, \mathbf{y}\}$ and covariance function $k(\cdot, \cdot)$, the posterior density $p(f_* | \mathbf{x}_*, \mathcal{D})$ at a test point $\mathbf{x}_*$ is shown to be [35]:

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

where $K \in \mathbb{R}^{n \times n}$ is a matrix with entries $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\mathcal{X}, \mathbf{x}_*) = [k(\mathbf{x}_1, \mathbf{x}_*), \ldots, k(\mathbf{x}_n, \mathbf{x}_*)]^\intercal$. 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]:

$$\operatorname{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = k(\mathbf{x}_i, \mathbf{x}_j) \tag{2}$$
$$\operatorname{cov}\left(\frac{\partial f(\mathbf{x}_i)}{\partial x_k}, f(\mathbf{x}_j)\right) = \frac{\partial}{\partial x_k} k(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$
$$\operatorname{cov}\left(\frac{\partial f(\mathbf{x}_i)}{\partial x_k}, \frac{\partial f(\mathbf{x}_j)}{\partial x_l}\right) = \frac{\partial^2}{\partial x_k \partial x_l} k(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

For our kernel choice we decided to use the square exponential kernel, a major reason for this was its differentiability. 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.

D. Line of Action

Similar to the work of [2], 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. 3. 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 $\Gamma = (\gamma_1(\cdot), \ldots, \gamma_m(\cdot))$, 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: Six example objects illustrating transparency, specularity, sensor noise that illustrate the need for shape uncertainty: (from top to bottom) tape, squirt bottle, stapler, loofa, marker, water bottle. Displayed from right to left are the HD color image, a point cloud observation from a Primesense Carmine, GPIS visualization technique from [?], the nominal shape on a 25×25 grid.
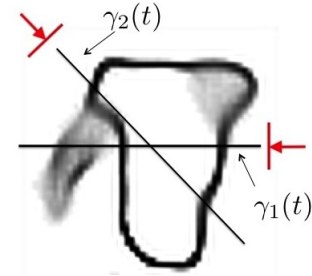

Fig. 3: Illustration of a grasp plan Γ composed of two lines of action, $\gamma_1(t)$ and $\gamma_2(t)$

E. Problem Definition

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

$$\Gamma^* \in \text{argmax}_{\Gamma \in G} E(Q(\Gamma)) \qquad (5)$$

with respect to a chosen grasp metric Q.

IV. MULTI-ARM BANDIT FOR GRASP SELECTION

While a standard approach to solving the problem in Eq. 5 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-arm bandit problem and forming a policy for selecting which grasp to sample.
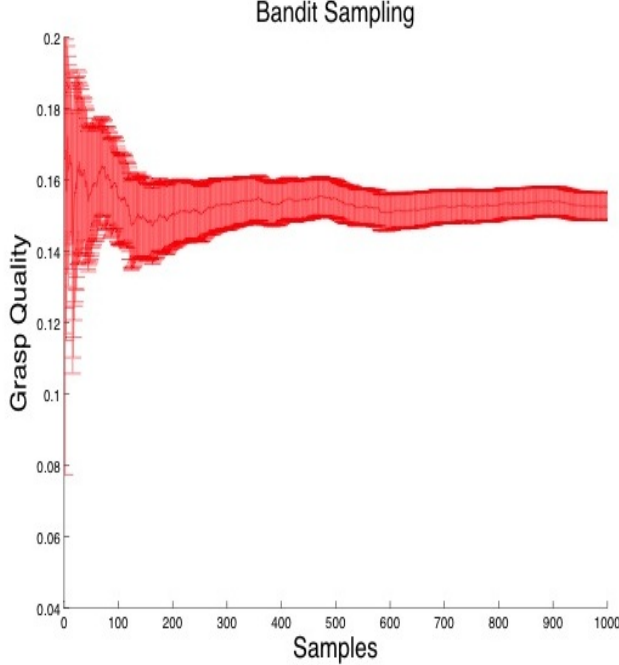

Fig. 4: Convergence of expected grasp quality, $E(Q(g))$, for a typical grasp. You can see it can take hundreds of evaluations to correctly estimate the expected grasp quality, thus it is important to intelligently allocate sampling resources

A. Our Implementation

In solving Eq. 5, 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. Thus, we are restricted in the type of bandit policies available to us, since most sophisticated polices require distributional information. [TODO: DOUBLE CHECK WITH DYLAN]

For our exploration policy, 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 [32]. 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 i that has been sampled n_i times and the samples have standard deviation σ_i is:

$$C_i := \frac{1.96\sigma_i}{\sqrt{n_i}}. \qquad (6)$$

Let $\hat{\mu}_i$ be the sample mean of the grasp quality for grasp plan i. Thus the confidence interval for grasp plan i is $[\hat{\mu}_i - C_i, \hat{\mu}_i + C_i]$ [15]. 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[8]. In future work, we will look at using other exploration strategies.

V. EVALUATING A GRASP

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

A. Grasp Metric

In their pioneering work over two decades ago Ferrari and Canny [38], 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[39], OpenGrasp[40] and Simox [41], 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 [2], [42].

B. Calculating the Expected Grasp Quality

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

$$E(Q(\Gamma)) = \int Q(g|S, \Gamma) p(S) dS \qquad (7)$$

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 [1], [8], [2]. See Fig. 5 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
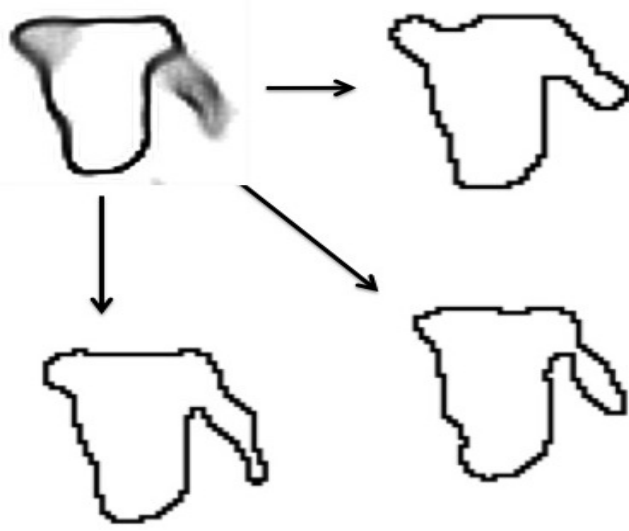

Fig. 5: Shape samples drawn from Eq. 8 on the object in the upper left corner. Given a shape sample we highlight the zero-crossing of the level set in black

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

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

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

Thus if the workspace is an $n \times n$ grid, the joint distribution is an n^2 multi-variate Gaussian, due to $sd : \mathbb{R}^2 \to \mathbb{R}$. Sampling from a Gaussian involves inverting the covariance matrix and inversion is in the naive way $O(n^3)$ [43]. 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:

$$p(g) = p(\mathbf{c}_1, ..., \mathbf{c}_m, \mathbf{n}_1, ..., \mathbf{n}_m | \Gamma, \mu(x), \Sigma(x)) \quad (9)$$

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

$$p(g) = \prod_{i=1}^{m} p(\mathbf{c}_i, \mathbf{n}_i | \gamma_i(t), \mu(x), \Sigma(x)) \quad (10)$$

We compute the expected grasp quality now as follows:

$$E(Q(\Gamma)) = \frac{1}{N} \sum_{i=1}^{N} Q(g_i), \quad g_i \sim p(g) \quad (11)$$

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

VI. 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 VI-A, we look at how to sample from $p(\mathbf{c}_i | \gamma_i(t), \mu(x), \Sigma(x))$. Then in section VI-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 VI-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:

$$sd(\gamma(t)) = 0 \quad (12)$$
$$sd(\gamma(\tau)) > 0, \ \forall \tau \in [a, t) \quad (13)$$

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

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

$$p(sd(\gamma(t)); \mu(t), \Sigma(t)) \ \forall t \in [a, b] \quad (14)$$

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. 3 and iterate from a to b until they reach a point that satisfies Eq. 12 and Eq. 13.
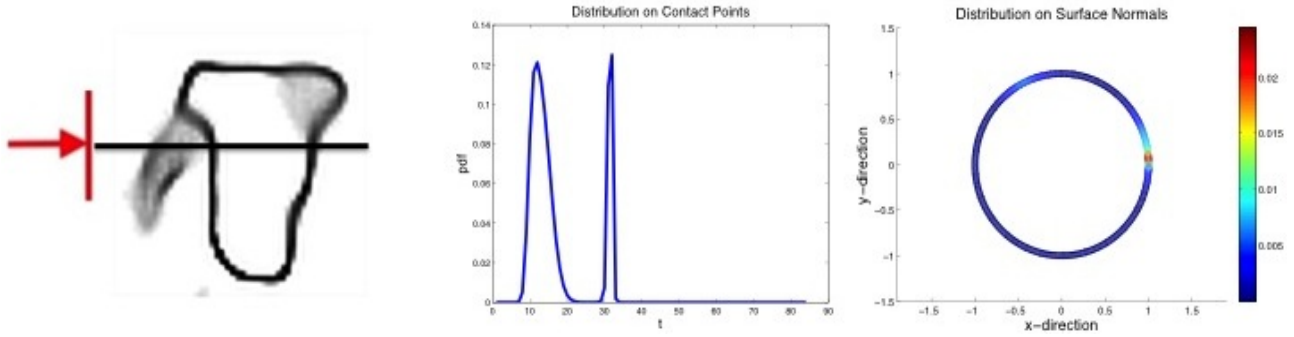
Fig. 6: (Left to Right): Line of action for a given gripper on an uncertain surface representing a measuring cup. Distribution $p(c)$ as a function of t, the position along the line of action $\gamma(t)$. The two modes correspond to the different potential contact points, either the handle or the base of the cup. Lastly, the distribution on the surface normals (inward pointing) along $\gamma(t)$ described by equation **??**.

B. Distribution on Surface Normals

Using Eq. 3 and Eq. 4, 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:

$$p(\mathbf{n}_i) = \int_a^b p\big(\mathbf{n}_i = \mathbf{v}|\mathbf{c}_i = \gamma(t)\big)p\big(\mathbf{c}_i = \gamma(t)\big)dt \qquad (15)$$

Grasp metrics such as Ferrari-Canny require $\mathbf{n}_i$ be normalized, or, equivalently, a member of the sphere $\mathcal{S}^{d-1}$ [38]. 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 theoretical 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:

$$\bar{z} = \frac{\int_\mathcal{W} x P(sd(x) < 0)dx}{\int_\mathcal{W} P(sd(x) < 0)dx} \qquad (16)$$

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. 7.
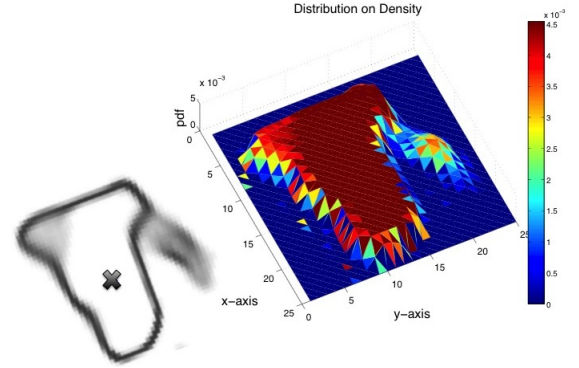

Fig. 7: Left: A surface with GPIS construction and expected center of mass (black X) Right: The distribution on the density of each point assuming uniform density

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. 10. This can be done by drawing samples from Eq. **??**. 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 $O(n^6)$.

VII. Experiments

For the experiments below we used common household objects shown in Fig. 2. 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 [**?**] 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. 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. 7 vs. grasps Eq. 11 was 0.004. After confirming the distributions converged close to the same value, we show the computational complexity in Fig. 8 of the two methods for evaluating 100 grasps on an $n \times n$ grid.
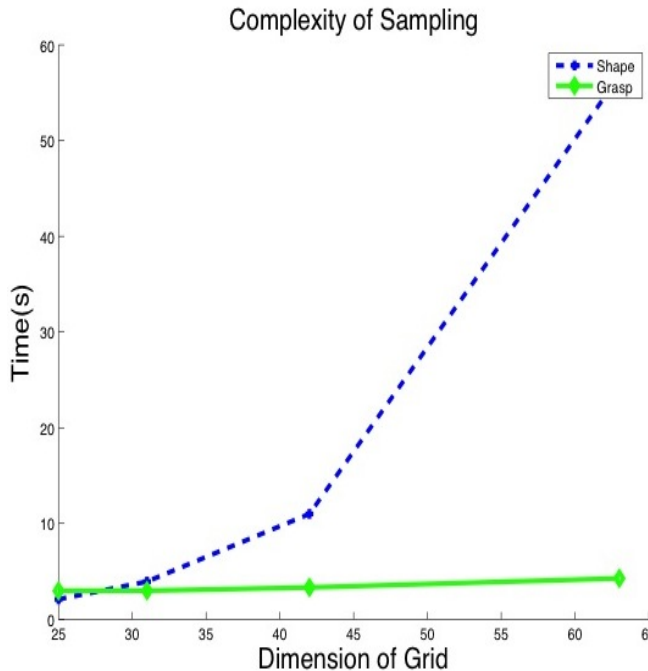

Fig. 8: Time it took to sample from 100 grasp distributions for a given resolution of the workspace. Blue line is sampling from $p(sd(\mathcal{R}))$ or shapes and Green is sampling from $p(g)$ or the calculated distribution on grasps. As you can see sampling from the calculated distributions scales much better.

B. Adaptive Sampling Technique

We consider the problem of selecting the best grasp out of a set G given a fix number of iterations I. For our experiments we look at selecting the best grasp out of a size of $|G| = 1000$ and $I = 7000$ or when the set of grasps is a single grasp. In Fig. 9, we plotted the simple regret, Eq. **??**, averaged over 20 runs and compare it to the naive Monte-Carlo method that randomly chooses a grasp plan to draw a sample from . The most interesting thing is that the regret is minimized an order of magnitude faster than the naive approach, thus motivating the use of including observations as you select samples.
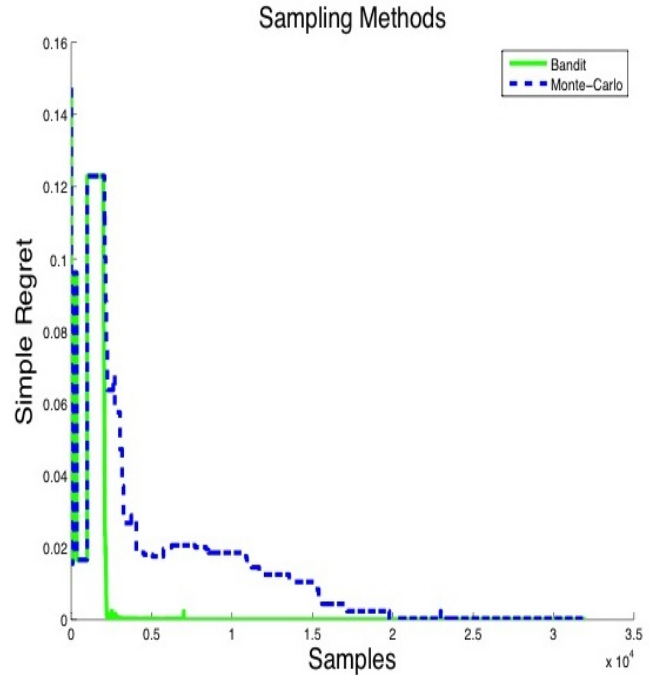

Fig. 9: Comparison of two sampling methods. Green is the adaptive bandit method that actively chooses which one to sample and Blue is the naive Monte-Carlo method. We measure their ability to converge in terms of simple regret Eq. **??** averaged over 100 runs. We had them determine the best grasp when $|G| = 1000$. As you can see our method converges much faster with an average of evaluation of 4 samples per grasp. We start the process though by deterministically evaluating every grasp 2 times to achieve a baseline, which explains the first part of the graph.

VIII. LIMITATIONS

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 [**?**].

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 [44].

Our budgeted multi-arm 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. Lastly, we are interested in investigating how well our algorithm works in parallel computing to be used in the cloud computing context. Furthermore, their is a large literature on formulating bandit policies and experimenting with them all will need to be done to see which one works best.

IX. FUTURE WORK

We have demonstrated how the multi-arm bandit problem can be used to sample grasps with shape uncertainty. However, one could imagine this approach to work with other forms of uncertainity like pose, center of mass, friction coefficient or motion.

We are interested in seeing how we can apply this to different shape representations other than GPIS like polygonal [1] or splines [2]. Furthermore, we want to extend this to 3D and see if our simplest bandit method still works well. There are many different methods for exploration policies, so it would be interesting to see how well they perform in 3D as well.

X. CONCLUSION

We demonstrate how formulation the problem of grasp selection as multi-arm bandit method enables much more efficeint sampling of grasp plans. We applied this to a new method for shape uncertainty, GPIS. While GPIS has been used before for very interesting applications [7],[?],[6] evaluating a grasp on it remained to computationally intensive if one sampled from the distribution on shape, $p(S)$, we present a an alternative distribution to sample from and demonstrate how it reduces computational complexity of sampling.

REFERENCES

[1] 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*, pp. 1106–1113, IEEE, 2012.

[2] 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*, pp. 1557–1563, IEEE, 2007.

[3] 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.

[4] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pp. 127–136, IEEE, 2011.

[5] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 303–312, ACM, 1996.

[6] 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)*, pp. 2845–2850, 2011.

[7] 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.

[8] B. Kehoe, D. Berenson, and K. Goldberg, "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*, pp. 576–583, IEEE, 2012.

[9] 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*, pp. 3493–3500, IEEE, 2013.

[10] K. Y. Goldberg and M. T. Mason, "Bayesian grasping," in *Robotics and Automation, 1990. Proceedings., 1990 IEEE International Conference on*, pp. 1264–1269, IEEE, 1990.

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

[12] 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*, pp. 557–562, IEEE, 2012.

[13] 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*, pp. 3258–3263, IEEE, 2012.

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

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

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

[17] 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.

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

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

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

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

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

[23] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[24] R. Agrawal, "Sample mean based index policies with o (log n) regret for the multi-armed bandit problem," *Advances in Applied Probability*, pp. 1054–1078, 1995.

[25] O. Cappé, A. Garivier, O.-A. Maillard, R. Munos, G. Stoltz, *et al.*, "Kullback–leibler upper confidence bounds for optimal sequential allocation," *The Annals of Statistics*, vol. 41, no. 3, pp. 1516–1541, 2013.

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

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

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

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

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

[31] D. F. Morrison, "Multivariate statistical methods. 3," *New York, NY. Mc*, 1990.

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

[33] D. Bergemann and J. Välimäki, "Bandit problems," tech. rep., Cowles Foundation for Research in Economics, Yale University, 2006.

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

[35] 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.

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

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

[39] 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.

[40] 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*, vol. 6472 of *Lecture Notes in Computer Science*, pp. 109–120. Springer Berlin / Heidelberg, 2010.

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

[42] 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.

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

[44] 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.