Incorporating Qualitative Information into Quantitative Estimation via Sequentially Constrained Hamiltonian Monte Carlo Sampling

Daqing Yi, Shushman Choudhury and Siddhartha Srinivasa

***Abstract*—In human-robot collaborative tasks, incorporating qualitative information provided by humans can greatly enhance the robustness and efficacy of robot state estimation. We introduce an algorithmic framework to model qualitative information as quantitative constraints on and between states. Our approach, named Sequentially Constrained Hamiltonian Monte Carlo, integrates Hamiltonian dynamics into Sequentially Constrained Monte Carlo sampling. We are able to generate samples that satisfy arbitrarily complex, non-smooth and discontinuous constraints, which in turn allows us to support a wide range of qualitative information. We evaluate our approach for constrained sampling qualitatively and quantitatively with several classes of constraints. SCHMC significantly outperforms the Metropolis-Hastings algorithm (a standard Markov Chain Monte Carlo (MCMC) method) and the Hamiltonian Monte Carlo (HMC) method, in terms of both the accuracy of the sampling (for satisfying constraints) and the quality of approximation. Compared to Sequentially Constrained Monte Carlo (SCMC), which supports similar kinds of constraints, our SCHMC approach has faster convergence rates and lower parameter sensitivity.**

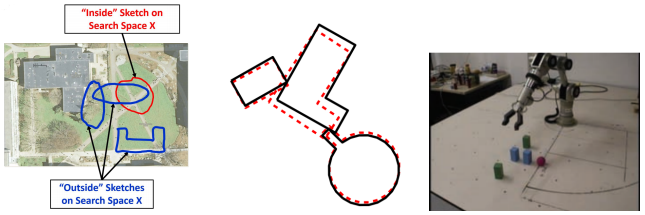
I. Introduction

We consider the problem of leveraging numerous forms of information for state estimation applications. A widely studied instance of this is the sensor fusion problem [10], which merges information from multiple sensors and sources [17] for estimating hidden states. We are interested in a related but different setting - incorporating information from humans.

In a collaborative human-robot team, both the human(s) and the robot(s) receive observations from the shared environment. There are several potential benefits to this, described in Section II. The sensing capabilities of the human complements that of a robot [12], often overcoming many of the robot's perceptual limitations [13] The human-supplied qualitative information helps enhance the robustness of estimation and reduce the uncertainty. For instance, information from experienced human rescuers has been used to more successfully predict the location of a target [1]. Furthermore, the human can provide knowledge about the physical world which facilitates corrections during state estimation [22]. Humans can provide qualitative descriptions [4] or high-level instructions [26] to help a robot with task execution.

Qualitative information from the human can be provided in several forms:

- *A general description of a state* that constrains the values of the state.
  - A human would highlight the "red" color in an environment of flickering lights.
  - A human would sketch a bounding perimeter for a desired target to be in (Figure 1a).
- *System Knowledge* that constrains the relationship between two or more states.
  - A human would know that two rigid objects should not collide (Figure 1b).
  - A human would know that a car cannot make a zero radius turn.
- *High-level abstraction* that defines a structure over the states.
  - A human would say "the blue cube is next to the red ball" in Figure 1c to refer to a picking position.
  - A human would say that the temperature falls after snowing while predicting temperature dynamics.



(a) A map sketch [1] (b) No collision [22] (c) Spatial relation [4]

Fig. 1: Various forms of qualitative human-furnished information that could be integrated into a state estimation framework. Our SCHMC approach can support all these kinds of constraints, as we shall demonstrate with our extensive examples.

The diverse nature of qualitative information, from sketches to numerical tolerances, makes it difficult to obtain a unified description. Approximations are usually resorted to so that numerical optimization can be applied to state estimation [1, 22].

Our key insight, described in Section III, is to integrate a wide range of qualitative information into quantitative models via constraints in a graphical structure. In this paper, we propose a general framework for this purpose. Our approach, shown in Section IV, follows a Sequentially Constrained Monte Carlo method [8] that allows for various kinds of constraints which define nonlinear, non-smooth and complex multi-modal problems. Also it uses adaptive Hamiltonian Monte Carlo [11] that supports efficiency in problem solving without requiring parameters for the steps, as needed for Sequential Monte Carlo.

We demonstrate and evaluate our approach through several interesting scenarios motivated by important applications,

All authors are with Robotics Institute, Carnegie Mellon University. daqingy@andrew.cmu.edu, shushmac@andrew.cmu.edu, siddh@cs.cmu.edu

outlined in Section V. Since our focus is on the mathematical details and implementation of the underlying algorithm, we use simulated examples of constrained sampling, filtering, and static state estimation that are easy to visualize. This also helps us highlight the versatility of our approach by demonstrating support for a number of different kinds of constraints.

II. Related Work

In a fusion framework, information is typically modeled as a random variable, so a common way of presenting human information is "soft data", using a probability distribution [2]. In a human-robot team, a robot can convert a qualitative description of the environment into a distribution to support a search task [16, 24]. Qualitative spatial relations from humans has been modeled into map generation for navigation [18]. A hand-sketched path and map has been used as a prior for a robot's simultaneous mapping and navigation [21].

Another approach of modeling qualitative information is as a "hard constraint". For planning a shared task, a human can generate a path shape as a topological constraint for the robot's planning and execution [25, 26]. Moreover, a lot of physics-based knowledge can be modeled as constraints in state estimation. The collision-free constraint for rigid bodies is used in correcting state estimation of object positions [22]. Hard constraints enforce the relationship between states, which reduces the uncertainty of state estimation [23], though it does increase the difficulty of problem solving.

In solving a constrained nonlinear optimization problem, sequential quadratic programming is a popular tool [22] that decomposes a problem into a sequence of subproblems and solves them iteratively. It requires the objective and the constraints to be twice differentiable, which greatly limits the support for constraints derived from qualitative information. For example, a hand sketch in Figure 1a can be difficult to convert into a mathematical description in a probabilistic inference framework. The solution is determined by the initial guess, and only local optimality is guaranteed.

One way to achieve global optimality is through Monte Carlo simulation. Markov Chain Monte Carlo [5](MCMC) is a computational technique that derives a set of samples to approximate a probability distribution by modeling a Markov Chain random walk. It recovers more information about a distribution than only estimating an optimal solution [5]. Sequential Monte Carlo [6] introduces a set of samples that form parallel Markov Chains from a sequence of bridging distributions for complex multimodal distributions, which applies to high dimensional spaces [20] and complex constraints [9]. Various transition kernels can be used in Sequential Monte Carlo [20]. Hamiltonian dynamics [19] have been introduced as a more efficient alternative to the random walk MCMC kernel for moving samples around. Therefore, we use an adaptive Hamiltonian Monte Carlo [11](HMC) kernel in a Sequentially Constrained Monte Carlo [8] framework for our overall approach.

III. Problem Statement

Assume we have a set of observations $O = \{o_1, \cdots, o_N\}$ that is associated with a set of states $X = \{x_1, \cdots, x_K\}$. In a problem that considers only quantitative information, state estimation can usually be defined as a MAP (Maximum Aposteriori Probability) problem, which is

$$\hat{X}^* = \arg\max_{\boldsymbol{X}} P(X \mid O). \tag{1}$$

We propose two types of qualitative information to model human information - *feature* and *relation*.

- **Feature** informs a property of a state. Let $f(x) \leq 0$ when the constraint is satisfied.
- **Relation** informs a conditional relationship between two states. Let $f(x_2 \mid x_1) \leq 0$ when the constraint is satisfied. Often, $f(x_2 \mid x_1)$ and $f(x_1 \mid x_2)$ are both equivalent to a mutual relation $f(x_1, x_2)$, which is satisfied when $f(x_1, x_2) \leq 0$.

Some examples of qualitative information are given below.

- *The color is red.* Assume the color is represented in HSV space and the H value $h \in [0, 180]$. The information "red color" indicates a feature that constrains the value of state h to be in $[0, 10] \cup [160, 180]$.
- *The object is on the left hand side.* Assuming the x-coordinate on the left hand side is all negative, this defines a feature that constrains the object location as $x \leq 0$.
- *Ball one and ball two are rigid.* Let s_1 be the estimated center of ball one with radius r_1. Let s_2 be the center of a ball with radius r_2. If the two balls cannot collide, we have $|s_1 - s_2| \geq r_1 + r_2$ as a constraint.
- *The point is inside a circle.* Assume the circle is a unit circle centered at the origin. This information implies that given y, x should satisfy $x - \sqrt{1-y^2} \leq 0 \wedge -x - \sqrt{1-y^2} \leq 0$. Similarly, given x, y should satisfy $y - \sqrt{1-x^2} \leq 0 \wedge -y - \sqrt{1-x^2} \leq 0$.
- *The object is a square.* Let s_1, s_2, s_3, s_4 be the estimated corners of a square on a plane and r be the estimated edge length of the square. By prior geometric knowledge, we have $|s_1 - s_2| = r$, $|s_2 - s_3| = r$, $|s_3 - s_4| = r$ and $|s_4 - s_1| = r$ as constraints.
- *A human sketches a region.* A human sketches a region on a map that indicates where a target is, as in Figure 1a. Let s be the estimated position of the target and R be a human-sketched region. This information implies a feature that requires $s \in R$.

Both features and relations define "hard constraints" on states to be estimated. In Figure 2, we use a graph structure to illustrate how quantitative and qualitative information are integrated. Each quantitative state is a node in the graph, while each qualitative state is an edge. A self-edge to a state x defines a constraint that the state x should satisfy. A directed edge from a state x_1 to another state x_2 defines a constraint that state x_2 should satisfy, given state x_1.

Assume that states x_1, x_2 and x_3 are estimated by observations o_1, o_2 and o_3 respectively. Qualitative information defines edges that connects nodes. $f^1(x_1)$ defines a feature

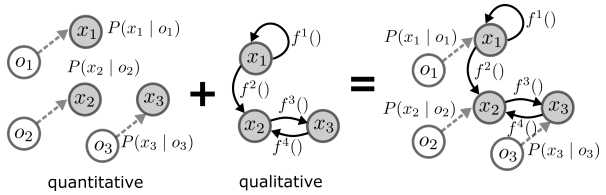


Fig. 2: Our underlying graph structure represents qualitative information via constraints on individual states and between multiple states.

constraint on x_1. $f^2(x_2 \mid x_1)$ defines a relation constraint on x_2 given x_1. $f^3(x_3 \mid x_2)$ and $f^4(x_2 \mid x_3)$ together define a relation constraint on x_2 and x_3.

Let $F = \{f^i\}_i$ denote the set of constraints on the set of states X. When all states in X satisfy the constraints, $F(X) \preceq 0$. We can reformulate the problem as

$$\hat{X}^* = \arg\max_{\boldsymbol{X}} P(X \mid O, F) = \arg\max_{\boldsymbol{X}^F} P(X \mid O, F), \quad (2)$$

where $\boldsymbol{X}^F = \{X \in \boldsymbol{X} \mid F(X) \leq 0\}$. The knowledge defines constraints, which makes a subset $\boldsymbol{X}^F$ of all possible states $\boldsymbol{X}$ valid. This reduces uncertainty in problem solving, which is measured by conditional mutual information - $I(X; F \mid O) = H(X \mid O) - H(X \mid O, F)$.

The layout in Figure 2 indicates that we could have the state estimations running separately in parallel, which generates an unconstrained distribution $P(X \mid O)$. The qualitative constraints represented by F are then introduced while querying states from $P(X \mid O, F)$. We rewrite $P(X \mid O, F)$ in terms of the unconstrained posterior $P(X \mid O)$ and constraints $\mathbb{1}_F(X)$.

$$\begin{aligned} P(X \mid O, F) &= \frac{P(X \mid O)\mathbb{1}_F(X)}{\int_{\boldsymbol{X}^F} P(X \mid O) dX} \\ &\propto P(X \mid O)\mathbb{1}_F(X), \end{aligned} \quad (3)$$

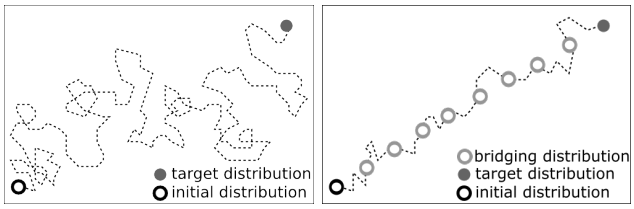
in which

$$\mathbb{1}_F(X) = \begin{cases} 1 & F(X) \leq 0 \\ 0 & F(X) > 0 \end{cases} \quad (4)$$

IV. Sequentially Constrained Hamiltonian Monte Carlo

Markov Chain Monte Carlo (MCMC) can compute $P(X \mid O, F)$ in low dimensional cases. Sequentially Constrained Monte Carlo [8] (SCMC) was proposed for high dimensional, multi-modal, discontinuous cases. Considering the efficiency issues of the inference problem, we introduce SCHMC, that adds Hamiltonian dynamics [19] to SCMC. This incorporation is non-trivial, because for Hamiltonian dynamics in the general MCMC, the performance is heavily dependent on selecting the number and length of the steps [19]. This property prevents it from fitting directly into a sequential Monte Carlo structure, because *bridging distributions* might require different parameters while selecting steps. Bridging distributions denote a sequence of distributions that transition from an unconstrained distribution to a fully constrained distribution. We use a kernel obtained from No-U-Turn sampling [11], that adaptively selects the step length and number for Hamiltonian dynamics. Moreover, the performance of SCMC depends on an adequate number of burn-in steps, while the fast convergence of No-U-Turn sampling makes SCHMC robust to the burn-in step selection. The MAP problem defined in Equation (2) can be solved by finding the sample with the highest posterior.

Instead of directly moving the samples toward the target distribution $P(X \mid O, F)$, SCMC constructs a sequence of bridging distributions. Each two consecutive distributions are similar enough that one can reach the other via a random walk. The sequence of bridging distributions guides the samples to move from the initial distribution to the target distribution, as shown in Figure 3.



(a) MCMC (b) SCMC

Fig. 3: SCMC uses bridging distributions to guide the initial distribution to the target, unlike MCMC which uses an unguided random walk.

A probit function is used to approximate the deformation from no constraint to an indicator function (hard constraint). We use $\Phi(-\tau C_F(X))$, in which Φ is an exponential function and $C_F(X)$ is a *deviation function*. The deviation function $C_F(X)$ measures how a state S deviates from a constraint F. It satisfies

- $F(X) \leq 0 \implies C_F(X) \leq 0$;
- $F(X) > 0 \implies C_F(X) > 0$.

We usually directly choose $C_F(X) = F(X)$. But choosing a $C_F(X)$ that provides local monotonicity can significantly improve convergence rate. As in Equation (5), as τ increases, $\Phi(-\tau C_F(X))$ gradually converges to $\mathbb{1}_F(X)$.

$$\lim_{\tau \to \infty} \Phi(-\tau C_F(X)) = \mathbb{1}_F(X). \quad (5)$$

Tuning τ from 0 to ∞ transforms the function shape from unconstrained to fully constrained, as shown in Figure 4 [1].

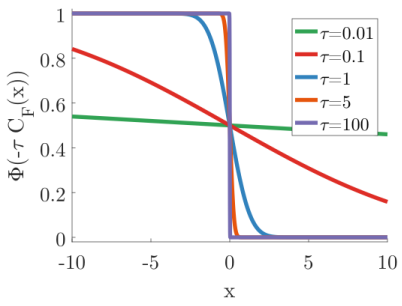


Fig. 4: The shapes of our probit function with different values of τ

Thus, a sequence of monotonically increasing τ, i.e. $\{\tau_t\}_{t=0}^T$ and $0 = \tau_0 < \tau_1 < \cdots < \tau_T = \infty$, determines

[1] In practice, how big τ needs to be to approximate ∞ depends on the limits of precision and the nature of the problem.

a sequence of bridging distributions $\{\pi_{\tau_t}(x)\}$ that transits from $P(x \mid O)$ to $P(x \mid O, F)$.

This sequence of bridging distributions is used by Sequentially Constrained Monte Carlo [8]. As mentioned before, applying Hamiltonian dynamics rather than random walks for transitioning between bridging distributions leads to Sequentially Constrained Hamiltonian Monte Carlo (SCHMC), outlined in Algorithm 1. We present the algorithm through the flow of evolutionary computation [6], which consists of initialization, promotion, correction, selection and mutation.

Algorithm 1 SCHMC $(P(X \mid O),\ F)$

INITIALIZATION: $X_0 \sim P(X \mid O)$, $\tau_0 \leftarrow 0$, $t \leftarrow 0$
while NOTCONVERGED$(X_t, P(X \mid O, F))$ **do**
  PROMOTION $\tau_{t+1} \leftarrow$ FINDNEXT(τ_t, X_t)
  CORRECTION: $W_{t+1} \leftarrow$ CALCWEIGHT(X_t)
  SELECTION: $X_t \leftarrow$ RESAMPLE(W_{t+1}, X_t)
  MUTATION: $X_{t+1} \leftarrow$ HAMILTONIANKERNEL$(X_t, \pi_{\tau_{t+1}}(x))$
  $t \leftarrow t+1$
return X_{t+1}

In INITIALIZATION, a set of samples is sampled from an unconstrained distribution $P(X \mid O)$. The samples are evolved iteratively until they converge to a fully constrained distribution $P(X \mid O, F)$. In PROMOTION, FINDNEXT() is called to calculate a new τ_{t+1} from the current τ_t and the sample set X_t. The new parameter τ_{t+1} implies a new bridging distribution $\pi_{\tau_{t+1}}$. In CORRECTION, weights of all the samples W_{t+1} are calculated by CALCWEIGHT(). The weight ω_n^{t+1} of a particle x_n^t is obtained by

$$\omega_n^{t+1} = \Phi\left((\tau_{t+1} - \tau_t) C_F(x_n^t)\right). \tag{6}$$

In SELECTION, RESAMPLE() is called to avoid particle degeneracy. Samples with high weights, i.e. high probabilities, are more likely to be resampled. In MUTATION, HAMILTONIANKERNEL() is called to move samples toward a new bridging distribution $\pi_{\tau_{t+1}}$. This runs till samples converge, which is checked by NOTCONVERGED(). In practice, convergence occurs when τ reaches a large enough value.

Effective Sampling Size (ESS) [20] measures the degeneracy of a set of samples.

$$\mathrm{ESS}_t = \frac{\left(\sum_{n=1}^{N} \omega_n^t\right)^2}{\sum_{n=1}^{N} (\omega_n^t)^2}, \tag{7}$$

in which ω_n^t is calculated by Equation (6). It is widely used in Sequential Monte Carlo to avoid the particle degeneracy problem, where most samples are located in a low probability region. In our case, a low ESS implies that many samples do not satisfy the constraint, therefore they have near zero probabilities. In order to guarantee that enough samples satisfy the constraint in the next bridging distribution, we use ESS_t to determine the promoted step length in PROMOTION. We choose a threshold of the effective sampling size θ_{ESS} required in the new bridging distribution, for instance, half of the number of samples $N/2$. The value of τ_{t+1} is chosen as one that satisfies $\mathrm{ESS}_t \geq \theta_{\mathrm{ESS}}$ in FINDNEXT(). RESAMPLE() is called to avoid particle degeneracy, following a common resampling procedure in Sequential Monte Carlo [7]. Resampling does not change the sample weights.

We use HAMILTONIANKERNEL to move samples toward a bridging distribution, following Hamiltonian Monte Carlo [19] (HMC). Hamiltonian Monte Carlo introduces dynamical evolution that avoids a random walk for greater efficiency. In order to sample from a random variable x defined by p.d.f $P(x)$, Hamiltonian dynamics is introduced where the system state is x and the momentum is an auxiliary variable q. The resulting system energy $H(x,q) = U(x) + K(q)$ determines the dynamical evolution of x and q, which is called Hamiltonian dynamics. $U(x)$ is the potential energy, which is determined by the state x. $K(q)$ is the kinetic energy, which is determined by the momentum q. The dynamics equations are represented by $\frac{dx}{dt} = \frac{\partial H}{\partial q}$ and $\frac{dq}{dt} = -\frac{\partial H}{\partial x}$. Because the Hamiltonian dynamics satisfies the property of *detailed balance* [19] that guarantees reversibility, it can be used in MCMC to generate samples (x, q) from a canonical distribution $\Phi(-H(x,q))$. If the potential energy is explicitly defined as $U(x) = -\log \pi(x)$, and the kinetic energy as $K(q) = \frac{1}{2} q^T M^{-1} q$, we can obtain samples for $\pi(x)$ by sampling from $\Phi(-H(x,q))$ and discarding q from each augmented point.

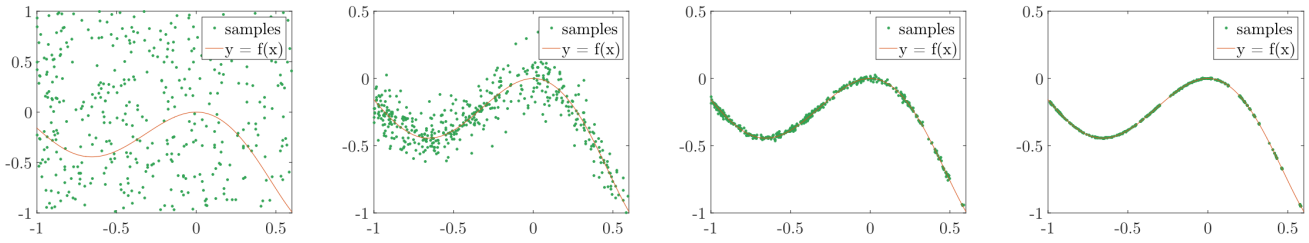
Algorithm 2 HAMILTONIANKERNEL $(x,\ \pi_{\tau_{t+1}}(x)\ )$

$q \sim N(0, M)$; $(x^0, q^0) \leftarrow (x, q)$
$q^0 \leftarrow q^0 - \frac{\epsilon}{2} \frac{\partial U(x^0)}{\partial x}$
for $m = 1, \cdots, M$ **do**
  $x^m \leftarrow x^{m-1} + \epsilon \frac{\partial K(q^{m-1})}{\partial q}$
  $q^m \leftarrow q^{m-1} - \epsilon * \frac{\partial U(q)}{\partial x}$
$q^M \leftarrow q^M - \frac{\epsilon}{2} \frac{\partial U(x^M)}{\partial x}$
if UNIFORM$([0,1])$ $<$ $\min\left(1, e^{H(x^M, q^M) - H(x^0, q^0)}\right)$ **then**
  $x \leftarrow x^M$
return x

It is known that the performance of HMC depends on selecting step number and step size [19] according to a target distribution. In order to guarantee performance in various bridging distributions, we use an adaptive HMC, the No-U-Turn sampler [11]. It is a variant of HMC that estimates a step size according to a given problem and adaptively selects a step number that guarantees ergodicity, which ensures that the samples avoid being trapped in a subset of the space [19].

V. EXPERIMENTS

We evaluate the performance of our method for sequentially constrained sampling, both quantitatively and qualitatively. Additionally, we demonstrate how it can be used for incorporating known information in the form of arbitrary constraints into filtering and state estimation problems. Note that while our work is strongly motivated by robotic applications that benefit from knowledge integration, the purpose of this paper is to present an algorithm for constrained sampling.



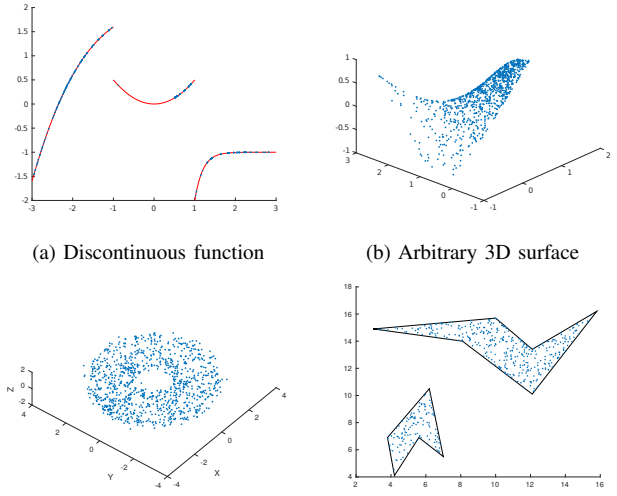
(a) $\tau \to 0$ (b) $\tau = 7.82$ (c) $\tau = 68.85$ (d) $\tau = 500$

Fig. 5: The SCHMC algorithm is used to sample along an arbitrary non-linear manifold with functional form $f(x) = 2x^5+3x^4-2x^3-4x^2+x \sin x$. The seed points are generated from a uniform distribution in the bounding area. As expected, as the value of τ increases and the weight of the constraint increases, the samples track the manifold more closely.

Therefore, we focus our quantitative evaluation on the performance for sampling. For the other applications, which have their own algorithms and task-specific performance criteria, we show how our method can be used in conjunction with such algorithms to accommodate arbitrary constraints.

A. Sequentially Constrained Sampling

SCHMC follows a sequence of bridging distributions that transit from an unconstrained distribution to a fully constrained distribution. As the constraint is gradually tightened by increasing τ, samples are driven toward a constraint manifold that follows the distribution. Figure 5 exemplifies how samples are moved toward an arbitrary non-linear manifold, starting from an unconstrained uniform distribution.



(a) Discontinuous function (b) Arbitrary 3D surface

(c) 3D torus (d) Multiple nonconvex polygons

Fig. 6: Various examples of constrained sampling using SCHMC. For Figure 6b the samples are drawn from a Gaussian distribution, while for the others they are drawn from a uniform distribution.

We use two *metrics* to evaluate the quality of our sampling procedure. The first metric indicates the *accuracy* of the sampling, i.e. how well the samples satisfy the constraints. For a given constraint $f(x) = 0$ which we want each sample to satisfy, the deviation of a given sample x' from the constraint, or the *error*, is $|f(x')|$. For a set of samples X obtained from some sampling procedure, we use the *root mean squared error* (RMSE) over X as our metric, which is basically

$$\text{RMSE}(X) = \sqrt{\frac{1}{|X|}\sum_{x' \in X} |f(x')|^2} \tag{8}$$

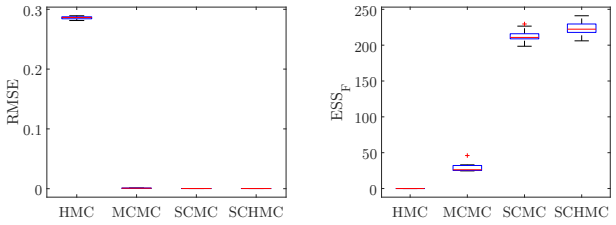
A lower RMSE implies that the set of points more closely matches the target distribution, so the lower the better.

The other metric indicates the *diversity* of the sampling, i.e. how well distributed the samples are over high-probability regions. We use the *Effective Sampling Size* (ESS), which was defined earlier in Equations (7) and (6). For this metric, rather than using the weights derived from the sampling, we strictly enforce the constraint. Therefore samples which do not satisfy the constraint have weight close to 0 and the ones that do have weight close to 1. A higher ESS implies that more of the samples satisfy the constraint and they are not distributed in low probability subspaces.

In our experiments, the samplings of MCMC and HMC converge well in visible low-dimensional problems. In high-dimensional problems, it is harder to converge to a target distribution as shown in Figure 3. Figure 7 shows the MRSE and ESS of samples generated from four different approaches, HMC, MCMC, SCMC and SCHMC. The problem is of an 8D multivariate normal distribution subject to a constraint as in Figure 13a. It is required that the Euclidean distances between subvectors of a state are constant, which defines a few disjoint feasible subspaces. HMC fails to converge well in such a discontinuous constrained high dimensional problem, which is manifested as high RMSE in Figure 7a. Though the samples by MCMC converge as well as those by SCMC and SCHMC, the diversity of samplings by MCMC is not good, which is shown by a low ESS in Figure 7b.

We evaluate the accuracy of SCHMC for constrained sampling over various classes of functions. For each case we have a function of the form $f(x) = 0$ and we compute the RMSE of 500 samples from an underlying uniform distribution. The results are shown in Figure 8. Following are the descriptions of each function class:

- *Polynomial* : $f([x,y]) = y - (5x^3 - 3x^2 + 4x - 1)$
- *Sinusoidal* : $f([x,y]) = y - (\sin 2x \cos 3x - 2\sin x \cos 4x)$
- *Exponential* : $f([x,y]) = y - e^{(3x^2-2)}$
- *Combined* : $f([x,y]) = y - (xe^{\sin x} + \cos(\sin x))$



(a) RMSE (b) ESS

Fig. 7: In an 8D problem subject to discontinuous constraints, HMC fails to converge to samples in feasible subspaces, which leads to high RMSE in Figure 7a. Furthermore, the diversity of sampling with MCMC is much worse than with SCMC and SCHMC. Thus the ESS of MCMC in Figure 7b is much lower than that of SCMC and SCHMC. The results are averaged over 10 different runs.

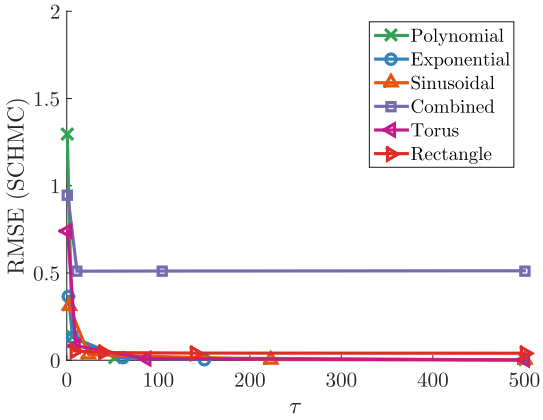


Fig. 8: As τ increases, the *root mean squared error* or the RMSE of our SCHMC approach, as defined in Equation (8), converges to a small number. The *combined* function needs a higher $\tau_{\max}$ to reach that level of convergence probably because it is more complex than the others.

- *Torus* : $f([x,y,z]) = \left(2 - \sqrt{x^2+y^2}\right)^2 + z^2 - 1$
- *Rectangle* : This uses the scenario of Figure 13b.

For instance, for the *Torus* problem, we define the deviation function mentioned earlier as

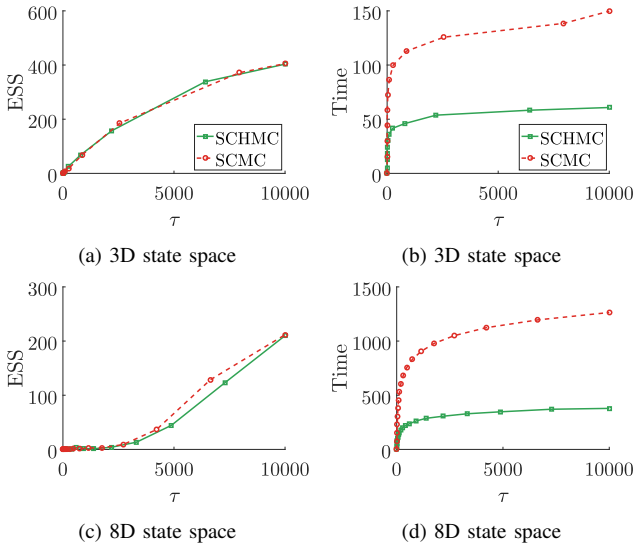
$$C_F([x,y,z]) = \left|\left(2 - \sqrt{x^2+y^2}\right)^2 + z^2 - 1\right| \quad (9)$$

Only when $\left(2 - \sqrt{x^2+y^2}\right)^2 + z^2 - 1 = 0$ are the samples constrained to the torus.

We use a 3D problem in Figure 6c and an 8D problem in Figure 13a to compare SCMC with SCHMC. As in Figure 9a and 9c, the ESS of both SCMC and SCHMC grow by almost the same ratio as τ increases. In the 3D problem, SCHMC takes less time to converge than SCMC. The efficiency improvement is more significant in the 8D problem.

B. Filtering

Markov Chain Monte Carlo techniques have been used in filtering applications, for tracking targets with complex state distributions [3, 14]. By modeling prior information as state constraints, SCHMC can be integrated into filtering problems [15]. For high-dimensional state spaces, rejection sampling may not be useful due to high rejection rate.



(a) 3D state space (b) 3D state space

(c) 8D state space (d) 8D state space

Fig. 9: Compared to SCMC, SCHMC achieves a similar ESS with quicker convergence to the constraint.

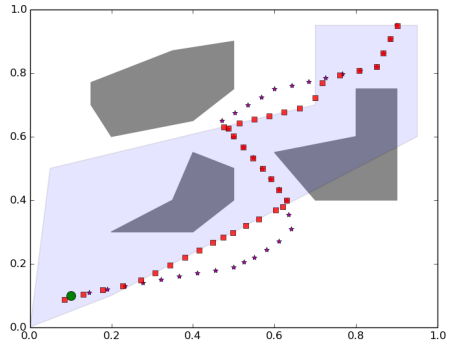


Fig. 10: An example of filtering with SCHMC. The purple trajectory is due to the noisy motion model, and it violates several of the system constraints. The corrected trajectory is the one with maximum aposteriori probability given the motion model trajectory, subject to all constraints.

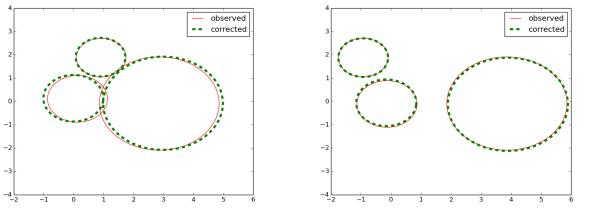
We demonstrate a simple application of filtering with SCHMC in Figure 10. We have a point robot moving in an $\mathbb{R}^2$ world with obstacles. There is a given motion model for the robot, with the following constraints:

- The robot stays within the blue zone.
- The robot has a minimum clearance from all obstacles.
- The robot has a maximum velocity, i.e. distance between successive waypoints, assuming unit timesteps

We choose a simple example with simplifying assumptions for convenience of implementation and ease of visualization. The usage of SCHMC for incorporating information as constraints generalizes to more complex problems in filtering (higher order state spaces, more complicated dynamics).

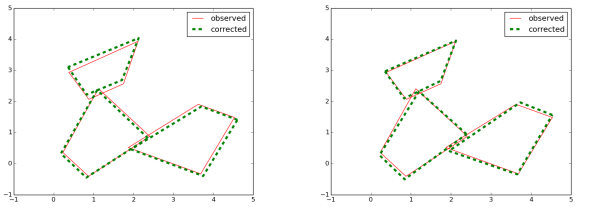
C. Pose Estimation

Another important application where prior information is useful is in estimating the pose of a system. The constraints can represent physics-based knowledge, geometric informa-



(a) Correction required (b) Correction not required

Fig. 11: Correctly estimated positions of ellipses with a non-collision constraint. Figure 11a shows how the estimation is corrected to satisfy the non-collision constraint. Figure 11b shows how the estimation need not be corrected as the non-collision constraint is already satisfied.



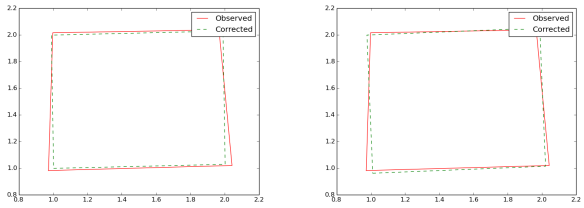
(a) Corrected centers and rotations. (b) Corrected endpoints.

Fig. 12: Correctly estimated positions of polygons with a non-collision constraint. In Figure 12a, if the shapes of the polygons are known, the centers and rotations of polygons are corrected to satisfy the constraint. In Figure 12b, if the shapes are unknown, end-points of polygons are corrected to satisfy the constraint. This is a much higher-dimensional problem (24 dimensions).

tion, rigid body assumptions and so on. We demonstrate a number of examples motivated by real-world applications. We work with state spaces of higher dimensions, but which can be visualized in 2D and 3D.

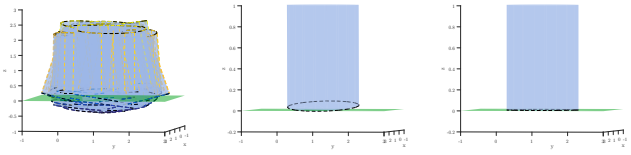
Our first set of examples involves estimating the pose of objects on a 2D plane. In Figure 11 we show a simple example of using the no-collision constraint to estimate ellipse positions. We then demonstrate how SCHMC can estimate the poses of multiple polygons in Figure 12, subject to a similar collision-free constraint. There are two variants depending on the outline of the polygon being known (Figure 12a) or unknown (Figure 12b). Besides non-collision constraints, we also show how SCHMC can support geometric knowledge-based constraints, as in Figure 13, such as the edge lengths of the polygon (Figure 13a) or that the polygon is a rectangle (Figure 13b).

Now we discuss estimating the 6-DOF pose of 3D objects subject to constraints from physics-based knowledge. This is a very relevant aspect of robust visual pose estimation for robotics applications. The first example, shown in Figure 14, estimates the pose of a single cylinder, subject to the constraint of being upright and on the surface of a plane. The second example, shown in Figure 15, estimates jointly the poses of a cylinder and a cuboid, subject to the constraints of the previous example, and additionally the constraint that they do not intersect each other.



(a) Edge lengths given (b) Object known to be rectangle

Fig. 13: Correctly estimated end-points of a polygon with a shape constraint. In Figure 13a, if the lengths of edges are given, the estimated end-points are corrected to satisfy edge-length constraints. In Figure 13b, if the polygon is known to be a rectangle, the estimations are corrected to satisfy a parallel edge constraint, which enforces the lengths of non-adjacent edges to be equal.



(a) Observed. (b) Averaged. (c) Constrained.

Fig. 14: Corrected pose estimation of a cylinder on a table. The left subfigure shows poses of cylinders from a set of noisy observations. The center subfigure shows an estimated pose by simply averaging all the observations. The right subfigure shows an estimated pose subject to the cylinder being upright and on the table.

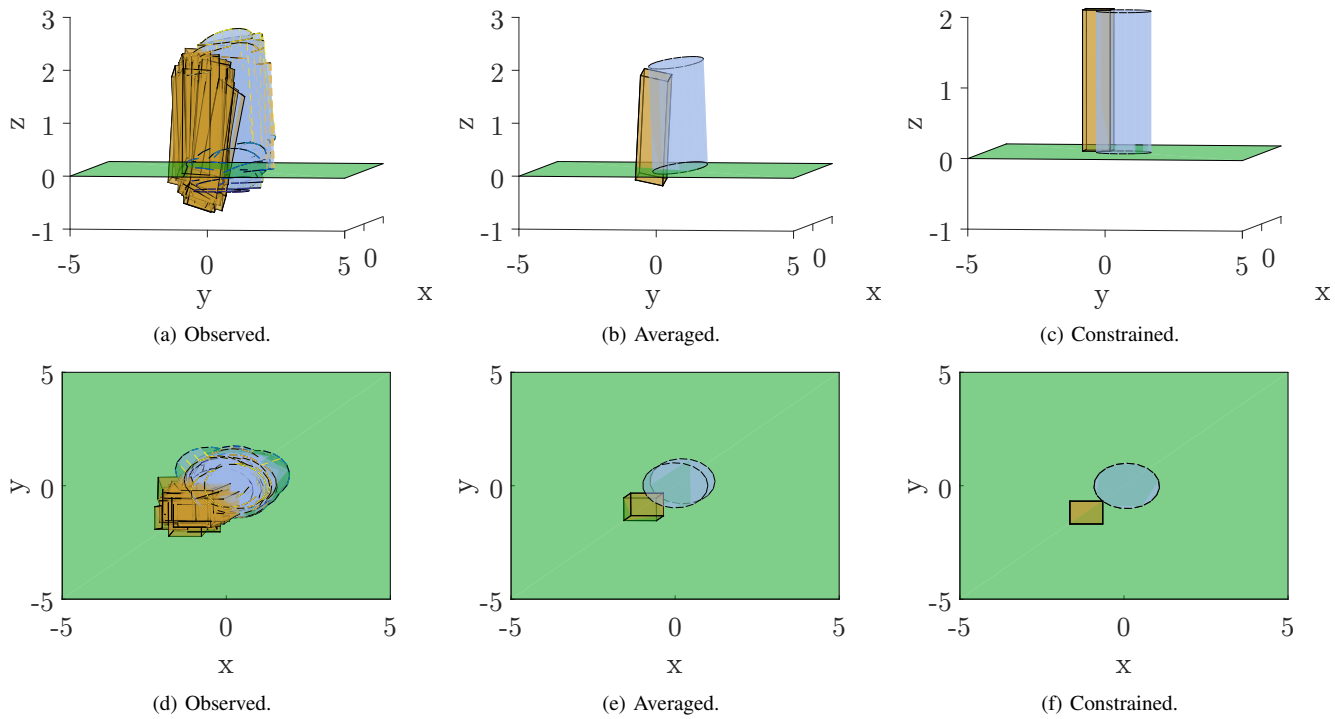
VI. Conclusion

In this paper, we present SCHMC for sampling from arbitrary distributions, subject to a wide range of constraints. This enables us to model human information into quantitative state estimation problems. SCHMC introduces Hamiltonian dynamics to a Sequentially Constrained Monte Carlo technique, in which a sequence of bridging distributions are created to guide samples from an unconstrained distribution to a fully constrained distribution. Hamiltonian dynamics is used in transiting between bridging distributions, which accelerates the convergence rate while maintaining sampling quality, especially in a high-dimensional space (as in state estimation problems). Also, the accelerated convergence rate makes SCHMC robust to the selection of burn-in steps for transiting between bridging distributions. We choose the No-U-Turn sampler to simulate the Hamiltonian dynamics so that parameters are adaptively tuned. The parameter-free nature of SCHMC makes it adaptive to different bridging distributions in the sampling process.

Our work focuses on the algorithm for constrained sampling and our experiments and evaluation is done keeping that in mind. Considering the original motivation of the paper, the primary direction of future work is to use SCHMC for modeling human information in a real-world human-robot collaborative task and demonstrate how the system performance improves as a result.

References

[1] N. Ahmed, M. Campbell, D. Casbeer, Y. Cao, and D. Kingston. Fully Bayesian learning and spatial reasoning with flexible human



(a) Observed. (b) Averaged. (c) Constrained.

(d) Observed. (e) Averaged. (f) Constrained.

Fig. 15: Corrected pose estimation of a cylinder and a cuboid on a table. 15a shows poses of cylinders and cuboids from a set of noisy observations. 15b shows estimated poses by simply averaging all the observations. 15c shows an estimated pose subject to the cylinder and cuboid being upright and on the table and they don't collide. Figures 15d–15f show corresponding overhead views.

sensor networks. In *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, pages 80–89. ACM, 2015.

[2] N. R. Ahmed, E. M. Sample, and M. Campbell. Bayesian multicategorical soft data fusion for human-robot collaboration. *IEEE Transactions on Robotics*, 29(1):189–206, 2013.

[3] M. Bocquel, H. Driessen, and A. Bagchi. Multitarget tracking with interacting population-based MCMC-PF. In *International Conference on Information Fusion (Fusion)*, pages 74–81. IEEE, 2012.

[4] M. Brenner, N. Hawes, J. D. Kelleher, and J. L. Wyatt. Mediating between qualitative and quantitative representations for task-orientated human-robot interaction. In *IJCAI*, pages 2072–2077, 2007.

[5] V. Chernozhukov and H. Hong. An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346, 2003.

[6] R. Daviet. Inference with Hamiltonian sequential Monte Carlo simulators. 2016.

[7] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[8] S. Golchi and D. A. Campbell. Sequentially constrained Monte Carlo. *Computational Statistics & Data Analysis*, 97:98–113, 2016.

[9] S. Golchi and J. L. Loeppky. Monte Carlo based designs for constrained domains. *arXiv preprint arXiv:1512.07328*, 2015.

[10] D. L. Hall and J. Llinas. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23, 1997.

[11] M. D. Hoffman and A. Gelman. The No-U-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[12] T. Kaupp, B. Douillard, F. Ramos, A. Makarenko, and B. Upcroft. Shared environment representation for a human-robot team performing information fusion. *Journal of Field Robotics*, 24(11-12):911–942, 2007.

[13] W. G. Kennedy, M. D. Bugajska, M. Marge, W. Adams, B. R. Fransen, D. Perzanowski, A. C. Schultz, and J. G. Trafton. Spatial representation and reasoning for human-robot collaboration. In *AAAI*, volume 7, pages 1554–1559, 2007.

[14] Z. Khan, T. Balch, and F. Dellaert. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE transactions on pattern analysis and machine intelligence*, 27(11):1805–1819, 2005.

[15] M. C. Koval, M. R. Dogar, N. S. Pollard, and S. S. Srinivasa. Pose estimation for contact manipulation with manifold particle filters. In *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pages 4541–4548. IEEE, 2013.

[16] L. Kunze, K. K. Doreswamy, and N. Hawes. Using qualitative spatial relations for indirect object search. In *IEEE Int. Conf. Robotics and Automation (ICRA)*, pages 163–168. IEEE, 2014.

[17] R. P. Mahler. *Statistical multisource-multitarget information fusion*. Artech House, Inc., 2007.

[18] M. McClelland, M. E. Campbell, and T. Estlin. Qualitative relational mapping for planetary rover exploration. In *AIAA Guidance, Navigation, and Control (GNC) Conference*, page 5037, 2013.

[19] R. M. Neal et al. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162, 2011.

[20] F. Septier and G. W. Peters. Langevin and Hamiltonian based sequential MCMC for efficient Bayesian filtering in high-dimensional spaces. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):312–327, 2016.

[21] D. C. Shah and M. E. Campbell. A qualitative path planner for robot navigation using human-provided maps. *I. J. Robotics Res.*, 32(13):1517–1535, 2013.

[22] L. L. Wong, L. P. Kaelbling, and T. Lozano-Perez. Collision-free state estimation. In *IEEE Int. Conf. Robotics and Automation (ICRA)*, pages 223–228. IEEE, 2012.

[23] L. L. Wong, L. P. Kaelbling, and T. Lozano-Pérez. Not seeing is also believing: Combining object and metric spatial information. In *IEEE Int. Conf. Robotics and Automation (ICRA)*, pages 1253–1260, 2014.

[24] D. Yi, M. A. Goodrich, and K. D. Seppi. MORRF*: Sampling-based multi-objective motion planning. In *IJCAI*, pages 1733–1741, 2015.

[25] D. Yi, M. A. Goodrich, and K. D. Seppi. Homotopy-aware RRT*: Toward human-robot topological path-planning. In *ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, pages 279–286. IEEE, 2016.

[26] D. Yi, T. M. Howard, M. A. Goodrich, and K. D. Seppi. Expressing homotopic requirements for mobile robot navigation through natural language instructions. In *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pages 1462–1468. IEEE, 2016.