

# Constrained Variational Inference via Safe Particle Flow

Yinzhuang Yi   Jorge Cortés   Nikolay Atanasov

**Abstract**—We propose a control barrier function (CBF) formulation for enforcing equality and inequality constraints in variational inference. The key idea is to define a barrier functional on the space of probability density functions that encode the desired constraints imposed on the variational density. By leveraging the Liouville equation, we establish a connection between the time derivative of the variational density and the particle drift, which enables the systematic construction of corresponding CBFs associated to the particle drift. Enforcing these CBFs gives rise to the safe particle flow and ensures that the variational density satisfies the original constraints imposed by the barrier functional. This formulation provides a principled and computationally tractable solution to constrained variational inference, with theoretical guarantees of constraint satisfaction. The effectiveness of the method is demonstrated through numerical simulations.

**Index Terms**—Estimation, Variational methods, Safety-critical control

## I. INTRODUCTION

**B**AYESIAN inference plays a key role in a variety of applications, including statistical learning [1], estimation theory [2], and motion planning [3]. In Bayesian inference problems, we start with a prior probability density function (PDF)  $p(\mathbf{x})$ , and given an observation  $\mathbf{z}$  and associated likelihood PDF  $p(\mathbf{z}|\mathbf{x})$ , we aim to compute a posterior PDF  $p(\mathbf{x}|\mathbf{z})$ , following Bayes’ rule. Traditional approaches to Bayesian inference include the Kalman filter [4], which relies on linear Gaussian assumptions, and its extension to nonlinear observation and transition models, the extended Kalman filter (EKF) [5]. As an alternative, particle filters [6] and sequential Monte Carlo methods [7] approximate the posterior using weighted particles instead of a parametric density.

Variational inference (VI) [8] is a formulation of Bayesian inference as an optimization problem with Kullback–Leibler (KL) divergence between a variational density  $q(\mathbf{x})$  and the posterior density  $p(\mathbf{x}|\mathbf{z})$  as the objective. Many effective methods [9] exist for VI. A particularly effective approach is particle-based VI [10]–[12], where the variational density is approximated by a finite set of particles that evolve according to a particle drift function. Examples of such methods include the Stein variational gradient descent [10], the particle flow particle filter [11], and diffusion-based VI [12].

In this paper, we consider a constrained VI problem where the posterior density must satisfy certain conditions, e.g., manifold constraints on orientation in robot state estimation

[13]. Constrained VI can be approached by requiring particle samples from the variational density to satisfy the constraints. Methods using this approach typically modify the particle drift to ensure constraint satisfaction [14]–[16]. For instance, equality constraints have been addressed using projection methods [15] and Lagrange multiplier formulations [14]. The projection method [15] is extended in [16] to handle multiple equality and inequality constraints. However, the inclusion of inequality constraints relies on slack variables, which leads to poor constraint satisfaction in practice. Instead of altering an established particle drift, one can also derive particle drift using a penalized objective that encodes the prescribed constraints [17], [18]. The penalized objective can be constructed either by augmenting the KL divergence with an additional penalty term [17] or by modifying the posterior density [17], [18]. A hybrid approach has been proposed in [19], where a particle drift is derived from a modified posterior that encodes inequality constraints, and subsequently adjusted to enforce equality constraints through a projection method similar to [15]. Nonetheless, penalization methods lack guarantees on exact constraint satisfaction. Moreover, all aforementioned methods enforce constraints only on individual particles and do not formally establish whether particle-wise constraint satisfaction guarantees variational-density-wise constraint satisfaction. To address this limitation, Chamon et al. [20] formulate a primal-dual approach, where the variational density and the dual variables associated with the constraints are updated simultaneously following the steepest descent/ascent direction of the Lagrangian under the Wasserstein metric. This method provides theoretical guarantees on the variational density asymptotically satisfying chance constraints on the expectations of the constraint functions, which is different from the deterministic constraints on the density function support studied here.

Our objective is to preserve the simplicity of modifying a desired particle drift while providing guarantees on constraint satisfaction for the variational density. This motivates using control barrier functions (CBFs), which provide a rigorous approach to enforce constraints on the evolution of a control system [21], [22]. The safe particle flow introduced here can be interpreted as a minimally modified instance of the gradient flow of the KL divergence in the space of PDFs. Recent developments in the CBF literature introduce the safe gradient flow [22], demonstrating the effectiveness of CBFs for enforcing constraints along the gradient flow dynamics in nonlinear optimization. CBF techniques have gained popularity in the control community due to their simplicity and formal guarantees for constraint satisfaction. For control-affine systems, CBF conditions take the form of linear constraints in the control input, enabling safe control synthesis via quadratic

We gratefully acknowledge support from ONR Award N00014-23-1-2353 and NSF FRR CAREER 2045945.

The authors are with the Contextual Robotics Institute, University of California San Diego, La Jolla, CA 92093, USA (e-mails: {yiyi, cortes, natanasov}@ucsd.edu).

programming. A comprehensive overview of CBF techniques and their application as safety constraints in quadratic programs is provided in [21]. These methods have been extended to functional spaces, illustrating the use of control barrier functionals for time-delayed system safety [23]. Despite their success in enforcing constraints within control systems, CBF techniques have not been used to enforce constraints in VI.

We propose a methodology that leverages CBF techniques to incorporate inequality constraints in VI in three key steps: (1) a barrier functional encodes the desired constraints on the variational density, yielding conditions on its time derivative. These are challenging to work with directly due to the infinite-dimensional nature of the density space; (2) the Liouville equation [24] relates the time derivative of the variational density to the deterministic particle drift, allowing us to instead formulate corresponding CBF constraints on the particle drift; (3) satisfying the CBF constraints on the deterministic particle drift guarantees that the variational density flow also satisfies the barrier constraints. The deterministic particle drift is essential for the validity of CBF techniques. The CBF-based flow formulation offers an anytime algorithm that ensures constraint satisfaction at any time during the particle evolution, unlike Lagrange multiplier-based methods that only guarantee asymptotic constraint satisfaction. Our method provides a principled and computationally tractable way of ensuring density-wise constraint satisfaction, thereby establishing a rigorous connection between safe particle flow and constrained VI.

## II. PROBLEM STATEMENT

Consider a Bayesian inference problem where  $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$  is a random variable of interest with prior PDF  $p(\mathbf{x})$ . Given a measurement  $\mathbf{z} \in \mathbb{R}^m$  with likelihood PDF  $p(\mathbf{z}|\mathbf{x})$ , the posterior PDF of  $\mathbf{x}$  conditioned on  $\mathbf{z}$  is determined by Bayes' theorem [25]:

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{z})}, \quad (1)$$

where  $p(\mathbf{z})$  is the marginal measurement PDF, computed as  $p(\mathbf{z}) = \int_{\mathcal{X}} p(\mathbf{z}|\mathbf{x})p(\mathbf{x}) d\mathbf{x}$ . Computing the posterior PDF in (1) is often intractable because  $p(\mathbf{z})$  may not have a closed form, except when the prior  $p(\mathbf{x})$  and the likelihood  $p(\mathbf{z}|\mathbf{x})$  are a conjugate pair. To calculate the posterior of non-conjugate prior and likelihood, approximation methods are needed.

This problem can be approached using VI methods [9, Ch. 10], which approximate the posterior  $p(\mathbf{x}|\mathbf{z})$  in (1) using a PDF  $q(\mathbf{x})$  with tractable expression, termed variational density. To perform the approximation, VI minimizes the KL divergence between the true posterior  $p(\mathbf{x}|\mathbf{z})$  and the variational density  $q(\mathbf{x})$ :

$$D_{KL}(q(\mathbf{x})||p(\mathbf{x}|\mathbf{z})) = \int_{\mathcal{X}} q(\mathbf{x}) \log \left( \frac{q(\mathbf{x})}{p(\mathbf{x}|\mathbf{z})} \right) d\mathbf{x}.$$

The variational density  $q(\mathbf{x})$  can be represented as a collection of weighted particles [10], leading to the particle-based VI considered in this paper.

This paper considers a constrained VI problem with *support constraints*: the variational density is restricted to take nonzero values only within a designated set, which we term safe. The

set  $\mathcal{S}$  is defined as the intersection of  $N$  inequality-constrained sets, each specified implicitly by a continuously differentiable function  $g_i : \mathcal{X} \rightarrow \mathbb{R}$ :

$$\mathcal{S} = \bigcap_{i \in \mathcal{I}} \mathcal{S}_i, \quad \mathcal{S}_i = \{\mathbf{x} \in \mathcal{X} \mid g_i(\mathbf{x}) \geq 0\},$$

where  $\mathcal{I} = \{1, \dots, N\}$ . Equality constraints specified by  $g^e(\mathbf{x})$  can be equivalently represented by a pair of inequality constraints  $g^e(\mathbf{x}) = 0 \Leftrightarrow [g^e(\mathbf{x}), -g^e(\mathbf{x})]^\top \geq \mathbf{0}$ . The safe set can encode, for instance, geometric constraints, such as manifold constraints in robot state estimation. We want to find a variational density that matches the Bayes' posterior as much as possible while satisfying the safety constraint  $q(\mathbf{x}) = 0$ , for all  $\mathbf{x} \in \mathcal{X} \setminus \mathcal{S}$ . This constraint is essential to ensure that the density is supported on the safe set. We make the following assumptions throughout the paper.

**Assumption II.1** (Feasibility). The state space  $\mathcal{X} \subset \mathbb{R}^n$  is bounded and the safe set  $\mathcal{S}$  is nonempty.

The above assumption ensures the existence of a feasible solution to the constrained VI problem. In addition, we make the following assumption on the variational densities.

**Assumption II.2** (Variational Density Family). The variational density  $q(\mathbf{x})$  belongs to the family  $\mathcal{P} = \{p(\mathbf{x}) \in \mathcal{L}^1(\mathcal{X}) \mid \int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x} = 1, p(\mathbf{x}) \geq 0\}$ , where  $\mathcal{L}^1(\mathcal{X})$  is the space of absolutely integrable functions on  $\mathcal{X}$  with respect to the Lebesgue measure.

Assumption II.2 ensures that the variational densities are valid PDFs, supported on the state space  $\mathcal{X}$ . We formulate the constrained VI problem as follows.

**Problem 1.** Find a variational density  $q(\mathbf{x})$  that solves the optimization problem:

$$\begin{aligned} \min_{q(\mathbf{x}) \in \mathcal{P}} \quad & D_{KL}(q(\mathbf{x})||p(\mathbf{x}|\mathbf{z})) \\ \text{s.t.} \quad & \int_{\mathcal{X} \setminus \mathcal{S}} q(\mathbf{x}) d\mathbf{x} = 0. \end{aligned} \quad (2)$$

The constraint ensures that  $q(\mathbf{x})$  has support strictly on  $\mathcal{S}$ .

## III. VARIATIONAL INFERENCE USING PARTICLE FLOW

Before considering the constrained formulation in Problem 1, we review a gradient flow method for solving unconstrained VI problems [26]. In this approach, an initial guess for the variational density is modified by following the steepest descent direction of the KL divergence functional. This yields a continuous-time trajectory  $q(\mathbf{x}; t)$  whose asymptotic limit is a solution to the VI problem.

### A. Gradient Flow

The tangent space  $T_q \mathcal{P}$  of the density family  $\mathcal{P}$  at a PDF  $q \in \mathcal{P}$  is [26]:

$$T_q \mathcal{P} = \left\{ \sigma(\mathbf{x}) \in \mathcal{L}^1(\mathcal{X}) \mid \int \sigma(\mathbf{x}) d\mathbf{x} = 0 \right\}.$$

The cotangent space  $T_q^* \mathcal{P}$  is the dual of  $T_q \mathcal{P}$ . We can introduce a bilinear map  $\langle \cdot, \cdot \rangle_{\mathcal{P}}$  as the duality pairing  $T_q^* \mathcal{P} \times$

$T_q\mathcal{P} \rightarrow \mathbb{R}$ . For any  $\psi \in T_q^*\mathcal{P}$  and  $\sigma \in T_q\mathcal{P}$ , the pairing can be identified as  $\langle \psi, \sigma \rangle_{\mathcal{P}} = \int_{\mathcal{X}} \psi(\mathbf{x}) \sigma(\mathbf{x}) d\mathbf{x}$ . The first variation of the KL divergence  $\frac{\delta D_{KL}(q||p)}{\delta q}$  with respect to  $q \in \mathcal{P}$  is an element of the cotangent space  $T_q^*\mathcal{P}$ , given by  $\frac{\delta D_{KL}(q||p)}{\delta q} = 1 + \log \frac{q}{p}$ . Given a metric tensor at  $q$ , denoted by  $M(q) : T_q\mathcal{P} \rightarrow T_q^*\mathcal{P}$ , we can express the Riemannian metric  $g_q : T_q\mathcal{P} \times T_q\mathcal{P} \rightarrow \mathbb{R}$  as  $g_q(\sigma_1, \sigma_2) = \langle M(q)\sigma_1, \sigma_2 \rangle_{\mathcal{P}}$ . The gradient of the KL divergence under the Riemannian metric, denoted by  $\nabla_q D_{KL}(q||p)$ , is defined as:  $g_q(\nabla_q D_{KL}(q||p), \sigma) = \langle \frac{\delta D_{KL}(q||p)}{\delta q}, \sigma \rangle_{\mathcal{P}}$ , for any  $\sigma \in T_q\mathcal{P}$ . Using the metric tensor, we can write  $\nabla_q D_{KL}(q||p) = M^{-1}(q) \frac{\delta D_{KL}(q||p)}{\delta q}$ . The gradient flow of the KL divergence with respect to this metric is:

$$\frac{\partial q(\mathbf{x}; t)}{\partial t} = -\nabla_q D_{KL}(q||p)|_{q=q(\mathbf{x}; t)}. \quad (3)$$

The convergence of the gradient flow to the optimum depends on the choice of Riemannian metric. For a detailed analysis of convergence properties under different metrics, we refer the reader to [26]. However, implementing the gradient flow directly is challenging because it is defined over the infinite-dimensional space of PDFs. To address this, the variational density  $q(\mathbf{x}; t)$  can be represented by a finite set of samples drawn from it, referred to as *particles*, and its evolution can be characterized through the evolution of these particles. The connection between the particle evolution and the gradient flow dynamics is formalized by the Liouville equation [24], which we review next.

### B. Liouville Equation and Particle Flow

Consider a random process  $\mathbf{x}(t) \in \mathbb{R}^n$  governed by the ordinary differential equation (ODE):

$$\frac{d\mathbf{x}(t)}{dt} = \phi(\mathbf{x}(t), t), \quad (4)$$

where  $\phi(\mathbf{x}(t), t)$  is the particle drift. Then, the PDF  $q(\mathbf{x}; t)$  of  $\mathbf{x}(t)$  evolves according to the Liouville equation [24]:

$$\frac{\partial q(\mathbf{x}; t)}{\partial t} = -\nabla_{\mathbf{x}} \cdot (q(\mathbf{x}; t) \phi(\mathbf{x}, t)), \quad (5)$$

where  $\nabla_{\mathbf{x}} \cdot$  denotes the divergence operator. The Liouville equation (5) establishes a mapping between the functional gradient in density space (3) and the particle drift  $\phi(\mathbf{x}(t), t)$ . Since the state space  $\mathcal{X}$  considered in this paper is a subset of  $\mathbb{R}^n$ , we impose the following assumption to ensure that the Liouville equation holds in our setting.

**Assumption III.1** (Conservation of Probability Mass). On the boundary of the state space, the particle drift  $\phi(\mathbf{x}, t)$  satisfies  $\langle \phi(\mathbf{x}, t), \hat{\mathbf{n}}(\mathbf{x}) \rangle_{\mathbb{R}^n} = 0$ , for all  $\mathbf{x} \in \partial\mathcal{X}$ , where  $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$  denotes the standard Euclidean inner product on  $\mathbb{R}^n$  and  $\hat{\mathbf{n}}(\mathbf{x})$  is the outward unit normal vector to  $\partial\mathcal{X}$  at  $\mathbf{x}$ .

This assumption ensures [27] that the trajectories  $p(\mathbf{x}; t)$  governed by the Liouville equation (5) remain in  $\mathcal{P}$ . Note that the particle drift must remain deterministic given our reliance on CBF techniques to ensure safety.

Now, consider a particle  $\mathbf{x}(t)$  sampled from the variational density  $q(\mathbf{x}; t)$ , whose evolution follows the gradient flow

dynamics in (3). The Liouville equation (5) is used to derive a corresponding particle drift  $\phi(\mathbf{x}, t)$  that induces the desired gradient flow for the variational density  $q(\mathbf{x}; t)$ , by solving:

$$\nabla_q D_{KL}(q||p)|_{q=q(\mathbf{x}; t)} = \nabla_{\mathbf{x}} \cdot (q(\mathbf{x}; t) \phi(\mathbf{x}, t)). \quad (6)$$

The evolution of the particle is then governed by (4), referred to as the *particle flow*. In this paper, we focus on the particle flow derived using the Stein Riemannian metric, originally introduced in [10]. Specifically, we consider the inverse metric tensor satisfying [26]

$$M^{-1}(q)\psi = -\nabla_{\mathbf{x}} \cdot \left( q(\mathbf{x}) \int_{\mathcal{X}} k(\mathbf{x}, \xi) q(\xi) \nabla_{\xi} \psi(\xi) d\xi \right),$$

where  $\psi \in T_q^*\mathcal{P}$  with  $k(\cdot, \cdot)$  denotes a positive definite kernel. The particle drift is obtained by solving (6):

$$\begin{aligned} \phi_d(\mathbf{x}, t) &= - \int_{\mathcal{X}} k(\mathbf{x}, \xi) q(\xi; t) \nabla_{\xi} \log \left( \frac{q(\xi; t)}{p(\xi, \mathbf{z})} \right) d\xi \\ &= \int_{\mathcal{X}} q(\xi; t) (\nabla_{\xi} k(\mathbf{x}, \xi) + k(\mathbf{x}, \xi) \nabla_{\xi} \log p(\xi, \mathbf{z})) d\xi \\ &\approx \frac{1}{M} \sum_{j=1}^M \left( \nabla_{\xi} k(\xi, \mathbf{x}) + k(\xi, \mathbf{x}) \nabla_{\xi} \log p(\xi, \mathbf{z}) \right) \Big|_{\xi=\mathbf{x}_j(t)}, \end{aligned} \quad (7)$$

where  $\{\mathbf{x}_j(t)\}_{j=1}^M \sim q(\mathbf{x}; t)$  are particles sampled from the variational density at time  $t$ . The second equality follows from integration by parts, while the last step uses Monte Carlo integration to approximate the expectation.

## IV. SAFE PARTICLE FLOW

The gradient flow of the KL divergence in (3) does not take the constraint in the optimization (2) into consideration. Directly modifying the gradient flow dynamics is challenging because of the infinite-dimensional nature of the space of PDFs and the difficulty of enforcing the constraint.

Our approach to deal with Problem 1 exploits the connection between the gradient flow dynamics and the particle drift established by the Liouville equation (6). In the forthcoming discussion, we first introduce a barrier functional in the space of PDFs and use it to formulate constraints on the variational density flow so that its continuous-time trajectory satisfies the constraint in (3) at all times. Using the Liouville equation, we then translate the constraints imposed on the variational density into equivalent constraints on the particle drift. Finally, we construct a safe particle drift by modifying a desired drift obtained from the unconstrained VI problem.

### A. Barrier Functions

We first review barrier functions in finite-dimensional vector spaces [28], [29]. Consider an autonomous system  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  in  $\mathbb{R}^n$  with state trajectories denoted by  $\mathbf{y}(t)$ . The safety of the system can be certified by ensuring that the trajectories remain within a safe set  $\mathcal{C} \subset \mathbb{R}^n$ . This is equivalent to showing that  $\mathcal{C}$  is forward-invariant [28].

To establish forward invariance, we introduce a continuously differentiable function  $b : \mathbb{R}^n \rightarrow \mathbb{R}$  that encodes the safe set as its zero-superlevel set,  $\mathcal{C} = \{\mathbf{y} \in \mathbb{R}^n \mid b(\mathbf{y}) \geq 0\}$ . Such a function is called a *barrier function*. A sufficient condition for

the forward invariance of  $\mathcal{C}$  is that the barrier function satisfies a differential inequality along system trajectories:  $\frac{db(\mathbf{y}(t))}{dt} + \alpha_b b(\mathbf{x}) \geq 0$ , where  $\alpha_b > 0$  is a positive constant. This ensures forward invariance of the safe set [28, Theorem 3.1].

### B. Barrier Functional Construction

Since the variational density evolves in the space of PDFs, rather than in a finite-dimensional vector space, we rely on the concept of *barrier functional* [23] to extend the forward invariance property to PDF trajectories. Based on the constraint in (2), the set of feasible densities is:

$$\mathcal{P}_s = \{p(\mathbf{x}) \in \mathcal{P} \mid \int_{\mathcal{X} \setminus \mathcal{S}} p(\mathbf{x}) d\mathbf{x} = 0\}. \quad (8)$$

We define the barrier functional  $\mathbf{h} : \mathcal{P} \rightarrow \mathbb{R}^N$ , whose  $i$ th component is given by

$$h_i(q(\mathbf{x}; t)) = - \int_{\mathcal{X} \setminus \mathcal{S}_i} g_i(\mathbf{x}) q(\mathbf{x}; t) d\mathbf{x}, \quad i \in \mathcal{I}. \quad (9)$$

We show next that the zero-level set of the barrier functional coincides with the set of feasible densities defined in (8).

**Lemma IV.1** (Consistent Barrier Functional). *The zero-level set of the barrier functional introduced in (9) satisfies  $\{p(\mathbf{x}) \in \mathcal{P} \mid \mathbf{h}(p(\mathbf{x})) = \mathbf{0}\} = \mathcal{P}_s$ .*

*Proof.* Based on the definition of the constraint set  $\mathcal{S}_i$ ,  $g_i(\mathbf{x}) < 0$  for all  $\mathbf{x} \in \mathcal{X} \setminus \mathcal{S}_i$ . Consequently,  $-g_i(\mathbf{x})p(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathcal{X} \setminus \mathcal{S}_i$ . Given the definition in (9), the condition  $h_i(p(\mathbf{x})) = 0$  implies that  $p(\mathbf{x}) = 0$  for all  $\mathbf{x} \in \mathcal{X} \setminus \mathcal{S}_i$ . Therefore, if  $\mathbf{h}(p(\mathbf{x})) = \mathbf{0}$ , then  $p(\mathbf{x}) = 0$  for all  $\mathbf{x} \in \bigcup_{i \in \mathcal{I}} (\mathcal{X} \setminus \mathcal{S}_i) = \mathcal{X} \setminus \bigcap_{i \in \mathcal{I}} \mathcal{S}_i = \mathcal{X} \setminus \mathcal{S}$ . Hence, we obtain  $\int_{\mathcal{X} \setminus \mathcal{S}} p(\mathbf{x}) d\mathbf{x} = 0$ .  $\square$

Based on Lemma IV.1, ensuring forward invariance of  $\mathcal{P}_s$  is equivalent to ensuring  $\mathbf{h}(p) = \mathbf{0}$ . Since  $\mathbf{h}(p) \geq \mathbf{0}$  for all  $p \in \mathcal{P}$  by definition, this is equivalent to ensuring  $\mathbf{h}(p) \leq \mathbf{0}$ . This yields the following barrier constraint:

$$\frac{d\mathbf{h}(q(\mathbf{x}; t))}{dt} + \alpha_h \mathbf{h}(q(\mathbf{x}; t)) \leq \mathbf{0}, \quad (10)$$

where  $\alpha_h > 0$  is a positive design parameter that determines the convergence rate of  $q(\mathbf{x}; t)$  to the safe set boundary [30].

### C. Safe Particle Flow

The constraint (10) can be used to derive corresponding constraints on the particle drift  $\phi(\mathbf{x}, t)$ , hence, establishing necessary conditions for the particle drift to render the feasible density set  $\mathcal{P}_s$  forward invariant.

As we show next, the conditions on the particle drift  $\phi(\mathbf{x}, t)$  can be expressed in terms of its Euclidean inner product with the gradient of the constraint functions  $\nabla_{\mathbf{x}} g_i(\mathbf{x})$ . This result facilitates our construction of a safe particle drift in the next section using CBF techniques.

**Theorem IV.2** (Safe Particle Flow). *Let  $\phi(\mathbf{x}, t)$  be a particle drift satisfying, for all  $i \in \mathcal{I}$ ,*

$$\int_{\mathcal{X} \setminus \mathcal{S}_i} q(\mathbf{x}; t) \langle \nabla_{\mathbf{x}} g_i(\mathbf{x}), \phi(\mathbf{x}, t) \rangle_{\mathbb{R}^n} d\mathbf{x} \geq \alpha_h h_i(q(\mathbf{x}; t)). \quad (11)$$

*Under Assumptions II.1, II.2 and III.1, the particle flow (4) ensures that the feasible density set  $\mathcal{P}_s$  is forward-invariant and exponentially stable.*

*Proof.* Assumption II.1 ensures that the feasible density set is nonempty. Under the regularity conditions stated in Assumption II.2, the time derivative of the  $i$ th component of the barrier functional is given by

$$\frac{dh_i(q(\mathbf{x}; t))}{dt} = \left\langle \frac{\delta h_i(q(\mathbf{x}; t))}{\delta q(\mathbf{x}; t)}, \frac{\partial q(\mathbf{x}; t)}{\partial t} \right\rangle_{\mathcal{P}}. \quad (12)$$

Observe that each barrier functional component in (9) can be expressed as an inner product between a scaled indicator function and the variational density:  $h_i(q(\mathbf{x}; t)) = - \langle g_i(\mathbf{x}) \mathbb{I}_{\mathcal{X} \setminus \mathcal{S}_i}(\mathbf{x}), q(\mathbf{x}, t) \rangle_{\mathcal{P}}$ . By the linearity of the inner product operator, we obtain:

$$\frac{\delta h_i(q(\mathbf{x}; t))}{\delta q(\mathbf{x}; t)} = -g_i(\mathbf{x}) \mathbb{I}_{\mathcal{X} \setminus \mathcal{S}_i}(\mathbf{x}). \quad (13)$$

Since the variational density evolution is governed by the Liouville equation, substituting (5) and (13) into (12) yields:  $\frac{dh_i(q(\mathbf{x}; t))}{dt} = \int_{\mathcal{X} \setminus \mathcal{S}_i} g_i(\mathbf{x}) \nabla_{\mathbf{x}} \cdot (q(\mathbf{x}; t) \phi(\mathbf{x}, t)) d\mathbf{x}$ . This expression can be simplified using Green's theorem [31], yielding

$$\begin{aligned} \frac{dh_i(q(\mathbf{x}; t))}{dt} &= \oint_{\Gamma_i} g_i(\mathbf{x}) q(\mathbf{x}; t) \langle \phi(\mathbf{x}, t), \hat{\mathbf{n}}_i(\mathbf{x}) \rangle_{\mathbb{R}^n} d\mathbf{x} \\ &\quad - \int_{\mathcal{X} \setminus \mathcal{S}_i} q(\mathbf{x}; t) \langle \nabla_{\mathbf{x}} g_i(\mathbf{x}), \phi(\mathbf{x}, t) \rangle_{\mathbb{R}^n} d\mathbf{x}, \end{aligned}$$

where  $\Gamma_i := \partial(\mathcal{X} \setminus \mathcal{S}_i)$  and  $\hat{\mathbf{n}}_i(\mathbf{x})$  is the outward unit normal vector to the boundary  $\Gamma_i$ . Since the boundary  $\Gamma_i$  satisfies  $\Gamma_i \subseteq \partial\mathcal{X} \cup \partial\mathcal{S}_i$ , and the boundary of the  $i$ th constraint set satisfies  $\partial\mathcal{S}_i \subseteq \partial\mathcal{X} \cup \{\mathbf{x} \in \mathcal{X} \mid g_i(\mathbf{x}) = 0\}$ . We have  $\Gamma_i \subseteq \partial\mathcal{X} \cup \partial\{\mathbf{x} \in \mathcal{X} \mid g_i(\mathbf{x}) = 0\}$ . By Assumption III.1, we have  $\langle \phi(\mathbf{x}, t), \hat{\mathbf{n}}_i(\mathbf{x}) \rangle_{\mathbb{R}^n} = 0$ , for all  $\mathbf{x} \in \partial\mathcal{X}$ . As a result,  $g_i(\mathbf{x}) \langle \phi(\mathbf{x}, t), \hat{\mathbf{n}}_i(\mathbf{x}) \rangle_{\mathbb{R}^n} = 0$  for all  $\mathbf{x} \in \Gamma_i$ , which yields  $\oint_{\Gamma_i} g_i(\mathbf{x}) q(\mathbf{x}; t) \langle \phi(\mathbf{x}, t), \hat{\mathbf{n}}_i(\mathbf{x}) \rangle_{\mathbb{R}^n} d\mathbf{x} = 0$ . The condition in (11) follows from the component-wise form of the constraint in (10). Forward-invariance of the feasible density set  $\mathcal{P}_s$  follows from [23, Theorem 3], while exponential stability is established via Grönwall's inequality [32].  $\square$

Having established conditions under which the particle drift renders the feasible density set  $\mathcal{P}_s$  forward invariant and exponentially stable, we next tackle the construction of a particle drift satisfying these conditions.

### D. Particle Drift Design

Inspired by CBF methods [21], we construct a safe particle drift  $\phi(\mathbf{x}, t)$  by modifying the desired Stein particle drift  $\phi_d(\mathbf{x}, t)$  in (7) that solves the unconstrained VI problem. We parameterize the particle drift  $\phi(\mathbf{x}, t)$  as follows:

$$\phi(\mathbf{x}, t) = \phi_d(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t), \quad (14)$$

where  $\mathbf{u} : \mathcal{X} \times [0, \infty) \rightarrow \mathbb{R}^n$  is an auxiliary control term introduced to modify the desired particle drift. The control term is chosen to satisfy the following condition for all  $i \in \mathcal{I}$ :

$$\nabla_{\mathbf{x}} g_i(\mathbf{x})^\top (\phi_d(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t)) + \alpha_g g_i(\mathbf{x}) \geq 0, \quad (15)$$



**TABLE I:** Runtime comparison of our safe PF, the projected PF [19], and CSVGD [16]. All methods employ the same ODE solver and integration time horizon.

Method	Exec. Time (s)	Avg. It. Time (s)	Total It.
Safe PF (ours)	9.922	0.060	164
Projected PF	9.558	0.043	218
CSVGD	176.293	0.187	941

where  $\alpha_g > 0$  is a positive constant. An input satisfying (15) can be obtained by solving the quadratic program:

$$\begin{aligned} \mathbf{u}(\mathbf{x}, t) = \underset{\mathbf{u} \in \mathbb{R}^n}{\operatorname{argmin}} \quad & \|\mathbf{u}\|_2^2 \\ \text{s.t.} \quad & \nabla_{\mathbf{x}} g_i(\mathbf{x})^\top (\phi_d(\mathbf{x}, t) + \mathbf{u}) + \alpha_g g_i(\mathbf{x}) \geq 0, \forall i \in \mathcal{I}. \end{aligned} \quad (16)$$

The optimization yields a minimally invasive control input that ensures satisfaction of the constraint (15). However, the input is well defined contingent upon the feasibility of (16). In this formulation, each constraint function  $g_i(\mathbf{x})$  can be considered as a CBF, and stacking them yields  $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_N(\mathbf{x})]^\top$ . Feasibility of (16) is guaranteed if  $\mathbf{g}(\mathbf{x})$  constitutes a valid vector CBF [22].

**Proposition IV.3** (Safe Particle Control). *Let  $\mathbf{g}(\mathbf{x})$  a valid vector CBF [22, Section II.C], the particle drift  $\phi(\mathbf{x}, t)$  defined by (14), with  $\mathbf{u}(\mathbf{x}, t)$  obtained by solving (16), satisfies (11) with  $\alpha_h = \alpha_g$ .*

*Proof.* The valid vector CBF condition guarantees the existence of a control input  $\mathbf{u}(\mathbf{x}, t)$  such that, for each constraint function  $g_i(\mathbf{x})$ , the particle drift satisfies  $\langle \nabla_{\mathbf{x}} g_i(\mathbf{x}), \phi(\mathbf{x}, t) \rangle_{\mathbb{R}^n} \geq -\alpha_g g_i(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{X}$ . By the definition of PDFs, we can multiply both sides of the inequality by  $q(\mathbf{x}; t)$  to obtain, for all  $\mathbf{x} \in \mathcal{X}$ ,  $q(\mathbf{x}; t) \langle \nabla_{\mathbf{x}} g_i(\mathbf{x}), \phi(\mathbf{x}, t) \rangle_{\mathbb{R}^n} \geq -\alpha_g q(\mathbf{x}; t) g_i(\mathbf{x})$ . Integrating both sides over  $\mathcal{X} \setminus \mathcal{S}_i$  preserves the inequality

$$\begin{aligned} \int_{\mathcal{X} \setminus \mathcal{S}_i} q(\mathbf{x}; t) \langle \nabla_{\mathbf{x}} g_i(\mathbf{x}), \phi(\mathbf{x}, t) \rangle_{\mathbb{R}^n} d\mathbf{x} \\ \geq \alpha_g \int_{\mathcal{X} \setminus \mathcal{S}_i} -q(\mathbf{x}; t) g_i(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

By the definition of the barrier functional (9), we have:  $\int_{\mathcal{X} \setminus \mathcal{S}_i} q(\mathbf{x}; t) \langle \nabla_{\mathbf{x}} g_i(\mathbf{x}), \phi(\mathbf{x}, t) \rangle_{\mathbb{R}^n} d\mathbf{x} \geq \alpha_g h_i(q(\mathbf{x}; t))$ , for all  $i \in \mathcal{I}$ , which coincides with (11) for  $\alpha_h = \alpha_g$ .  $\square$

The key steps of the proposed method are summarized in Algorithm 1. Instead of the Stein particle drift (7), our method can be formulated for other desired particle drifts, provided the associated gradient flow minimizes the KL divergence. Such drifts can be obtained by solving (6) with the KL divergence gradient computed under different Riemannian metrics [26].

## V. EVALUATION

We evaluate the proposed safe particle flow method on a Bayesian estimation problem with one equality constraint and one inequality constraint. We compare the proposed method with the projection-based approaches [16], [19], and demonstrate that our method achieves better approximation accuracy while strictly satisfying the imposed constraints.

## Algorithm 1 Safe Particle Flow

---

**Require:** Particles  $\{\mathbf{x}_j(0)\}_{j=1}^M$ , joint density  $p(\mathbf{x}, \mathbf{z})$ , and observation  $\tilde{\mathbf{z}}$

**Output:** Particles  $\{\mathbf{x}_j(T)\}_{j=1}^M$  that approximate the solution to (2)

- 1: **function**  $\mathbf{f}(\{\mathbf{x}_j(t)\}_{j=1}^M, t)$
- 2:   **for** each particle  $\mathbf{x}_j(t)$  **do**
- 3:      $\phi_d(\mathbf{x}_j(t), t) \leftarrow$  Evaluate (7) with  $p(\mathbf{x}, \tilde{\mathbf{z}})$  and  $\{\mathbf{x}_k(t)\}_{k=1}^M$  at  $(\mathbf{x}_j(t), t)$
- 4:      $\mathbf{u}(\mathbf{x}_j(t), t) \leftarrow$  Solve (16) with  $\phi_d(\mathbf{x}_j(t), t)$  at  $(\mathbf{x}_j(t), t)$
- 5:      $\phi(\mathbf{x}_j(t), t) \leftarrow \phi_d(\mathbf{x}_j(t), t) + \mathbf{u}(\mathbf{x}_j(t), t)$
- 6:   **return**  $\{\phi(\mathbf{x}_j(t), t)\}_{j=1}^M$
- 7: **while** ODE solver running **do**
- 8:    $\{\mathbf{x}_j(T)\}_{j=1}^M \leftarrow$  SolveODE( $\mathbf{f}(\{\mathbf{x}_j(t)\}_{j=1}^M, t)$ ) with initialization  $(\{\mathbf{x}_j(0)\}_{j=1}^M, t=0)$  and termination time  $T$
- 9: **return**  $\{\mathbf{x}_j(T)\}_{j=1}^M$

---

We consider the state space  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\|_\infty \leq 10^3\}$ . The prior density is a truncated Gaussian supported on  $\mathcal{X}$  with  $\hat{\mathbf{x}} = [0, 0]^\top$  and  $P = \begin{bmatrix} 15 & -5 \\ -5 & 15 \end{bmatrix}$ . The likelihood function is a Gaussian density  $p(\mathbf{z}|\mathbf{x}) = p_{\mathcal{N}}(\mathbf{z}; H(\mathbf{x}), R)$  with  $H(\mathbf{x}) = \|\mathbf{x}\|$ ,  $R = 1$ , and  $\mathbf{x}^* = [14.7, -10.1]^\top$ , where  $\mathbf{x}^*$  denotes the true value of  $\mathbf{x}$  used to generate the observation. The following constraints are imposed on the Bayesian estimation problem:

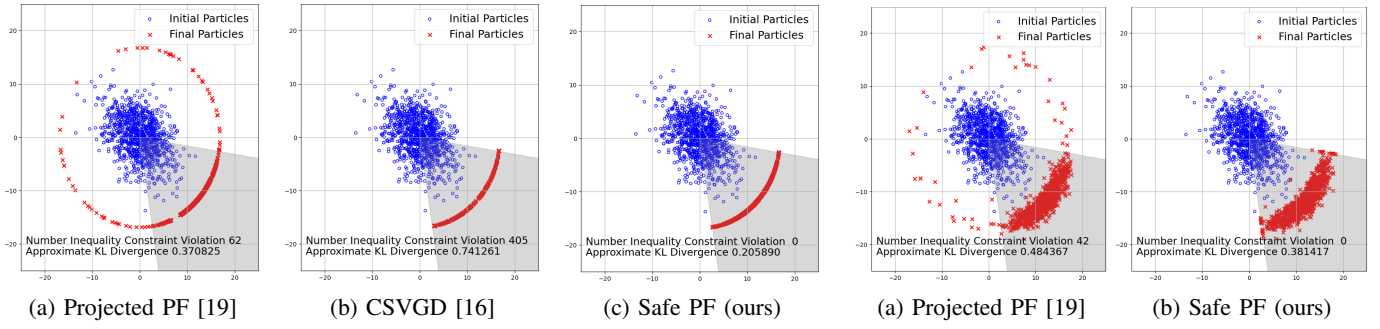
$$\begin{aligned} g(\mathbf{x}) &= \pi/5 - \arccos(\mathbf{d}^\top \mathbf{x} / \|\mathbf{x}\|_2) \geq 0 \\ g^e(\mathbf{x}) &= \|\mathbf{x}\|^2 - r^2 = 0, \end{aligned} \quad (17)$$

where  $\mathbf{d} = \frac{1}{\sqrt{2}}[1, -1]^\top$  and  $r = 15.8$ . The equality constraint enforces that the density is supported on a circle of radius  $r$ . The inequality constraint encodes a field-of-view restriction, requiring the density to lie within a cone centered along direction  $\mathbf{d}$  with half-angle  $\frac{\pi}{5}$ . All methods are initialized with particles  $\{\mathbf{x}_j(0)\}_{j=1}^{10^3}$  drawn from the prior. The desired particle drift is given by (7) with an RBF kernel of bandwidth 3.

The results considering both equality and inequality are shown in Fig. 1c. For reference, we also show in Fig. 1a the results obtained using the projected particle flow (PF) [19] and in Fig. 1b the results of constrained Stein variational gradient descent (CSVGD) [16]. The projected PF satisfies the equality constraint but violates the inequality constraint. CSVGD achieves better inequality constraint satisfaction, as the particles violating the inequality constraint remain close to the safe region. Our safe PF satisfies both constraints and achieves a low KL divergence estimate. To exclude interference between equality and inequality constraints, we repeated the experiment with only the inequality constraint. The results are shown in Fig. 2. The projected PF fails to satisfy the inequality constraint, as shown in Fig. 2a. In contrast, our safe PF satisfies the inequality constraint while achieving good convergence, as shown in Fig. 2b. Our method has computational efficiency comparable to [19], while being significantly more efficient than [16], as illustrated in Tab. I.

## VI. CONCLUSION

We have introduced a novel safe particle flow method to satisfy constraints in VI problems. We have established that



**Fig. 1:** Comparison of our safe PF with CSVGD [16] and the projected PF [19]. For each method, the desired particle drift is the Stein particle drift (7). The region satisfying the inequality constraint in (17) is shown in gray. The initial and final particles are shown as blue dots and red crosses, respectively.

**Fig. 2:** Comparison of safe PF with the projected PF [19], considering only the inequality constraint in (17). The initial and final particles are shown as blue dots and red crosses, respectively.

the constraints on the variational density can be equivalently reformulated as constraints on the particle drift. Combining ideas from safety control and the dynamical systems approach to algorithms, we have shown how to design a particle drift satisfying those constraints by solving a convex quadratic program. Our method proposed a simple yet efficient way to construct a safe particle flow while providing formal guarantees for constraint satisfaction for the variational density. Future work will focus on improving the efficiency of our method to enable real-time and high-dimensional applications, such as state estimation on manifolds and trajectory optimization.

## REFERENCES

- [1] M. E. Khan and H. Rue, “The Bayesian learning rule,” *Journal of Machine Learning Research*, vol. 24, no. 281, pp. 1–46, 2023.
- [2] T. D. Barfoot, *State Estimation for Robotics*. Cambridge University Press, 2017.
- [3] M. Mukadam, J. Dong, X. Yan, F. Dellaert, and B. Boots, “Continuous-time Gaussian process motion planning via probabilistic inference,” *The International Journal of Robotics Research*, vol. 37, no. 11, pp. 1319–1340, 2018.
- [4] R. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [5] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Courier Corporation, 2005.
- [6] N. J. Gordon, D. J. Salmond, and A. F. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” in *IEE Proceedings F Radar and Signal Processing*, vol. 140, 1993, pp. 107–113.
- [7] R. Neal, “MCMC using Hamiltonian dynamics,” in *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011, pp. 113–162.
- [8] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, “An introduction to variational methods for graphical models,” *Machine Learning*, vol. 37, pp. 183–233, 1999.
- [9] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [10] Q. Liu and D. Wang, “Stein variational gradient descent: A general purpose Bayesian inference algorithm,” *Advances in Neural Information Processing Systems*, vol. 29, 2016.
- [11] F. Daum and J. Huang, “Nonlinear filters with log-homotopy,” in *SPIE Signal and Data Processing of Small Targets*, vol. 6699, 2007, pp. 423–437.
- [12] T. Geffner and J. Domke, “Langevin diffusion variational inference,” in *International Conference on Artificial Intelligence and Statistics*, 2023, pp. 576–593.
- [13] T. Lee, “Spacecraft attitude estimation with a single magnetometer using matrix Fisher distributions on  $SO(3)$ ,” in *AIAA Scitech Forum*, 2019.
- [14] E. Heiden, C. E. Denniston, D. Millard, F. Ramos, and G. S. Sukhatme, “Probabilistic inference of simulation parameters via parallel differentiable simulation,” in *International Conference on Robotics and Automation (ICRA)*, 2022, pp. 3638–3645.
- [15] R. Zhang, Q. Liu, and X. Tong, “Sampling in constrained domains with orthogonal-space variational gradient descent,” *Advances in Neural Information Processing Systems*, pp. 37108–37120, 2022.
- [16] T. Power and D. Berenson, “Constrained Stein variational trajectory optimization,” *IEEE Transactions on Robotics*, vol. 40, pp. 3602–3619, 2024.
- [17] G. Tabor and T. Hermans, “Constrained Stein variational gradient descent for robot perception, planning, and identification,” *arXiv preprint arXiv:2506.00589*, 2025.
- [18] M. Gurbuzbalaban, Y. Hu, and L. Zhu, “Penalized overdamped and underdamped Langevin Monte Carlo algorithms for constrained sampling,” *Journal of Machine Learning Research*, vol. 25, no. 263, pp. 1–67, 2024.
- [19] K. J. Craft and K. J. DeMars, “Nonlinear particle flow for constrained Bayesian inference,” in *AIAA Scitech Forum*, 2024, p. 0428.
- [20] L. F. Chamon, M. R. Karimi, and A. Korba, “Constrained sampling with primal-dual Langevin Monte Carlo,” *Advances in Neural Information Processing Systems*, pp. 29285–29323, 2024.
- [21] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, “Control barrier function based quadratic programs for safety critical systems,” *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2017.
- [22] A. Allibhoy and J. Cortés, “Control-barrier-function-based design of gradient flows for constrained nonlinear programming,” *IEEE Transactions on Automatic Control*, vol. 69, no. 6, pp. 3499–3514, 2024.
- [23] A. K. Kiss, T. G. Molnar, A. D. Ames, and G. Orosz, “Control barrier functionals: Safety-critical control for time delay systems,” *International Journal of Robust and Nonlinear Control*, vol. 33, no. 12, pp. 7282–7309, 2023.
- [24] A. Wibisono, V. Jog, and P.-L. Loh, “Information and estimation in Fokker-Planck channels,” in *IEEE International Symposium on Information Theory*, 2017, pp. 2673–2677.
- [25] T. Bayes, “LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S.” *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, 1763.
- [26] Y. Chen, D. Z. Huang, J. Huang, S. Reich, and A. M. Stuart, “Gradient flows for sampling: Mean-field models, Gaussian approximations and affine invariance,” *arXiv preprint arXiv:2302.11024*, 2023.
- [27] G. Crispin, *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Berlin Heidelberg, 2009.
- [28] F. Blanchini, “Set invariance in control,” *Automatica*, vol. 35, no. 11, pp. 1747–1767, 1999.
- [29] S. Prajna and A. Jadbabaie, “Safety verification of hybrid systems using barrier certificates,” in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2004, pp. 477–492.
- [30] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, “Control barrier functions: Theory and applications,” in *European Control Conference (ECC)*, 2019, pp. 3420–3431.
- [31] J. J. Duistermaat and J. A. C. Kolk, *Multidimensional Real Analysis II: Integration*. Cambridge University Press, 2004.
- [32] T. H. Gronwall, “Note on the derivatives with respect to a parameter of the solutions of a system of differential equations,” *Annals of Mathematics*, vol. 20, no. 4, pp. 292–296, 1919.