Duality between density function and value function with applications in constrained optimal control and Markov Decision Process

Yuxiao Chen and Aaron D. Ames

***Abstract*— Density function describes the density of states in the state space with some initial state distribution. Its evolution follows the Liouville Partial Differential Equation (PDE). We show that the density function is the dual of the value function in the optimal control problems and strong duality holds. By utilizing the duality, constraints that are hard to enforce in the primal value function optimization such as safety constraints in robot navigation, traffic capacity constraints in traffic flow control can be posed on the density function, and the constrained optimal control problem can be solved with a primal-dual algorithm that alternates between the primal and dual optimization. The primal optimization follows the standard optimal control algorithm with a perturbation term generated by the density constraint, and the dual problem solves the Liouville PDE to get the density function under a fixed control strategy and updates the perturbation term. We show examples in robot navigation and traffic control to demonstrate the capability of the proposed formulation.**

I. Introduction

The problem of optimal control is one of the most well-studied problems in control. Due to Bellman's principle of optimality [3], dynamic programming has been the standard tool for solving optimal control problems, both in continuous state space and discrete state space, like in the case of a Markov Decision Process (MDP). The key tool used in dynamic programming is called a value function, which is the optimal cost-to-go at a given state and time, and the optimal control strategy is derived from the value function. However, some constrained optimal control problems are hard to cast as a constrained value function optimization problem, such as problems with safety constraint or convergence rate constraint. We show that the density function, first proposed by Rantzer [12], [14], is the dual of the value function in many different optimal control formulations, and some constrained optimal control problems can be conveniently written as an optimization over the density function. The idea of duality is not a new one, such as the concept of co-state in classic optimal control [11], the occupation measure approach in [9], [10], [15], [16]. However, the setup for occupation measure is typically finite-horizon and the computation is done with moment programming formulated as Semidefinite programming; while the approach in this paper uses a primal-dual algorithm and turn the computation to the HJB PDE of the primal problem.

The original use of the density function in [14] was as a dual to the Lyapunov function to prove stability of nonlinear systems. Since the density function follows the Liouville equation, which is a PDE and hard to enforce, the computation method for the density function has been limited to analytical method (propose one and validate by hand) and Sum of Squares programming [12]. Liouville equation is also directly used to formulate optimal control problem and analytical solution can be found for linear systems, as shown in [4]. We show in this paper that instead of viewing the density function as a certificate of stability, it actually has physical meaning as the distribution of states, and is the dual of the value function in optimal control. Besides, we propose an ODE approach to compute the density function, and on top of that a primal-dual algorithm that solves optimal control problems with density constraint.

Yuxiao Chen Aaron Ames are with the Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA, 91106, USA. Emails: {chenyx, ames}@caltech.edu

Nomenclature For the remainder of the paper, $\mathbb{N}$ denotes the set of natural numbers, $\mathbb{N}_+$ denotes the positive natural number, $\mathbb{R}$ denotes the set of real numbers. Given a dynamic equation $\dot{x} = F(x)$, $\Phi_F(x_0, T)$ denotes the flow map of the dynamics with initial state x_0 and horizon T. $\langle a, b\rangle_{\mathcal{X}} = \int_{\mathcal{X}} a(x) \cdot b(x) dx$ denotes the inner product of two functions a and b. $\mathbf{0}$ denotes a vector of all zeros or a function that is always zero, depending on the context. $\mathbb{1}_S$ denotes the indicator function of a set S.

II. Background review and problem setup

In this section, we review the concept of density function and optimal control, and formally define the problem to solve. We will review the optimal control formulation with continuous state and input space and also Markov decision process, where the state space and input space are discrete.

A. Optimal control and value function

There are numerous results in optimal control, we review the setting with continuous state and input space and continuous time. The standard formulation is the following:

$$\min_u \int_0^T C\left(x\left(t\right), u\left(t\right)\right) dt + D\left(x\left(T\right)\right) \text{ s.t. } \dot{x} = F\left(x, u\right), \tag{1}$$

where $x \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state, $u \in \mathcal{U} \subseteq \mathbb{R}^m$ is the control input, $\dot{x} = F(x, u)$ is the dynamics described as an ODE, $T \in \mathbb{R}_+ \cup \{+\infty\}$ is the horizon of the problem, $C : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ is the running cost function and $D : \mathcal{X} \to \mathbb{R}$ is the terminal cost function. We further assume that

$$\mathcal{U} := \{u \mid g(u) \leq 0\} \tag{2}$$

for some function g.

Remark 1. *For simplicity, we only consider time-invariant dynamics. It should be straightforward to extend the results to the case with time-varying dynamics.*

The Pontryagin Maximum Principle (PMP) [11] gives necessary conditions for the optimality of the solution, and perhaps the most classic tool for solving the optimal control problem is dynamic programming [3], which utilize the principle of optimality, and formulate the problem as a Hamilton-Jacobi-Bellman PDE:

$$\begin{cases} \frac{\partial V}{\partial t} + \min_{u \in \mathcal{U}} \{\nabla V(x,t) \cdot F(x,u) + C(x,u)\} = 0 \\ V(x,T) = D(x) \end{cases}, \quad (3)$$

where

$$V(x_0,t) = \min_{u \in \mathcal{U}} \left\{ \int_t^T C(x(t), u(t))\, dt + D(x(T)) \right\}, \quad \text{s.t.} x(t) = x_0, \dot{x} = F(x,u) \quad (4)$$

is the optimal cost-to-go of the optimal control problem for an initial condition x_0 at time t. Once the value function is known, the optimal policy is then

$$u^\star(x) = \arg\min_{u \in \mathcal{U}} \{\nabla V(x,t) \cdot F(x,u) + C(x,u)\}. \quad (5)$$

B. Markov decision process

Another relevant problem is the Markov decision process (MDP), which is a 4-tuple (S, A, P_a, R_a) with

- S is a finite set of states,
- A is a finite set of actions (sometimes the action at state s is limited to $A_s \subseteq A$),
- $P_a(s,s') = \mathbb{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$ is the transition probability from s to s' under action a,
- $R_a(s,s')$ is the reward associated with the transition from s to s' under action a.

An MDP solves for the optimal policy that maximizes the discounted cumulative reward

$$\sum\nolimits_{t=0}^{\infty} \gamma^t R_{a_t}(s_t, s_{t+1}), \quad (6)$$

where γ is the discount factor. One can also formulate an MDP that minimizes the cumulative cost, as is shown in Section IV-B.

A policy π is a mapping from state to the action space. A policy is deterministic if it maps a state to a deterministic action, it is stochastic if it maps a state to a distribution over multiple actions. A policy is stationary if it does not change with time. It can be proved that for a finite MDP with the reward function defined in (6), there always exists a stationary deterministic policy [13].

MDP can also be solved by dynamic programming, where it appears as the value iteration method:

$$\pi(s) = \arg\max_a \left\{ \sum_{s'} P_a(s',s) \left(R_a(s,s') + \gamma V(s')\right) \right\}$$
$$V(s) = \sum_{s'} P_{\pi(s)}(s,s') \left(R_{\pi(s)}(s,s') + \gamma V(s')\right), \quad (7)$$

C. Density function for dynamic systems

On the other hand, density function was proposed by Anders Rantzer in [14] as a dual to Lyapunov function. The density function $\rho : \mathcal{X} \times [0,T] \to \mathbb{R}$ can be understood as the measure of state concentration in the state space. Given the dynamics $\dot{x} = F(x)$, the evolution of density function follows the Liouville PDE:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \cdot F) = \phi(t,x,\rho), \quad \rho(x,0) = \rho_0(x), \quad (8)$$

where $\phi : [0,\infty) \times \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ is the supply function, $\phi(t,x_0,\rho(x_0,t)) > 0$ means a source, i.e., states with initial condition $x(t) = x_0$ appears at x_0 with intensity $\phi(t,x_0,\rho(x_0,t)) > 0$, and $\phi(t,x_0,\rho(x_0,t)) < 0$ denotes a sink, i.e. states exit the system with intensity $\phi(t,x_0,\rho(t,x_0))$ at x_0, time t. We allow ϕ to depend on ρ to allow more flexible characterization of the supply.

The Liouville PDE can be transformed and solved as an ODE since

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \cdot F) = \left.\frac{d\rho}{dt}\right|_{\dot{x}=F(x)} + (\nabla \cdot F)\rho = \phi. \quad (9)$$

This implies that we can integrate the following ODE to get the density function alone the trajectory of the dynamic system $\dot{x} = F(x)$ as

$$\begin{bmatrix} \dot{x} \\ \dot{\rho} \end{bmatrix} = \begin{bmatrix} F(x) \\ \phi(t,x,\rho) - \nabla \cdot F(t,x)\,\rho \end{bmatrix}. \quad (10)$$

With this, we can evaluate the density function at any state x_T, any time T with the following two step procedure:

- First, solve the reverse ODE of $\dot{x} = -F(x)$ with initial condition x_T to get $\Phi_{-F}(x,T) = \Phi_F(x,-T)$.
- Then, solve the extended ODE in (10) with initial condition $[\Phi_F(x,-T), \rho_0(\Phi_F(x,-T))]^\intercal$ to time T.

Assumption 1. *$\mathcal{X}$ is forward invariant under all possible dynamics considered in the optimal control problem.*

Remark 2. *Assumption 1 could be achieved if some barrier function intervention is implemented on $\partial\mathcal{X}$, see [2], [5] for example.*

For a stationary supply function, i.e. ϕ only depending on x and ρ, one would hope that there exists a stationary density function that any initial condition converges to, this is not always the case, but we provide sufficient condition for the convergence.

Define the extended dynamics in (10) as $\overline{F}$. Given a stationary supply function ϕ and an initial density function ρ_0, from the two step procedure shown above, we have

$$\rho(x,t) = \Phi_{\overline{F}}([\Phi_F(x,-t), \rho_0(\Phi_F(x,-t))]^\intercal, t)_{\downarrow\rho}, \quad (11)$$

where $\downarrow\rho$ means the projection of $[x,\rho]^\intercal$ to ρ.

Theorem 1. *Given a stationary supply function ϕ and an initial density function ρ_0, assume that there exists a ρ_s such that*

$$\forall x \in \mathcal{X}, \frac{\partial \rho_s}{\partial t} = \phi(x,\rho_s) - \nabla \cdot (\rho_s \cdot F) = 0.$$

For any $x \in \mathcal{X}$, if there exists $T \geq 0$ such that $\forall t \geq T, \rho_0(\Phi_F(x,-t)) = \rho_s(\Phi_F(x,-t))$, then $\forall t \geq T, \rho(x,t) = \rho_s(x)$.

Proof. The proof follows from the fact that

$$\begin{aligned}\forall t \geq T, \rho\left(x,t\right) &= \Phi_{\bar{F}}([\Phi_F\left(x,-t\right), \rho_0\left(\Phi_F\left(x,-t\right)\right)]^\intercal, t)_{\downarrow\rho} \\ &= \Phi_{\bar{F}}([\Phi_F\left(x,-t\right), \rho_s\left(\Phi_F\left(x,-t\right)\right)]^\intercal, t)_{\downarrow\rho} \\ &= \rho_s\left(x\right)\end{aligned} \tag{12}$$

■

With Theorem 1 and Assumption 1, if there exists a $T > 0$ such that $\Phi_T(x,-T) \notin \mathcal{X}$, then clearly

$$\rho_0(\Phi_T(x,-T)) = 0 = \rho_s(\Phi_T(x,-T)),$$

therefore the density at x converges to the stationary density ρ_s in finite time T.

Lemma 1. *Under Assumption 1, if the system reaches a stationary density distribution,*

$$\int_{\mathcal{X}} \phi dx = 0.$$

Proof. Under Assumption 1, we have

$$\int_{\mathcal{X}} \nabla \cdot \left(\rho \cdot F\right) dx = \int_{\partial\mathcal{X}} \rho F \cdot \overrightarrow{n}\, ds = 0. \tag{13}$$

Then

$$\int_{\mathcal{X}} \frac{\partial \rho}{\partial t} dx = \int_{\mathcal{X}} \phi - \nabla \cdot \left(\rho \cdot F\right) dx = \int_{\mathcal{X}} \phi dx = 0.$$

■

D. Density function for MDP

Similarly, one can define the density over states in an MDP $\rho : S \to \mathbb{R}^N$, where $N = |S|$ is the cardinality of S. For a given policy π, let P^π denote the transition probability matrix:

$$P^\pi_{ij} = \mathbb{P}\left(s_{t+1} = s_j \mid s_t = s_i, a \sim \pi\left(s_i\right)\right). \tag{14}$$

Given an initial density ρ_0 over states, the evolution of the density under π follows

$$\rho_{t+1} = \gamma (P^\pi)^\intercal \rho_t + \phi\left(\rho_t\right), \tag{15}$$

where $\phi : \mathbb{R}^N \to \mathbb{R}^N$ is the supply function.

Remark 3. *Here we do not restrict $\mathbf{1}^\intercal \rho_t = 1$ as in the case of probability distribution. The probability distribution can be viewed as a special case of density with $\mathbf{1}^\intercal \rho_t = 1$ and $\phi = 0$.*

Remark 4. *The idea of the dual variable in MDP has been studied, for example, in [1], and recently in [6], but they differ from the density function discussed in this paper in that the density function has a physical meaning rather than simply being the dual variable. The evolution of density function is governed by not only the Liouville equation, but also the supply function and the initial condition. Therefore, we can pose constraints on the density function with physical meaning.*

III. Duality in optimal control

In this section, we show the duality relationship between the value function and the density function for the optimal control problem with continuous state space and input space.

Depending on the setup of the optimal control problem, the supply function ϕ will take different forms. We present several setups, but only present the detail of one setup, that is, the optimal control problem with terminal condition.

For clarity, for the remainder of the paper, we call the value function optimization the primal problem, and the density function optimization the dual problem.

A. Duality in optimal control with terminal condition

Consider the optimal control problem with terminal condition $x \in \mathcal{X}_f$, where $\mathcal{X}_f$ is the destination.

Assumption 2. *For simplicity, we assume that $\mathcal{X}_f$ is a compact set and all state in the state space $\mathcal{X}$ can reach $\mathcal{X}_f$ in finite time.*

We consider a supply function $\phi = \phi_+ + \phi_-$, where ϕ_+ is a stationary nonnegative supply function that only depends on x, encoding the information of the distribution of new states coming into the state space, and ϕ_- takes the following form:

$$\phi_-\left(x,t,\rho\right) = \begin{cases} 0, & x \notin \mathcal{X}_f \\ -\delta\left(t\right)\rho, & x \in \partial\mathcal{X}_f \\ -\phi_+(x), & x \in int(\mathcal{X}_f), \end{cases} \tag{16}$$

where $\delta(t)$ is a Dirac delta function at t. This means that the density function immediately becomes zero once the state enters $\mathcal{X}_f$.

An example of this setup is the robot navigation problem where the initial position of the robot follows the distribution ϕ_+ and the goal is to reach the destination. Another example is the mail collection problem where the destination is the post office and the distribution of mail collection pops up following the distribution of ϕ_+.

Remark 5. *Note the difference between ϕ_+ and ρ_0. ϕ_+ specifies how new states enter the system over time, while ρ_0 specifies the initial distribution of the states at $t = 0$.*

One can clearly formulate an optimal control problem for individual initial conditions, but instead we look at the overall cost of the whole system, similar to [14]. Given a control strategy $u = \mathbf{u}(x)$, let $V^{\mathbf{u}}$ be the cost-to-go associated to $\mathbf{u}$ for a given state, then the overall cost rate over time is

$$J = \langle \phi_+, V^{\mathbf{u}} \rangle_{\mathcal{X}}. \tag{17}$$

By Bellman's principle of optimality, we know that the optimal value function of each state is independent of the state trajectory before the it reaches that state, which implies that there exists a pure state feedback law $\mathbf{u}^\star$ that minimizes the overall cost J, and is determined by the following

equation:

$$\begin{aligned} J^\star =& \left\langle \phi_+, V^{\mathbf{u}^\star} \right\rangle_{\mathcal{X}} \\ \text{s.t.} & \mathbf{u}^\star(x) = \arg\min_{u \in \mathcal{U}} \ \nabla V \cdot F(x,u) \\ & C + \nabla V \cdot F = 0 \\ & V|_{\mathcal{X}_f} = D. \end{aligned} \tag{18}$$

Note that this is simply a inner product of the optimal value function and the positive supply function. Since $V^{\mathbf{u}^\star}$ is completely determined by the equality constraint, we leave out the optimization sign.

Alternatively, if the density function reaches a stationary distribution ρ_s, the overall cost rate can also be represented as

$$J = \langle C, \rho_s \rangle_{\mathcal{X}} + \langle D, -\phi_- \rangle_{\mathcal{X}}, \tag{19}$$

where the first part represents the overall running cost and the second part represents the terminal cost.

This means that instead of thinking of the value function for each x, we can think about the stationary density distribution ρ_s. The following optimization solves for the optimal overall cost:

$$\begin{aligned} & \min_{\rho_s, \mathbf{u}} \langle \rho_s, C \rangle_{\mathcal{X}} - \langle \phi_-, D \rangle_{\mathcal{X}} \\ & \text{s.t.} \nabla \cdot (\rho_s \cdot F(x, \mathbf{u}(x))) = \phi, \\ & \quad \forall x \in \mathcal{X}, \mathbf{u}(x) \in \mathcal{U}, \ \rho_s(x) \geq 0, \end{aligned} \tag{20}$$

Theorem 2. *The optimization in* (20) *and* (18) *are dual to each other and if there exists optimal solutions to both problems, there is no duality gap.*

Proof. We show one direction, from (20) to (18), and the other direction is similar. The Lagrangian is formulated as

$$\begin{aligned} \mathcal{L} = & \langle \rho_s, C \rangle_{\mathcal{X}} - \langle \phi_-, D \rangle_{\mathcal{X}} + \langle \mu, \phi - \nabla \cdot (\rho_s \cdot F) \rangle_{\mathcal{X}} \\ & - \langle \lambda_0, -\rho_s \rangle_{\mathcal{X}} - \langle \lambda_1, g \circ \mathbf{u} \rangle_{\mathcal{X}}, \end{aligned} \tag{21}$$

where $\mu : \mathcal{X} \to \mathbb{R}$, $\lambda_0 : \mathcal{X} \to \mathbb{R}_+$ and $\lambda_1 : \mathcal{U} \to \mathbb{R}_+$ are the Lagrange multipliers. First notice that

$$-\langle \phi_-, D \rangle_{\mathcal{X}} = \langle \delta \rho_s, D \rangle_{\partial \mathcal{X}_f} + \langle \phi_+, D \rangle_{int(\mathcal{X}_f)}$$

Then by Assumption 1, we use the adjoint condition:

$$\langle \mu, \nabla \cdot (\rho_s \cdot F) \rangle_{\mathcal{X}} = -\langle \nabla \mu, \rho_s \cdot F \rangle_{\mathcal{X}} = -\langle \rho_s, \nabla \mu \cdot F \rangle_{\mathcal{X}}. \tag{22}$$

The Lagrangian then becomes

$$\begin{aligned} \mathcal{L} = & \left\langle \rho_s, C + \mathbb{1}_{\partial \mathcal{X}_f} \delta D + \nabla \mu F + \lambda_0 \right\rangle_{\mathcal{X}} \\ & + \langle \phi_+, D \rangle_{int(\mathcal{X}_f)} + \langle \mu, \phi \rangle_{\mathcal{X}} - \langle \lambda_1, g \circ \mathbf{u} \rangle_{\mathcal{X}} \end{aligned} \tag{23}$$

The Kuhn-Karush-Tucker (KKT) condition reads

Stationarity condition:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \rho_s} = & \quad C + \mathbb{1}_{\partial \mathcal{X}_f} \delta D + \nabla \mu F + \lambda_0 = \ 0 \\ \frac{\partial \mathcal{L}}{\partial u} = & \quad \rho_s \left(\frac{\partial C}{\partial u} + \nabla \mu \frac{\partial F}{\partial u} + \lambda_1 \frac{\partial g}{\partial u} \right) = \ 0 \end{aligned} \tag{24}$$

Complementary slackness:

$$\mu \cdot (\phi - \nabla \cdot (\rho_s \cdot F)) = \lambda_0 \cdot \rho_s = \lambda_1 \cdot g(u) = 0 \tag{25}$$

This implies that when $\rho_s > 0$, i.e. for area in $\mathcal{X}$ with nonzero density,

$$\begin{aligned} \mathbf{u}^\star(x) = & \arg\min_{g(u) \leq 0} \ C + \nabla \mu \cdot F, \\ & C + \delta D \mathbb{1}_{\partial \mathcal{X}_f} + \nabla \mu F = 0, \end{aligned} \tag{26}$$

which directly come from the stationarity condition and utilized the fact that $\rho_s > 0 \to \lambda_0 = 0$. Furthermore, since $\rho_s = 0$ inside $\mathcal{X}_f$, the optimal input then can be picked arbitrarily from $\mathcal{U}$, therefore μ has to be constant within $int(\mathcal{X}_f)$. let $\mu_0 = \mu(x) \ |_{int(\mathcal{X}_f)}$. Note that by Lemma 1, at stationary density, $\int_{\mathcal{X}} \phi dx = 0$, which implies

$$\langle \mu - \mu_0, \phi \rangle_{\mathcal{X}} = \langle \mu, \phi \rangle_{\mathcal{X}} - \langle \mu_0, \phi \rangle_{\mathcal{X}} = \langle \mu, \phi \rangle_{\mathcal{X}}.$$

Therefore, we can replace μ with $\mu - \mu_0$ and both the KKT condition and the value of the Lagrangian remain unchanged. Without loss of generality, we can assign $\mu_0 = 0$. Then μ satisfies

$$\begin{aligned} & \mu = \begin{cases} 0, & x \in int(\mathcal{X}_f) \\ D(x), & x \in \partial \mathcal{X}_f \end{cases}, \\ & \forall x \notin \mathcal{X}_f, \nabla \mu \cdot F = -C \end{aligned} \tag{27}$$

Define

$$V = \mu + \mathbb{1}_{int(\mathcal{X}_f)} D, \tag{28}$$

then we have

$$\begin{aligned} & V|_{\mathcal{X}_f} = D, \\ & \forall x \notin \mathcal{X}_f, \nabla \mu \cdot V = -C, \\ & \mathbf{u}^\star(x) = \arg\min_{u \in \mathcal{U}} \ C + \nabla V \cdot F, \end{aligned} \tag{29}$$

which is exactly the solution of the optimal control problem in (18).

Besides, from (23), if such an solution to the optimal problem exists, the dual objective becomes

$$d^\star = \max_{\lambda_0, \lambda_1, \mu} \min_{\rho_s, \mathbf{u}} \mathcal{L} = \langle \phi_+, D \rangle_{\mathrm{int}(\mathcal{X}_f)} + \langle \phi, \mu \rangle_{\mathcal{X}}. \tag{30}$$

Since $\phi_- \ |_{\mathcal{X} \setminus \mathcal{X}_f} = 0$,

$$d^\star = \langle \phi_+, V \rangle_{\mathcal{X}}, \tag{31}$$

which shows that there is no duality gap. ∎

B. Density function in several other forms of optimal control

Consider an infinite horizon optimal control problem with the following cost function:

$$V(x) = \int_0^\infty e^{-\kappa \tau} C\left(x\left(\tau\right), u\left(\tau\right)\right) d\tau, \tag{32}$$

where κ is the discount factor. In this case, the negative supply function takes the following form:

$$\forall x \in \mathcal{X}, \phi_-(x) = -\kappa \rho. \tag{33}$$

The primal optimal control problem is the following:

$$\begin{aligned} J^\star = & \langle V, \phi_+ \rangle \\ \text{s.t.} & C + \nabla V \cdot F - \kappa V = 0 \\ & u^\star(x) = \arg\min_{u \in \mathcal{U}} \ C + \nabla V \cdot F. \end{aligned} \tag{34}$$

The corresponding density optimization takes the form

$$\begin{aligned}&\min_{\rho_s,\mathbf{u}}\langle\rho_s, C\rangle_{\mathcal{X}}\\ &\text{s.t.}\nabla\cdot(\rho_s\cdot F(x,\mathbf{u}(x)))=\phi_+-\kappa\rho_s,\\ &\quad\forall x\in\mathcal{X}, g(\mathbf{u}(x))\le 0,\ \rho_s(x)\ge 0,\end{aligned}\tag{35}$$

Another setup is a fixed horizon optimal control problem. In this case, there is no supply function or stationary density, but instead a initial distribution of the states ρ_0, and the cost function is defined as

$$V(x)=\int_0^T C\left(x\left(\tau\right),u\left(\tau\right)\right)d\tau+D(x(T)).\tag{36}$$

The primal optimal control problem is the following:

$$\begin{aligned}J^\star=&\langle V\left(0,\cdot\right),\rho_0\rangle_{\mathcal{X}}\\ &\text{s.t.}\frac{\partial V}{\partial t}+C+\nabla V\cdot F=0\\ &\mathbf{u}^\star(t,x)=\arg\min_{u\in\mathcal{U}} C+\nabla V\cdot F\\ &V(T,\cdot)=D,\end{aligned}\tag{37}$$

and we can show that the dual problem to this is

$$\begin{aligned}&\min_{\rho,\mathbf{u}}\ \langle\rho,C\rangle_{\mathcal{X}\times[0,T]}+\langle\rho(T,\cdot),D\rangle_{\mathcal{X}}\\ &\text{s.t.}\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\cdot F)=0,\\ &\quad\rho\left(0,\cdot\right)=\rho_0,\ \rho\ge 0\end{aligned}\tag{38}$$

C. Density function for MDP

Similar to the optimal control problem in continuous state and input space, there is duality relationship between the density function and the value function in MDP.

The overall reward for a MDP is the following:

$$J=\langle\phi,V\rangle_S,\ V\left(s\right)=\sum\nolimits_{t=0}^{\infty}\gamma^t R_{a(t)}\left(s\left(t\right),s\left(t+1\right)\right).\tag{39}$$

The Liouville equation for stationary density is derived from (15) as

$$\rho=\gamma(P^\pi)^\intercal\rho+\phi.\tag{40}$$

The dual optimization is then

$$\begin{aligned}&\max_{\rho,\pi}\sum_i^N\rho^i\sum_j^N P^{ij}_{\pi(s^i)}R^{ij}_{\pi(s^i)}\\ &\text{s.t.}\gamma\sum_j^N P^{ji}_{\pi(s^j)}\rho_j-\rho^i+\phi^i=0.\end{aligned}\tag{41}$$

Theorem 3. *The primal problem in* (39) *and dual problem in* (41) *are dual to each other with no duality gap.*

Proof. Starting with the dual problem in (41). The Lagrangian is then formulated as

$$\begin{aligned}\mathcal{L}&=\sum_i^N\rho^i\sum_j^N P^{ij}_{\pi(s^i)}R^{ij}_{\pi(s^i)}+\sum_i^N\mu^j\left(\gamma\sum_j^N P^{ji}_{\pi(s^j)}\rho^j-\rho^i+\phi^i\right)\\ &=\sum_i^N\rho^i\left(\sum_j^N P^{ij}_{\pi(s^i)}\left(\gamma\mu^j+R^{ij}_{\pi(s^i)}\right)-\mu^i\right)+\sum_i^N\mu^i\phi^i\end{aligned}\tag{42}$$

Replacing μ^i with V^i, the KKT condition implies

$$\begin{aligned}V^i&=\sum_j^N P^{ij}_{\pi(s^i)}\left(\gamma V^j+R^{ij}_{\pi(s^i)}\right)\\ \pi\left(s^i\right)&=\arg\max_{a\in A}\sum_j^N P^{ij}_a\left(\gamma V^j+R^{ij}_a\right),\end{aligned}\tag{43}$$

which is the optimality condition for the value function, and it's easy to check that when an optimal solution for the primal problem exists, there is no duality gap, i.e.,

$$\min_\rho\mathcal{L}=\sum_i^N\phi^i V_i.\tag{44}$$

■

IV. Constrained optimal control enforced with density function

With density function, it is convenient to pose some constrained optimal control problems that are hard to pose with value function. Here we list a few

- In an optimal control problem, preventing the state to enter the dangerous area $\mathcal{X}_d$,
- In a robot navigation problem solved as a finite-horizon optimal control problem, enforcing a lower bound on the proportion of states that reaches the destination at the end of the horizon,
- In an traffic assignment problem solved as an MDP, enforcing upper bounds on road sections to prevent congestion.

All of the above mentioned problems can be posed as constrained optimization of the density function, and we present a primal-dual algorithm to solve it.

We will show two examples, the first one is the optimal control problem with a destination, and the constraint is that the state never enter a dangerous area $\mathcal{X}_d$. The second example is an MDP with upper bounds on the density of some states.

A. Optimal control with safety constraint

For the optimal control problem with a terminal condition, the unconstrained version is studied in Section III-A. Although the safety constraint is hard to impose on the value function, it is very convenient to impose it on the density formulation. The constrained optimization of density function is

$$\begin{aligned}&\min_{\rho_s,\mathbf{u}}\langle\rho_s,C\rangle_{\mathcal{X}}-\langle\phi_-,D\rangle_{\mathcal{X}}\\ &\text{s.t.}\nabla\cdot(\rho_s\cdot F(x,\mathbf{u}(x)))=\phi,\\ &\quad g(\mathbf{u}(x))\le 0,\ \rho_s(x)\ge 0\\ &\quad\rho_s\mid_{\mathcal{X}_d}\le\rho^{\max},\end{aligned}\tag{45}$$

where $\rho^{\max}$ is the tolerance, and it takes the value 0 if the constraint is absolute.

This optimization on density function may be hard to solve, but one can use a primal-dual algorithm and solve

the primal value function problem instead. With this extra safety constraint, the Lagrangian becomes

$$\begin{aligned}\mathcal{L} = &\langle \rho_s, C\rangle_{\mathcal{X}} - \langle \phi_-, D\rangle_{\mathcal{X}} + \langle \mu, \phi - \nabla\cdot(\rho_s \cdot F)\rangle_{\mathcal{X}} \\ &-\langle \lambda_0, -\rho_s\rangle_{\mathcal{X}} - \langle \lambda_1, g\circ \mathbf{u}\rangle_{\mathcal{X}} + \langle \rho_s - \rho^{\max}, \sigma \mathbb{1}_{\mathcal{X}_d}\rangle_{\mathcal{X}},\end{aligned} \tag{46}$$

where $\sigma : \mathcal{X} \to \mathbb{R}_+$ is the Lagrange multiplier associated with the safety constraint. The primal problem then becomes

$$\begin{aligned} J^\star = &\left\langle \phi_+, V^{\mathbf{u}^\star}\right\rangle_{\mathcal{X}} \\ \text{s.t.} &\mathbf{u}^\star(x) = \arg\min_{u\in\mathcal{U}} \ \nabla V \cdot F \\ &C + \sigma \mathbb{1}_{\mathcal{X}_d} + \nabla V \cdot F = 0 \\ &V|_{\mathcal{X}_f} = D. \end{aligned} \tag{47}$$

The only difference from the unconstrained case is the perturbation term σ on the running cost within $\mathcal{X}_d$. The primal-dual algorithm for the constrained optimal control is then the following:

Algorithm 1 Primal-dual algorithm for optimal control with safety constraint

1: $\sigma(0) \leftarrow \mathbf{0}$, $k = 0$
2: **do**
3: Solve (47) with $\sigma(k)$, get $\mathbf{u}^\star$.
4: Estimate stationary density ρ_s under $\mathbf{u}^\star$.
5: $\sigma(k+1) \leftarrow \max\{\mathbf{0}, \sigma(k) + \alpha\left((\rho_s - \rho^{\max})\mathbb{1}_{\mathcal{X}_d}\right)\}$.
6: $k \leftarrow k+1$
7: **while** $\|\max(\mathbf{0}, \rho_s - \rho^{\max})\mathbb{1}_{\mathcal{X}_d}\|_\infty > \epsilon$
8: **return** $\mathbf{u}^\star, \rho_s, V$

$\alpha > 0$ is the step size and $\epsilon > 0$ is the tolerance on the complementary slackness condition. The algorithm iterates between the optimal value function problem, which solves the optimal value function and control strategy, and the dual problem, which computes the stationary density function and updates the running cost perturbation σ. It terminates if a feasible solution that is close enough to the optimum (assessed by the complementary slackness condition) is found.

We use a robot navigation problem as example, where the robot follows a simple 2D kinetic model:

$$\begin{aligned}\dot{x} &= u_1 \\ \dot{y} &= u_2.\end{aligned} \tag{48}$$

The destination is a small ball around the origin, and $\mathcal{X} = [-2, 2]\times[-2, 2]$. The input bound $\mathcal{U} = \{u \mid \|u\| \le 0.5\}$, the positive supply ϕ_+ is plotted in Fig. 1 and the red circled area is $\mathcal{X}_d$.

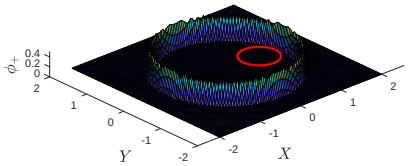

Fig. 1: positive supply function ϕ_+

The primal-dual algorithm terminates after 4 iterations, and a comparison of the result with and without the safety constraint is shown in Fig. 2.

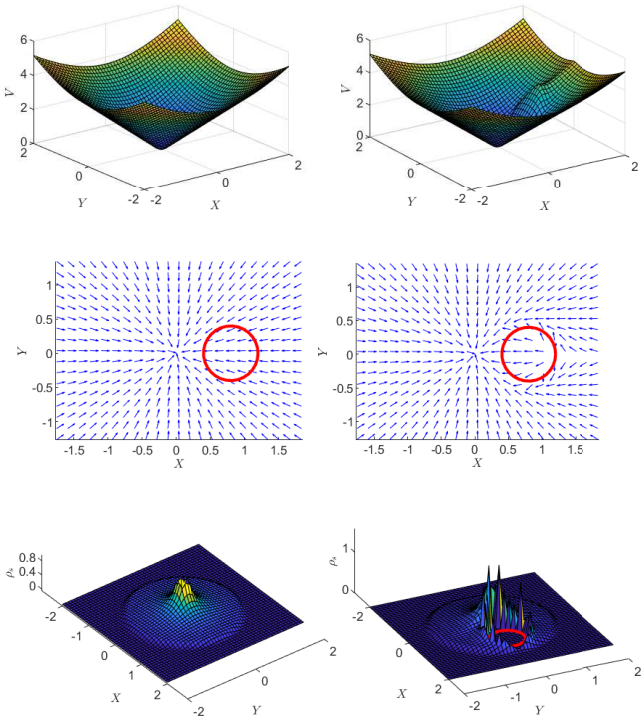

Fig. 2: Comparison of the optimal control problem solutions

The left side shows the value function, the phase portrait and the stationary density ρ_s for the unconstrained case, and the right side shows the constrained case. From the density plot, we see zero density within the danger area $\mathcal{X}_d$.

The value function of the constrained case has a small "bump" around $\mathcal{X}_d$, which is sufficient to steer all state around $\mathcal{X}_d$ and satisfy the safety constraint.

B. MDP with state density constraint

In this section, we present an application of the density function on constrained MDP. We propose a primal-dual algorithm that can not only solve MDP with density constraint, but also multiple MDPs with aggregate density constraint. We illustrate the method with a traffic control example from

[8], where the task is to control the macroscopic traffic flow for the area shown in Fig. 3. The area is divided into $N = 7$ regions, and for each region, the traffic capacity is governed by the Macroscopic Fundamental Diagram (MFD) of traffic flow [7]. The idea is that when the vehicle density is low, the traffic flow rate increases with vehicle density; when the vehicle density is larger than a threshold, congestion starts to form and the flow rate decreases with vehicle density. We call the turning point the critical density.

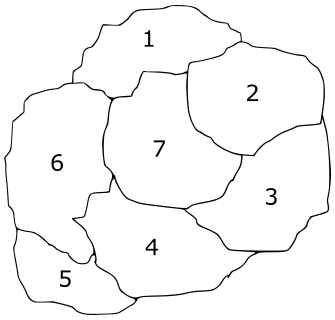

Fig. 3: The map of the traffic control area

It is assumed that each region in Fig. 3 has a critical density $\rho_i^{\max}$, and the task of traffic control is to minimize the cost while keeping the density of each region below the critical density.

We assume that for each vehicle, the transition cost only depends on the state. For example, a vehicle wants to get to region 1 from region 5, if it takes the path 5-4-7-1, then the total cost is $C_5 + C_4 + C_7 + C_1$.

Remark 6. *In this example, γ is chosen to be 1. However, we keep the γ in the derivation to allow for cases with $\gamma < 1$.*

In fact, this problem cannot be solved as a single MDP since the vehicles have different destinations. It is solved as 7 MDPs with 7 different destinations. Let $\phi_{+,i}^j$ denote the traffic demand from s_i to s_j, and we assume that all $\phi_{+,i}^j$ are given. The action is simply to choose which neighboring region to visit next. A stochastic policy would determine the transition probability matrices of the MDPs.

The density optimization problem is formulated as follows:

$$\begin{aligned} \min_{\pi,\rho} \sum\nolimits_{i=1}^{N} C_i \sum\nolimits_{j=1}^{N} \rho_i^{\pi_j} \\ \text{s.t.} \forall j \in \{1,...,N\}, \rho^{\pi_j} = \gamma (P^{\pi_j})^\intercal \rho^{\pi_j} + \phi^j \\ \forall i \in \{1,...,N\}, \sum\nolimits_{j=1}^{N} \rho_i^{\pi_j} \le \rho_i^{\max}, \end{aligned} \tag{49}$$

where π_j is the strategy for the traffic demand with destination s_j, which determines the transition probability matrix P^{π_j}. $P_{i,k}^{\pi_j}$ denotes the transition probability from s_i to s_k under π_j. $\rho^{\pi_j} \in \mathbb{R}^N$ is the traffic density vector with destination s_j under π_j. Similarly, we denote V^{π_j} as the value function vector under policy π_j with destination s_j. The negative supply ϕ_-^j is defined as

$$\phi_{-,i}^j = \begin{cases} 0, & i \neq j \\ -\rho_i^{\pi_j}, & i = j \end{cases}. \tag{50}$$

Claim 1. *The Liouville equation for this negative supply vector is equivalent to the following modified equation:*

$$\rho^{\pi_j} = \gamma (\overline{P}^{\pi_j})^\intercal \rho^{\pi_j} + \phi_+^j, \tag{51}$$

where $\overline{P}^{\pi_j}$ is obtained by modifying the j-th row of P^{π_j} to be all zero.

Proof. For s_j, since it's the destination, $P_{jj}^{\pi_j} = 1$ and the rest of the j-th row are zero. Subtract both sides of the Liouville equation in (49) by $\rho_j^{\pi_j}$, we get (51). ■

With (51), ρ^{π_j} can be conveniently calculated as

$$\rho^{\pi_j} = \left(I - \gamma (\overline{P}^{\pi_j})^\intercal\right)^{-1} \phi_+^j. \tag{52}$$

The equivalent primal value function formulation of the same problem is the following:

$$\begin{aligned} J = \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{N} \phi_{+,i}^j V_i^{\pi_j^\star} \\ \text{s.t.} \pi_j^\star = \arg\min_{\pi_j} P^{\pi_j} V^{\pi_j} \\ V^{\pi_j^\star} = \gamma P^{\pi_j^\star} V^{\pi_j^\star} + C + \sigma, \end{aligned} \tag{53}$$

where $\sigma \in \mathbb{R}^N$ is the perturbation term generated by the density constraint, $V^{\pi_j^\star} \in \mathbb{R}^N$ is the value function for traffic with destination s_j under the optimal policy $\pi_j^\star$.

We develop a primal-dual algorithm based on the Lagrangian of the MDP problem shown in (42):

Algorithm 2 Primal-dual algorithm for constrained MDP

$\sigma(0) \leftarrow \mathbf{0}$, $k \leftarrow 0$, $\forall j \in \{1,...,N\}, P^{\pi_j}(0) \leftarrow I$
do
  for $j \in \{1,...,N\}$ **do**
    for $i \in \{1,...,N\}$ **do**
      if $i = j$ **then**
        $V_i^{\pi_j} = C_i, P_i^{\pi_j} = \mathbf{0}$
      else
        $P_i^{\pi_j}(k+1) \leftarrow \mathbf{Proj}_{\Delta_j}\left(P_i^{\pi_j}(k) - \alpha \rho_i^{\pi_j} \gamma V^{\pi_j}\right)$
      end if
    end for
    $\rho^{\pi_j} = (I - (\gamma \overline{P}^{\pi_j}(k+1))^\intercal)^{-1} \phi_+^j$
    $V^{\pi_j} = (I - \gamma P^{\pi_j}(k+1))^{-1}(C + \sigma(k))$
  end for
  $\rho_c \leftarrow \sum_{j=1}^{N} \rho^{\pi_j}$
  $\sigma(k+1) \leftarrow \{\mathbf{0}, \sigma(k) + \beta(\rho_c - \rho^{\max})\}$
  $k \leftarrow k+1$.
while $\neg(\rho_c \le \rho^{\max})$ **or** $\max_j \|P^{\pi_j}(k+1) - P^{\pi_j}(k)\| \ge \epsilon$
return P^{π_j},ρ^{π_j}, V^{π_j}

$\alpha > 0$ and $\beta > 0$ are step sizes for the policy update and σ update. $\mathbf{Proj}$ is the projection operator, and Δ_i is the

probability simplex for s_i, defined as

$$\Delta_i = \left\{ P \in \mathbb{R}_+^{1\times N} \mid \sum\nolimits_j P_j = 1, (j \notin \mathcal{N}_i) \to (P_j = 0) \right\}, \tag{54}$$

where $\mathcal{N}_i$ is the neighbor set of s_i. The projection is done by solving the following quadratic programming:

$$\mathbf{Proj}_{\Delta_i} P_{des} = \arg\min_{P\in\Delta_i} \|P - P_{des}\|_2^2 . \tag{55}$$

As an example, we pose density constraint only on region 7, since it's in the center of the map and likely the most popular route to take. The comparison of the density distribution with and without the density constraint is shown in Fig. 4 in color difference. The left plot is the cumulative density in the unconstrained case where a very high density appears in region 7; the right plot is the cumulative density in the constrained case where the density in region 7 is diverted into other regions.

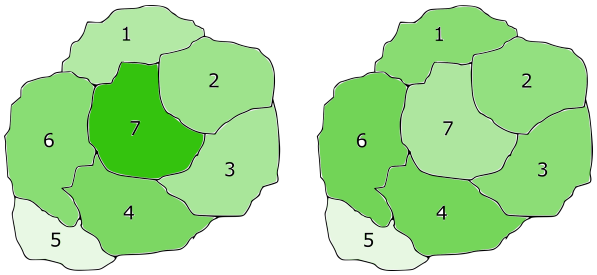

Fig. 4: Comparison of the cumulative density in constrained and unconstrained MDP

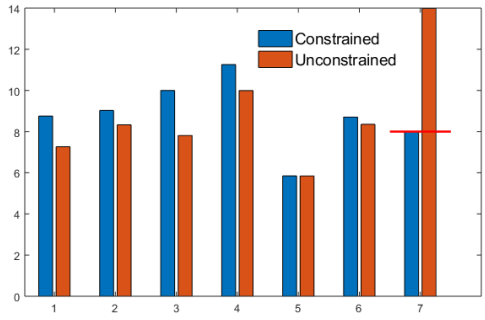

Fig. 5: Bar plot of the density in constrained and unconstrained cases

Fig. 5 shows the bar plot in the two cases, where the red line on the 7-th column shows the upper bound of the cumulative density of region 7.

The state cost is set as $C = [1.2, 1.2, 1.4, 1.1, 1, 1.6, 0.8]^T$. Under such C, the overall cost for the unconstrained case is 71.05, and 79.21 for the constrained case.

In the unconstrained case, the optimal strategy is usually deterministic (except special cases where V has equal entries) since it can be proved that there always exists a deterministic policy [13] that is optimal for a stationary MDP; however, in the constrained case, this is not the case. A simple example is a MDP with only two states, one has larger reward, but with a constraint on the density of it. Then the optimal policy is obviously a stochastic policy that barely satisfies the density constraint. Similarly, in the traffic example, the optimal policy for the constrained case renders the density at s_7 exactly at $\rho_7^{\max}$.

V. Conclusion

In this paper, we present the density function as the dual of the value function in both optimal control and Markov decision process. Some constraint such as safety constraint and density constraint can then be formulated as an optimization on the density function. Then a primal-dual algorithm is proposed to solve the optimal control problem with constraint on density function. We demonstrate the capability of the formulation with two examples, one on robot navigation and one on macroscopic traffic control. We plan to extend this work to the model-free reinforcement learning setting, where the density function cannot be computed directly from the model, and has to be estimated. Moreover, we plan to analyze the convergence of the primal-dual algorithm and improve the convergence rate.

References

[1] E. Altman. *Constrained Markov decision processes*, volume 7. CRC Press, 1999.

[2] 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*, 62(8):3861–3876, 2017.

[3] R. Bellman. *Dynamic programming*. Courier Corporation, 2013.

[4] R. W. Brockett. Optimal control of the liouville equation. *AMS IP Studies in Advanced Mathematics*, 39:23, 2007.

[5] Y. Chen, H. Peng, J. Grizzle, and N. Ozay. Data-driven computation of minimal robust control invariant set. In *Decision and Control (CDC), 2018 IEEE 57th Annual Conference on. IEEE*, 2018.

[6] B. Dai, A. Shaw, N. He, L. Li, and L. Song. Boosting the actor with dual critic. *arXiv preprint arXiv:1712.10282*, 2017.

[7] N. Geroliminis and J. Sun. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, 45(3):605–617, 2011.

[8] A. Kouvelas, M. Saeedmanesh, and N. Geroliminis. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B: Methodological*, 96:26–45, 2017.

[9] J. B. Lasserre, D. Henrion, C. Prieur, and E. Trélat. Nonlinear optimal control via occupation measures and lmi-relaxations. *SIAM journal on control and optimization*, 47(4):1643–1666, 2008.

[10] A. Majumdar, R. Vasudevan, M. M. Tobenkin, and R. Tedrake. Convex optimization of nonlinear feedback controllers via occupation measures. *The International Journal of Robotics Research*, 33(9):1209–1230, 2014.

[11] L. S. Pontryagin. *Mathematical theory of optimal processes*. Routledge, 2018.

[12] S. Prajna, P. A. Parrilo, and A. Rantzer. Nonlinear control synthesis by convex optimization. *IEEE Transactions on Automatic Control*, 49(2):310–314, 2004.

[13] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[14] A. Rantzer. A dual to lyapunov's stability theorem. *Systems & Control Letters*, 42(3):161–168, 2001.

[15] P. Zhao, S. Mohan, and R. Vasudevan. Control synthesis for nonlinear optimal control via convex relaxations. In *2017 American Control Conference (ACC)*, pages 2654–2661. IEEE, 2017.

[16] P. Zhao, S. Mohan, and R. Vasudevan. Optimal control for nonlinear hybrid systems via convex relaxations. *arXiv preprint arXiv:1702.04310*, 2017.